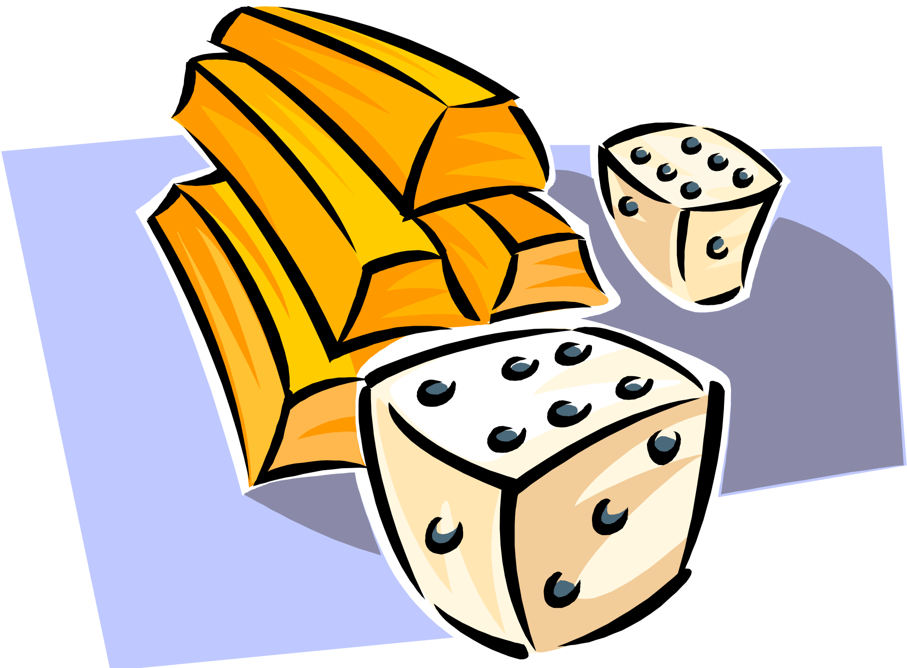
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
АВТОМОБИЛЬНО-ДОРОЖНЫЙ ИНСТИТУТ
ГОСУДАРСТВЕННОГО ВЫСШЕГО УЧЕБНОГО ЗАВЕДЕНИЯ
«ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

ДонНТУ

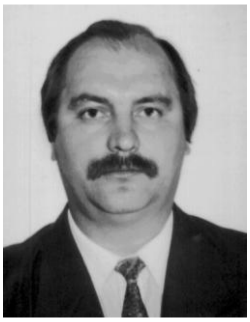
Полуянов В.П.

Статистика



Горловка
2002

УД7К 311
1Л4



Полуянов В.П. Основы статистики: учебное пособие / Автомобильно - дорожный институт. – Горловка, 2001. – с.

Учебное пособие содержит текст лекций, программу курса, задания для практических занятий, экзаменационные вопросы по дисциплине. Конспект лекций излагает теоретические основы одного из направлений современного естествознания, которыми, как минимум, должен владеть будущий специалист по экономике.

Учебное пособие предназначено для студентов автомобильно-дорожного института всех форм обучения по всем экономическим специальностям.

Рецензенты:

Утверждено в качестве учебного пособия редакционно-издательским советом автомобильно-дорожного института.

ISBN

СОДЕРЖАНИЕ

Тема 1. *Статистика как общественная наука и её информационная база*

1. ***Теоретические основы статистики. Связь статистики с другими науками.***
2. ***Предмет статистики как общественной науки.***
3. ***Основные стадии статистического исследования. Разделы статистической науки.***
4. ***Специфические приемы и методы статистического изучения явлений общественной жизни***
5. ***Современная организация статистики в Украине и ее задачи.***

1. Теоретические основы статистики. Связь статистики с другими науками

Термин «статистика» (нем. Statistik, от итал. stato, позднелат. status - государство), применяется для обозначения:

1) вида общественной деятельности, направленной на получение, обработку и анализ информации, характеризующей количеств, закономерности жизни общества во всём её многообразии (технико-экономические, социально-экономические, социально-политические явления, культура) в неразрывной связи с её качественным содержанием. В этом смысле понятие Статистика совпадает с понятием статистического учёта.

2) Отрасли общественных наук (и соответствующие ей учебные дисциплины), в которой излагаются общие вопросы измерения и анализа массовых количественных отношений и взаимосвязей.

В более узком смысле слова Статистика рассматривается как совокупность данных о каком-либо явлении или процессе, отрасли общественных наук; и как некоторая обобщающая характеристика явления.

Этот термин был введен в научный оборот в XVIII веке немецким ученым Готфридом Ахенвелем.

Развитие статистики тесно связано со становлением и развитием государственности, когда потребовались сбор и обработка массовых данных, используемых в государственном управлении. В дальнейшем, по мере совершенствования математического аппарата, статистические методы находят широкое применение во многих отраслях знаний.

По форме проявления причинных связей законы природы и общества делятся на два класса: детерминированные и статистические.

Детерминированные законы позволяют точно указать состояние явления в любой заданный промежуток времени. Статистические законы определяют это состояние не однозначно, а лишь с некоторой вероятностью, являющейся объективной мерой возможности реализации заложенных в прошлом тенденций изменения. Статистические законы охватывают все стороны материальных и общественных явлений. Выявлением и определением параметров статистических законов и занимается статистика.

Место статистики в системе наук определяется, прежде всего, ее предметом. Как общественная наука она связана с экономической теорией, отраслевыми экономиками, экономикой предприятий.

Статистика тесно взаимодействует в единой системе с оперативно-техническим и бухгалтерским учетом. Она использует единичные данные как составляющие массовых общественных явлений для получения обобщающих сведений в масштабе всей экономики.

Статистика по своему предмету отличается от математики и, изучая количественную сторону явлений, широко использует методы математической науки, прежде всего теорию вероятностей и математическую статистику.

Теория вероятностей как раздел математики, изучает свойства массовых случайных событий, способных многократно повторяться при воспроизведении определенного комплекса условий. Основное свойство любого случайного события, независимо от его природы, - мера, или вероятность его осуществления.

Математическая статистика опирается на теорию вероятностей и ее задача состоит в том, чтобы по ограниченным данным (выборке) восстановить с определенной степенью достоверности характеристики, присущие генеральной совокупности, т.е. всему мыслимому набору данных, описывающему изучаемое явление.

Из теории вероятностей в самостоятельные отрасли науки выделились теория случайных процессов, теория массового обслуживания, теория информации, эконометрическое моделирование и др. Одной из важнейших сфер приложения теории вероятностей является экономика, в которой используется эконометрическое моделирование, регрессионный анализ, трендовые и сглаживающие модели и другие методы.

Статистика – это наука, которая изучает принципы и методы работы с массовыми числовыми данными. Социально-экономические явления могут быть охарактеризованы определенной совокупностью показателей, которые обрабатываются с помощью различных статистических методов, что и входит в сферу изучения статистики.

2. Предмет статистики как общественной науки.

Статистика – наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной, количественное выражение закономерностей общественного развития.

Объект статистического изучения и анализа - различные явления и процессы общественной жизни.

Предмет статистики являются размеры и количественные соотношения между массовыми общественными явлениями, закономерности их формирования, развития, взаимосвязи. Т.е. статистика изучает количественную сторону массовых общественных явлений, однако количественная характеристика явления неразрывно связана с его качественным содержанием.

Статистика, рассматривая массовые общественные явления, позволяет сделать прогноз их развития, направление и скорость изменения количественных характеристик.

С вопросом о предмете статистики связаны понятия статистической закономерности и статистической совокупности.

Статистическая закономерность - это повторяемость, последовательность и порядок в массовых процессах. Выявить и количественно измерить статистическую закономерность можно с помощью закона больших чисел, в основу которого положена массовость и причинная зависимость явлений. Отсюда следует, что статистическая закономерность свойственна только некоторой совокупности, а не отдельному объекту.

Статистическая совокупность - это определенное множество элементов, объединенных условиями существования и развития. Состав элементов и способ их объединения определяют структуру совокупности. Совокупность представляет собой не механическое объединение элементов, а упорядоченную систему, каждый элемент которой представляет собой единство общего и единичного, необходимого и случайного.

Отображая характер действия объективных законов развития общества в конкретных условиях пространства и времени, статистические закономерности выявляются различным образом. Статистика изучает следующие ***группы закономерностей***:

- развития (динамики) явлений;
- распределения элементов совокупности;
- структурных сдвигов;
- связи между явлениями.

3. Основные стадии статистического исследования. Разделы статистической науки.

Любое статистическое исследование последовательно проходит ***три этапа***:

1. сбор первичного статистического материала, характеризуемый регистрацией фактов или опросом респондентов;
2. собранные данные подлежат систематизации и группировке, характеристики отдельных элементов переходят в обобщающие показатели в форме абсолютных, относительных или средних величин.
3. анализ вариации, динамики, взаимосвязей.

На этапе обобщения данных массового наблюдения (второй этап) элементы совокупности классифицируют по определенным признакам. Упорядоченную таким образом статистическую совокупность называют статистическим рядом. В зависимости от способа классификации различают ряды ***распределения*** и ряды ***динамики***. Ряд распределения - это результат классификации, группировки элементов совокупности в статике (на определенный момент или за определенный интервал времени). Динамический ряд составляется из значений статистических показателей во времени (по периодам или моментам времени) и описывает динамику развития массового процесса.

Статистика - многоотраслевая наука, состоящая из отдельных разделов или отраслей, которые, будучи ее самостоятельными частями, тесно связаны между собою. Обычно рассматривают четыре составных части статистики:

1. ***общая теория статистики***, которая рассматривает категории статистической науки, а также совместные для любых массовых явлений методы и средства анализа;
2. ***экономическая статистика***, изучающая явления и процессы, происходящие в экономике, разрабатывающая систему экономических показателей и методы изучения экономики страны или региона как одного целого;
3. ***отраслевые статистики*** (промышленная, финансовая, социальной инфраструктуры и т.д.) разрабатывают содержание и методы исчисления показателей, которые отражают особенности каждой отдельной отрасли;
4. ***социальная статистика***, предметом которой является изучение социальных условий и характера труда, уровня жизни, доходов, потребления материальных благ и услуг населением.

Иногда встречается и более подробная классификация разделов статистической науки:

1. статистическая теория;
2. социально-экономическая статистика;
3. социальная статистика;
4. экономическая статистика;
5. функциональная статистика;
6. отраслевая статистика;
7. региональная статистика;
8. статистика предприятия;
9. история статистической науки;
10. направления в современной статистике.

Являясь общественной наукой, статистика изучает действие объективных экономических законов в реальной жизни.

4. Специфические приемы и методы статистического изучения явлений общественной жизни

Статистическая методология – это совокупность взаимосвязанных между собой общих правил, специальных методов и приёмов.

Теоретической основой предмета статистики являются положения социально-экономической теории и принципы диалектического метода познания.

Основной метод статистики – диалектический метод познания всех явлений в их взаимозависимости и взаимообусловленности.

К ***специфическим методам*** статистики относят:

1. массовое наблюдение;
2. сводка и группировка;

3. обобщающие характеристики.

Статистическая методология - это комплекс специальных, свойственных только статистике методов, средств исследования. Она основывается на общефилософских (диалектическая логика) и общенаучных (сравнения, анализ, синтез) принципах.

В соответствии с принципами ***диалектической логики*** статистика любое общественное явление рассматривает не изолировано, а во взаимосвязи с другими, выявляет факторы, которые обуславливают вариацию значений признаков в границах совокупности, оценивает эффекты влияния факторов и величину причинно-следственных связей.

Статистическая методология основывается на сочетании ***анализа*** и ***синтеза***. Рассматривая совокупности элементов, статистика, с одной стороны, определяет схожие черты и отличия, объединяет элементы в группы, выделяя при этом разные типы и формы явлений, а с другой - обобщает информацию как по отдельным группам (типам), так и по совокупности в целом.

5.***Современная организация статистики в Украине и ее задачи.***

Статистический анализ массовых явлений и процессов ***-необходимое звено*** в системе управления экономикой и государством в целом. При помощи статистики осуществляется «обратная связь», то есть поток информации идет от объекта к субъекту управления - руководству предприятий, объединений, в территориальные, отраслевые и центральные органы власти. Без достоверной, всесторонней и своевременной информации невозможны эффективные управленческие решения.

Одной из ***главных задач*** статистики является изучение на основе статистических данных происходящих в стране социально-экономических и научно-технических процессов и явлений, экономический анализ материалов и представление Президенту, Правительству, органам исполнительной власти официальной статистической информации по актуальным вопросам экономического и социального развития страны, выполнения государственных и региональных программ по решению важнейших народнохозяйственных проблем.

На основе статистической информации правительство разрабатывает свою экономическую и социальную политику, оценивает ее результаты, составляет экономические прогнозы. Статистическая информация обеспечивает подготовку двухсторонних и многосторонних экономических соглашений между государствами. Статистика дает информацию для решения региональных задач, для предпринимательской деятельности - об уровне цен на товары в разных регионах, объемах реализации товаров, условиях кредитования, уровне и темпах инфляции, занятости и т.д.; наконец, в той или иной степени статистика нужна каждому из нас для принятия решений по выбору стратегии поведения.

Работа органов государственной статистики в Украине осуществляется в соответствии с Законом Украины «***О Государственной статистике***», введенным в действие постановлением Верховного Совета от 17 сентября ***1992*** г. Статья 3 этого Закона определяет ***главные задачи*** государственной статистики:

- реализация государственной политики в области статистики;
- сбор, разработка, обобщение и всесторонний анализ статистической информации о процессах, происходящих в экономической и социальной жизни Украины и ее регионов;
- разработка и внедрение статистической методологии, которая базируется на результатах научных исследований, международных стандартах и рекомендациях;
- обеспечение достоверности, объективности, оперативности, стабильности и целостности статистической информации;
- обеспечение доступности, гласности и открытости сводных статистических данных в пределах действующего законодательства.

Статья 4 определяет, что в ***систему органов государственной статистики*** в Украине входят: Министерство статистики Украины (с 1997 г. - Государственный Комитет по статистике (Госкомстат)), органы государственной статистики в Республике Крым, областях, районах и городах, составляющие единую систему органов государственной статистики Украины. Кроме органов государственной статистики, государственной статистикой занимаются также министерства и ведомства Украины, другие юридические лица для выполнения заданий, входящих в их компетенцию в соответствии с утвержденными формами государственной статистической отчетности. Объемы ведомственной (отраслевой) статистической отчетности определяются министерствами и ведомствами самостоятельно по согласованию с органами государственной статистики.

Основные функции всех статистических органов состоят в сборе, обработке, анализе и представлении данных в удобном пользователю виде. Статистические службы обязаны оперативно предоставлять информацию органам управления.

Все статистические органы, имеют внутреннюю структуру: отделы (а в Госкомстате Украины - управления) статистики предприятий, сельского хозяйства, капитального строительства и т.д.

Организация ***международной статистики*** осуществляется статистическими службами Организации Объединенных Наций (ООН), специализированных учреждений (МОТ, ВОЗ и др.) и других международных организаций - Организации экономического сотрудничества и развития (ОЕСД), Европейского сообщества (ЕС), Международного валютного фонда (МВФ), Мирового банка и т.д. Деятельность статистических служб этих организаций включает разработку международных стандартов, обеспечивающих сравнимость статистических показателей разных стран, осуществление международных сопоставлений, публикацию данных по группам стран, регионам и миру в целом.

Координация деятельности статистических служб стран - членов СНГ осуществляется созданным в 1992 г. ***статистическим комитетом Содружества Независимых Государств***. Публикуются статистические сборники по странам СНГ и другим государствам ближнего зарубежья.

Начало ***реформирования*** системы государственной статистики в Украине в соответствии с потребностями рыночных преобразований положил Указ Президента Украины от ***23 мая 1992*** года «***О переходе Украины к общепринятой в международной практике системе учета и статистики***». Исходя из необходимости интеграции Украины в мировые экономические отношения и перестройки системы экономической информации, учета и статистики в соответствии с международными стандартами, этот Указ предусматривал создание Координационного совета по вопросам организации перехода Украины на международную систему учета и статистики, утверждая персональный состав этого Совета. Основными ***задачами Координационного Совета*** определялись следующие:

- реализация мер по перестройке системы экономико-статистических показателей;
- внедрение национального счетоводства, методов бухгалтерского учета, единой системы классификации и кодировки технико-экономической информации с учетом международных стандартов и рекомендаций, включая системы каталогизации и штриховой кодировки продукции.

Кабинет Министров Украины постановлением ***№326 от 4 мая 1993 г.*** утвердил концепцию построения национальной системы статистики и Государственную программу перехода на международную систему учета и статистики.

Концепция предусматривала решение проблем интеграции методологической базы отечественной статистики с общемировыми стандартами осуществить в три этапа:

- ***Подготовительный*** этап, который определяет организационные, методологические и нормативно-технические принципы согласования действующей статистической практики с рекомендациями международных статистических организаций. Для этого предусматривается:
  - изучение и определение путей согласования экономической терминологии и действующих классификаций во всех отраслях статистики с международными и прежде всего классификации отраслей народного хозяйства с Международной классификацией всех видов экономической деятельности;
  - пересмотр действующих форм первичного, бухгалтерского и статистического учета и отчетности, методов их ведения, сбора, обработки и использования;
  - централизация разработки всех статистических показателей, необходимых на общегосударственном уровне;
  - выполнение экспериментальных расчетов полной системы показателей в соответствии с международными стандартами;
  - переподготовка специалистов, занятых в общенациональных экономических органах.
- ***Переходный*** этап, на котором внедряются в практику международные статистические рекомендации на основании методологических разработок, выполненных национальными экономическими учреждениями, научно-

исследовательскими институтами, творческими коллективами, учеными, а также осуществляется переподготовка работников учета во всех отраслях народного хозяйства.

- ***Заключительный*** этап, во время которого достигается интеграция всех сфер статистической деятельности и расширение прямых и постоянных контактов статистических органов Украины с статистическими подразделениями ООН и ее специализированными органами, другими международными и национальными организациями.

Реорганизация системы учета и отчетности потребовало решение целого ряда задач:

- унифицировать статистическую отчетность независимо от отраслей народного хозяйства, видов деятельности и форм собственности, отказаться от отчетности, которая характеризует внутреннюю производственную деятельность предприятий. Одновременно с этим возникает потребность статистического наблюдения новых хозяйственных формирований - концернов, ассоциаций, консорциумов, бирж, коммерческих банков, индивидуальных и фермерских хозяйств и т.д.
- изменить порядок сбора и разработки статистической отчетности: сокращение сплошной отчетности и значительное расширение выборочных обследований, переписей, разовых учета, специальных расчетов. Это касается в первую очередь мелких и средних предприятий, особенно частных.
- выбрать оптимальную периодичность сбора информации. В новых условиях интерес центральных органов государственной исполнительной власти на некоторую информацию за очень короткие периоды (недельная, декадная) уменьшается. Однако заинтересованность в такой информации возрастает у предприятий, которым нужно реагировать не только на долговременные тенденции, а и на краткосрочные сигналы рынка.
- разработать систему статистических показателей экономической конъюнктуры и создать специальное периодическое издание для обеспечения оперативного анализа совершения экономических реформ и интеграции национальных статистических показателей с международными публикациями по этим вопросам.
- определить принципы доступа общественности к информации, углубить информированность населения о социально-экономическом состоянии страны через статистические показатели. Безусловная защита должна иметь коммерческая тайна отдельных хозяйственных единиц. В связи с этим надо определить круг данных, не являющихся коммерческой тайной.

Обязательным условием внедрения этих мероприятий является переход на международные классификации и в первую очередь адаптация к Международной стандартной отраслевой классификации всех видов экономической деятельности и к Гармонизированной системе кодирования товаров.

Одним из приоритетных направлений создания национальной статистики является ***пересмотр и совершенствование макроэкономических показателей*** в соответствии с международными стандартами. Это предусматривает внедрение системы национальных счетов (СНС), рассматривающих экономику как единое целое, без принципиальной разницы между производством материальных благ и предоставлением услуг, дающих общую оценку результатам деятельности всего народного хозяйства как с материально-вещественной, так и с финансовой точки зрения для всех хозяйственных единиц на всех уровнях функционирования экономического механизма.

Все это требует с одной стороны пересмотра ***системы бухгалтерского учета*** на уровне хозяйственных единиц и финансовых учреждений, интеграции бухгалтерской и финансовой отчетности, а с другой стороны - оптимального сочетания отраслевых и межотраслевых статистик, охвата ими полного круги объектов наблюдения при разработке стоимостных и натуральных показателей.

В этом же Постановлении в соответствии с концепцией была утверждена и ***Государственная программа перехода Украины на международную систему учета и статистики***. В программе по каждому направления определялись сроки реализации и скоординированы действия министерств и ведомств, местных статистических органов, концернов, ассоциаций, общественных организаций, научно-исследовательских и учебных учреждений, привлекаемых к ее выполнению.

Программой признавались следующие приоритетные направления на пути перехода к принятой в международной практике системе учета и статистики:

- приведение в соответствие с международными стандартами общеметодологических основ государственной статистики и системы показателей;
- создание системы национальных счетов, включая разработку межотраслевых балансов;
- реорганизация первичного, бухгалтерского и банковского учета, статистики государственных финансов и финансов предприятий, кредитно-денежной статистики;
- разработка платежного баланса;
- приведение в соответствие с международными стандартами расчетов уровней инфляции, в частности показателей статистики цен потребителей, изготовителей, экспорта и импорта, а также статистики занятости населения и безработицы;
- внедрение ведения Государственного реестра учетных единиц;
- разработка единой системы классификации и кодирования технико-экономической информации;
- внедрение штрихового кодирования;
- подготовка и переподготовка кадров учета и статистики;
- развертывание широкого международного сотрудничества в области статистики и учета.

Первый этап реформирования, был в основном завершен в 1997 году. Реализация второго этапа осуществлялась в течение 1998 – 2002 гг.

Контрольные вопросы к теме 1.

1. Что означает термин статистика?

2. Что является предметом статистического изучения?
3. В чём заключается связь предмета статистики с другими науками и в чём их различие?
4. Что является теоретической основой предмета статистики?
5. Какой обобщающий подход к исследованию общественных явлений?
6. Перечислите специфические методы предмета статистики.
7. Назовите разделы статистической науки и дайте их краткую характеристику.
8. Каковы задачи государственной статистики в условиях перехода к рыночным условиям?
9. Охарактеризуйте современную организацию статистики в Украине.
10. Дайте характеристику основных законодательных актов 1992-1998 г.г. по вопросам реорганизации системы государственной статистики в Украине.

Тема 2. Статистическое наблюдение

1. *Понятие статистического наблюдения*

Статистическое наблюдение - это тщательно спланированная, научно организованная регистрация массовых данных о любых социально-экономических явлениях и процессах. Его характерной чертой является характер и массовость данных, а также способ их получения.

Различают первичное и вторичное статистическое наблюдение. ***Первичное наблюдение*** - это регистрация исходных данных непосредственно от объекта наблюдения. Примером может быть текущий учет количества зарегистрированных браков и разводов в соответствующем учреждении; опрос населения относительно отношения к процессу приватизации имущества. ***Вторичное наблюдение*** - это сбор ранее зарегистрированных и обработанных статистических данных, к примеру материалов банковских отчетов, результатов аудиторской проверки, итогов биржевых торгов.

Статистические данные - это массовые системные количественные характеристики социально-экономических явлений и процессов. Массовость отличает статистические данные от других, поскольку благодаря переходу от отдельных к массовым фактам, можно выявить общую закономерность, устранив влияние случайных причин.

К статистическим данным предъявляются определенные требования:

- ***достоверность;***
- ***полнота данных*** как по объему, так и по их содержанию;
- ***своевременность;***
- ***сравнимость***, в том числе и по единицам измерения;
- ***доступность данных.***

Статистическое наблюдение осуществляется в ***три этапа***:

- подготовка наблюдения;
- регистрация статистических данных;
- формирование базы данных.

Самым ответственным этапом является подготовка статистического наблюдения. На этом этапе решаются основные ***методологические вопросы:*** что и как будет изучаться и на какие вопросы должны быть получены ответы. На этом же этапе решаются ***организационные вопросы:*** кто, где, когда проводит наблюдение и что для этого необходимо. Таким образом, на первом этапе составляется подробный ***план статистического наблюдения,*** содержащий решение методологических и организационных вопросов. В соответствии с поставленной целью определяют круг конкретных задач, которые решают в результате наблюдения. Разработка ***программно-методологических вопросов*** плана наблюдения состоит в научно-практическом обосновании и определении сути явления, условий его формирования и проявления. Кроме того, разрабатывается система признаков, которые характеризуют явление, принимается в расчет возможность их количественной обработки и проверки на точность.

Цель наблюдения - получение статистических данных, которые служат основанием для обобщенной характеристики состояния и развития явления или процесса с определением соответствующей закономерности. Конечной целью наблюдения является подготовка управленческих решений и соответствующих мероприятий. Цель обследования определяет его объект.

Сложную задачу представляют определение объекта и единицы наблюдения. ***Объект наблюдения*** – статистическая совокупность единиц изучаемого явления. Четкое определение сути и границ объекта позволяет предотвратить разные трактовки результатов обследования. Для этого применяются цензы. ***Ценз*** - это набор количественных ограничительных признаков. Совокупность состоит из отдельных единиц. Единицей совокупности может выступать человек, факт, предмет и т.п. ***Единица наблюдения*** – первичная ячейка, от которой получают необходимую информацию. Каждая единица имеет множество признаков, которые зарегистрировать нельзя. Поэтому надо выбрать главное.

После определения носителей признаков и источников информации составляется программа наблюдения. ***Программа наблюдения*** - это перечень вопросов, на которые надо получить ответы в результате наблюдения. Содержание и количество вопросов формируются согласно с целью наблюдения и реальными возможностями его проведения (денежными, трудовыми затратами и временем регистрации). Составление программы – ответственная задача, так как успех наблюдения на 90% зависит о чёткости её составления.

Главными требованиями при ***определении признаков*** является простота формулировки, четкая трактовка и однозначность. Поскольку признаки по форме выражения могут быть не только количественные, но и качественные, то, подбирая шкалу измерения, предпочтение отдается не только более информативным признакам, но и признакам с более широкими возможностями статистической обработки. На основании сформированного перечня признаков разрабатывается статистический инструментарий.

Статистический инструментарий - это набор статистических формуляров, а также инструкций и разъяснений относительно проведения статистического наблюдения, регистрации данных.

Статистический формуляр - это учетный документ единого образца, содержащий адресную характеристику объекта наблюдения и статистические данные о нем. Статистическими формулярами являются отчеты, переписные и опросные листки, бланки документов, анкеты.

Некоторые формуляры содержат также краткую инструкцию по их заполнению.

В процессе составления анкет используют вопросы закрытого, полузакрытого и открытого типов. К ***закрытому типу*** относятся вопросы, к которым прилагается полный перечень ответов, а респонденты должны только выбрать нужный. В ***полузакрытом типе*** дополнительно к предлагаемому перечню ответов имеется возможность дать свой собственный ответ. Вопросы ***открытого типа*** нуждаются в самостоятельном формулировании ответа.

При подготовке вопросов анкет одновременно готовятся и макеты таблиц, предназначенные для заполнения информацией, полученной в результате статистического наблюдения. Это позволяет установить взаимосвязь отдельных вопросов.

Программой наблюдения предусматривается также определение ***вида и способа регистрации данных***. Преимущественно вид и способ наблюдения зависят от его цели, сути объекта наблюдения, объема и степени точности ожидаемых результатов. Важными обстоятельствами есть также финансовое и трудовое обеспечение обследования, фактор времени.

Непременным условием программ наблюдения является ***преемственность*** их содержания, единиц наблюдения, методик исчисления. Это необходимо для сравнения результатов наблюдения во времени и в пространстве.

Важным вопросом подготовки статистического наблюдения является ***обеспечение точности данных регистрации.*** Точность результатов достигается, с одной стороны, соответствующей ***системой контроля,*** а другой, - тщательно отработанным ***механизмом сбора*** данных и практическим опытом в этой работе. Такой опыт формируется во время ***пробных,*** так называемых ***пилотажных*** обследований - небольших по объему, имеющих целью испытать, уточнить программу наблюдения, повысить качество разработки организационных вопросов.

Вторую составную частью плана наблюдения представляет система организационных вопросов. ***Организационные вопросы*** – это:

- место проведения обследования;
- время проведения обследования;
- организации и персонал, привлеченные к обследованию;
- материально-техническое обеспечение;
- система гарантии точности результатов.

Организационные и программно-методологические вопросы тесно связаны между собой и зависят от цели и условий обследования.

Прежде всего выясняется, на какой орган возложена ответственность за проведение обследований, их подготовку. В зависимости от масштабности объекта наблюдения, а также заинтересованности в результатах обследования, можно выделить следующие группы органов:

- центральные органы государственной статистики, а именно Госкомитет статистики Украины и его региональные отделения - государственные обследования на макроуровне;
- статистические отделы министерств и ведомств - государственные обследования локального по тематике характера;
- специальные институты, агентства, международные организации - обследования, основанные на изучении общественного мнения или мотивации, поведения и оценки отдельных субъектов общественно-экономической жизни.

К таким институтам относятся: Институт социологии НАН Украины, филиал Института общественного мнения Геллапа, Международный институт социологии, Международная организация Труда (МОТ), Международный центр исследования проблем тенденций деятельности экономического характера, Генеральное агентство в делах экономики и финансов при комитете ЕС.

- аналитические отделы отдельных экономических структур (предприятий, организаций, фирм, банков, бирж, страховых обществ) - обследования на микроуровне.

Дальше необходимо обосновать ***места обследования*** - пункты, в которых находятся единицы наблюдения и осуществляется регистрация данных. ***Время наблюдения*** (объективное время) - это время, к которому относится данное наблюдение. Если объектом наблюдения является ***процесс,*** то выбирается интервал времени, за который накапливаются данные. Если объектом наблюдения является ***определенное состояние,*** то выбирается ***критический момент*** - момент времени, на который осуществляется регистрация данных. ***Период наблюдения*** (субъективное время) - время, в течение которого регистрируются данные.

Проведение любого статистического наблюдения нуждается в соответствующем материально-техническом обеспечении: вычислительной и множительной технике, транспортных средствах, средствах печати, статистическом инструментарии и рекламных носителях.

На ***втором этапе*** осуществляется непосредственный сбор статистических данных. Этот процесс кропотливый, нуждается в четком взаимодействии, координации и оперативности всех исполнительных служб. Чем короче время сбора, тем более своевременной и дешевой будет информация. От качества сбора зависят также точность, полнота и достоверность данных.

Третий этап предусматривает контроль и накопление данных наблюдения, а также организацию их хранения. На этом этапе отрабатывается система оперативного доступа и поиска необходимых данных. Это очень важно в условиях централизованного наблюдения, когда накапливаются огромные объемы полезной и дорогой информации, а найти ее и использовать очень тяжело, а иногда и невозможно.

2. Формы и виды наблюдения

Существует ***две формы наблюдения: отчётность и специальные обследования***. Отчётность даёт в современных условиях основную массу информации, но не всю. Поэтому приходиться прибегать к специальным обследованиям. Вся отчетность строго устанавливается государственным комитетом статистики по формам и временным периодам.

По видам наблюдения делятся:

- по полноте охвата единиц совокупности (сплошное, не сплошное)
- по моменту наблюдения (текущее, периодическое, единовременное)

При ***сплошном*** наблюдении обследуются все единицы совокупности.

Несплошное наблюдение в свою очередь подразделяется на:

- обследование основного массива,
- анкетное,
- монографическое,
- выборочное.

Обследование основного массива заключается в сборе информации о наиболее крупных единицах совокупности. Например, регистрация цен на рынках только в крупных городах Украины.

Анкетное наблюдение – это вид несплошного наблюдения и одновременно – способ получения информации. Оно состоит в том, что разработанная анкета распространяется определённому кругу лиц и после заполнения возвращается статистическим (или другим) органам. Заполнение анкеты – добровольное и анонимное.

Монографическое наблюдение - обследование нескольких единиц или небольшого числа совокупности, но с особой тщательностью и глубиной.

Выборочное наблюдение строится на принципах теории вероятностей и позволяет строго математически обрабатывать материалы. Этот вид наблюдения позволяет исследовать только часть совокупности, а полученные результаты с помощью специальных методов распространить с определенной долей вероятности на всю совокупность. В настоящее время широко распространяется для сбора информации.

Текущее наблюдение ведется систематически по мере возникновения явлений. Например, регистрация рождаемости, смертности, учет выручки реализации, явки и неявки работников на работу и т. д.

Периодическое наблюдение проводится через определенные, обычно одинаковые промежутки времени. Например, результаты сессии студентов вузов.

Единовременное наблюдение. Проводится один раз, по мере надобности.

3. Способы наблюдения

Выделяют три способа ***регистрации данных***:

- непосредственное наблюдение,
- документальное наблюдение,
- опрос.

Непосредственное наблюдение осуществляется путём регистрации изучаемых единиц на основе непосредственного осмотра, подсчёта, взвешивания и т. д. Например, изучение интенсивности пассажиропотоков на городском транспорте, осмотр железнодорожных вагонов на крупных станциях пути следования.

Специально организованные наблюдения - это форма наблюдения, которая охватывает сферы жизни и деятельности. К числу таких наблюдений относятся: переписи, учет, специальные обследования, опрос.

Перепись - это сплошное или выборочное наблюдение массовых явлений с целью определения их размера и состава на определенную дату. Перепись осуществляется периодически (как правило, с равным интервалом) или единовременно.

Учет - сплошные наблюдения массовых явлений, которые основываются на данных обзора, опросы и документальных записей.

Специальные обследования - выборочные наблюдения отдельных массовых явлений согласно с определенной тематикой, выходящей за границы отчетности. Они могут быть периодическими или единовременными.

Начала возрождаться прежняя форма наблюдения - ***статистический реестр*** - это список или перечень единиц определенного объекта наблюдения с указанием необходимых признаков, который составляется и обновляется во время постоянного наблюдения. В планах госстатистики намечено составление ***единых государственных реестров:*** населения, субъектов деятельности (предприятий, организаций разной формы собственности), домашних хозяйств, земельного фонда, технологий.

Реестр населения - это поименный перечень жителей региона, который регулярно пересматривается и обновляется.

Реестр предприятий и организаций - это перечень субъектов всех видов экономической деятельности с указанием их реквизитов и основных показателей. Этот реестр дает возможность наладить единое информационное пространство, в которое входят субъекты рынка.

Документальный способ. Данные берутся из документов, на этом способе основана отчётность. ***Отчетность*** - это форма наблюдения, при которой каждый субъект деятельности регулярно представляет свои данные в государственные органы статистики и ведомства в виде документов (отчетов) специально утвержденной формы.

Отчетность характеризуется такими свойствами, как обязательность, систематичность, достоверность.

Обязательность - сдача отчетов обязательна для всех зарегистрированных субъектов деятельности с соблюдением унифицированной формы, утвержденного перечня показателей, с указанием реквизитов подотчетного субъекта (названия, адреса, фамилии и подписи ответственного лица, даты составления отчета).

Систематичность предусматривает регулярное, своевременное составление и сдачу отчетности в утвержденные сроки.

Достоверность - данные, приведенные в отчетности, обязаны быть подлинными и исключать любые искажение (сокрытия и приписки). За достоверность предъявленных данных субъекты деятельности несут юридическую ответственность.

Отчетность составляется на основании первичных данных оперативного и бухгалтерского учета. В зависимости от уровня утверждения и назначения отчетность подразделяется на внешнюю и внутреннюю. ***Внешняя*** утверждается и собирается органами госстатистики, министерствами и ведомствами. ***Внутренняя*** - разрабатывается самим субъектом деятельности для собственных оперативных, управленческих и аналитических потребностей.

По периодичности сдачи отчетность делится на ***периодическую*** и ***годовую***. Периодическая отчетность (месячная, квартальная, полугодовая) охватывает показатели текущей деятельности субъектов, годовая - подводит итоги финансово-производственной деятельности субъектов за год.

Опрос - это, как правило, выборочное наблюдение мнений, мотивов, оценок, которые регистрируются со слов респондентов.

Исключением является сплошной опрос всего населения - ***референдум*** - массовое волеизъявление относительно принципиальных социально-политических и экономических вопросов.

Опрос осуществляется следующими способами:

- устный,
- анкетный или саморегистрация,
- корреспондентский.

Устный опрос используется в переписях населения, в социологических исследованиях. Этот способ более точен, но дорогостоящий.

Саморегистрация или анкетный способ. Он наиболее дешёвый, но менее точен. Широко используется при переписях бюджета семей, службами занятости.

Корреспондентский способ состоит в том, что органы статистики рассылают специальные бланки отдельным подобранным лицам и периодически получают заполненные материалы в установленные сроки. Например, сведения о потребительских «корзинах» разных социальных групп, доходы и расходы семей и др.

При любом способе проведения статистическое наблюдение проводится пассивно: статистика хочет как можно точнее зарегистрировать данные без какого-либо влияния на наблюдаемый процесс. Принципиально иным способом собирания данных является ***эксперимент***. В этом случае статистику принадлежит активная роль: он должен не только наблюдать, а полностью контролировать ситуацию, планировать эксперимент и реализовать свой план. Эксперимент позволяет выявить влияние каких-либо установленных ограничений или нагрузок на поведение людей. Например, влияние на скорость реакций человека пребывания без сна в течение одних, двух, трех суток. Эксперимент традиционно входил в круг методов биологической, медицинской статистики, приложений статистического метода в естественных науках. В настоящее время все большее распространение получают идеи «социального эксперимента».

4. Ошибки статистического наблюдения

Существует два вида ошибок:

- ошибки репрезентативности;
- ошибки регистрации.

Ошибки ***репрезентативности*** возникают при выборочном наблюдении.

Ошибки ***регистрации*** образуются вследствие неправильного установления фактов в процессе наблюдения или ошибочной их записи.

Ошибки регистрации делятся на случайные и систематические.

Случайные ошибки могут быть вызваны разными причинами (невниманием, низкой квалификацией работника, усталостью и т.д.) и выражаются в пропусках, описках, занесением данных в не те графы таблицы. Они не влияют на результат исследования, так как с помощью законов теории вероятностей их можно снивелировать.

Систематические ошибки имеют однонаправленность и делятся на преднамеренные и непреднамеренные.

Преднамеренные ошибки возникают в результате умышленного искажения информации (занижение прибыли фирмы, личных доходов).

Непреднамеренные ошибки совершаются без всякого умысла. Например, людям свойственно округлять цифры к 0 или к 5.Чтобы исправить искажение информации применяют сглаживание и выравнивание.

Собранные в результате статистического наблюдения данные подлежат контролю. Контроль данных наблюдения одновременно является методологическим и организационным вопросом. Если его рассматривать как способ ***предупреждения ошибок,*** то он ближе к методологии. Если контроль - это ***выявление и исправление ошибок,*** то это скорее организационное мероприятие.

Контроль - это прежде всего проверка данных обследований на их полноту и достоверность. Контроль полноты данных осуществляется, как правило, визуально, при проверке наличия данных по всем единицами и позициями. Данные на достоверность проверяют средствами логического и арифметического контроля.

Логический контроль - это проверка совместимости данных сравнением взаимозависимых признаков. Логический контроль устанавливает только наличие ошибки, а не ее величину. Определить размер ошибки и исправить ее можно в ходе ***арифметического контроля,*** то есть проверки прямым или побочным перерасчетом зарегистрированных данных.

Меры борьбы с ошибками:

- логический и арифметический контроль;
- использование выборочного наблюдения вместо сплошного;
- точная редакция вопросов программы и разработка инструкций.

***Контрольные вопросы к теме* 2.**

1. Дайте определение статистического наблюдения? В чем его сущность?
2. Какие две организационные формы наблюдения Вы знаете?
3. Какие вопросы составляют программно-методологическую часть плана, а какие относятся к вопросам организационной части плана?
4. Охарактеризуйте виды несплошного наблюдения?
5. Какие виды наблюдения по моменту регистрации Вы знаете?
6. Перечислите способы опроса для получения информационных данных?
7. Что представляют собой случайные и систематические ошибки регистрации?
8. Перечислите способы контроля для проверки данных?
9. Какие изменения происходят в статистической отчетности предприятий на современном этапе?

Задания для контроля

1. Определите, какой из приведенных примеров является статистическим наблюдением:
- регистрация количества безработных в службе занятости;
- оценка стоимости квартиры, выставленной на аукционе;
- определение потребления электроэнергии в отдельном домашнем хозяйстве;
- опрос мнения экспертов относительно перспектив развития бизнеса в регионе;

2. Определите, что является объектом следующих наблюдений:
- обследование коммерческих банков по вопросам их инвестиционной деятельности;
- обследование супермаркетов по оценке спроса населения на импортные продукты питания;
- обследование фермерских хозяйств с целью определения обеспеченности техникой.

3. Определите единицу совокупности и единицу наблюдения в приведенных примерах:
- мониторинг продажи компенсационных сертификатов на аукционах;
- изучение мнения пользователей платных стоматологических услуг в государственных и негосударственных медицинских учреждениях;
- обследование финансовой деятельности трастовых компаний.

4. Определите основной перечень вопросов программы следующих наблюдений:
- рынка туристических услуг;
- доходности объектов недвижимости;
- мнения потребителей рекламного продукции.

5. Определите объективное и субъективное время следующих наблюдений:
- обследования курса цен на акции по данным биржевых торгов, которые происходят 4 раза в месяц;
- учет численности зачисленных в вузы студентов на начало учебного года;
- учет остатков кредиторской задолженности банков на конец каждого года.

6. Какие организационные формы наблюдения целесообразно применить при обследовании:
- количества всех совместных предприятий и их реквизитов;
- платежеспособности клиентов страховых компаний;
- показателей работы нотариальных контор?

7. Определите вид и способ следующих наблюдений:
- оценка экспертов относительно качества введенного в действие жилья;
- подведение итогов приватизации объектов;
- обследование социально-демографического состава незанятого населения, обратившегося в службу занятости.

Тема 3. Сводка и группировка

1. ***Понятие сводки и группировки***
2. ***Виды группировок***
3. ***Группировочные признаки***
4. ***Понятие статистических рядов и таблиц***

1. Понятие сводки и группировки

Переход от единичного к общему происходит в статистике осуществляется через сводку. Сводка и группировка является второй стадией статистического исследования.

Под ***сводкой*** в *узком смысле* слова понимается подсчет общих итогов совокупности. Под ***сводкой*** в *широком смысле* слова понимают разделение совокупности на группы и подсчет итогов как по группам, так и в целом по совокупности.

Сущность ***статистической сводки*** состоит в классификации и агрегировании материалов наблюдения. Элементы совокупности по определенным признакам объединяются в группы, классы, типы, а информация о них агрегируется как в границах групп, так и в целом по совокупности. Основное назначение сводки состоит в таком агрегировании наблюдаемых массовых явлений и процессов, при котором можно выявить их типичные черты и закономерности.

Сводка является основой дальнейшего анализа статистической информации. По сводным данным рассчитываются обобщающие показатели, выполняется сравнительный анализ, а также анализ причин, по которым показатели одной группы отличаются от показателей других групп, а также изучаются взаимосвязи между признаками.

При составлении статистической сводки выполняется следующая ***последовательность работ***:

- разработка программы систематизации и группирования данных;
- обоснование системы показателей, характеризующих как отдельные группы, так и совокупность в целом;
- проектирование макетов таблиц, в которых представляются результаты сводки;
- определение технологических схем обработки информации, типов ЭВМ, программного обеспечения;
- подготовка данных к обработке на ЭВМ, формирование автоматизированных банков данных;
- непосредственная обработка данных для составления сводки, обобщение данных, расчет системы показателей.

Программа систематизации и группировки данных предусматривает *выбор* группировочных признаков и правил формирования групп. Разработка программы, как и обоснование системы показателей, зависит от цели исследования, сути изучаемого явления, особенностей совокупности, степени вариации группировочных признаков.

Сводка бывает ***централизованная*** и ***децентрализованная***. ***Централизованная*** сводка предполагает обработку материала в одном месте. ***Децентрализованная*** сводка – поэтапная обработка собранных данных. Так, годовая отчетность промышленных предприятий городов области поступает в отдел промышленности комитета государственной статистики каждой области. Там их обрабатывают, анализируют и составляют единый областной отчет по промышленности, который направляется в Госкомстат.

Результаты статистической сводки представляются в форме статистических таблиц, макеты которых разрабатываются вместе с программой обработки данных.

2. Виды группировок

Разделение совокупностей на группы, однородные в том или ином понимании, связано с такими терминами как ***систематизация, типология, классификация, группирование***. Традиционно распределение выполняют по схеме: из множества признаков, которые описывают явление, выбирают разделительные, а потом совокупность подразделяют на группы и подгруппы в соответствии со значениями этих признаков.

Главный принцип любого распределения основывается на двух положениях:

- в один класс, группу объединяются элементы в какой-то степени схожие между собою;
- степень сходства между элементами, которые относятся в один класс, значительно выше, чем между элементами, относящимися к разным классам.

При выполнении группировки необходимо решить:

- что взять за основу группирования;
- сколько групп, позиций необходимо выделить;
- как разделить группы.

Основой группировки может быть какой-либо атрибутивный или количественный признак, который имеет качественно отличные градации. Такой признак называют ***группировочным***. В зависимости от сложности массового явления (процесса) и цели исследования группировочных признаков может быть один, два и более. Группировка по одному признаку называется ***простой***, по двум и более – ***комбинационной***.

В статистической практике широко используются разграничение совокупностей по атрибутивным признакам - ***классификации*** и ***номенклатура***. Они разрабатываются международными и национальными статистическими органами и рекомендуются как статистический стандарт. Преимущественно это многоступенные классификации с подробной номенклатурой групп и подгрупп, с четко определенными требованиями и условиями отнесения элементов совокупности в ту или иную группу.

В странах с развитой рыночной экономикой используется целый набор международных статистических классификаций. К ним относятся отраслевая классификация видов экономической деятельности, стандартная классификация занятий, стандартная торговая классификация и др. Разновидностью классификаций являются товарные номенклатуры, например Брюссельская таможенная номенклатура.

В качестве примера рассмотрим более подробно основные принципы построения Общего Классификатора "Отрасли Народного Хозяйства Украины" (***ОКОНХ***), который является составной частью единой системы классификации и кодирования технико-экономической и статистической информации. ОКОНХ вмещает в себя такие понятия как отрасль, подотрасль, виды деятельности. Как известно, в международной классификации именно вид экономической деятельности используется как классификационный признак хозяйственных субъектов, независимо от форм собственности и организационно-правовых форм ведения хозяйства.

ОКОНХ определяет отраслевой принцип построения народного хозяйства. Кодами ОКОНХ индексируются предприятия, которые относятся к той или иной отрасли народного хозяйства. Для отнесения субъекта к соответствующей категории учета, необходимо определить такое понятие, как основной вид деятельности. Кроме основного вида деятельности следует рассмотреть также второстепенные и вспомогательные виды деятельности.

Существует несколько критериев отбора основного вида деятельности:

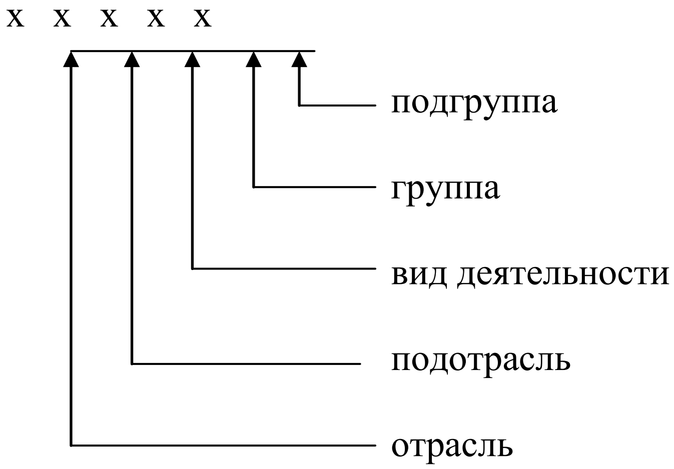
- по количеству занятых работников;
- по удельному весу в общем объеме производства продукции (товаров, услуг);
- по удельному весу в полученной общей прибыли.

Максимальное значение (в %) согласно выбранному критерию определяет ***основной вид деятельности*** хозяйственного субъекта. ***Второстепенным видом деятельности*** является любая другая деятельность, часть которой согласно выбранному критерию отбора меньше части основного вида деятельности. Каждый хозяйственный субъект может иметь несколько второстепенных видов деятельности. ***Вспомогательный вид деятельности*** - это деятельность, направленная исключительно на достижение главной задачи - содействовать выполнению основного вида деятельности хозяйственного субъекта.

В ОКОНХ отрасли народного хозяйства распределяются на сферу материального производства и непроизводственную сферу. К сфере материального производства относятся отрасли, которые определяют виды деятельности, создающие материальные блага в форме продукции, энергии, в форме перемещения грузов, хранения продуктов, сортировки, упаковки и других функций, которые являются продолжением производства в сфере обращения, в частности - торговля и общая коммерческая деятельность по обеспечению функционирования рынка. Остальные виды деятельности образовывают в своей совокупности непроизводственную сферу.

В ОКОНХ используется иерархическая классификация. Структура кода - пять цифровых разрядов.

Признаком распределения на всех степенях иерархии является подотрасль или вид деятельности. Каждый из последующих уровней дает видам деятельности более глубокую детализацию. Общая структура кода имеет такой вид:



Классификации имеют устойчивый характер и обеспечивают сравнимость данных в пространстве и времени.

Наряду с классификациями для выявления определенных вопросов конкретного исследования используют ***группировки***.

На группировки в статистическом анализе возлагаются определенные функции, в частности:

- изучение структуры и структурных сдвигов;
- определение типов социально-экономических явлений, выделение однородных групп и подгрупп;
- выявление взаимосвязей между признаками.

Согласно этими функциям группировки делятся на три вида: ***структурные, типологические, аналитические***.

При анализе явления необходимо изучать его структуру и происходящие структурные сдвиги. С этой целью используют структурные группировки, которые характеризуют распределение единиц совокупности по удельному весу к общему итогу. Примером структурной группировки может служить группировка по видам доходов и расходов бюджета Донецкой области в 1998 г. (табл. 1).

Таблица 1.

Исполнение доходной и расходной части
бюджета Донецкой области за 1998 год

	млн. грн.	**в % к общей сумме**
Доходы – всего,	**1557,8**	**100,0**
в том числе:		
Налог на добавленную стоимость	госбюджет	-
Налог на прибыль предприятий и организаций	609,0	39,1

Акцизный сбор	госбюджет	-
Подоходный налог с граждан	454,7	29,2
Взносы в Фонд ликвидации последствий Чернобыльской катастрофы	госбюджет	-
Поступления в Пенсионный фонд	госбюджет	-
Налог с владельцев транспортных средств	18,7	1,2
Платежи за использование природных ресурсов	184,0	11,8
Прочие налоги и платежи	291,4	18,7
Расходы – всего,	**1503,5**	**100,0**
в том числе:		
На управление	33,9	2,3
На правоохранительную деятельность	19,7	1,3
На образование	339,1	22,6
На охрану здоровья	372,7	24,8
На социальную защиту населения и Социальное обеспечение	260,9	17,3
На жилищно-коммунальное хозяйство	178,8	11,9
На культуру и искусство	30,2	2,0
На строительство	31,1	2,1
На транспорт и дорожное хозяйство	172,0	11,4
На выплаты, связанные с ликвидацией последствий Чернобыльской катастрофы	госбюджет	-
Выплаты Пенсионного фонда	госбюджет	-
Прочие расходы	65,1	4,3

Еще один пример структурной группировки – распределение населения Донецкой области по месту жительства характеризуется данными, приведенное в табл. 3.2., из которой видно, что в городах проживает подавляющее число жителей области и на одного сельского жителя приходится 9 городских. Каждая группа характеризуется также количеством и удельным весом населения трудоспособного возраста, что позволяет сделать более углубленный анализ.

Структурная группировка позволяет изучать интенсивность вариации группировочного признака.

Таблица 3.2

Распределение населения Донецкой области в 1998 г.

Место проживания	Всего населения, тыс. чел	Из них в трудоспособном возрасте	
		тыс. чел.	проц.
Города	4 492,1	2 601,9	57,9
Сельская местность	495,2	252,8	51,1
Всего	4 987,3	2 854,7	57,2

Весьма часто на основе структурной группировки изучают динамику структуры некоторой совокупности. Рассмотрим пример. На основании данных о динамике отраслевой структуры промышленной продукции, приведенной в табл. 3.3., рассчитаем показатель среднего абсолютного изменения отраслевой структуры промышленной продукции в Украине с 1985 г. (год начала перестройки) по 1998 г

Таблица 3.3.

Распределение промышленной продукции по отраслям

		1985	1998	W_1-W_0	$(W_1\text{-}W_0)^2$
	Вся промышленность	100,0	100,0		
	в том числе:				
1	электроэнергетика	3,2	16,5	13,3	176,89
2	топливная промышленность	7,2	11,6	4,4	19,36
3	черная металлургия	12,6	22,9	10,3	106,09
4	химическая и нефтехимическая промышленность	6,1	6,4	0,3	0,09
5	машиностроение и металлообработка	28,4	15,1	13,3	176,89
6	деревообрабатывающая и целлюлозно-бумажная промышленность	2,8	1,7	1,1	1,21
7	промышленность строительных материалов	3,6	3,3	0,3	0,09
8	стеклянная и фарфорофаянсовая промышленность	0,5	0,6	0,1	0,01
9	легкая промышленность	11,6	1,5	10,1	102,01
10	пищевая промышленность	18,7	14,9	3,8	14,44
11	мукомольно-крупяная и комбикормовая промышленность	3,0	2,0	1,0	1,00
	Сумма			58,0	598,08
	Показатель			5,3	7,37

Приведено по: Статистичний щорічник України за 1998 рік / Держкомстат України; За ред О.Г. Осауленко. - К.: Техніка, 1999. - 576 с.(стр. 113)

Расчет данного показателя производится по следующей формуле:

$$d_{w_1-w_0}=\frac{\sum_{i=1}^{k}\left|w_{1i}-w_{0i}\right|}{k}$$

Кроме того, возможен расчет сводного показателя абсолютных структурных сдвигов на основе формулы среднего квадратического отклонения:

$$s_{w_1-w_0}=\sqrt{\frac{\sum_{i=1}^{k}(w_{1i}-w_{0i})^2}{k}}$$

При отсутствии структурных сдвигов значения этих показателей равны нулю, их значение тем больше, чем значительнее изменения удельных весов групп. При выполнении указанных расчетов необходимо, чтобы число групп во всех сравниваемых периодах оставалось неизменным.

Типологическая группировка - это распределение качественно неоднородной совокупности на классы, социально-экономические типы, однородные группы. Основное назначение такой группировки состоит в идентификации типов. Выбор группировочного признака и количественных межгрупповых границ основывается на всестороннем теоретическом анализе сути явления, его характерных черт и особенностей формирования в конкретных условиях времени и пространства.

Типологическая группировка характеризует разделение исследуемого общественного явления на социально-экономические типы, классы. Например, деление населения на социальные группы, фермерские хозяйства – по производственному направлению, отрасли, производящие средства производства и отрасли, производящие предметы потребления. В отличие от ***структурной группировки***, которая на первый план выдвигает ***количественные различия***, типологическая группировка использует прежде всего ***качественные различия.***

В качестве примера, в учебнике по статистике[1, с. 33-34] приводится на материале России за 1992-1993 г.г. анализ тенденции обнищания населения. При этом степень обнищания оценивалась по методике ООН как отношение среднедушевого дохода отдельных домашних хозяйств к среднедушевому доходу по стране в целом независимо от абсолютного размера последнего. Граница бедности определялась на уровне 40-50% общего среднедушевого дохода.

Таблица 3.4

Степень обнищания населения России в течение 1992-1993 гг.

Доход, % до черты бедности	Идентификация групп	% к итогу		Структурные изменения, пунктов
		1992	1993	
50 и менее	Очень бедные	6,5	9,4	+2,9
51-100	Бедные	21,4	18,4	-3,0
101 и более	Не бедные	72,1	72,2	+0,1
Итого	х	100	100	х

По данным табл. 3.4 в 1992 году за границей бедности находилось 6,5% населения, в 1993 году удельный вес этой группы возрос до 9,4%, то есть на 2,9 пункта.

Углубленный анализ процесса трансформации групп можно выполнить при помощи динамичных группировок - таблиц перехода. На основе «шахматной» таблицы перехода (табл. 3.5) можно сделать взвод о том, что первая группа (очень бедных) не была устойчивой. В 1993 г. 32.4% численности этой группы перешло во вторую группу и 40.1% - в третью. Одновременно первая группа пополнилась за счет перехода из второй и третьей групп. Во второй группе сохранилось только 26.3% численности 1992 г., в третьей группе - 78.7%.

Таблица 3.5

Характеристика процесса обнищания населения России в течение 1993 года

Доход, в % к границе бедности в 1992 г.	Доход, в % к границе бедности в 1993 г.			Доля группы, %, 1992 г.
	50 и меньше	51-100	101 и больше	
50 и меньше	27,5	32,4	40,1	6,5
51-100	13,9	26,3	59,8	21,4
101 и больше	6,4	14,9	78,7	72,1
Доля группы, %, 1993 г.	9,4	18,4	72,2	100,0

Как видно из приведенных примеров, деление группировок на структурные и типологические достаточно условно. Как структурные, так и типологические группировки характеризуют структуру совокупности, определяют характерные черты и особенности, но отличаются уровнем качественных различий между группами.

Аналитические группировки выявляют связи между явлениями. С их помощью можно выявить наличие и направление связи между признаками, из которых один рассматривается как результативный, а другой (другие) как фактор (факторы), влияющие на результат (значение результативного признака). Вывод о наличии связи можно сделать на основе распределения двух взаимосвязанных признаков в соответствии с характером распределения частот. Так, по данным табл. 3.6 намерение изменить профессию ближайшем времени имеют преимущественно неудовлетворенные условиями работы(15 из 22) и, наоборот, среди довольных большинство (26 из 35) совсем не планирует изменять свой профессиональный статус.

Таблица 3.6

Распределение молодых работников по степени удовлетворенности условиями работы

Степень удовлетворенности условиями труда	**Имеете ли Вы намерение изменить профессию**			
	Да, в ближайшее время	***Да, в перспективе***	***Нет***	***Всего***

Удовлетворен	0	20	26	**46**
Не определился	7	18	9	**34**
Не удовлетворен	15	5	0	**20**
Всего	**22**	**43**	**35**	**100**

Если результативный признак количественный, для каждой группы по факторному признаку можно определить среднее значение результативного признака. При наличии связи между признаками групповые средние результативного признака систематически изменяются от группы к группе (увеличиваются или уменьшаются). Если связь между признаками линейная, то возможен расчет показателя средней силы связи:

$$b_{yx} = \frac{\bar{y}_m - \bar{y}_1}{x'_m - x'_1},$$

где $\bar{y}_m$, $\bar{y}_1$ - средние значения результативного признака в последней и первой группах соответственно; x'_m, x'_1 - середины интервалов (или средние значения) факторного признака в последней и первой группах. В случае прямой связи значение показателя больше нуля, обратной – меньше нуля. Если связь между признаками нелинейная, то показатель средней силы связи в большинстве случаев не имеет значения.

Взаимосвязь между результативным и факторными признаками может быть продемонстрирована применением комбинационной группировки. Однако, аналитическая группировка подробнее и выразительнее описывает эту взаимосвязь. Еще один пример аналитической группировки приведен в табл. 3.7. В данной группировке факторным признаком является срок сбора урожая озимой пшеницы после наступления полной зрелости зерна. Каждая группа характеризуется удельным весом посевной площади, с которой собран урожай, и средней урожайностью (результативный признак). По данным табл. 3.7, чем больший срок сбора урожая, тем меньшая урожайность и больше потери зерна. Разница урожайности между первой и третьей группами составляет 20 ц/га.

Таблица 3.7

Зависимость урожайности озимой пшеницы от срока сбора

Срок уборки	**Площадь уборки, %**	**Урожайность, ц/га**
Своевременно	30	42
С незначительным опозданием	50	36
Со значительным опозданием	20	22
В целом	**100**	**35**

Наиболее сложным вопросом теории группировок является выбор группировочного признака.

3. Группировочные признаки

Признаки в статистике – это свойства, характерные черты или особенности явлений, которые выражаются статистическими величинами. Признаки, положенные в основании группировки, называются группировочными. Они бывают ***атрибутивными*** (качественными) и ***количественными***. Примером атрибутивного признака может быть: пол, социальное положение, формы обучения в вузе и т.д.

Количественные делятся на прерывные (***дискретные***) и непрерывные (***интервальные***). Дискретный признак выражается определенным значением, например, балл успеваемости, число детей в семье и т.д. Интервальный признак может принимать любые значения в определенных пределах «от ... до ...».

Величина интервала – это разность между максимальным и минимальным значениями признака в каждой интервальной группе. Интервалы бывают ***равные*** и ***неравные***, ***закрытые*** и ***открытые***. Открытые интервалы имеют одну какую-нибудь обозначенную границу в первой и последней интервальной группе. Закрытые интервалы в обязательном порядке имеют как верхнюю, так и нижнюю границы интервала.

Вопрос о числе групп и границах интервала для каждого вида группировок должен решаться по-разному.

Величину интервала можно определить по формуле :

$$i = \frac{x_{\max} - x_{\min}}{k}$$, где

i – величина интервала ;

x_{max}, x_{min} – максимальное и минимальное значение признака в совокупности;

k – число групп.

Обычно число групп (k) предопределяет сам дискретный признак. Либо величину интервала определяют по формуле Стерджесса:

$$i = \frac{x_{\max} - x_{\min}}{1 + 3{,}332 \cdot \lg n},$$

где n – численность всей совокупности.

По знаменателю данного выражения можно определить число групп.

В практике используют три формальных способа формирования групп: ***равные*** интервалы, ***кратные*** интервалы и ***равные частоты***. В структурных и аналитических группировках обычно используют первый способ. При широком диапазоне вариации признака и неравномерном распределении его значений используют неравные интервалы, но сформированные по принципу кратности. Принцип равных частот используется в основном в аналитических группировках.

При определении числа групп обычно придерживаются следующих правил:

- в одну группу не должно попадать свыше половины всех единиц совокупности;

- в средних группах должно быть больше единиц совокупности, чем в крайних.

4. Понятие статистических рядов и таблиц

Конечные результаты сводки или группировки представляются в виде статистических рядов или таблиц.

Статистический ряд представляет собой наиболее простую форму представления статистического материала.

Одной из разновидностью структурных группировок являются ***ряды распределения*** Ряд распределения характеризует состав или структуру статистической совокупности по определенному признаку. Элементами ряда распределения являются значения признака x_i и соответствующие этим значениям частоты f_i. В зависимости от группировочного признака ряды распределения делятся на ***атрибутивные*** и ***вариационные***.

Вариацией значений какого-либо признака совокупности называется различие его значений у разных единиц совокупности в один и тот же период или момент времени. ***Вариационным*** рядом называется упорядоченное распределение единиц совокупности по значениям признака и подсчет числа единиц с тем или иным признаком. ***Атрибутивным*** рядом называется ряд, в построение которого осуществлено по атрибутивному (качественному) признаку.

Значение группировочного признака называется ***вариантой***. Каждой варианте соответствует определенная частота или доля. ***Частоты*** показывают, сколько раз повторяются отдельные варианты, а ***доли*** характеризуют их удельный вес в совокупности, то есть это относительные частоты.

Существуют три формы вариационного ряда: ***ранжированный ряд, дискретный ряд*** *и* ***интервальный ряд***. ***Ранжированный ряд*** – это перечень отдельных единиц совокупности в порядке возрастания или убывания изучаемого признака. ***Дискретный вариационный ряд*** представляет собой совокупность конкретных значений варьирующего признака и числа единиц совокупности с данным значением признака. ***Интервальный вариационный ряд*** состоит из интервалов признаков, значения которого изучаются и числа единиц совокупности, попадающих в данный интервал (частот) или долей этого числа от общей численности совокупности (частостей).

Статистический ряд - наиболее простая форма представления собранного материала. Его удобнее изучать при помощи графического метода.

Интервальные ряды изображают на графике с помощью ***гистограмм***, а дискретные – с помощью ***полигона распределения***. Гистограмма представляет собой столбиковую диаграмму, в которой основания столбиков расположены по оси абсцисс соответственно интервалам варьирующего признака, а высоты столбиков – частоты, соответствующие масштабу оси ординат. ***Полигон распределения*** представляет собой такое изображение дискретного вариационного ряда, при котором по оси абсцисс находятся значения признака (или середины интервалов), а по оси ординат – соответствующие частоты (частости) данного признака. Преобразованной формой вариационного ряда является ***ряд накоп-***

ленных частот. Это ряд значений числа единиц совокупности данного и всех меньших интервалов. Такой ряд называется ***кумулятивным***. Можно построить кумулятивное распределение «не меньше, чем», а можно «больше, чем». В первом случае график кумулятивного распределения называется ***кумулятой***, а во втором – ***огивой***.

Таблица – компактное изображение собранного материала. При помощи таблиц проще провести сравнение и анализировать сводные данные.

По логическому содержанию статистическая таблица рассматривается как «статистическое предложение». ***Подлежащим*** его есть объект исследования: перечень элементов совокупности, их группы, отдельные территориальные единицы или временные интервалы. Конечно подлежащее размещают в левой части, в названии строк. ***Сказуемое таблицы*** - это система показателей, которые характеризуют предмет как объект исследования. Сказуемое формирует в логической последовательности верхние заголовки таблицы.

Таблица состоит из следующих элементов :

- заглавие таблицы, где отражено ее основное содержание;
- подлежащие, то о чем говорится в данной таблице, располагается в левой части таблицы по строкам;
- сказуемое – показатели, характеризующие подлежащее. Располагается в правой части таблицы по графам;
- сетка – пересечение горизонтальных и вертикальных линий;
- содержание – цифровая характеристика, которая заносится в клетки таблицы.

По построению подлежащего различают статистические таблицы ***простые, групповые*** и ***комбинационные***.

Подлежащее простой таблицы простой таблицы представляет собой перечень отдельных единиц совокупности или дат. В последнем случае ряд принято называть хронологическим. Примером такой таблицы является ведомость на получение стипендии.

В групповой таблице подлежащим является группировка по какому-либо признаку, например, группы студентов вуза по формам обучения.

В комбинационных таблицах каждую группу разбивают на подгруппы по одному или нескольким другим признакам. Например, группы студентов вуза разбить не только по формам обучения, но и по полу, возрасту.

По построению сказуемого различают простые и комбинированные. При простой разработке, каждая графа сказуемого отдельно друг от друга характеризует подлежащее. При комбинированной таблице показатели сказуемого разрабатываются в сочетании друг с другом.

Размещение подлежащего и сказуемого подчинено принципа компактного и рационального изложения материала, его анализа.

Необходимо придерживаться определенных правил технического оформления таблиц.

- Таблица должна содержать только ту информацию, которая непосредственно характеризует объект исследования. Надо избегать лишней, второстепенной информации.
- Название таблицы, заголовки строк и граф имеют быть четкими, лаконичными, без сокращения. В названии указывается объект, его временные и географические признаки. Если названия отдельных граф (строк) повторяются, имеют одинаковые сроки или одинаковое содержание, то их целесообразно объединить совместным заголовком.
- В верхних и боковых заголовках указывают единицы измерения с использованием общепринятых сокращения (т, кВт, грн. и т.д.), иногда для них отводится отдельная графа. Если единица измерения одинакова для всех данных таблицы, ее указывают над таблицей.
- Строки и графы целесообразно нумеровать. При этом графу с названием подлежащего обозначают литерой алфавита, другие графы - номерами. Это дает возможность раскрыть методику расчета показателей сказуемого таблицы. Отдельные блоки таблицы можно разделить двойными линиями.
- Информация, размещаемая в строках (графах) таблицы, прежде всего групповой или комбинационной, обобщается итоговой строкой «Всего» или «В целом по совокупности», который завершает статистическую таблицу; если итоговая строка размещается первой, то детализация ее представляется при помощи словосочетания «в том числе». При этом можно представлять перечень не всех, а только определяющих составляющих.
- Числа, по возможности, необходимо округлять, в пределах одной и той же строки или графы обязательно с одинаковой степенью точности.
- Отсутствие данных в таблице обозначается в соответствии с причинами:

  – если клетка таблицы, прежде всего итоговая, не может быть заполненная, относится знак «х»;

  – когда сведения о явлении отсутствуют, ставится три точки«...» или «дан. от.»;

  – отсутствие самого явления обозначается тире (« - »);

  – очень малые числа записываются (0.0) или (0.00).

Если нужная дополнительная информация, определенные уточнения цифровых данных, в таблицу добавляется примечание.

Макет статистической таблицы - это комбинация горизонтальных строк и вертикальных граф, на пересечении которых располагаются клетки. Левые боковые и верхние клетки предназначены для словесных заголовков - перечня составляющих совокупности и системы показателей, остальные - для числовых данных. Основное содержание таблицы указывается в ее названии (рис. 3.8).

Название таблицы

Наименования строк	Верхние заголовки				
А	1	2	3	4	…

Боковые заголовки					
Итоговая строка					

Рис. 3.8. Макет статистической таблицы

На практике используются разные по степени автоматизации и типами ЭВМ технологические схемы обработки первичных данных. Одинаковыми для всех являются две операции: кодирования данных и переноса их с документов на технические носители информации, например на магнитные диски.

Коды - это условные идентификаторы признаков. Централизованно разработаны единые классификаторы, оформленные в виде словарей, обеспечивают однозначность кодов. Сформированы автоматизированные банки данных дают возможность многократного использования информации, при этом углубляется анализ, уменьшается количество ошибок.

***Контрольные вопросы к теме* 3.**

1. Какие виды сводки Вы знаете? Дайте их краткую характеристику.
2. В чем состоит значение метода группировок в статистическом анализе собранной информации?
3. Назовите виды группировок и дайте краткую характеристику каждого вида? Приведите примеры.
4. Какие вопросы решаются при построении групп и выборе группировочного признака?
5. Дайте характеристику интервалов.
6. В чем заключается особенность группировок при изучении взаимосвязей между признаками?
7. Укажите виды группировок: а) распределение брачных пар (жениха и невесты) по возрасту; б) распределение банков по уровню ликвидности и рейтинговой оценке их финансового положения; в) распределение фирм по длительности оборота оборотных средств с определением среднего уровня рентабельности капитала в каждой группе.
8. Как определить количество групп?
9. Что такое полигон и гистограмма, для чего они применяются?
10. Как построить кумуляту и где используется данный график ?
11. Какие виды статистических таблиц Вы знаете?
12. Каких правил следует придерживаться при построении таблиц?

По результатам опроса имеются следующие данные. Применены следующие коды самооценки склонности к риску:

1 - риск избегаю вообще;
2 - готов рискнуть с определенными гарантиями;
3 - люблю рисковать

№ п/п	Возраст, лет	Склонность к риску	№ п/п	Возраст, лет	Склонность к риску
1	48	1	16	45	3
2	29	2	17	23	2
3	43	2	18	61	2
4	32	2	19	33	2
5	28	2	20	42	1
6	62	1	21	27	3
7	35	1	22	37	2
8	22	3	23	65	1
9	56	3	24	40	2
10	50	2	25	49	2
11	25	3	26	57	2
12	67	1	27	19	3
13	21	3	28	70	2
14	39	1	29	36	2
15	28	2	30	52	3

Постройте ряды распределения респондентов:
А) по возрасту, выделив три группы:
1) до 30 лет;
2) 30-60;
3) 60 и старшие;
Б) по склонности к риску, выделив отдельные группы;
В) По возрасту и склонности к риску.
Результаты группировок представьте в виде статистических таблиц. Сделайте выводы.

По приведенным данным постройте аналитическую группировку, которая бы описывала качество работы ткацких станков в зависимости от их технического состояния (0 - прошел планово-предупредительный ремонт, 1 - нуждается в ремонте). Результаты приведите в виде таблицы.

№ п/п	Техническое состояние станков	Число обрывов нити на 100 кв. м. ткани	№ п/п	Техническое состояние станков	Число обрывов нити на 100 кв. м. ткани
1	1	73	11	1	71
2	1	76	12	0	66
3	0	68	13	0	65
4	0	64	14	1	73
5	1	70	15	1	72
6	0	67	16	0	69
7	1	79	17	1	71

8	1	75	18	0	72
9	0	73	19	1	70
10	1	70	20	1	73

Тема 4. Средние величины в социально-экономическом анализе

1. *Понятие метода средних величин*

Статистика изучает массовые явления и процессы, и каждое из них может быть охарактеризовано некоторыми индивидуальными значениями статистического показателя. Однако для массовых явлений характерно и другое свойство которое заключается в некоторой близости отдельных характеристик. Взаимодействие элементов совокупности приводит к ограничению вариации хотя бы части их свойств. Эта объективно присущая для всех явлений тенденция обуславливает широкое применение в практике статистических исследований ***метода средних величин***.

Метод средних величин – один из самых распространенных методов расчета статистических показателей. Средняя величина является обобщающей мерой варьирующего признака в статистической совокупности. Показатель в форме средней характеризует уровень признака в расчете на единицу совокупности. Значение признака каждого элемента объединяет в себе как совместные для всей совокупности типовые черты, так и свойственные только этому элементу индивидуальные особенности. Абстрагируясь от индивидуальных особенностей отдельных элементов, можно выявить то общее, типовое, что свойственно всей совокупности.

Именно в средней взаимно компенсируются индивидуальные отличия элементов и обобщаются типовые черты. Типовое значение средней непосредственно связано с однородностью совокупности. Средняя характеризует типовой уровень только при условии, что совокупность качественно однородная. В неоднородной совокупности она не отражает реалий.

Таким образом, можно дать следующее определение ***средней – это обобщающий показатель, выражающий типичный уровень количественного варьирующего признака качественно однородных единиц совокупности.***

Несмотря на то, что главное значение средних величин состоит в их обобщающей функции, в практике статистических исследований используют и другие ее свойства. ***Если средняя величина обобщает качественно однородные значения признака, то она является типической характеристикой признака в данной совокупности***. Однако неправильно сводить роль средних величин только к характеристике типичных значений признаков в однородных по данному признаку совокупностях. На практике значительно чаще современная статистика использует средние величины, обобщающие явно неоднородные явления, как, например, урожайность всех зерновых культур по территории всей страны. Еще более ясна нетипичность такого среднего показателя, как произведенный национальный доход в среднем на душу населения. Такие показатели называются ***системные средние***.

Системные средние могут характеризовать как пространственные или объектные системы, существующие одномоментно (государство, отрасль, регион, планета Земля и т.п.), так и динамические системы, протяженные во времени (год, десятилетие, сезон и т.п.). Типическая средняя может обобщать системные средние для однородной совокупности, или системная средняя может обобщать типические средние для единой, хотя и неоднородной, системы. При этом даже типическая средняя не является раз и навсегда данной, неизменной характеристикой. «Типичность» любой средней величины - понятие относительное, ограниченное как в пространстве, так и во времени.

Связь определяющего свойства с элементами совокупности описывается функцией $f(x_1,x_2,...x_n)$, которое выражает определенное математическое действие над эмпирическими значениями признака (суммирования, умножение, возведение в степень, извлечение корня) и определяет вид средней. Так, при суммировании значений признака определяющее свойство обеспечивает средняя арифметическая, при умножении - средняя геометрическая и т. д.

При исчислении средних в социально-экономических исследованиях необходимо четко осознать ***определяющее свойство совокупности*** и ***логическую формулу средней***. Числитель логической формулы средней представляет собой объем значений варьирующего признака (определяющее свойство), а знаменатель - объем совокупности. Как правило, определяющее свойство - это реальная абсолютная или относительная величина, которая имеет самостоятельное значение в анализе. В каждом конкретном случае для реализации логической формулы используется определенный вид средней, в частности: средняя арифметическая; средняя гармоническая; средняя геометрическая; средняя квадратичная и т. д.

Таким образом, при использовании метода средних величин для типической средней необходимо соблюдение двух условий:

- средние следует вычислять только для качественно-однородной совокупности ;
- реальность показателей, входящих в расчет средней.

В зависимости от характера исходной информации средняя любого вида может быть простой или взвешенной. Обозначается средняя символом $\bar{x}$ (черта над символом означает среднее из индивидуальных значений) и имеет ту же самую единицу измерения, что и признак.

2. Средняя арифметическая

Самым распространенным видом средней, применяемой в социально-экономическом анализе, является средняя арифметическая.

Средняя арифметическая простая:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}\text{,где}$$

$\bar{x}$ – средняя величина,

x – индивидуальные значения признака отдельных единиц совокупности,

n – численность совокупности.

Простая средняя арифметическая используется в расчете популярного фондового индекса Доу-Джонса, для определения среднего остатка оборотных средств по балансу, среднегодовой численности населения и др.

Средняя арифметическая взвешенная:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} fi}\text{, где}$$

f – частота.

Данная формула широко применяется в расчетах экономических показателей. Например, надо определить среднюю заработную плату работника АО, имея данные по отдельным филиалам АО (табл. 4.1.):

Таблица 4.1

Сведения о заработной плате

Филиалы АО	Заработная плата работника, в гривнах	Численность работников
1	260	20
2	220	10
3	200	15

$$\bar{x} = \frac{260 \cdot 20 + 220 \cdot 10 + 200 \cdot 15}{45} = 231{,}11 \text{ грн.}$$

Средняя арифметическая величина может быть дробным числом, если даже индивидуальные значения признака могут принимать только целые значения (дискретный признак). Ничего «предосудительного» для метода средних в этом не заключено; из сущности средней не вытекает, что она обязана быть реальным значением признака, которое могло бы встретиться у какой-либо единицы совокупности.

Если при группировке значения осредняемого признака заданы интервалами, то при расчете средней арифметической величины в качестве значения признака в группах принимают середины этих интервалов, т.е. исходят из гипотезы о равномерном распределении единиц совокупности по интервалу значений признака. Для открытых интервалов в первой и последней группе, если таковые есть, значения признака надо определить экспертным путем исходя из сущности, свойств признака и совокупности.

Свойства арифметической средней величины

Знание некоторых математических свойств средней арифметической полезно как при ее использовании, так и при ее расчете.

1. Сумма отклонений индивидуальных значений признака от его среднего значения равна нулю.

Доказательство:

$$\sum_{i=1}^{n}(x_i-\bar{x})=(x_1-\bar{x})+(x_2-\bar{x})+\ldots+(x_n-\bar{x})=$$

$$=x_1+x_2+\ldots+x_n-n\bar{x}=\sum_{i=1}^{n}x_i-n\frac{\sum_{i=1}^{n}x_i}{n}=0$$

Примечание. Для средней взвешенной сумма взвешенных отклонений равна нулю.

2. Если каждое индивидуальное значение признака умножить или разделить на постоянное число, то и средняя увеличится или уменьшится во столько же раз.

Доказательство:

$$\frac{\sum_{i=1}^{n}(x_i:c)}{n}=\frac{\frac{x_1}{c}+\frac{x_2}{c}+\ldots+\frac{x_n}{c}}{n}=\frac{\frac{x_1+x_2+\ldots+x_n}{c}}{n}=\frac{x_1+x_2+\ldots+x_n}{n}:c=\bar{x}:c.$$

Вследствие этого свойства индивидуальные значения признака можно сократить в с раз, произвести расчет средней и результат умножить на с.

3. Если к каждому индивидуальному значению признака прибавить или из каждого значения вычесть постоянное число, то средняя величина возрастет или уменьшится на это же число.

Доказательство:

$$\frac{\sum_{i=1}^{n}(x_i+c)}{n}=\frac{(x_1+c)+(x_2+c)+\ldots+(x_n+c)}{n}=\frac{\sum_{i=1}^{n}x_i+nc}{n}=\bar{x}+c.$$

Это свойство полезно использовать при расчете средней величины из многозначных и слабоварьирующих значений признака.

4. Если веса средней взвешенной умножить или разделить на постоянное число, средняя величина не изменится.

Доказательство:

$$\frac{\sum_{i=1}^{n}x_i\frac{f_i}{c}}{\sum_{i=1}^{n}\frac{f_i}{c}}=\frac{(\sum_{i=1}^{n}x_if_i):c}{(\sum_{i=1}^{n}f_i):c}=\bar{x}.$$

Используя это свойство, при расчетах следует сокращать веса на их общий сомножитель либо выражать многозначные числа весов в более крупных единицах измерения.

5. Сумма квадратов отклонений индивидуальных значений признака от средней арифметической меньше, чем от любого другого числа.

Доказательство. Составим сумму квадратов отклонений от переменной а:

$$f(a) = \sum_{i=1}^{n} (x_i - a)^2.$$

Чтобы найти экстремум этой функции, нужно ее производную по а приравнять нулю:

$$\frac{df}{da} = 2\sum_{i=1}^{n} (x_i - a)(-1) = 0.$$

Отсюда имеем:

$$\sum_{i=1}^{n} (a - x_i) = 0; \;\; a\sum_{i=1}^{n}(1) - \sum_{i=1}^{n}(x_i) = 0; \; na = \sum_{i=1}^{n} x_i; \;\; a = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

Таким образом, экстремум суммы квадратов отклонений достигается при $a = \bar{x}$. Так как логически ясно, что максимума функция не может иметь, этот экстремум является минимумом.

3. *Средняя гармоническая*

Средняя гармоническая простая вычисляется по формуле:

$$\bar{x} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Средняя гармоническая взвешенная вычисляется по следующей формуле:

$$\bar{x} = \frac{\sum_{i=1}^{n} M_i}{\sum_{i=1}^{n} \frac{M_i}{x_i}}, \text{ где}$$

М – вес или готовое произведение X*f=M

Средняя гармоническая взвешенная применяется тогда, когда отсутствуют действительные носители признака.

Данная формула применяется при построении средне гармонического индекса, когда нет возможности использовать агрегатную форму построения индексов, для определения изменения цены по минимальной продовольственной «корзины». Например, имеются данные, представленные в табл. 4.2.

Таблица 4.2

Оборот фирмы

Фирма	Продолжительность оборота, в днях	Средний остаток сырья, тыс. грн.

1	20	200
2	30	900

Исчислить среднюю продолжительность оборота средств.

$$\bar{x} = \frac{200 + 900}{\frac{200}{20} + \frac{900}{30}} = 27{,}5 \text{ дней}$$

Средняя гармоническая может быть также использована, например, при вычислении средней скорости автомобиля. Допустим автомобил половину пути двигался со скоростью 40 км/час, а вторую – 60 км/час. Тогда средняя скорость:

$$\bar{x} = \frac{2}{\frac{1}{40} + \frac{1}{60}} = 48 \text{км/час}$$

Пример 3.

Средняя выработка продукции на одного рабочего за период характеризуется следующими данными:

Таблица 4.3

Данные о производстве продукции в цехах

Бригада	Цех №1		Бригада	Цех №2	
	Выработка продукции за день, шт.	Число рабочих, чел.		Выработка продукции за день, шт.	Объем выработанной за день продукции, шт.
1	20	8	4	38	418
2	30	11	5	36	432
3	35	16	6	20	140

Определить среднедневную выработку рабочих по каждому цеху.

Решение

1. Среднюю выработку рабочих по первому цеху определим по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} fi} = \frac{20 * 8 + 30 * 11 + 35 * 16}{8 + 11 + 16} = 30$$

2. Среднюю выработку рабочих по второму цеху определим по формуле средней гармонической:

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}} = \frac{418 + 432 + 140}{\frac{418}{38} + \frac{432}{36} + \frac{140}{20}} = 33$$

4. Средняя геометрическая

Средняя геометрическая простая вычисляется по формуле:

$$\bar{x} = \sqrt[n]{x_1 x_2 .. x_n}$$

По этой формуле вычисляется биржевой индекс «Файнэшл Таймс», сложные проценты на рынке ценных бумаг, среднегодовой коэффициент роста.

Средняя геометрическая взвешенная вычисляется по формуле:

$$\bar{x} = \sqrt[\sum_{i=1}^{k} f_k]{\Pi(x^{f_k})}$$

где

$\Pi(x^f)$ – произведение

f – продолжительность отрезков времени.

Например: Первые два года фирма увеличила выручку от реализации на 10% или в 1,1 раза ежегодно, за следующие 3 года она увеличивала ежегодно на 20% или в 1,2раза. Чему равен среднегодовой темп роста выручки от реализации за весь период работы фирмы?

$$\bar{x} = \sqrt[5]{(1{,}1^2)(1{,}2^3)} \approx 1{,}16 \text{ или } 116\%$$

5. Средняя хронологическая

Средняя хронологическая вычисляется по формуле:

$$\bar{x} = \frac{\frac{x_1}{2} + x_2 + x_3 + \ldots + \frac{x_n}{2}}{n-1}, \text{ где}$$

n- число дат

$x_1, x_2 \ldots, x_n$ – уровни ряда

Если значение признака приводится на конкретную дату, то для расчета среднего уровня ряда применяется средняя хронологическая. Например, имеются остатки незавершенного производства по фирме: на 1.01-200 тыс. грн., 1.02-180 тыс. грн., 1.03-160 тыс. грн., 1.04-190 тыс. грн. Определить средний остаток незавершенного производства за первый квартал.

$$\bar{x} = \frac{\frac{200}{2} + 180 + 160 + \frac{190}{2}}{4-1} = 178{,}3 \text{ тыс. грн.}$$

6. Скользящая средняя

Суть расчета средней скользящей состоит в том, что по эмпирическим данным вычисляют скользящие средние, которые получают из сумм показателей, последовательно сдвигающихся на одну дату по ряду. Затем подвижные суммы следует разделить на число дат, принятых за период сглаживания. Таким образом, получаем скользящую среднюю.

Произведем сглаживание данных о курсе акций (цифры условные) при помощи трехдневной средней. Найдем среднее значение для первых трех уровней:

(80+72+99)/3=84 и отнесем его ко второму дню. Затем отбросим первый член ряда, прибавим четвертый день и рассчитаем для третьего дня: (77+99+120)/3=97 и т.д. Получим следующие средние скользящие (табл. 4.4):

Таблица 4.4

Определение средней скользящей

День	Цена акции, грн.	Скользящая средняя
	80	-
	72	84
	99	97
	120	103
	90	92
	66	76
	72	-

7. *Структурные средние (мода и медиана)*

К структурным средним относятся мода и медиана. Мода (M_0) – наиболее часто встречающееся значения признака в совокупности. Например: 2, 4, 3, 3, 3, 3, 1, 5. Мода – 3. Обычно встречаются ряды с одним модальным значением признака. В противном случае ряд называется бимодальным или мультимодальным.

В интервальном вариационном ряду модальным интервалом считается интервал с наибольшей частотой. В этом случае для вычисления моды используется следующая формула:

$$Mo = x_0 + \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} i$$

где: x_0 - нижняя граница модального интервала;

f_{Mo} - частота в модальном интервале;

f_{Mo-1} - частота в интервале, предшествующем модальному;

f_{Mo+1} - частота в интервале, следующим за модальным;

i – величина интервала.

Медиана (M_e) – середина ранжированного статистического ряда. Медиана делит ряд на две равные части. Вначале определяют порядковый номер медианы:

N_{me}=(n+1)/2,

где n – объем ряда (число единиц в ряду).

Если ряд состоит из четного числа членов, то медиана определяется как полусумма двух срединых вариант. Так, дан ряд 10, 20, 30, 40, 50, ..., 80

N_{me}=(8+1)/2=4,5, M_e=(40+50)/2=45.

В интервальном вариационном ряду для нахождения медианы применяется формула:

$$M_e = x_0 + \frac{\frac{\sum_{j=1}^{k} f_j}{2} - f'_{Me-1}}{f_{Me}} i$$

где: Me – медиана

x_0 - нижняя граница интервала, в котором находится медиана;

f'_{Me-1} - накопленная частота в интервале, предшествующем медианному;

f_{Me} - частота в медианном интервале;

i – величина интервала;

k – число групп.

В дискретном вариационном ряду медианой считается значение признака в той группе, в которой накопленная частота превышает половину числа единиц совокупности.

Аналогично медиане вычисляются значения признака, делящие совокупность на 4 равные части (по числу единиц). Эти величины называют квартилями и обозначают буквой Q со значком номера квартиля. Q_2 совпадает с медианой. Первый и третий квартили рассчитываются по следующим формулам:

$$Q_1 = x_0 + \frac{\frac{\sum_{j=1}^{k} f_i}{4} - f'_{Q1-1}}{f_{Q1}} i$$

$$Q_3 = x_0 + \frac{\frac{3\sum_{j=1}^{k} f_i}{4} - f'_{Q3-1}}{f_{Q3}} i$$

Значения признака, делящие ряд на 5 равных частей, называются квинтилями, на десять – децилями, на 100 – перцентилями.

8. Степенные средние

Средние арифметическая, гармоническая, геометрическая, квадратическая могут быть представлены в виде некоторой системе величин, вычисленных из степенной средней

$$\overline{x} = \left(\frac{\sum_{i=1}^{n} x_i^k}{n} \right)^{\frac{1}{k}}$$, где

К – показатель степени.

Изменяя показатель степени К, получают согласно формуле степенной средней разные виды средних:

$$k=1 \rightarrow \bar{x}=\frac{\sum_{i=1}^{n} x_i}{n}$$ – среднюю арифметическую

$$k=-1 \rightarrow \bar{x}=\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$ – среднюю гармоническую

$$k=2 \rightarrow \bar{x}=\sqrt{\frac{\sum_{i=1}^{n} x^2}{n}}$$ – среднюю квадратическую.

Чем больше показатель степени К, тем больше значение средней:

$$X_{\text{гар}} < X_{\text{геом}} < X_{\text{ар}} < X_{\text{кв}}$$

Указанное соотношение называется правилом мажорантности.

9. Примеры решения задач

Пример 1.

Приведем пример расчета показателей, характеризующих структуру. Допустим имеется некоторое условное распределение хозяйств области по урожайности зерновых культур.

Таблица 4.5

Распределение хозяйств области по урожайности зерновых культур

Группы хозяйств по урожайности, ц/га x_j	Число хозяйств f_j	Середина интервала, ц/га x_j'	$x_j' f_j$	Накопленная частота f_j'
10-15	6	12,5	75,0	6
15-20	9	17,5	157,5	15
20-25	20	22,5	450,0	35
25-30	41	27,5	1127,5	76
30-35	26	32,5	845,0	102
35-40	21	37,5	787,5	123
40-45	14	42,5	595,0	137
45-50	5	47,5	237,5	142
50-55	1	52,5	52,5	143
Итого	143		4327,5	

$$\bar{x}=\frac{\sum x_i' f_i}{\sum f_i}=\frac{4327{,}5}{143}\approx 30{,}3 \text{ ц/га}$$

$$M_0=25+\frac{(41-20)}{(41-20)+(41-26)}\cdot 5=27{,}9 \text{ ц/га}$$

$$M_e = 25 + \frac{71{,}5 - 35}{41} \cdot 5 = 29{,}5 \text{ ц/га}$$

$$Q_1 = 25 + \frac{(35{,}75 - 35)}{41} \cdot 5 = 25{,}09 \text{ ц/га}$$

$$Q_3 = 35 + \frac{(107{,}25 - 102)}{21} \cdot 5 = 36{,}25 \text{ ц/га}$$

Пример 2.

На початку 90-х років середня тривалість шлюбу в регіоні становила 8 років. Останнім часом кількість розлучених пар зросла, особливо інтенсивно серед молодих сімей (тривалість шлюбу менше 5 років). Як змінилась середня тривалість шлюбу?

Решение.

Обозначим:

а – продолжительность брака молодых семей;

в – продолжительность брака остальных семей;

m – удельный вес молодых семей;

(1-m) – остальных семей;

Δm – изменение удельного веса молодых семей

Тогда средняя продолжительность брака:

$a * m + b * (1-m) = 8$

$a * (m-\Delta m) + b * (1 - m + \Delta m) = C;$

$C-8 = a * (m-\Delta m) + b * (1 - m + \Delta m) - (a * m + b * (1-m)) =$

$= \cancel{a * m} - a * \Delta m \cancel{+ b} - \cancel{b * m} + b * \Delta m - \cancel{a * m} \cancel{- b} + \cancel{b * m} =$

$= - a * \Delta m + b * \Delta m$

$C = 8 + \Delta m * (b - a)$.

Пример 3.

У 1997 р. середній вік засуджених становив 29 років. За 1998 - 2001 рр. загальне число засуджених зросло на 5%, а питома вага молоді - на 10%. Як змінився середній вік засуджених?

Решение.

а – средний возраст молодежи;

в – средний возраст остальных;

m – удельный вес молодежи;

(1-m) – удельный вес остальных;

Δm – изменение удельного веса молодежи.

Тогда средний возраст осужденных:

$a * m + b * (1-m) = 29$

$a * (m+0,1) + b * (1 – m – 0,1) = C;$

$C-29= a * (m + 0,1) + b * (1 – m – 0,1) – (a * m + b * (1-m))=$

$= \cancel{a * m} + a * 0,1 \cancel{+ b} – \cancel{b * m} - b * 0,1 - \cancel{a * m} \cancel{- b} + \cancel{b * m} =$

$= a * 0,1 - b * 0,1$

$C=29+ 0,1 * (a-b).$

Пример 4.

Комерційна плата за навчання у державних навчальних закладах нижча, ніж у приватних. У поточному році розмір плати в усіх навчальних закладах не змінився, проте збільшилась частка тих, хто навчається у приватних вузах. Як змінився середній розмір плати за навчання?

Решение.
а – оплата обучения в государственных институтах;
в – оплата обучения в коммерческих институтах;
m – удельный вес обучающихся в государственных институтах;
(1-m) – удельный вес остальных;
Δm – изменение удельного обучающихся в частных институтах.

Тогда средняя стоимость обучения:
$a * m + b * (1-m);$

$a * (m-\Delta m) + b * (1 – m + \Delta m);$

$a * (m-\Delta m) + b * (1 – m + \Delta m) – (a * m + b * (1-m))=$

$= \cancel{a * m} - a * \Delta m \cancel{+ b} – \cancel{b * m} + b * \Delta m - \cancel{a * m} \cancel{- b} + \cancel{b * m} =$

$= - a * \Delta m + b * \Delta m = \Delta m * (b - a).$

Контрольные вопросы к теме 4

1. Что представляет собой средняя величина? Каковы условия ее применения?
2. Какие виды средних величин чаще всего применяются в экономических расчетах?

3. Как исчисляется средняя арифметическая простая и взвешенная? В каких случаях она применяется?
4. В каких случаях применяется средняя гармоническая и как она вычисляется?
5. Для каких целей применяется средняя геометрическая?
6. Дайте характеристику средней хронологической.
7. Объясните расчет средней скользящей.
8. Что представляет собой степенная средняя, как она связана с другими средними?
9. В чем заключается правило мажорантности?
10. Дайте понятие моды и медианы.
11. Почему среднюю рассматривают как типовой уровень признака в совокупности? Какие виды средних используют в статистическом анализе?
12. Назовите основные свойства средних величин.
13. Изменится ли средняя, если частоты заменить долями?
14. Производство и реализация фарфорофаянсовой посуды одним из предприятий характеризуется следующими данными:

Сервизы	Изготовлено комплектов	Часть продукции, реализованной на экспорт, %
Чайный	220	50
Кофейный	50	60
Столовый	130	40

Определите средний процент реализованной на экспорт продукции и обоснуйте выбор средней.

15. Продукция, которую изготавливает малое предприятие, имеет разный уровень рентабельности реализации. В апреле месяце этот показатель составлял:

Продукт	Рентабельность реализации	Прибыль, млн. ден. ед.
A	12	600
B	7	140
C	21	630

Определите средний уровень рентабельности реализации и обоснуйте выбор средней.

16. За 4 квартал фирмой реализовано продукции на 21.6 млрд. грн. Остатки оборотных средств составляли, млн. грн.:
    На 1 января - 1400;
    На 1 февраля - 1550;
    На 1 марта - 1270;
    На 1 апреля - 1600.
    Определите: 1) среднеквартальный остаток оборотных денежных средств; 2) количество оборотов оборотных денежных средств и продолжительность одного оборота. Объясните, какие использованы средние и почему.
17. Укажите сферу использования обобщающих, интегральных оценок.

18. Что характеризует многомерная средняя? Какую аналитическую функцию она выполняет в статистическом анализе?

Тема 5. ***Показатели вариации***

1. ***Размах вариации***

Вариацией признака называется его изменение у единиц совокупности. Вариация порождается комплексом разнообразных условий, фактов, воздействующих на элементы совокупности. Изучение вариации имеет большое практическое и научное значение. Оно осуществляется при помощи показателей вариации.

Размах вариации – разность между максимальным и минимальным значением признака изучаемой совокупности.

$W=X_{max}-X_{min}$

где: W – размах вариации,

X_{max} – наибольшее значение признака,

X_{min} – наименьшее значение признака.

Величина размаха вариации вычисляется по крайним точкам ряда, которые в наибольшей степени подвержены случайным изменениям. Поэтому принято считать более надежным показателем инквартильный размах. Он вычисляется как разность между третьим и первым квартилем.

$$R=Q_3-Q_1$$

где: Q_1- первый квартиль,

Q_3 – третий квартиль.

В качестве показателя силы вариации используют среднее линейное отклонение:

$$\alpha = \frac{\sum_{j=1}^{k} \left| x_j^{'} - \bar{x} \right| f_j}{\sum_{j=1}^{k} f_j}.$$

2. ***Дисперсия и среднее квадратическое отклонение***

Показатель среднего квадратического отклонения рассчитывается по следующей формуле:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}.$$ Для дискретного ряда

$$\sigma = \sqrt{\frac{\sum_{j=1}^{k}(x_j^{'} - \bar{x})^2 f_j}{\sum_{j=1}^{k} f_j}}.$$ Для интервального ряда

Квадрат среднего квадратичного отклонения называется дисперсией.

Дисперсия – средний квадрат отклонений индивидуальных значений признака от их средней величины.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n},$$ где

σ^2 (сигма)	– дисперсия;
x_i	– значение признака отдельных единиц совокупности;
$\bar{x}$	– среднее значение всех единиц совокупности;
n	– численность совокупности.

Для вариационного ряда дисперсия вычисляется по следующей формуле:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{n} f_i}$$

Дисперсия признака равна разности между средним квадратом значений признака и квадратом их средней, т.е.

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 = \overline{x^2} - (\bar{x})^2,$$

а для вариационного ряда

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}\right)^2 = \overline{x^2} - (\bar{x})^2$$

Доказательство:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + (\bar{x})^2) = \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i\bar{x} + \sum_{i=1}^{n}\bar{x}^2$$

заменим $\sum_{i=1}^{n} x_i = n\bar{x}$ и разделим все слагаемые на n:

$$\sigma^2 = \frac{\sum_{j=1}^{m} x_i^2}{n} - 2(\bar{x})^2 + (\bar{x})^2 = \overline{x^2} - (\bar{x})^2$$

Размах вариации, среднее квадратическое отклонение являются именованными величинами. Чем меньше эти показатели, тем совокупность более однородна, тем более типичной будет средняя величина.

Дисперсия альтернативного признака вычисляется по формуле:

$$\sigma^2 = pq,$$

где: p – доля единиц обладающих данным признаком,
q – доля единиц, не обладающих данным признаком.

$$q = 1 - p.$$

Среднее квадратическое отклонение альтернативного признака:

$$\sigma = \sqrt{pq}$$

3. *Коэффициенты вариации и дифференциации*

Рассмотренные ранее показатели вариации, кроме дисперсии, числа именованные и поэтому их неудобно использовать для сопоставления вариации различных признаков. Для оценки интенсивности вариации и сравнения ее в разных совокупностях, а тем более для разных признаков, используют относительные показатели. Одним из таких показателей является ***коэффициент вариации***:

$$\upsilon = \frac{\sigma \cdot 100\%}{\bar{x}}.$$

Совокупность считается однородной, если коэффициент вариации не превышает критического значения 33%.

Используются также показатели относительного размаха вариации $\rho = R : \bar{x}$ и относительного отклонения по модулю - $m = \alpha : \bar{x}$.

Коэффициент дифференциации:

$$K_V = \frac{Q_3 - Q_1}{Q_3 + Q_1},$$

где:

Q_1 – первый квартиль;
Q_3 – третий квартиль.

$$\upsilon \approx 1{,}5 K_v$$

Пример. Допустим имеются данные по обследованию обеспеченности жильем некоторых домохозяйств города:

Жилая площадь на одного члена домохозяйства, м^2	Количество домохозяйств, f_j	x'_j	$x'_j f_j$	Накопленная частота f'_j
До 5	17	4	68	17
5-7	39	6	234	56
7-9	51	8	408	107
9-11	42	10	420	149
11-13	29	12	348	178
13-15	15	14	210	193
15 и больше	7	16	112	200
Всего	200		1800	

Расчет обобщающих характеристик вариации

x'_j	f_j	$x^1_j - \bar{x}$ ($\bar{x} = 9$)	$\lvert x'_j - \bar{x}\rvert f_j$	$(x'_j - \bar{x})^2$	$(x'_j - \bar{x})^2 f_j$
4	17	-5	85	25	425
6	39	-3	117	9	351
8	51	-1	51	1	51
10	42	1	42	1	42
12	29	3	87	9	261
14	15	5	75	25	375
16	7	7	49	49	343
Всего	200		506		1848

Среднюю рассчитаем по формуле среднеарифметической взвешенной:

$$\bar{x} = \frac{\sum_1^7 x'_j f_j}{\sum_1^7 f_j} = \frac{4*17 + 6*39 + 8*51 + 10*42 + 12*29 + 14*15 + 16*7}{200} = 9.$$

Среднее линейное отклонение:

$$\alpha = \frac{\sum_{j=1}^{k} \lvert x'_j - \bar{x}\rvert f_j}{\sum_{j=1}^{k} f_j} = \frac{506}{200} = 2{,}53 \text{ м}^2$$

Дисперсия:

$$\sigma^2 = \frac{\sum_1^7 (x'_j - \bar{x})^2 f_j}{\sum_1^7 f_j} = \frac{1848}{200} = 9{,}24 \text{ м}^2$$

Среднее квадратичное отклонение:

$$\sigma = \sqrt{\sigma^2} = \sqrt{9{,}24} = 3{,}04 \text{ м}^2$$

Коэффициент вариации:

$$\upsilon = \frac{\sigma \cdot 100\%}{\bar{x}} = \frac{3{,}04}{9} = 0{,}338$$

4. *Моменты распределения и показатели его формы*

Для дальнейшего изучения характера вариации используются средние значения различных степеней отклонений различных средних величин признака от его средней арифметической. Эти показатели получили название центральных моментов распределения порядка, соответствующего степени, в которую возводятся отклонения. По несгруппированным данным центральные моменты определяются следующим образом:

$$(\mu_1) \to \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n} = 0$$

$$(\mu_2) \to \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \overline{x^2} - (\bar{x})^2 = \sigma^2$$

$$(\mu_3) \to \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n}$$

$$(\mu_4) \to \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$$

По сгруппированным данным:

$$(\mu_1) \to \frac{\sum_{j=1}^{k}(x_j' - \bar{x})f_j}{\sum_{j=1}^{k} f_j} = 0$$

$$(\mu_2) \to \frac{\sum_{j=1}^{k}(x_j' - \bar{x})^2 f_j}{\sum_{j=1}^{k} f_j} = \overline{x^2} - (\bar{x})^2 = \sigma^2$$

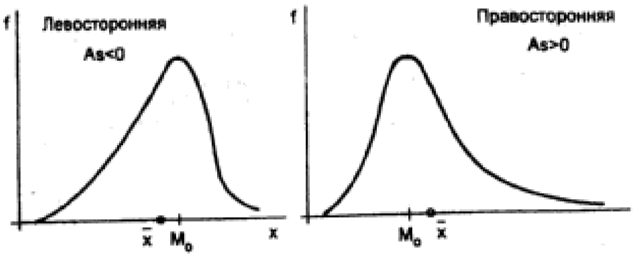
$$(\mu_3) \to \frac{\sum_{j=1}^{k}(x_j' - \bar{x})^3 f_j}{\sum_{j=1}^{k} f_j}$$

$$(\mu_4) \to \frac{\sum_{j=1}^{k}(x_j^{'}-\bar{x})^4 f_j}{\sum_{j=1}^{k} f_j}$$

На основе момента третьего порядка строится показатель, характеризующий степень ассиметричности распределения – коэффициент ассиметрии:

$$A_s = \frac{\mu_3}{\sigma^3}.$$

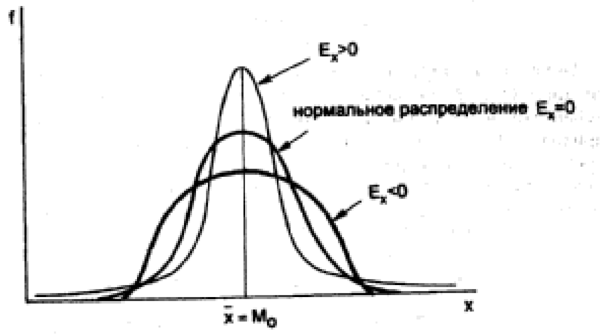
Распределения с сильной правосторонней и левосторонней ассиметрией показаны на рисунке:



На основе момента четвертого порядка строится показатель эксцесса:

$$E_x = \frac{\mu_4}{\sigma^4}$$

Сравнение рядов с одинаковой силой вариации и одинаковой ассиметрией, но разными показателями эксцесса изображено на рисунке:



5. Виды и взаимосвязь дисперсий

Дисперсия занимает особое место в статистическом анализе социально-экономических явлений. На ней основан один из методов статистического анализа – дисперсионный анализ. В отличии от других характеристик вариации благодаря своим математическим особенностям она является важным элементов и других статистических методов.

Свойства дисперсии:

- если все варианты увеличить или уменьшить на одно и то же число, то дисперсия не изменится;
- если все варианты увеличить или уменьшить в А раз, то дисперсия увеличится или уменьшиться в A^2 раз.
- Если частоты заменить долями, то дисперсия не изменится.

Если совокупность разбита на группы по какому-либо признаку, то для любого признака можно вычислить дисперсию как в целом по совокупности, так и в каждой группе. При этом различают групповую и общую среднюю, а отклонения индивидуальных значений от общей средней состоит из двух частей:

$$(x_i - \bar{x}) = (x_i - \bar{x}_j) + (\bar{x}_j - \bar{x});$$

где: x_i - индивидуальное значение признака;

$\bar{x}_j$ - средняя по некоторой группе в совокупности (групповая средняя);

$\bar{x}$ - средняя по совокупности.

Обобщающими характеристиками этих отклонений являются общая, групповая и межгрупповая дисперсия.

Общая дисперсия характеризуется вариацией признака относительно общей средней:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}.$$

Групповая дисперсия характеризуется вариацией относительно групповой средней:

$$\sigma_j^{\ 2} = \frac{\sum_{j=1}^{m_j}(x_j - \bar{x}_j)^2}{m_j}$$

Средняя из групповых дисперсий:

$$\overline{\sigma^2} = \frac{\sum_{j=1}^{m}\sigma_j^2 f_j}{\sum_{j=1}^{m} f_j}$$

Межгрупповая дисперсия:

$$\delta^2 = \frac{\sum_{j=1}^{m}(\bar{x}_j - \bar{x})^2 f_j}{\sum_{j=1}^{m} f_j}$$

Существует правило сложения дисперсий:

Общая дисперсия равна сумме средней из групповых дисперсий и межгрупповой дисперсии:

$$\sigma^2 = \overline{\sigma^2} + \delta^2$$

Отношение межгрупповой дисперсии к общей дисперсии называется корреляционным отношением:

$$\eta^2 = \frac{\delta^2}{\sigma^2}$$

Контрольные вопросы к теме 5

1. По какой формуле можно вычислить размах колеблемости?
2. Какой показатель называется дисперсией и по каким формулам его можно рассчитать?
3. По каким формулам вычисляется среднее квадратическое отклонение для количественного и альтернативного признаков?
4. Как называется относительный показатель колеблемости и по каким формулам его можно вычислить?
5. Как вычислить коэффициент дифференциации?
6. Дайте понятие квартили и как вычислить инквартильный размах?
7. Как взаимосвязаны показатели вариации и для чего эти взаимосвязи используются?
8. Что представляет собой правило сложения дисперсий и в чем его практическое значение?

Тема 6. Выборочный метод

1. *Выборочный метод в статистике*

Выборочное наблюдение – это такой вид несплошного наблюдения, при котором исследуется не вся совокупность, а только ее определенная часть. При этом исследуемая часть совокупности называется ***выборочной***, а вся совокупность называется ***генеральной***. Статистические характеристики выборочной совокупности рассматриваются как оценка с определенной степенью достоверности характеристик генеральной совокупности.

Имеется несколько ***причин*** необходимости и целесообразности использования выборочного метода в статистике:

- повышение точности данных, поскольку уменьшение числа единиц наблюдения уменьшает ошибки регистрации;
- при проведении выборочных исследований достигается относительная экономия материальных, трудовых и финансовых ресурсов, а также существенным образом сокращаются затраты времени на сбор данных;
- выборочное наблюдение незаменимо в том случае, когда сам процесс наблюдения связан с порчей объекта (изучение качества продукции).

Объективной гарантией того, что проведение выборочного исследования позволит получить достоверные выводы по всей генеральной совокупности, является соблюдение следующих принципов и ***правил проведения выборочного обследования***:

- обязательное определение границ генеральной совокупности;
- разработка программы наблюдений и инструкций;
- определение списка единиц выборочной совокупности;
- установление допустимого размера погрешности и определение объема выборки;
- определение вида выборочного наблюдения;
- установление сроков проведения наблюдения;
- определение потребности в кадрах для проведения выборочного наблюдения, их подготовка;
- оценка точности и достоверности данных выборки, определение порядка их распространения на генеральную совокупность.

Статистические данные могут быть представлены как выборочные даже в том случае, если они таковыми непосредственно не являются. Например:

если данные сплошного наблюдения рассматриваются как один из возможных вариантов проявления некоторого процесса;

если совокупность содержит малое число единиц;

если используются результаты эксперимента, рассматриваемого как некоторая выборка из бесконечно большого числа экспериментов.

Поскольку выборочная совокупность не абсолютно точно отражает состав генеральной совокупности, то и статистические оценки выборочной совокупности не абсолютно точно совпадают с оценками генеральной совокупности. Отличие состоит в ошибке наблюдения и в ошибке выборки:

выборочная оценка = генеральный параметр ± ошибка наблюдения ± ошибка выборки.

2. Виды выборки. Способы отбора.

Для того, чтобы по выборочной совокупности делать какие-то выводы о генеральной совокупности, выборочная совокупность должна быть ***репрезентативной***, т.е. достаточно точно отражать свойства генеральной совокупности. Репрезентативность выборки может быть обеспечена при соблюдении объективности отбора данных. Возможны три ***способа отбора***:

- случайный;
- по определенной схеме;
- сочетание первого и второго способа.

Виды выборки:

- типическая (расслоенная, стратифицированная, районированная) – выборка производится из генеральной совокупности, предварительно разделенной на типы (слои, страты);
- серийная (гнездовая) – единицей отбора является не единица наблюдения, а серия (гнездо) единиц совокупности;
- повторная или бесповторная;
- многоступенчатая, которая используется при изучении социально-экономических объектов сложной структуры, когда на каждой ступени используются разные единицы отбора – от более крупных до единицы наблюдения на последней ступени;
- многофазовая, включающая определенное количество, каждая из которых имеет программу наблюдения со своей степенью подробности.

От случайной выборки следует отличать квотный отбор, когда выборка конструируется из единиц определенных категорий (квот). Выборка может быть нерепрезентативной, даже если она формируется в соответствии с известными пропорциями генеральной совокупности, но отбор производится без определенной схемы.

3. Ошибки выборки

Ошибка выборки представляет собой разность между значением показателя, полученного по выборке, и генеральным параметром. Принято среднюю, относительную величину и дисперсию в генеральной совокупности обозначать соответственно как μ, π, σ^2, а в выборочной совокупности - $\bar{x}$, ρ, s^2.

Тогда ошибка репрезентативности выборочной средней равна:

$$\mu_{\bar{x}} = \bar{x} - \mu,$$

выборочной относительной величины:

$$\mu_p = p - \pi,$$

дисперсии:

$$\mu_{\sigma^2} = s^2 - \sigma^2.$$

Если представить, что было произведено бесконечно большое число выборок равного объема из одной и той же генеральной совокупности, то показатели этих выборок образовали бы ряд возможных значений: выборочных средних величин, дисперсии, относительных величин. Для каждой отдельной выборки характерна своя ошибка репрезентативности. Следовательно, можно было бы построить ряды распределения выборок по величине ошибки репрезентативности для каждого показателя. В таких распределениях улавливается тенденция в концентрации ошибок около центрального значения. Число выборок с той или иной величиной ошибки репрезентативности может быть симметрично или ассиметрично относительно этого центрального значения. При бесконечно большом числе выборок получается кривая частот, которая называется ***кривой выборочного распределения***. Свойства таких распределений используются для получения статистических заключений, установления вероятности той или иной величины ошибки репрезентативности.

Рассмотрим выборочное распределение средней величины. С увеличением числа выборок средняя всех выборок будет приближаться к генеральной средней, т.е. $\bar{x} \to \mu$, а $S_{\bar{x}} \to 0$.

Средняя ошибка выборочной средней при повторном отборе определяется по формуле:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n-1}} = \frac{\sigma}{\sqrt{n-1}};$$

При бесповторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n-1}\left(\frac{N-n}{N-1}\right)}$$

Из формулы следует, что средняя ошибка выборки тем больше, чем больше вариация в генеральной совокупности, и тем меньше, чем больше объем выборки.

Если значения n достаточно велики (n>100), то приведенные формулы средней ошибки выборочной средней можно использовать в следующем виде:

При повторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}};$$

При бесповторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$$

Ошибка конкретной выборки может принимать различные значения, но отношение ее к средней ошибке практически не превышает ±3, если величина n достаточно большая (>100).

Средняя ошибка выборочной средней показывает возможные отклонения характеристик выборочной совокупности от характеристик генеральной совокупности. В практическом смысле имеет значение нахождение предельной ошибки выборки. Зная среднюю ошибку выборки, можно найти границы, за которые не выйдет величина конкретной ошибки выборки. Однако это можно утверждать только с определенной степенью вероятности.

Доверительный уровень вероятности – это такой уровень вероятности, который устанавливают при определении границ, в которых находятся параметры генеральной совокупности.

Чаще всего принимают доверительную вероятность равной 0,95; 0,954; 0,997 или 0,999. Доверительный уровень вероятности 0,95 означает, что только в 5 случаях из 100 ошибка может выйти за установленные границы; вероятности 0,954 - в 46 случаях из 1000, при 0,997 - в 3 случаях, а при 0,999 - в 1 случае из 1000.

Уровень значимости (существенности) – это относительное количество ошибочных выводов в общем количестве выводов. Уровень значимости дополняет надежный уровень вероятности до 1.

Предельная ошибка выборки – это такая ошибка выборки, которая исчислена с заданным уровнем надежной вероятности.

Отношение ошибки конкретной выборки к средней ошибке выборочной средней ***называется нормированным отклонением*** и обозначается как t:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}.$$

В статистической литературе нормированное отклонение еще называется коэффициентом надежности или коэффициентом кратности.

Предельная ошибка выборки определяется по формуле:

$$\Delta = tS_{\bar{x}}.$$

С учетом приведенных формул:

При повторном отборе:

$$\Delta_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}};$$

При бесповторном отборе:

$$\Delta_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$$

В этом случае искомый надежный интервал для оценки генеральной средней по выборочной средней определяется по формуле:

$$\mu = \bar{x} \pm \Delta_{\bar{x}}$$

Вероятность отклонения выборочных средних от генеральной средней при бесконечно большом числе наблюдений ($n \to \infty$) определяется законом нормального распределения Лапласа-Гаусса:

$$\Phi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

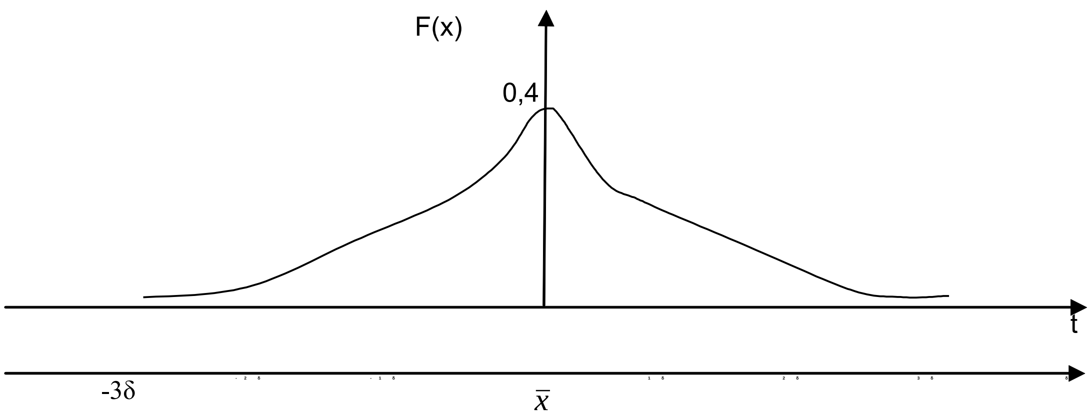
где: $f(t)$ - плотность вероятности;

σ - среднее квадратичное отклонение значений переменной x от средней в генеральной совокупности.

Так как средняя нормированных отклонений t=0, дисперсия $\sigma^2 = 1$, т.е. $\sigma = 1$, то выражение уравнения Лапласа-Гаусса может быть записано в следующем виде:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Это уравнение называется стандартным уравнением нормальной кривой. График кривой нормального распределения вероятностей имеет вид:



Площадь, ограниченная кривой и всей осью абсцисс равна 1.

Ординаты на графике соответствуют вероятности при том или ином значении t. Чтобы определить вероятность значений в некотором интервале от t_1 до t_2, следует найти отношение между частью площади кривой, заключенной между t_1 и t_2, ко всей площади кривой.

Уравнение Лапласа - Гаусса предполагает непрерывное изменение t и неограниченное возрастание n. Поэтому площадь нормальной кривой, заключенную между ординатами t_1 и t_2 определяют, интегрируя функцию ($f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$).

Имеются таблицы, которые содержат значения вероятностей для нормированных отклонений t или для интервалов от t_1 до t_2. Эти таблицы содержит пропорциональные доли площадей, заключенных между ординатами, соответствующими $\pm\, t$. Зная нормированное отклонение t, можно определить вероятность или на основе определенной вероятности установить величину t.

На пересечении строк и граф таблицы находится значение вероятности $F(t)$, соответствующее данному значению t. Для краткости записи в таблице приводятся только десятичные знаки вероятности, следовательно, к табличному значению $F(t)$ надо приписывать ноль целых.

При определении нормированного отклонения t по таблице «Значение интеграла вероятностей» необходимо принять определенный уровень вероятности суждения о точности данной выборки, для чего задаются уровнем доверительной вероятности.

Наиболее часто используют следующие значения нормированного отклонения и соответствующие им вероятности:

t	1,00	1,96	2,00	2,58	3,00
Ф(t)	0,6827	0,9500	0,9545	0,9901	0,9973

Относительная ошибка выборки определяется по формуле:

$$\nu_\mu = \frac{S_{\bar{x}}}{\bar{x}} * 100\%$$

В случае ***типического*** отбора всю генеральную совокупность разделяют на группы по изучаемому признаку, а затем из каждой группы случайным образом отбирают необходимое количество единиц. При этом необходимо обеспечить принцип пропорционального представительства каждой группы. Для этого используют показатель численности, или дисперсий, или средних квадратичных отклонений. Это дает возможность устранить межгрупповую дисперсию. Поэтому ошибка выборки определяется только средней из групповых дисперсий, которая меньше от общей дисперсии на величину межгрупповой дисперсии. Ее называют остаточной дисперсией.

Тогда средняя ошибка при типическом отборе определяется по следующим формулам:

При повторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2_{ост}}{n}};$$

При бесповторном отборе:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2_{ост}}{n}\left(1 - \frac{n}{N}\right)}$$

Суть ***серийного отбора*** заключается в том, что отбираются не отдельные единицы совокупности, а серии таких единиц. При этом общее число серий рассматривают как численность генеральной совокупности, а количество отобранных серий как объем выборочной совокупности.

При серийном отборе устраняется влияние внутригрупповой вариации, поэтому при расчете ошибки выборки используют только вариацию между сериями.

4. Задачи, решаемые при применении выборочного метода

При применении выборочного метода возникает три основные задачи:

- определение возможного предела ошибок репрезентативности, гарантированного с заданной вероятностью;
- определение объемов выборки для получения требуемой точности результатов с заданной вероятностью;
- определение вероятности того, что ошибка выборки не превысит допустимой погрешности.

4.1. определение возможного предела ошибок репрезентативности, гарантированного с заданной вероятностью.

Чтобы вычислить ошибку выборки при принятой доверительной вероятности, нужно рассчитать величину средней ошибки $s_{\bar{x}}$. Формула для ее определения включает дисперсию признака в генеральной совокупности σ^2, которая, как правило, неизвестна. Может быть определена только выборочная дисперсия s^2. Доказано, что соотношение между σ^2 и s^2 определяется следующим равенством:

$$\sigma^2 = s^2 \frac{n}{n-1}$$

Если n велико, то сомножитель $n/(n-1) \approx 1$ и можно принять выборочную дисперсию в качестве оценки величины генеральной дисперсии. Подставив предыдущее выражение в формулу средней ошибки выборочной средней, получим:

$$s_{\bar{x}} = \sqrt{\frac{s^2}{n-1}} = \frac{s}{\sqrt{n-1}} \quad \text{или} \quad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Соответственно $\Delta_{\bar{x}} = ts_{\bar{x}} = t\dfrac{s}{\sqrt{n}}$

<u>Пример</u>. Для определения скорости расчетов с кредиторами предприятий одного треста была проведена случайная выборка 50 платежных документов, по которым средний срок перечисления денег оказался равен 28,2 дня со стандартным отклонением 5,4 дня. Определим средний срок прохождения всех платежей в течение данного года с доверительной вероятностью F(t) = 0,95. Тогда по таблице значения интеграла вероятности t = 1,96; скорректированная дисперсия:

$$\sigma^2 = s^2\frac{n}{n-1} = 5{,}4^2\frac{50}{50-1} = 29{,}755\,.$$

Средняя ошибка выборки:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{29{,}755}{50}} = \sqrt{0{,}5971} = \pm 0{,}77$$

Отклонение выборочной средней от генеральной с вероятностью 0,95 составит $\Delta_x = ts_{\bar{x}} = 1{,}96 * 0{,}77 = 1{,}51$ дня.

Δ называется доверительной ошибкой выборки или предельной ошибкой выборки. Рассчитав величину $\Delta_{\bar{x}}$, можно записать следующее неравенство:

$$28{,}2\text{-}1{,}5 \le \mu \le 28{,}2\text{+}1{,}5$$

$$26{,}7 \text{ дня} \le \mu \le 29{,}7 \text{ дня}$$

Таким образом, с вероятностью 0,95 можно утверждать, что средняя продолжительность расчетов предприятия данного треста с кредиторами составляет не менее 26,7 дня и не более 29,7 дня.

Ошибка для альтернативного признака (или для выборочной относительной величины, или для доли) определяется аналогично. Дисперсия относительной величины по данным выборки:

$$s^2 = pq = p(1-p)$$

Средняя ошибка выборочной доли:

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

Предельная ошибка выборочной доли с принятой доверительной вероятностью имеет вид:

$$\Delta_p = ts_p = t\sqrt{\frac{p(1-p)}{n}}$$

<u>Пример</u>. По данным выборочного изучения 100 платежных документов предприятия одного треста оказалось, что в 6 случаях сроки расчетов с кредиторами были превышены. С вероятностью 0,954 требуется установить доверительный интервал доли платежных документов треста без нарушения сроков:

$$q = \frac{6}{100} = 0{,}06 \text{ или } 6\%,\ p=0{,}94$$

$$s_p = \sqrt{\frac{0,94*0,06}{100}} = \pm 0,024$$

$$\Delta_p = 2*0,024 = \pm 0,048$$

Генеральная доля платежных документов π, не выходящих за установленные сроки, с вероятностью 0,954 находится в интервале:

$$0,892 \le \pi \le 0,988 \text{ или } 89,2\% \le \pi \le 98,8\%$$

4.2. определение объемов выборки для получения требуемой точности результатов с заданной вероятностью

Объем выборки рассчитывается на стадии проектирования выборочного обследования. Так как при повторном отборе

$$\Delta = t\sqrt{\frac{\sigma^2}{n}}, \text{ то } n = \frac{t^2\sigma^2}{\Delta^2}$$

где: Δ - допустимая погрешность, которая задается исследователем исходя из требуемой точности результатов проектируемой выборки;

t - табличная величина, соответствующая заданной доверительной вероятности F(t), с которой будут гарантированы оценки генеральной совокупности по данным выборочного исследования;

σ^2 - генеральная дисперсия.

В зависимости от вида выборки, ее объем определяется по следующим формулам:

собственно-случайная и механическая выборка:

$$n = \frac{t^2\sigma^2}{\Delta^2} \text{ (повторный отбор);}$$

$$n = \frac{t^2\sigma^2 N}{\Delta^2 N + t^2\sigma^2} \text{ (бесповторный отбор);}$$

типическая выборка:

$$n = \frac{t^2\overline{\sigma}^2}{\Delta^2} \text{ (повторный отбор);}$$

$$n = \frac{t^2\overline{\sigma}^2 N}{\Delta^2 N + t^2\overline{\sigma}^2} \text{ (бесповторный отбор);}$$

серийная выборка:

$$n = \frac{t^2\delta^2}{\Delta^2} \text{ (повторный отбор);}$$

$$n = \frac{t^2\delta^2 R}{\Delta^2 R + t^2\delta^2} \text{ (бесповторный отбор);}$$

Генеральная дисперсия, как правило, неизвестна. Поэтому используют в расчетах ее оценку. Можно использовать результаты прошлых исследований той же совокупности, если ее структура и условия развития достаточно стабильны. Можно пользоваться следующим способом:

$$\sigma \approx \frac{1}{3}\bar{x}$$

или:

$$\sigma \approx \frac{1}{5}(x_{\max} - x_{\min}).$$

Для относительной величины принимают максимальную величину дисперсии 0,25.

При расчете n не следует гнаться за большими значениями t и малыми значениями Δ, так как это приведет к увеличению объема выборки, а следовательно, к увеличению затрат средств, труда и времени.

4.3. Определение возможного предела ошибок репрезентативности, гарантированного с заданной вероятностью

Решение этой задачи заключается в отыскании *F(t)* на основе формулы предела ошибки выборки:

$$\Delta = t\sqrt{\frac{s^2}{n}} \Rightarrow t = \frac{\Delta}{\sqrt{\frac{s^2}{n}}}$$

где Δ- допустимый размер погрешности оцениваемого показателя;

s^2 - дисперсия показателя, рассчитанная по данным выборочного наблюдения;

n - объем проведенной выборки.

Пример. При выборочной проверке качества электроламп средняя продолжительность горения составила 1420 часов ($\bar{x}$ = 1420 ч) при $\sigma = 61{,}03$ и объеме выборки 100 электроламп. Определить вероятность того, что средний срок горения электроламп будет меньше гарантированных 1410 часов.

Учитывая среднюю продолжительность горения по выборке допустимая погрешность составит 1410 - 1420 = - 10 ч.

Следовательно:

$$10 = t\sqrt{\frac{61{,}03^2}{100}}, \text{ отсюда } t = 1{,}64.$$

Соответствующая доверительная вероятность 0,899. Это означает, что из 100 ламп 90 будут иметь параметры в заданных пределах, и только у 10 лам характеристики выйдут из этого предела. Т.к. нас интересуют только лампы, у которых характеристики хуже нижней границы, и не интересуют те, у которых они лучше верхней, то можно сделать вывод, что из 100 ламп 5 могут гореть менее 1410 ч.

5. Малая выборка

Таблицы интеграла вероятностей используются для выборок большого объема из бесконечно большой генеральной совокупности. Но уже при *n* < 100 получается несоответствие между табличными данными и вероятностью предела; при $n < 30$ погрешность становится значительной. Несоответствие вызывается главным образом характером распределения единиц генеральной совокупности. При большом объеме выборки особенность распределения в генеральной совокупности не имеет значения, так как распределение отклонений выборочного показателя от генеральной характеристики при большой выборке всегда оказывается нормальным.

В выборках небольшого объема $n < 30$ характер распределения генеральной совокупности сказывается на распределении ошибок выборки. Поэтому для расчета ошибки выборки при небольшом объеме наблюдения (уже менее 100 единиц) отбор должен проводиться из совокупности, имеющей нормальное распределение.

Теория малых выборок разработана английским статистиком В. Госсетом (писавшим под псевдонимом Стьюдент) в начале XX в. В 1908 г. им построено специальное распределение, которое позволяет и при малых выборках соотносить t и доверительную вероятность $F(t)$. При $n > 100$ таблицы распределения Стьюдента дают те же результаты, что и таблицы интеграла вероятностей Лапласа, при $30 < n < 100$ различия незначительны. Поэтому практически к малым выборкам относят выборки объемом менее 30 единиц (безусловно большой считается выборка с объемом более 100 единиц).

Использование малых выборок в ряде случаев обусловлено характером обследуемой совокупности. Так, в селекционной работе «чистого» опыта легче добиться на небольшом числе делянок. Производственный и экономический эксперимент, связанный с экономическими затратами, также проводится на небольшом числе испытаний.

В случае малой выборки только для нормально распределенной генеральной совокупности могут быть рассчитаны и доверительные вероятности, и доверительные пределы генеральной средней.

*Плотность вероятностей распределения **Стьюдента** описывается функцией*

$$f(t,n) = B_n(1+\frac{t^2}{n-1})^{-\frac{n}{2}}$$

где: t - текущая переменная;

n - объем выборки;

B - величина, зависящая лишь от n.

Распределение Стьюдента имеет только один параметр: *d.f.* -число степеней свободы (иногда обозначается *k*).

Это распределение, как и нормальное, симметрично относительно точки $t = 0$, но оно более пологое. При увеличении объема выборки, а следовательно, и числа степеней свободы распределение Стьюдента быстро приближается к нормальному. Число степеней свободы равно числу тех индивидуальных значений признаков, которыми нужно располагать для определения искомой характеристики.

Так, для расчета дисперсии должна быть известна средняя величина. Поэтому при расчете дисперсии $d.f. = n - 1$.

Таблицы распределения Стьюдента публикуются в двух вариантах:

1. аналогично таблицам интеграла вероятностей приводятся значения t и соответствующие вероятности $F(t)$ при разном числе степеней свободы;
2. значения t приводятся для наиболее распространенных доверительных вероятностей 0,90; 0,95 и 0,99 или для 1 - 0,9 = 0,1, 1 - 0,95 = 0,05 и 1 - 0,99 = 0,01 при разном числе степеней свободы.

Малые выборки широко используются для решения задач, связанных с испытанием статистических гипотез, особенно гипотез о средних величинах.

6. *Статистическая проверка гипотез (общие понятия)*

Оценку генерального параметра получают на основе выборочного показателя с учетом ошибки репрезентативности.

В статистике существует и другой прием, когда в отношении свойств генеральной совокупности выдвигается некоторая гипотеза о величине средней, дисперсии, характере распределения, форме и тесноте связи между переменными. Проверка гипотезы осуществляется на основе выявления согласованности эмпирических данных с гипотетическими (теоретическими). Если расхождение между сравниваемыми величинами не выходит за пределы случайных ошибок, гипотезу принимают. При этом не делается никаких заключений о правильности самой гипотезы, речь идет лишь о согласованности сравниваемых данных. Основой проверки статистических гипотез являются данные случайных выборок. При этом безразлично, оцениваются ли гипотезы в отношении реальной или гипотетической генеральной совокупности. Последнее открывает путь применения этого метода за пределами собственно выборки: при анализе результатов эксперимента, данных сплошного наблюдения малой численности. В этом случае рекомендуется проверить, не вызвана ли установленная закономерность стечением случайных обстоятельств, насколько она характерна для того комплекса условий, в которых находится изучаемая совокупность.

Особенно часто процедура проверки статистических гипотез применяется для оценки существенности расхождений сводных характеристик отдельных совокупностей (групп): средних, относительных величин. Такого рода задачи, как правило, возникают в социальной статистике. Трудоемкость статистико-социологических исследований приводит к тому, что почти все они строятся на несплошном учете.

Статистической гипотезой называется предположение о свойстве генеральной совокупности, которое можно проверить, опираясь на данные выборки. Обозначается гипотеза буквой H от латинского слова hypothesis. Так, может быть выдвинута гипотеза о том, что средняя в генеральной совокупности равна некоторой величине $H: \mu = a$, или о том, что генеральная средняя больше некоторой величины $H: \mu > b$.

Различают ***простые*** и ***сложные*** гипотезы. Простая гипотеза однозначно характеризует значение случайной величины. Например, $H: \mu = a$. *Сложная* гипотеза состоит из конечного или бесконечного числа простых гипотез, при этом указывается некоторая область вероятных значений параметра. Например, $H: \mu > b$. Эта гипотеза состоит из множества простых гипотез $H: \mu = c$, где c - любое число, большее b.

Гипотезы о параметрах генеральной совокупности называются ***параметрическими***, *о* распределениях - ***непараметрическими***.

Гипотеза о том, что две совокупности, сравниваемые по одному или нескольким признакам, не отличаются, называется ***нулевой гипотезой*** (или нуль-гипотезой). Она обозначается H_0. При этом предполагается, что действительное различие сравниваемых величин равно нулю, а выявленное по данным отличие от нуля носит случайный характер. Например, $H_0: \mu_1 = \mu_2$ и т.д.

Нулевая гипотеза отвергается тогда, когда по выборке получается результат, который при истинности выдвинутой нулевой гипотезы маловероятен. При этом принимают некоторую границу невозможного или маловероятного, которую обычно принимают $\alpha = 0{,}05$, т.е. 5%, или 0,01; 0,001.

Статистическим критерием называют определенное правило, устанавливающее условия, при которых проверяемую нулевую гипотезу следует либо отклонить, либо не отклонить. Критерий проверки статистической гипотезы определяет, противоречит ли выдвинутая гипотеза фактическим данным или нет.

Проверка статистических гипотез складывается из следующих этапов:

- формулируется в виде статистической гипотезы задача исследования;
- выбирается статистическая характеристика гипотезы;
- выбираются испытуемая и альтернативная гипотезы на основе анализа возможных ошибочных решений и их последствий;
- определяются область допустимых значений, критическая область, а также критическое значение статистического критерия (t, F, χ^2) по соответствующей таблице;
- вычисляется фактическое значение статистического критерия;
- проверяется испытуемая гипотеза на основе сравнения фактического и критического значений критерия, и в зависимости от результатов проверки гипотеза либо отклоняется, либо не отклоняется.

При проверке гипотез по одному из критериев возможны *два ошибочных решения:*

1. неправильное отклонение нулевой гипотезы: ошибка 1-го рода;
2. неправильное принятие нулевой гипотезы: ошибка 2-го рода.

В то время, как фактически нулевая гипотеза верна (1) и нулевая гипотеза не верна (2), принимают два ошибочных решения: 1) нулевая гипотеза отклоняется и принимается альтернативная гипотеза; 2) нулевая гипотеза не отклоняется. Возможные решения представлены в табл. 6.1.

Таблица 6.1

Возможные выводы при проверке гипотез

Решение по критерию	Фактически	
	H_0 верна	H_0 не верна
H_0 отклоняется	Ошибка 1-го рода	Правильное решение
H_0 не отклоняется	Правильное решение	Ошибка 2-го рода

Если, например, установлено, что новое минеральное удобрение лучше, хотя на самом деле его действие не отличается от старого, то это ошибка 1-го рода. Если мы решили, что оба вида удобрений одинаковы, то допущена ошибка 2-го рода.

Вероятности, соответствующие неверным решениям, называются ***риском 1*** и ***риском 2***. *Риск 1* равен вероятности ошибки α *(уровню значимости), риск 2* равен вероятности ошибки β. Поскольку α всегда больше нуля, то всегда есть риск ошибки β. При заданных α и объеме выборки *n* значение β будет тем больше, чем меньше принятое α. Если *n* велико, то α и β могут быть сколь угодно малыми, т.е. решения будут более обоснованными. При малом объеме выборки и малом α возможность установить фактически существующие различия малы.

Обычно задают значение α и пытаются сделать возможно малым β. Вероятность 1 - β называется ***мощностью критерия:*** чем она больше, тем меньше вероятность ошибки второго рода.

Альтернативная гипотеза H_1 может быть сформулирована по-разному в зависимости от того, какие отклонения от гипотетической величины нас особенно беспокоят: положительные, отрицательные либо и те, и другие. Соответственно альтернативные гипотезы могут быть записаны как

$$H_1 : \mu \succ a$$

$$H_1 : \mu \prec a$$

$$H_1 : \mu \neq a$$

От того, как формулируется альтернативная гипотеза, зависят границы критической области и области допустимых значений.

Критической областью называется область, попадание значения статистического критерия в которую приводит к отклонению H_0. Вероятность попадания значения критерия в эту область равна принятому уровню значимости.

Область допустимых значений дополняет критическую область. Если значение критерия попадает в область допустимых значений, это свидетельствует о том, что выдвинутая гипотеза H_0 не противоречит фактическим данным (H_0 не отклоняется).

Точки, разделяющие критическую область и область допустимых значений, называются ***критическими точками*** или ***границами критической области***. В зависимости от формулировки альтернативной гипотезы критическая область может быть ***двухсторонняя*** или ***односторонняя*** (левосторонняя либо правосторонняя).

Если вычисляемое значение критерия попадает в критическую область, нулевая гипотеза отклоняется, она противоречит фактическим данным.

Одна из важнейших задач анализа вариационных рядов заключается в выявлении закономерности распределения и определении ее характера. Основной путь в выявлении закономерности распределения - построение вариационных рядов для достаточно больших совокупностей. Большое значение для выявления закономерностей распределения имеет правильное построение самого вариационного ряда: выбор числа групп и размера интервала варьирующего признака.

Когда мы говорим о характере, типе закономерности распределения, то имеем в виду отражение в нем общих условий, определяющих вариацию. При этом речь всегда идет о распределениях качественно однородных явлений. Общие условия, определяющие тип закономерности распределения, познаются анализом сущности явления, тех его свойств, которые определяют вариацию изучаемого признака. Следовательно, должна быть выдвинута какая-то научная гипотеза, обосновывающая определенный тип теоретической кривой распределения.

Под теоретической кривой распределения понимается графическое изображение ряда в виде непрерывной линии изменения частот в вариационном ряду, функционально связанного с изменением вариантов (значений признака). Теоретическое распределение может быть выражено аналитически - формулой, которая связывает частоты вариационного ряда и соответствующие значения признака. Такие алгебраические формулы носят название ***законов распределения.***

Большое познавательное значение имеет сопоставление фактических кривых распределения с теоретическими.

Гипотезы о распределениях заключаются в том, что выдвигается предположение о том, что распределение в генеральной совокупности подчиняется какому-то определенному закону. Проверка гипотезы состоит в том, чтобы на основании сравнения фактических (эмпирических) частот с предполагаемыми (теоретическими) частотами сделать вывод о соответствии фактического распределения гипотетическому распределению.

Под гипотетическим распределением необязательно понимается нормальное распределение. Может быть выдвинута гипотеза о биномиальном распределении, распределении Пуассона и т.д. Причина частого обращения к нормальному распределению в том, что в этом типе распределения выражается закономерность, возникающая при взаимодействии множества случайных причин, когда ни одна из них не имеет преобладающего влияния. Закон нормального распределения лежит в основе многих теорем математической статистики, применяемых для оценки репрезентативности выборок, при измерении связей и т. д. В социально-экономической статистике нормальное распределение встречается редко, но сравнение с ним важно для выяснения степени и характера отклонения от него фактического распределения.

7. *Оценка существенности различия средних*

Для оценки существенности различия двух средних предназначен критерий t-Стьюдента, исчисляемый по формуле:

$$t_{расч} = \frac{\bar{x}_2 - \bar{x}_1}{\sigma_{разн}}, \text{ где}$$

$t_{расч}$ – фактическое значение критерия t-Стьюдента,
$\bar{x}_1, \bar{x}_2$ –средние двух совокупностей,
$\sigma_{разн}$ – показатель колеблемости разности.

Показатель колеблемости разности может быть исчислен так:

$$\sigma_{разн} = \sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}}, \text{ где}$$

σ_1^2 – дисперсия признака в первой совокупности,
σ_2^2 – дисперсия признака во второй совокупности,
n_1, n_2 – объем совокупностей,
$(n_1\text{-}1)$, $(n_2\text{-}1)$ – число степеней свободы

Если объемы совокупности велики, то приведенную формулу упрощают, т.е. в знаменателях дробей ставят просто n_1 и n_2. Исчисленное с помощью формулы фактическое значение критерия – $t_{расч}$. сравнивают с его критическим значением – $t_{кр}$. Если совокупность большого объема, то величина $t_{кр}$ подчиняется закону нормального распределения и берется из таблиц интегральной функции нормального распределения F(t) или функции Лапласа Ф(t). Если расчеты ведутся по малой совокупности (n<30), то величина $t_{кр.}$ подчиняется распределению Стьюдента и ее критическое значение берется из таблиц S(t)-распределения.

Если в ходе сравнения окажется, что $t_{расч} \le t_{кр}$, то нулевую гипотезу считают не противоречащей данным наблюдения. При $t_{расч} > t_{кр}$ нулевую гипотезу отвергают в пользу одной из альтернативных гипотез.

Рассмотрим расчет показателей вариации по следующим ее данным об успеваемости в двух группах студентов (табл. 4.4):

Таблица 4.4

Успеваемость студентов

Группы по полу	Число студентов в группе	Успеваемость отдельных студентов в баллах
Юноши	6	5; 5; 4,6; 4,5; 3,2; 3
Девушки	6	4,2; 4,1; 4; 4; 3,9; 3,8

В какой группе успеваемость выше; являются ли различия в средней успеваемости двух групп студентов существенными?

	Юноши	Девушки
W	2,0	0,4
σ^2	0,6617	0,0167

σ	0,813	0,129
V	19,4%	3,2%

$\bar{x}_1$ (*юноши*) $= 4{,}2$; $\bar{x}_2$ (*девушки*) $= 4{,}0$;

$\delta_{разн} = 0{,}37$; $t_{расч} = \dfrac{4{,}0 - 4{,}2}{0{,}37} = 0{,}54$.

$t_{кр}$=2,57 при числе степеней свободы (6-1=5) при 5%-ом уровне значимости.

Так как $t_{расч} < t_{кр}$ (0,54<2,57), то различия в средней успеваемости двух групп студентов не существенны.

8. Примеры решения задач

Пример 1.

Для формирования партии зерна с целью ее дальнейшей обработки в зерносушилке отобрали из прибывающих от комбайнов машин 10 проб на влажность. Были получены следующие результаты:

18,1	18,4
17,4	17,5
18,3	19,0
17,0	18,7
17,7	17,9

Определить среднюю влажность зерна, среднюю и предельную ошибку выборочной средней влажности зерна, доверительный интервал, в которых находится влажность зерна в генеральной совокупности.

Надежный уровень вероятности принять равным 0,9545.

№ п/п	Влажность зерна, %	Квадрат влажности	№ п/п	Влажность зерна, %	Квадрат влажности
	x_i	x_i^2		x_i	x_i^2
1	18,1	327,6	6	18,4	338,6
2	17,4	302,8	7	17,5	306,3
3	18,3	334,9	8	19,0	361,0
4	17,0	289,0	9	18,7	349,7
5	17,7	313,3	10	17,9	320,4
			Всего	180,0	3243,5

Решение.

1. Средняя влажность зерна:

$$\bar{x} = \frac{180}{10} = 18{,}0$$

2. Выборочная дисперсия

$$\sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 = \frac{3243{,}5}{10} - 18{,}0^2 = 0{,}346$$

3. Генеральная дисперсия

$$S^2 = \sigma^2 \frac{n}{n-1} = 0{,}346 \frac{10}{10-1} = 0{,}384$$

4. Средняя ошибка

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{0{,}384}{10}} = 0{,}2\%$$

5. Предельная ошибка

$$\Delta = t\mu = 2{,}0 * 0{,}2 = 0{,}4$$

6. Интервальная оценка

$$18{,}0 - 0{,}4 < \mu < 18{,}0 + 0{,}4$$

Пример 2

В институте работает 500 человек профессорско-преподавательского состава. По результатам выборочного исследования 50 человек установлено, что 10% из них имеют стаж работы менее 10 лет. Отбор проводили способом случайной бесповторной выборки.

Определить с надежной вероятностью 0,954 среднюю и граничную ошибки выборочной доли, а также доверительный интервал, в котором находится удельный вес преподавателей со стажем работы менее 10 лет в генеральной совокупности.

Решение.

1. Предельная ошибка выборочной доли

$$\Delta = 2\sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0{,}1(1-0{,}1)}{50}\left(1-\frac{50}{500}\right)} = 0{,}08$$

2. Доверительный интервал

Нижняя граница – 0,02

Верхняя граница – 0,18

Пример 3.

Оценка существенности среднего различия (в отличие от оценки существенности различия средних)

Проводится исследование влияния на урожайность некоторого удобрения. При этом использовали 10 участков одинакового размера. На 5 контрольных участках удобрение не вносилось, а на 5 опытных вносилось. Участки подобраны парами по критерию равенства всех остальных условий. Необходимо проверить статистическую гипотезу относительно среднего различия между парами взаимосвязанных наблюдений в генеральной совокупности.

Наблюдение	Опытный участок	Контрольный участок	Разность	Квадрат разности
1	13,6	10,5	3,1	9,61
2	16,2	12,4	3,8	14,44
3	17,9	15,3	2,6	6,76
4	13,5	11,7	1,8	3,24
5	14,8	13,1	1,7	2,89
Всего	76,0	63,0	13,0	36,94
Средняя	15,2	12,6	2,6	

Т.к. наблюдения двух выборок попарно связаны (зависимые выборки), то проверяют не различие между средними, а среднее значение различия между парами наблюдений.

Сформулируем нулевую и альтернативную гипотезы.

$H_0 : \bar{d} = 0$

$H_1 : \bar{d} \neq 0$

Уровень значимости примем 0,05.

Определим среднюю разность:

$$\bar{d} = \frac{13{,}0}{5} = 2{,}6$$

Определим дисперсию средней разности:

$$S^2 = \frac{36{,}94}{5} - 2{,}6^2 = 0{,}628$$

Дисперсия в генеральной совокупности:

$$\sigma^2 = S^2 \frac{n}{n-1} = 0{,}628 \frac{5}{5-1} = 0{,}785$$

Вычислим среднюю ошибку средней разности

$$\mu_{\bar{d}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{0{,}785}{5}} = 0{,}396$$

Рассчитаем фактическое значение критерия Стьюдента

$$t_{факт} = \frac{|\bar{d}|}{\mu_{\bar{d}}} = \frac{2{,}6}{0{,}396} = 6{,}566$$

Табличное значение критерия Стьюдента 2,78.

Поскольку фактическое значение больше критического, то нулевую гипотезу следует отклонить. Это означает, что величина средней разности средних урожайностей по двум парам выборок является существенной.

Пример 4.

Для определения средней заработной платы рабочих завода была проведена 20% бесповторная выборка (по цехам) с отбором единиц пропорционально численности групп. Результаты выборки следующие:

Цех	Объем выборки	Средняя заработная плата	Среднее квадратическое отклонение
1	120	873	30
2	100	886	80
3	180	900	60
Всего	400		

С вероятность 0,997 (т.е. t=3) определить пределы, в которых будет находиться заработная плата всех рабочих завода.

Решение

Находим выборочную среднюю заработную плату:

x=888,4

Средняя из групповых дисперсий – 3490

$\Delta = 3\sqrt{\frac{3490}{400}(1-0{,}2)}$ =7,9

Отсюда генеральная средняя

880,5<888,4<886,4

Пример 5.

При проверке непараметрических гипотез ставится задача о соответствии фактического распределения некоторому заранее заданному распределению. В качестве таких распределений чаще всего используется нормальное распределение, функция которого имеет вид:

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

Теоретические частоты находятся по следующей формуле:

$$f_m(t) = \frac{Nh}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

N - сумма частот ряда

h – размер интервала

В целом ряде случаев, если вариационный ряд представляет собой распределение по дискретному признаку, где по мере увеличения значений признака частоты резко уменьшаются и где средняя арифметическая близка по значению к дисперсии, используют в качестве теоретического распределение Пуассонна, функция которого имеет вид:

$$P_ч = \frac{a^x e^{-a}}{x!}$$

Px – вероятность наступлений некоторых отдельных значений x

a – средняя арифметическая уровней ряда

Теоретические частоты находятся по следующей формуле:

$$f_m(t) = NP_x$$

N – общее число единиц ряда.

Используются следующие критерии согласия:

Критерий Пирсона:

$$\chi^2 = \sum \frac{(f - f_m)^2}{f_m}$$

Фактическое значение сравнивается с теоретическим. Нулевая непараметрическая гипотеза принимается, если фактическое значение критерия меньше теоретического.

Критерий Романовского

$$\frac{\left|\chi^2 - k\right|}{\sqrt{2k}}$$

χ^2 - фактическое значение критерия Пирсона

k – число степеней свободы, которое находится как число групп в ряду минус единица и минус число параметров эмпирического распределения, использованных для нахождения теоретических частот.

Нулевая непараметрическая гипотеза принимается, если значение критерия меньше 3.

Критерий Колмогорова

$$\lambda = d\sqrt{N}$$

d – максимальная величина расхождения между накопленными фактическими и теоретическими частостями.

N – сумма всех частот.

Если используются не частности, а частоты, то формула принимает вид:

$$\lambda = \frac{D}{\sqrt{N}}$$

Фактическое значение сравнивается с теоретическим. Нулевая непараметрическая гипотеза принимается, если фактическое значение критерия меньше теоретического.

Интервал			f_i	$x_i f_i$	$x^2_i f_i$	Нормированное отклонение	f(t)	Теоретически частоты
Нижняя граница	Верхняя граница	Средина						
120	130	125	1	125	15625	-2,81	0,0077	
130	140	135	8	1080	145800	-2,04	0,0502	
140	150	145	27	3915	567675	-1,27	0,1792	2
150	160	155	58	8990	1393450	-0,49	0,3532	5
160	170	165	56	9240	1524600	0,28	0,3839	5
170	180	175	34	5950	1041250	1,05	0,2301	3
180	190	185	14	2590	479150	1,82	0,0761	1
190	200	195	2	390	76050	2,59	0,0139	
Итого			200	32280	5243600			20

Средняя 161,4

Дисперсия 168

Среднее квадратичное отклонение 13

$$\frac{Nh}{\sigma}=\frac{200*10}{13}=154$$

Критерий Пирсона

Фактические частоты	Теоретические частоты	Разность частот	Квадрат разности	Нормированный квадрат
1	1	0	0	0
8	8	0	0	0
27	28	-1	1	0,0357
58	55	3	9	0,1636
56	59	-3	9	0,1525
34	35	-1	1	0,0286
14	12	2	4	0,3333
2	2	0	0	0
200	200			0,7138

Для рассматриваемого примера k=5. Табличное значение при уровне значимости 0,05 составляет 11,07.

Так как полученное значение меньше критического, то можно считать случайным расхождение между теоретическими и фактическими частотами. Выдвинутая гипотеза о близости эмпирического распределения к нормальному принимается.

Критерий Романовского $\frac{|0,71-5|}{\sqrt{2*5}}=1,4$ меньше 3, что также подтверждает ранее сделанный вывод.

Критерий Колмогорова

Фактические частоты	Теоретические частоты	Накопленные частоты		Абсолютная разность накопленных частот
		Фактические частоты	Теоретические частоты	
1	1	1	1	0
8	8	9	9	0
27	28	36	37	1
58	55	94	92	2
56	59	150	151	1
34	35	184	186	2
14	12	198	198	0
2	2	200	200	0
200	200	Максимум		2

$\frac{2}{\sqrt{200}} = 1,1414$. По таблице значений критерия Колмогорова находим, что данному значению соответствует вероятность 1,0000. Следовательно, вполне можно полагать, что расхождения между теоретическими и фактическими частотами носят случайный характер.

Пример 6.

Основным источником информации о жизненном уровне населения в целом и отдельных слоев, его характеристики по структуре доходов и расходов, а также данные относительно потребления товаров и услуг в зависимости от уровня материальной обеспеченности, численного состава домохозяйств и по другим социально-экономическим аспектам является выборочное обследование условий жизни домохозяйств.

Обследование в Украине проводится с 1999 г., осуществляется на постоянной основе, базируется на международных стандартах и отвечает современной социально-демографической ситуации в Украине.

Домохозяйство - совокупность лиц, которые совместно проживают в одном жилом помещении или его части, обеспечивают себя всем необходимым для жизни, ведут общее хозяйство, полностью или частично объединяют и расходуют средства. Эти лица могут находиться в родственных отношениях или в других отношениях. Домохозяйство может состоять из одного лица (статья 1 Закона Украины "О Всеукраинской переписи населения").

Домохозяйства принимают участие в обследовании один год.

Обследования проводится ежеквартально во всех регионах страны и репрезентует все население, за исключением военнослужащих срочной службы, лиц, которые находятся в местах лишения свободы, лиц, которые постоянно проживают в домах-интернатах, домах для лиц преклонного возраста, общежитиях (кроме семейных), а также маргинальных прослоек населения. Кроме того, при формировании выборочной совокупности были также исключенные территории, которые не могут быть обследованы в связи с радиоактивным загрязнением (зоны отчуждения и обязательного отселения).

В 2003 г. начальный объем выборочной совокупности обследования составлял 12534 домохозяйств. На протяжении года приняли участие в обследовании 9716 домохозяйств (78,6% отобранных адресов за исключением нежилых помещений). Итоги выборочного обследования распространяются с применением процедуры статистического взвешивания на генеральную совокупность - все население страны.

В основу исследования дифференциации домохозяйств по уровню материальной обеспеченности положены показатели затрат, как более надежные по сравнению с данными относительно доходов, которые респонденты склонны занижать.

Денежные расходы домохозяйств включают: расходы на покупку продуктов питания, алкогольных и табачных изделий, непродовольственных товаров, на оплату услуг, расходы, связанные с ведением личного подсобного хозяйства, денежную помощь родственникам и другим лицам, затраты на покупку недвижимости, на строительство, на приобретение акций, сертификатов, валюты прирост вкладов в банковских учреждениях, алименты, налоги (кроме налога на доход), сборы, взносы и др.

Совокупные расходы состоят из денежных расходов, а также стоимости потребленных домохозяйством и подаренных ему родственниками и другими лицами продуктов питания, полученных из личного подсобного хозяйства и в порядке самозаготовок, суммы льгот и безналичных субсидий на оплату жилищно-коммунальных услуг и электроэнергии, суммы безналичных льгот на проезд в транспорте, на оплату телефона, путевок в санаторно-курортные учреждения, на покупку лекарства, лечения, зубопротезирования и т.п.. Стоимость подаренных домохозяйством родственникам и другим лицам продуктов питания, полученных из личного подсобного хозяйства, включается в совокупные расходы в связи с постоянным характером этой помощи.

Условие.

Сгруппированные результаты выборочного исследования домохозяйств по уровню средних денежных расходов на одного человека в 2003 г. приведены в таблице:

Средние расходы на человека в месяц, грн.	удельный вес
до 60,0	3,1
60,1-90,0	7,0
90,1-120,0	10,3
120,1-150,0	11,6
150,1-180,0	11,3
180,1-210,0	11,2
210,1-240,0	9,9
240,1-270,0	7,0
270,1-300,0	5,7
300,1-360,0	8,2
свыше 360,0	14,7
Все население	100,0

Проверить гипотезу о соответствии эмпирического распределения нормальному.

Решение.

1. Определим параметры выборочной совокупности.

Т.к. известно, что обследовано 9716 домохозяйств, то от удельного веса перейдем к частотам. При этом восстановим равенство интервалов путем объединения последних интервалов.

Средние расходы на человека в месяц, грн.	Частота	Расчет средней		Расчет дисперсии
x_i	f_i	x_i'	f_i*x_i'	$f_i*(x_i')^2$
до 60,0	301	45,1	13575,10	612237,01
60,1-90,0	680	75,1	51068,00	3835206,80
90,1-120,0	1001	105,1	105205,10	11057056,01
120,1-150,0	1127	135,1	152257,70	20570015,27
150,1-180,0	1098	165,1	181279,80	29929294,98
180,1-210,0	1088	195,1	212268,80	41413642,88
210,1-240,0	962	225,1	216546,20	48744549,62
240,1-270,0	680	255,1	173468,00	44251686,80
270,1-300,0	554	285,1	157945,40	45030233,54
свыше 300,0	2225	315,1	701097,50	220915822,25
Итого	9716		1964711,60	466359745,16

Средняя 202,21
Дисперсия 7108,63294
Среднее квадратичное отклонение 84,31

Определим теоретические частоты нормального распределения и критерия Пирсона.

Фактическая частота	Нормированное отклонение	Значение функции нормального распределения	Теоретическая частота	Разница частот	Квадрат разницы частот
f_i	t	f(t)	ft		
301	-1,86	0,0703	486	-185	34213,8
680	-1,51	0,1280	443	237	56340,9
1001	-1,15	0,2055	710	291	84419,7
1127	-0,80	0,2906	1005	122	14959,1
1098	-0,44	0,3621	1252	-154	23666,7
1088	-0,08	0,3975	1374	-286	81964,5
962	0,27	0,3845	1329	-367	134917
680	0,63	0,3277	1133	-453	205110
554	0,98	0,2461	851	-297	88018,2
2225	1,34	0,1628	1126	1099	1208656
9716			9708		

Нормированное отклонение:

$$t = \frac{x_i^{'} - \bar{x}}{\delta}$$

Значение функции нормального распределения

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Теоретическая частота

$$ft = 9716 \frac{30}{\delta} f(t)$$

Взвешенный квадрат разницы

$$\frac{(f_i - ft)^2}{ft}$$

Контрольные вопросы к теме 6

1. В чем состоят причины необходимости и целесообразности использования выборочного метода в статистике?
2. В чем состоят основные принципы и правила проведения выборочного обследования.
3. В чем состоят различия основных видов выборки.
4. Дайте основные понятия *статистической гипотезы.*
5. Каким показателем можно оценить существенное различие средних? Дайте характеристику этого показателя.

Тема 7. Методы изучения связи между явлениями и их использование для управления социально-экономическими процессами

1. Виды связей между явлениями

По характеру зависимости одного явления от другого различают только два типа связи:

- функциональная (полная), при которой каждому значению одного аргумента соответствует одно вполне определенное значение функции;
- статистическая, при которой каждому значению аргумента соответствует ряд распределения, т.е. множество значений функции.

Все известные в науке в настоящее время связи попадают под определение статистической связи. Под него попадает и определение функциональной связи. В этом случае функциональную связь можно рассматривать как частный случай статистической связи, когда значениям одной переменной соответствуют «распределения» значений второй переменной, состоящие из одного или нескольких значений и имеющих вероятность равную единице.

Корреляционной называют важнейший частный случай статистической связи, состоящий в том, что разным значениям одной переменной соответствуют различные средние значения другой переменной. С изменением значения признака х закономерным образом изменяется и среднее значение признака у, в то время как в каждом отдельном случае значение признака у (с различными вероятностями) может принимать множество различных значений.

Связь будет статистической, но не будет корреляционной, если с изменением значения признака х закономерным образом изменяется какая-либо другая статистическая характеристика признака у (показатели вариации, ассиметрии, эксцесса и т.д.). Статистическая связь между двумя признаками (переменными величинами) предполагает, что каждый из них имеет случайную вариацию индивидуальных значений относительно средней величины. Если же такую вариацию имеет лишь один из признаков, а значения другого являются жестко детерминированными, то говорят лишь о регрессии, но не о статистической (тем более корреляционной) связи.

Корреляционная связь между признаками может возникать разными путями:

- Важнейший путь - ***причинная зависимость*** результативного признака (его вариации) от вариации факторного признака. В этом случае один из признаков выступает как независимая переменная (фактор), а другой - как зависимая переменная (результат).
- Совершенно иная интерпретация необходима при изучении корреляционной связи между двумя следствиями общей причины.
- Третий путь возникновения корреляции - взаимосвязь признаков, каждый из которых и причина, и следствие.

При изучении корреляционной связи преследуются следующие цели:

- измерение тесноты связи признаков между собой;
- расчет параметров уравнения, которое выражает связь средних значений зависимой переменной со значениями независимой переменной.

По направлению различают связь прямую и обратную. Если с увеличением аргумента «х» функция «у» увеличивается, то такая связь называется прямой связью и наоборот, когда признаки варьируют независимо друг от друга – это отсутствие связи.

По аналитическому выражению корреляционную связь различают как прямолинейную и криволинейную.

Если точки корреляционного поля располагаются приблизительно по прямой, то говорят о прямолинейной зависимости. Если точки корреляционного поля расположены так, что можно провести через них параболу, гиперболу или другую кривую, то говорят о криволинейной зависимости. Самый простой способ установления вида зависимости заключается в построении так называемых диаграмм рассеивания. В этом случае графики можно построить вручную или с помощью пакетов прикладных программ, таких, как «Statgraf», «Excel», «Microstat».

Для ориентировочной оценки тесноты связи пользуются непараметрическими показателями статистики. К ним относят: коэффициент корреляции знаков, коэффициент корреляции рангов, коэффициент ассоциации, коэффициент взаимной сопряженности.

Корреляционно-регрессионный анализ применяется также и при решении задач, которые не имеют формального математического характера, а носящих содержательный характер:

1. ***Выделение важнейших факторов,*** влияющих на результативный признак (т.е. на вариацию его значений в совокупности). Эта задача решается в основном на базе мер тесноты связи факторов с результативным признаком.
2. ***Оценка хозяйственной деятельности по эффективности использования имеющихся факторов производства***. Эта задача решается путем расчета для каждой единицы совокупности тех величин результативного признака, которые были бы получены при средней по совокупности эффективности использования факторов и сравнения их с фактическими результатами производства.
3. ***Прогнозирование возможных значений результативного признака при задаваемых значениях факторных признаков.*** Такая задача решается путем

подстановки ожидаемых, или планируемых, или возможных значений факторных признаков в уравнение связи и вычисления ожидаемых значений результативного признака. Приходится решать и обратную задачу: вычисление необходимых значений факторных признаков для обеспечения планового или желаемого значения результативного признака в среднем по совокупности. Эта задача обычно не имеет единственного решения в рамках данного метода и должна дополняться постановкой и решением оптимизационной задачи на нахождение наилучшего из возможных вариантов ее решения (например, варианта, позволяющего достичь требуемого результата с минимальными затратами).

4. ***Задача подготовки данных, необходимых в качестве исходных для решения оптимизационных задач.*** Например, для нахождения оптимальной структуры производства в районе на перспективу исходная информация должна включать показатели производительности на предприятиях разных отраслей и форм собственности. В свою очередь, эти показатели могут быть получены на основе корреляционно-регрессионной модели либо на основании тренда динамического ряда (а тренд - это тоже уравнение регрессии).

2. Коэффициент корреляции знаков Г.Фехнера

Этот коэффициент был разработан немецким ученым Г. Фехнером. Его расчет основан на сопоставлении знаков отклонений от средней, т.е. вычисляют по средней арифметической простой среднее значение факторного признака (x) и результативного фактора (y). Затем каждое эмпирическое значение факторного и результативного признаков сравнивают со средним $\left(x-\bar{x}\right)$ и $\left(y-\bar{y}\right)$, проставляют отклонения в виде знаков.

Коэффициент корреляции знаков определяется по формуле:

$$K_{\phi} = \frac{\Sigma_C - \Sigma_H}{\Sigma_C + \Sigma_H}$$, где

Σ_C – сумма совпадений знаков,

Σ_H – сумма несовпадений знаков.

Коэффициент корреляции знаков колеблются в пределах от -1 до +1. Чем ближе коэффициент к единице, тем теснее связь. Если все знаки отклонения разные, то коэффициент знаков будет равен -1, что говорит об обратной связи между факторами.

Пример.

Вычислим коэффициент корреляции знаков Фехнера на основании данных о потреблении картофеля в семьях с разным доходом на одного члена семьи в Украине в 1998 г[1]. Соответствующие данные и необходимые расчеты приведены в табл. 7.1.

[1] Статистичний щорічник України за 1998 рік / Держкомстат України. - К.: Техніка, 1999. - 576 с.

По данным таблицы имеем: количество совпадений знаков – 7, несовпадений – 1.

Таблица 7.1

Пример расчета коэффициента Фехнера

Доход на одного члена семьи, грн		Количество семей, %	Потребление, кг.			
x_j	x_j'	d_j	y_j	$x_j' d_j$	$x_j' - \bar{x}$	$y_j - \bar{y}$
До 30	15	0,8	69,8	0,1	-100,2	-64,8
30,1-60,0	45	11,9	111,2	5,4	-70,2	-23,4
60,1-90,0	75	25,3	129,9	19,0	-40,2	-4,7
90,1-120,0	105	23,3	146,3	24,5	-10,2	11,7
120,1-150,0	135	15,0	152,7	20,3	19,8	18,1
150,1-180,0	165	9,2	152,2	15,2	49,8	17,6
180,1-210,0	195	5,8	152,7	11,3	79,8	18,1
более 210,0	225	8,7	161,9	19,6	109,8	27,3
Итого		100,0	1076,7	115,2		

Средний доход на одного члена семьи – 115,2 грн.
Среднее потребление картофеля на одного члена семьи – 134,6 кг.
Величина связи:

$$K_\phi = \frac{7-1}{7+1} = \frac{6}{8} = 0{,}75.$$

Коэффициент Фехнера грубый показатель тесноты связи, не учитывающий величину отклонений признаков от средних значений, но он может служить некоторым ориентиром в оценке интенсивности связи. В приведенном примере он указывает на тесную связь признаков.

3. Коэффициент корреляции рангов Спирмена.

Этот показатель предложил английский психолог Ч. Спирмен (1863-1945). Он вычисляется не по первичным данным, а по рангам (порядковым номерам). Если проранжировать совокупность по двум признакам, связь между которыми изучается, то полное совпадение рангов означает максимально тесную прямую связь, а полная противоположность рангов - максимально тесную обратную связь. Ранжировать оба признака необходимо в одном и том же порядке: либо от меньших значений признака к большим, либо наоборот.

Коэффициент корреляции рангов определяется по следующей формуле:

$$\rho = 1 - \frac{6\sum_{j=1}^{n} d_j^2}{n^3 - n}, \text{ где}$$

n – число наблюдений,

d – разность между порядковыми номерами парных членов (d=Nx-Ny).

Коэффициент корреляции рангов колеблется от -1 до +1.

Пример. Вычислим коэффициент Спирмена по данным табл. 7.1. Расчет приведен в табл. 7.2.

Таблица 7.2.

Пример расчета коэффициента Спирмена

Доход на одного члена семьи, грн		Количество семей, %	Потребление, кг.	Ранг по доходам	Ранг по потреблению	Разность рангов	
x_j	x_j'	d_j	y_j	p_x	p_y	$d = p_x - p_y$	d_j
до 30	15	0,8	69,8	1	1	0,0	
30,1-60,0	45	11,9	111,2	2	2	0,0	
60,1-90,0	75	25,3	129,9	3	3	0,0	
90,1-120,0	105	23,3	146,3	4	4	0,0	
120,1-150,0	135	15,0	152,7	5	6,5	-1,5	2,
150,1-180,0	165	9,2	152,2	6	5	1,0	
180,1-210,0	195	5,8	152,7	7	6,5	0,5	0,
более 210,0	225	8,7	161,9	8	8	0,0	
Сумма							3

$$r_{p_x p_y} = 1 - \frac{6 * 3{,}5}{8^3 - 8} = 0{,}958$$

Если среди значений признаков *x* и *y* встречается несколько одинаковых, образуются *связанные ранги, т.*е. одинаковые средние номера, то в этом случае коэффициент Спирмена вычисляется по формуле:

$$r_{p_x p_y} = 1 - \frac{6\sum d^2 - A - B}{\sqrt{(n^3 - n - 12A)(n^3 - n - 12B)}}$$

где:

$$A = \frac{1}{12}\sum_j (A_j^3 - A_j);$$

$$B = \frac{1}{12}\sum_k (B_k^3 - B_k)$$

j – номера связок по порядку для признака x

A_j – число одинаковых рангов в j-ой связке по x

k – номера связок по порядку для признака y

B_k - число одинаковых рангов в k-ой связке по y

Таким образом, для вышеприведенного примера имеем:

$$B = \frac{1}{12}(2^3 - 2) = 0,5$$

$$r_{p_x p_y} = 1 - \frac{6*3,5 - 0,5}{\sqrt{(8^3 - 8)(8^3 - 8 - 12*0,5)}} = 0,959$$

Для проверки значимости этого коэффициента рассчитывается стандартная ошибка:

$$\sigma_\rho = \sqrt{\frac{1-\rho^2}{n-2}},$$

где

n – число наблюдений,

ρ^2 – квадрат коэффициента корреляции рангов (коэффициент детерминации).

Статистика t при этом имеет вид: $t = \frac{\rho}{\sigma_\rho}$, где ρ – проверяемый коэффициент рангов. При этом t должно оказаться более 3, чтобы с достаточной уверенностью можно было бы сказать, что связь не случайна.

По приведенному примеру:

$$\sigma_h = \sqrt{\frac{1-0,959^2}{8-2}} = 0,1156$$

$$t = \frac{0,959}{0,1156} = 8,297$$

Таким образом, рассчитанная связь не является случайной.

4. Коэффициент корреляции рангов М. Кендалла

М. Кендалл предложил в качестве меры связи между переменными x и y использовать коэффициент, получивший название «коэффициент корреляции рангов Кендалла». Он вычисляется по формуле:

$$r = \frac{2*S}{n(n-1)}$$

Для расчета *S* ряды *x* и *y* ранжируются в порядке возрастания. Затем по ряду *y* считается число последующих рангов, превышающих данный ($П_i$) и число последующих рангов, ниже данного ($Н_i$). Затем определяется *S* как:

$$S = \sum_{i=1}^{n}(П_i - Н_i)$$

Для примера рассчитаем коэффициент корреляции рангов Кендалла по следующим данным (табл. 7.3):

Таблица 7.3

Расчета коэффициента корреляции Кендалла

Ранги по x	Ранги по Y		s
1	1	15-0	15
2	3	13-1	12
3	5	11-2	9
4	4	11-1	10
5	7	9-2	7
6	9	7-3	4
7	10	6-3	3
8	8	6-2	4
9	2	7-0	7
10	13	3-3	0
11	12	3-2	1
12	11	3-1	2
13	6	3-0	3
14	14	2-0	2
15	16	0-1	-1
16	15		
	Сумма		78

$$r = \frac{78}{\frac{1}{2}*16*15} = 0{,}65$$

Коэффициент корреляции рангов Кендалла изменяется в пределах от –1 до +1. При достаточно большом числе наблюдений между коэффициентами корреляции рангов Спирмена и Кендалла существует следующее соотношение:

$$\rho = \frac{3}{2} r\,.$$

Существенность коэффициента корреляции рангов Кендалла проверяется по формуле:

$$t_{расч} = \sqrt{\frac{2(2n+5)}{9n(n-1)}}$$

Если в ряду имеются такие значения, для которых невозможно установить разные ранги, то для них вычисляется средний ранг, который приписывается каждому значению. Такой средний ранг может быть и числом дробным. Формула вычисления коэффициента в этом случае выглядит следующим образом:

$$r = \frac{S}{\sqrt{(\frac{n(n-1)}{2} - A)(\frac{n(n-1)}{2} - B)}}\text{, где:}$$

$$A = \frac{\sum_{i=1}^{q} t_{ki}(t_{ki}-1)}{2}, \quad B = \frac{\sum_{i=1}^{q} t_{li}(t_{li}-1)}{2}$$

5. Коэффициент согласия Кендалла

В связи с использованием экспертных оценок в различных отраслях знаний, получил распространение ранговый коэффициент согласия Кендалла (коэффициент конкордации), который вычисляется для оценки степени тесноты связи между несколькими признаками по следующей формуле:

$$\omega = \frac{12 * S}{m^2 (n^3 - n)}, \text{ где}$$

m – количество строк,

n – количество столбцов,

S – сумма квадратов отклонений сумм по строкам от их общего среднего значения.

$$S = \sum_{j=1}^{n} (A_j - \overline{A})^2;$$

$$A_j = \sum_{i=1}^{m} r_{ij};$$

$$\overline{A} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij}}{n}.$$

где r_{ij} - ранг *i*-того эксперта по *j*-тому параметру.

Изменяется данный коэффициент в пределах от 0 до 1.

Для определения значимости коэффициента конкордации пользуются критерием Пирсона, который определяется по формуле:

$\chi^2 = (n-1)m \cdot \omega$ при степени свободы $\upsilon = n-1$.

По таблицам распределения χ^2 определяют критическое значение χ^2, а затем расчетный показатель χ^2 сопоставляют с критическим. Если $\chi^2_{расч} > \chi^2_{кр}$, то коэффициент конкордации – значим.

Вычислим показатель конкордации на следующем примере:

Пяти экспертам было предложено проранжировать семь факторов, влияющих на уровень качества некоторого признака (табл. 7.4):

Таблица 7.4

Результаты ранжирования

Эксперт	Факторы						
	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	1	2	6	4	7	3	5
2	1	2	7	6	3	5	4
3	7	1	6	4	2	5	3
4	3	1	5	6	4	7	2

5	1	2	6	4	5	7	3	
Сумма рангов	13	8	30	24	21	27	17	$\Sigma = 140$
Отклонения от средней суммы рангов	-7	-12	10	4	1	7	-3	$\bar{x} = 20$
Квадрат отклонения	49	144	100	16	1	49	9	$\Sigma = 368$

$$\omega = \frac{12 \cdot 368}{5^2 (7^3 - 7)} = 0{,}526$$

$$\chi^2 = 5 * 0{,}526 * (7 - 1) = 15{,}78$$

При числе степеней свободы равном 7-1=6 по таблице значений критерия Пирсона при уровне значимости 0,05 определяем $\chi^2_{кр} = 12{,}59$. Т.к. 15,78>12,59, то можно сделать вывод, что согласованность во мнениях экспертов не случайна.

6. Линейный коэффициент парной корреляции и коэффициент детерминации

Наиболее совершенным показателем степени тесноты связи является линейный коэффициент парной корреляции, предназначенный для характеристики прямолинейной связи между *x* и *y*. Его значение может изменяться от 0 до ±1. Коэффициент парной корреляции служит показателем интенсивности линейной связи. Он вычисляется по следующей формуле:

$$r_{\bar{y}_x} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}.$$

На практике чаще используется преобразованная формула Пирсона:

$$r_{\bar{y}_x} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right) \cdot \left(\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right)}}.$$

При использовании этой формулы отпадает необходимость вычислять отклонения индивидуальных значений признаков от средней величины, что позволяет более точно определять коэффициент линейной корреляции, исключив искажения при округлении средних величин и среднего квадратического отклонения.

Ниже приведен расчет коэффициента корреляции по данным показателям оборачиваемости оборотных средств и уровням рентабельности (табл. 7.5):

Таблица 7.5

Расчет коэффициента парной корреляции

Количество оборотов x	Уровень рентабельности y	xy	x^2	y^2
1	17,0	17,0	1	289,0
2	20,0	40,0	4	400,0
3	25,0	75,0	9	625,0
4	28,0	112,0	16	784,0
5	34,0	170,0	25	1156,0
Итого: 15	124,0	414,0	55	3254,0

Отсюда

$$r_{\bar{y}_x} = \frac{414 - \dfrac{15 \cdot 124}{5}}{\sqrt{\left(55 - \dfrac{15^2}{5}\right) \cdot \left(3254 - \dfrac{124^2}{5}\right)}} = \frac{414 - 372}{\sqrt{10 \cdot (3254 - 3075{,}2)}} = \frac{42}{\sqrt{1788}} = \frac{42}{42{,}3} = 0{,}99$$

Однако сама величина коэффициента корреляции еще не доказывает достоверность связи между исследуемыми факторами, поэтому его требуется проверить на значимость. Существует множество способов оценки существенности линейного коэффициента корреляции.

Значимость коэффициента корреляции можно проверить при помощи статистики t, которая имеет следующую формулу:

$$t = \frac{r_{\bar{y}_x}}{\sigma_{r_{\bar{y}_x}}} > 3 \text{ раз},$$

где:

$l_{\bar{y}_x}$ – проверяемый линейный парный коэффициент корреляции,

$\sigma_{r_{\bar{y}_x}}$ – стандартная ошибка коэффициента корреляции.

При малом числе наблюдений $\sigma_{t_{\bar{y}_x}}$ вычисляется по формуле:

$\sigma_{r_{\bar{y}_x}} = \sqrt{\dfrac{1 - r^2_{\bar{y}_x}}{n-2}}$, а при большом числе наблюдений $\sigma_{r_{\bar{y}_x}} = \dfrac{1 - r_{\bar{y}_x}}{\sqrt{n-1}}$.

Применительно к нашему примеру:

$$\sigma_{r_{\bar{y}_x}} = \sqrt{\frac{1 - 0{,}99^2}{5-2}} = \sqrt{\frac{0{,}0199}{3}} = 0{,}084, \qquad t = \frac{0{,}99}{0{,}084} = 11{,}8.$$

Следовательно, можно сделать вывод о значимости коэффициента корреляции и связь между факторами не случайна.

Квадрат коэффициента парной корреляции (r^2) называется ***коэффициентом детерминации***. Он показывает, какой удельный вес занимает факторный признак в общей совокупности всех признаков, влияющих на результативный показатель.

Коэффициент линейной корреляции используется при наличии линейной зависимости, при нелинейной зависимости может использоваться такой показатель, как эмпирическое корреляционное отношение (η). Он изменяется от нуля до единицы. Расчет эмпирического корреляционного отношения осуществляется по формуле:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$$, где

σ^2 – общая дисперсия, определяемая без поправок на смещение:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_{общ})^2}{n}$$

δ^2 – межгрупповая дисперсия, определяемая:

$$\delta^2 = \frac{\sum_{i=1}^{m}(\bar{y}_i - \bar{y}_{общ})^2}{m}$$, где $\bar{y}_i$ – среднее значение каждой i-той группы,

$\bar{y}_{общ}$ – общая средняя для всех групп,

m – число групп.

Проверка значимости эмпирического корреляционного отношения определяется при помощи статистики *t*. Стандартная ошибка η определяется как:

$$\sigma_\eta = \sqrt{\frac{1-\eta^2}{n-m}}$$, где *n* – число наблюдений, *m* – число групп.

7. *Метод анализа категоризированных переменных и критерий хи-квадрат*

В случае альтернативной вариации, т.е. в том случае если вариация исчерпывается только двумя возможностями, выявлении связи некоторого показателя с выделенным фактором рекомендуется использовать метод анализа категоризированных переменных - таблиц сопряженности признаков.

Если обозначить фактор через А, а результативный признак через Т, то таблица сопряженности будет выглядеть следующим образом (табл.7.6.):

Таблица 7.6.

Таблица сопряженности

Признак	Т	$\bar{Т}$	Итого
А	a	b	a+b
$\bar{А}$	c	d	c+d
Итого	a+c	b+d	n=a+b+c+d

Признаки считаются независимыми, т.е. между признаками связь является корреляционной, а не функциональной, если соблюдается следующее условие:

$$\frac{a}{(a+c)} \neq \frac{b}{(b+d)} \neq \frac{(a+b)}{n};$$

Для категоризированных переменных коэффициент связи (или его еще называют коэффициент ассоциации) Q, коэффициент коллигации Y, коэффициент контингенции $К_к$, дисперсия коэффициента связи σ_Q^2, дисперсия коэффициента коллигации σ_Y^2, коэффициент Пирсона χ^2 (с поправкой на непрерывность) вычисляются по следующим формулам:

$$Q = \frac{ad-bc}{ad+bc};$$

$$Y = \frac{1-\sqrt{\frac{bc}{ad}}}{1+\sqrt{\frac{bc}{ad}}};$$

$$K_k = \frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

$$\sigma_Q^2 = \frac{1}{4}(1-Q^2)(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d});$$

$$\sigma_Y^2 = \frac{1}{16}(1-Y^2)(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d});$$

$$\chi^2 = \frac{n(|ad-bc|-\frac{n}{2})^2}{(a+b)(a+c)(b+d)(c+d)}.$$

В указанной выше формуле определения χ^2 введена так называемая поправка Йетса. Она применяется, если объем выборки менее 40. При большем объеме выборки χ^2 определяется по следующей формуле:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)};$$

Для проверки значимости коэффициента связи и коэффициента коллигации может использовать статистика t. При этом расчетное значение статистики *t* получают по формулам:

$$t = \frac{Q}{\sqrt{\sigma_Q^2}};$$

$$t = \frac{Y}{\sqrt{\sigma_Y^2}}$$

Наличие связи считается достаточно доказанным, если t>2.

Достоверность результатов может быть проверена и с помощью критерия Пирсона по указанной выше формуле (с поправкой Йетса или без нее). Фактическое значение сравнивается с табличным для заданного уровня достоверности (значимости) и числа степеней свободы (а в случае рассматриваемой нами четырехпольной таблицы число степеней свободы равно 1).

Пример. Производится исследование влияния пола на отношение к работе. Для этого исследовано 100 работников, из которых 12 мужчин и 88 женщин (табл. 7.7):

Таблица 7.7.

Пример расчета четырехпольной таблицы

Пол	Отношение к работе		
	довольны	недовольны	итого
мужской женский	10 60	2 28	12 88
Всего	70	30	100

$$Q = \frac{ad - bc}{ad + bc} = \frac{10*28 - 2*60}{10*28 + 2*60} = \frac{160}{400} = 0,4$$

$$\sigma_Q^2 = \frac{1}{4}(1 - Q^2)(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}) = \frac{1}{4}(1 - 0,4^2)(\frac{1}{10} + \frac{1}{2} + \frac{1}{60} + \frac{1}{28}) = 0,137$$

$$t = \frac{Q}{\sqrt{\sigma_Q^2}} = \frac{0,4}{\sqrt{0,137}} = 1,08$$

Поскольку наличие связи считается достаточно доказанным, если t>2, то в данном примере связь статистически значимой признать нельзя.

Рассчитаем фактическое значение критерия Пирсона:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)} = \frac{100(10*28 - 2*60)^2}{70*30*12*88} = 1,154$$

Согласно таблице «Процентные точки распределения χ^2» для ν=1 и Q=5%, $\chi^2_{кр} = 3,84$.

В нашем примере $\chi^2_{факт} < \chi^2_{кр}$. Таким образом наличие связи между признаками не подтвердилось на достаточном уровне значимости, т.е. согласно проведенного исследования статистически значимой связи между полом и отношением к работе обнаружить не удалось.

8. Уравнение регрессии

Термин ***регрессия*** означает, что по известным значениям независимой переменной производят прогноз неизвестных значений зависимой переменной.

Важным этапом регрессионного анализа является определение типа функции, которая характеризует зависимость между факторами. Чаще всего используют следующие типы функций:

$\bar{y}_x = ax + b$ – линейную;

$\bar{y}_x = \frac{a}{x} + b$ – гиперболическую;

$\bar{y}_x = ax^2 + bx + c$ – параболическую;

$\bar{y}_x = x^a b$ – степенную;

$\bar{y}_x = a^x b$ – показательную.

Параметр a называется коэффициентом регрессии, b – свободным членом.

Для нахождения параметров a и b уравнения регрессии используют чаще всего метод наименьших квадратов.

Параметры уравнения прямой a и b можно определить по следующим формулам: $a = \dfrac{\Sigma xy - \dfrac{\Sigma x \cdot \Sigma y}{n}}{\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}}$, или $a = r_{\bar{y}_x} \cdot \dfrac{\sigma_y}{\sigma_x}$ $\quad b = \bar{y} - a\bar{x}$.

Коэффициент регрессии показывает, на сколько в среднем изменяется величина зависимого (результативного) признака (y) при изменении независимого (факторного) признака (x) на единицу.

9. Коэффициент эластичности

Коэффициент эластичности исчисляется по формуле:

$$Э = a \cdot \frac{\bar{x}}{\bar{y}},$$

где:

a – коэффициент регрессии,

$\bar{x}$ и $\bar{y}$ – средние значения коррелируемых переменных.

Коэффициент эластичности показывает на сколько процентов изменится величина результативного признака (y) при изменении признака-фактора (x) на один процент.

В отличии от коэффициента регрессии коэффициенты эластичности можно сравнивать.

Изучение корреляционной связи одновременно между тремя и более связанными между собой признаками носит название множественной (многофакторной) корреляции. Совокупный коэффициент корреляции характеризует совокупное влияние сразу всех выявленных факторов.

Широкое распространение пакетов прикладных программ позволяет существенно расширить диапазон использования методов корреляционно-регрессивного анализа.

При проведении корреляционно-регрессионного анализа необходимо учитывать «ложную» корреляцию. Она может возникнуть при простом совпадении течения явлений во времени (табл. 7.8).

Таблица 7.8

Пример «ложной» корреляции

Годы	Рост ребенка, x	Высота дерева, y	N_x	N_y
1995	0,8	2,0	1	1
1996	0,9	3,0	2	2
1997	1,0	4,0	3	3

Если вычислить коэффициент корреляции рангов, то он равен $\rho=1$. Часто возникает «ложная» корреляция при сопоставлении двух экономических показателей, у которых общий делитель. Примером может служить определение корреляционной связи между показателем технической вооруженности труда и производительностью труда.

Контрольные вопросы к теме 7

1. Какая связь называется функциональной, а какая связь – корреляционной?
2. Какие бывают виды связи по направлению, аналитическому выражению?
3. Как используются диаграммы рассматривания для характеристики выбора связи между явлениями?
4. Какие показатели являются мерой тесноты связи между двумя признаками?
5. Как оценить значимость коэффициентов корреляции?
6. Что характеризует коэффициент детерминации?
7. Какие методы непараметрической статистики вы знаете для изменения связи?
8. Какие показатели используют для измерения тесноты связи между качественными признаками?
9. Что характеризуют параметры уравнения регрессии?
10. Что такое «ложная» корреляция?

Тема 8. Моделирование и анализ динамики социально-экономических явлений

1. Общая характеристика рядов динамики

Ряды динамики – это статистические данные, характеризующие изменения явлений во времени. В зависимости от характера изучаемого явления различают три вида динамических рядов: моментные, интервальные и ряды средних.

Моментными рядами называются ряды статистических величин, показывающие изменения явления на определенную дату (момент времени). Примером могут служить данные счетов бухгалтерского учета, приведенные на 1-ое число каждого месяца. Данные моментных рядов суммированию не подлежат.

Интервальные ряды – это ряды статистических величин, показывающие изменения явления за определенные временные промежутки (периоды времени). Интервальные ряды можно суммировать. Например, производство продукции предприятием за 1 квартал характеризуется следующими данными: январь – 300 тыс. грн., февраль – 305 тыс. грн., март – 306 тыс. грн., всего – 911 тыс. грн.

Ряды, характеризующие изменения средних уровней изучаемого явления во времени, называются рядами **средних** величин. Например, ряд, приведенный в табл. 8.1.

Таблица 8.1

Среднемесячная заработная плата в Украине, грн.

1995 г.	1996 г.	1997 г.	1998 г.	1999 г.
73	126	143	153	178

Ряды средних величин суммировать бессмысленно.

При статистическом изучении динамики необходимо различать понятие ***тенденции*** и понятие ***колеблемости***. При изучении динамическое ряда усредненное значение показателей ряда проявляется в динамике как тенденция. В то же время отдельные уровни ряда отклоняются от основной тенденции, испытывая колебания. Для некоторых рядов наблюдаются сезонные колебания - регулярные повторения из года в год.

Тенденция динамики ряда связана с действием долгосрочных причин. Она может измениться, но только в силу изменения степени воздействия этих долгосрочных причин. Колебания отдельных уровней ряда связаны с действием краткосрочных причин. В связи с названными характеристиками тенденции и колеблемости их необходимо четко различать между собой. Смешение тенденции и колеблемости может привести к неверным выводам.

2. Показатели динамики рядов

Динамические ряды характеризуются при помощи показателей, определяющих ***характер***, ***интенсивность***, ***направление***. Наиболее простым показателем, характеризующим динамический ряд, является показатель ***среднего уровня ряда***. Он определяется в зависимости от вида динамического ряда. Если ряд моментный с равными промежутками между датами, то используют формулу средней хронологической.

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + \frac{y_n}{2}}{n-1}$$, где

$y_1, y_2, \ldots, y_n$ – фактические значения уровней моментного ряда,

n – число уровней.

Средний уровень интервального ряда с равными промежутками между датами определяют по средней арифметической простой:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$, где

$\bar{y}$ – средний уровень,

y – фактические значения уровней ряда,

n – число уровней.

Если промежутки между датами не равные, то средний уровень ряда вычисляют по средней арифметической взвешенной.

Различают цепные и базисные показатели динамического ряда. Цепные показатели получают, если каждый уровень последующий сравнивают с предыдущим. Базисные показатели получают, если каждый уровень последующий сравнивают с первоначальным уровнем, принятый за базу сравнения.

Показатель, показывающий скорость роста – ***абсолютный прирост***. Он определяется как разность между двумя сравниваемыми уровнями и показывает на сколько один уровень больше (меньше) по сравнению со сравниваемым уровнем.

$\Delta_i = y_i - y_{i-1}$ – цепной, где

Δ_i – абсолютный прирост,

y_i – текущий уровень,

y_{i-1} – предшествующий уровень ряда.

$\Delta_i = y_i - y_0$ – базисный, где y_0 – начальный уровень ряда.

Средний абсолютный прирост $(\overline{\Delta})$ рассчитывается по формуле:

$$\overline{\Delta} = \frac{\sum \Delta_i}{n}, \text{ где}$$

$\sum \Delta_i$ – сумма цепных приростов,

n – число приростов.

$$\overline{\Delta} = \frac{y_n - y_0}{n-1}, \text{ где}$$

y_n – конечный уровень ряда,

y_0 – начальный уровень ряда,

n – число уровней.

Средний абсолютный прирост показывает на сколько в среднем увеличивается (уменьшается) изучаемое явление.

Коэффициент роста ($К_Р$) – определяется как отношение двух сравниваемых уровней и показывает, во сколько раз один уровень больше (меньше) по отношению со сравниваемым.

$$K_P = \frac{y_i}{y_{i-1}} \text{ – цепной,} \qquad K_P = \frac{y_i}{y_0} \text{ – базисный.}$$

Если коэффициенты роста выражают в процентах, то они называются ***темпами роста*** ($Т_Р$):

$$T_p = K_p \cdot 100\%.$$

Темп прироста показывает на сколько процентов увеличился (уменьшился) текущий уровень по сравнению с базисным, принятым за 100%.Темп прироста можно рассчитать по данным о темпе роста. Для этого надо от темпа роста вычесть 100 или от коэффициента роста – единицу, тогда получим коэффициент прироста.

$$Т_{ПР}=Т_Р-100\%.$$

Среднегодовой коэффициент роста $(\overline{K}_p)$ определяется по средней геометрической:

$$\overline{K}_P = \sqrt[n]{K_1 \cdot K_2 \cdot \ldots \cdot K_n}, \text{ где}$$

$K_1K_2...K_n$ – произведение цепных коэффициентов роста,

n – число коэффициентов роста.

$$\overline{K}_p = \sqrt[n]{K_{баз}}, \text{ где}$$

$К_{баз}$ – базисный коэффициент роста за весь период,

n – число лет.

$K_{баз} = K_1 \cdot K_2 \cdot \ldots \cdot K_n$ или $K_{баз} = \frac{y_n}{y_0}$, где y_n и y_0 – конечный и первоначальный уровни ряда.

Средний коэффициент роста показывает, во сколько раз в среднем за рассмотренный период изменились уровни динамического ряда.

Средний темп роста $\overline{T}_P$ определяется по следующей формуле: $\overline{T}_P = \overline{K}_P \cdot 100\%$.

Абсолютное значение одного процента прироста (А%):

$$A\% = \frac{\Delta_i}{T_{np}} = \frac{y_{i-1}}{100}$$, где

Δ_i – цепной абсолютный прирост,

$T_{пр}$ – темп прироста в процентах,

y_{i-1} – предшествующий уровень ряда.

Данный показатель позволяет правильно оценить темп прироста при сопоставлении его с показателем абсолютного прироста.

Пункты роста – это разность базисных темпов роста (в %):

$$\frac{y_i}{y_0} \cdot 100\% - \frac{y_{i-1}}{y_0} \cdot 100\% = \frac{y_i - y_{i-1}}{y_0} \cdot 100\%$$, где

y_i – текущий уровень ряда,

y_{i-1} – предшествующий уровень ряда,

y_0 – первоначальный уровень ряда, принятый за базу сравнения.

Цепные и базисные характеристики динамики находятся в тесной взаимосвязи. Так, сумма абсолютных приростов равна конечному базисному:

$$\sum_{t=1}^{n} \Delta_t = \sum_{t=1}^{n} (y_t - y_{t-1}) = y_n - y_0$$

Произведение цепных индексов роста равно конечному базисному:

$$k_1 k_2 ... k_n = \prod_{1}^{n} k_t = K_n = \frac{y_n}{y_0}$$

Рассмотрим расчет вышеуказанных показателей по данным, представленным в табл. 8.2.

Таблица 8.2

Динамика внешнеторгового оборота страны, млн. дол.

Год	Внешний торговый оборот, млн. дол.	Абсолютный прирост, млн. дол.		Темп роста, %		Темп прироста, %		Абсолютное значение 1% прироста, млн. дол.	Пункты роста, %
		базисный	цепной	базисный	цепной	базисный	цепной		
1994	1040,2	–	–	–	–	–	–	–	–

1995	1120,3	80,1	80,1	107,7	107,7	7,7	7,7	10,4	7,7
1996	2654,0	1613,8	1533,7	255,2	236,9	155,2	136,9	11,2	147,5
1997	2731,8	1691,6	77,8	262,7	103,0	162,7	3,0	26,5	7,5

$$\bar{y} = \frac{7546{,}3}{4} = 1886{,}6 \text{ млн. дол.}$$

$$\bar{\Delta} = \frac{1691{,}6}{3} = 563{,}87 \text{ млн. дол.}$$

В течение 1994-1997 гг. в среднем внешнеторговый оборот увеличился на 563, 87 млн. дол.

$$\overline{K_P} = \sqrt[3]{262{,}7} = 1{,}375$$

$$\overline{T_P} = 1{,}375 \cdot 100\% = 137{,}5\%$$

$$\overline{T_{\text{пр}}} = 37{,}5\%$$

В среднем внешнеторговый оборот увеличивался ежегодно на 37,5% по стране.

Для сравнения тенденции развития нескольких явлений событий, отраженных в соответствующих динамических рядах, применяется способ ***приведения рядов к одному основанию***. Для этого показатели каждого ряда исчисляют в виде базисных темпов роста, как это сделано в табл. 8.3.

Таблица 8.3

Внешняя торговля

Квартал	Экспорт, млн. $	Импорт, млн. $	В процентах к 1 кварталу	
			экспорт	импорт
1	61,1	35,0	100,0	100,0
2	61,0	53,3	99,9	152,3
3	57,2	57,1	93,7	163,2
4	77,9	39,7	127,5	113,5

По данным табл. 8.3 видна тенденция роста импорта по отрасли.

Для выявления и характеристики основной тенденции ряда используют различные методы ***сглаживания*** и ***аналитического выравнивания*** динамических рядов.

Суть сглаживания состоит в укрупнении временных интервалов и заменой первичного ряда рядом средних величин. Это позволяет взаимоуравновесить (сгладить) колебания и более четко выразить тенденцию. Для выполнения сглаживания применяют ступенчатые и скользящие средние. Ступенчатые средние получаются путем вычисления средних с определенным шагом, например, первых трех членов ряда, 4-6, 7-9 и т.д. Скользящая средняя вычисляется следующим образом: по формуле средней арифметической находится значение средней для *l* первых нечетных членов ряда. Каждое последующее значение вычисляется для *l* следующих членов ряда, среди которых убирается первый из предыдущего расчета средней и добавляется следующий. Метод скользящей средней устраняет колебания и четко показывает тренд ряда. В статистической практике могут также использоваться двойные сглаживания на основе скользящей средней, применяются взвешенные скользящие средние.

Однократное аналитическое выравнивание неполно освобождает параметры тренда от влияния колеблемости, и при сильных колебаниях они могут быть сильно искажены, что в приведенном примере случилось с параболой. Для дальнейшего исключения искажающего влияния колебаний на параметры тренда следует применить ***метод многократного скользящего выравнивания***. Этот прием состоит в том, что параметры тренда вычисляются не сразу по всему ряду, а скользящим методом, сначала за первые m периодов времени или моментов, затем за период от 2-го до $m + 1$, от 3-го до $(m + 2)$-го уровня и т.п.

Применение методики скользящего многократного выравнивания возможно только при достаточно большом числе уровней ряда, как правило, 15 и более. Смысл многократного скользящего выравнивания в том, что при последовательных сдвигах базы расчета параметров на концах ее и в середине окажутся разные уровни с разными по знаку и величине отклонениями от тренда. Поэтому при одних сдвигах базы параметры будут завышаться, при других - занижаться, а при последующем усреднении значений параметров по всем сдвигам базы расчета произойдет дальнейшее взаимное погашение искажений параметров тренда колебаниями уровней.

Многократное скользящее выравнивание не только позволяет получить более точную и надежную оценку параметров тренда, но и осуществить контроль правильности выбора типа уравнения тренда. Если окажется, что ведущий параметр тренда, его константа при расчете по скользящим базам не беспорядочно колеблется, а систематически изменяет свою величину существенным образом, значит, тип тренда был выбран неверно, данный параметр константой не является.

Аналитическое выравнивание состоит в нахождении такой формы функции, ординаты которой были наиболее близки к фактическим значениям ряда. Именно в результате аналитического сглаживания выявляется функция, наиболее близко описывающая тенденцию ряда.

3. Методы выявления тенденций.

При выявлении типа тенденции прежде всего необходимо установить:

- были ли условия развития объекта достаточно однородными в изучаемый период;
- какой характер действия основных факторов развития;
- не произошло ли существенное качественное изменение условий развития объекта внутри изучаемого периода времени.

Например, при техническом перевооружении предприятия, скорее всего общая тенденция его развития не сохранится. Поэтому в подобных случаях прибегают к «периодизации» ряда, т.е. его дроблению на отдельные подпериоды: до реконструкции, во время таковой (если она длительна) и после освоения новой технологии. Чем крупнее изучаемая система, чем больше факторов влияют на динамику изучаемого признака, тем реже возможны резкие, скачкообразные изменения в ряду динамики.

При выборе типа трендовой функции используют анализ цепных характеристик интенсивности динамики ряда, так, как это показано в табл. 8.5.

Таблица 8.5

Соответствие типа выбранной функции цепным характеристикам динамического ряда

Цепные характеристики интенсивности динамики ряда	Рекомендуемый тип трендовой функции
Цепные абсолютные приросты относительно стабильны и не имеют четкой тенденции к росту или уменьшению.	Линейная
Цепные темпы роста относительно стабильны и не имеют четкой тенденции к росту или уменьшению.	Парабола
Цепные темпы приросты относительно стабильны и не имеют четкой тенденции к росту или уменьшению.	Экспоненциальная
Прочие варианты	Иные функции.

В статистическом прогнозировании используются следующие основные типы уравнений тренда:

Линейная форма тренда:

$$\bar{y} = a + bt$$

где:

$\bar{y}$ - уровни, освобожденные от колебаний, выровненные по прямой;

a - начальный уровень тренда в момент или период, принятый за начало отсчета времени t;

b - среднегодовой абсолютный прирост (среднее изменение за единицу времени); константа тренда.

Параболическая форма тренда:

$$\bar{y} = a + bt + ct^2$$

где: с - квадратический параметр, равный половине ускорения; константа параболического тренда. Под ***ускорением*** понимается разность между абсолютным изменением за данный период и абсолютным изменением за предыдущий период:

$\Delta_i^{'} = \Delta_i - \Delta_{i-1}$ Остальные обозначения прежние.

Экспоненциальная форма тренда:

$$\bar{y} = ak^t$$

где k - темп изменения в разах; константа тренда.

Логарифмическая форма тренда:

$$\bar{y} = a + b\log t$$

Логарифмическая парабола:

$$y = ab^t c^{t^2}$$

Тренд в форме степенной кривой:

$$\bar{y} = at^b$$

где b - константа тренда.

Гиперболическая форма тренда:

$$\bar{y} = a + \frac{b}{t}$$

Логистическая форма тренда:

$$\bar{y} = \frac{y_{\max} - y_{\min}}{e^{a+bt} + 1} + y_{\min}$$

Каждый тип уравнения тренда при анализе и прогнозировании развития экономических явлений выражает те или иные качественные свойства развития. Основные рекомендации по применению той или иной функции тренда, в зависимости от особенностей развития тех или иных явлений, приведены в табл. 8.6.

После теоретического исследования особенностей разных форм тренда необходимо обратиться к фактическому ряду динамики, так

Таблица 8.6.

Соответствие типа тренда основным качественным характеристикам развития явления

Тип трендовой функции	Особенности применения трендовой функции в зависимости от качественных характеристик развития явления	Примеры применения
Линейный	Отражает тенденцию изменений при действии множества разнообразных факторов, изменяющихся различным образом по разным закономерностям. Равнодействующая этих факторов при взаимном погашении особенностей отдельных факторов (ускорение, замедление, нелинейность) часто выражается в примерно постоянной абсолютной скорости изменения, т.е. в прямолинейном тренде.	Тенденции динамики урожайности для масштаба области, республики, крупного региона, страны в целом.
Параболический	Выражает ускоренное или замедленное изменение уровней ряда с постоянным ускорением. Параболическая форма тренда с отрицательным ускорением ($c < 0$) приводит со временем не только к приостановке роста уровня, но и к его снижению со все большей скоростью. Такой характер развития может быть свойствен производству устаревшей продукции, ликвидируемой отрасли сельского хозяйства на предприятии (ферме) и т.п.	Ускоренный характер развития можно ожидать при наличии важных факторов прогрессивного развития (прогрессирующее поступление нового высокопроизводительного оборудования, увеличение среднесуточного прироста живого веса поросят с возрастом и т.п.). Ускоренное возрастание может также происходить в период после снятия каких-то сдерживающих развитие преград - ограничений в распределении дохода, в уровне оплаты труда, при повышении цены реализации на дефицитную продукцию.

Продолжение табл. 8.6

Экспоненциальный с показателем степе-	Выражает тенденцию ускоренного и все более ускоряющегося возрастания уровней. При росте по экспоненте абсолютный	Такой характер свойствен размножению организмов при отсутствии ограничения со стороны среды: кормов,

ни больше 1.	прирост пропорционален достигнутому уровню. Рост любого объекта по экспоненциальному закону может продолжаться только небольшой исторический период времени, ибо ресурсы для любого процесса развития всегда встретят ограничения.	пространства, хищников, болезней.
Экспоненциальный с показателем степени меньше 1.	Экспоненциальный тренд означает тенденцию постоянно все более замедляющегося снижения уровней динамического ряда.	Тенденция может быть присуща динамике трудоемкости продукции, удельных затрат топлива, металла на единицу полезного эффекта (на 1 кВт ч, на 1 $м^2$ жилой площади и т.д.) при технологическом прогрессе.
Логарифмический	Пригоден для отображения тенденции замедляющегося роста уровней при отсутствии предельного возможного значения.	Пригоден для отображения роста спортивных достижений, роста производительности агрегата по мере его освоения и совершенствования, повышения продуктивности скота или вообще эффективности системы при ее совершенствовании без качественных, коренных преобразований.
Степенной	Пригоден для отображения изменений с разной мерой пропорциональности изменений во времени.	Жестким условием является обязательное прохождение через начало координат.

продолжение табл. 8.6

Гиперболический	Подходит для отображения тенденции, процессов, ограниченных предельным значением уровня.	Предельный коэффициент полезного действия двигателя, предел 100%-ной грамотности населения и т.п.
Логистический	Отражает тенденцию от ускоряющегося роста к равномерному (вогнутость) и от равномерного роста посреди периода к	Используется для описания процесса насыщения потребительского рынка каким-то новым товаром: сначала мед-

	замедляющемуся (выпуклость). Подходит для отображения развития в течение длительного периода, проходящего все фазы.	ленный, но все ускоряющийся рост, затем рост равномерный.

как не всегда можно надежно установить, какой должна быть форма тренда из теоретических соображений. По фактическому динамическому ряду тип тренда устанавливают на основе графического изображения, на основе статистической проверки гипотезы о постоянстве параметра тренда и другими способами.

После логического анализа динамического ряда и выбора гипотезы о соответствующей форме тренда, необходимо подтвердить реальность (или возможность) использования данной гипотезы. Наиболее обоснованным приемом выявления тренда является проверка статистической гипотезы о постоянстве того или иного показателя динамики. Рассмотрим этот прием по данным табл. 8.7.

В первую очередь проверяется гипотеза о наиболее простой - линейной форме уравнения тренда, т.е. о несущественности различий цепных абсолютных изменений. По данным, приведенным в табл. 8.7. имеем 12 абсолютных изменений скользящей средней, которая хотя и сгладила сильные колебания уровней ряда, но как видим, ее абсолютные изменения далеко не одинаковы. Разбиваем эти 12 цепных приростов на два подпериода: по 6 приростов в каждом, и для каждого подпериода вычисляем среднюю $\overline{\Delta}_k$, среднее квадратическое отклонение как оценку генерального среднего квадратического отклонения с учетом потери одной степени свободы вариации, S

$$S_{\Delta k} = \sqrt{\frac{\sum_{j=1}^{6}(\Delta_j - \overline{\Delta})^2}{6-1}}$$

и среднюю ошибку среднего изменения $m_{\Delta k}$:

Таблица 8.7

Проверка гипотезы о линейном тренде индекса цен (1990 г. = 100%)

Годы	Уровни y_i	Средние скользящие за 5 лет, $\overline{y}_i$	Цепные приросты, $\overline{y}_i - \overline{y}_{i-1}$
1979	105		
1980	111		
1981	110	110,0	
1982	106	113,8	+3,8
1983	118	114,2	+0,4
1984	124	110,6	-3,6
1985	113	107,6	-3,0
1986	92	105,8	-1,8

1987	91	103,6	-2,2
1988	109	101,0	-2,6
1989	113	101,4	+0,4
1990	100	101,4	0,0
1991	94	98,0	-3,4
1992	91	95,8	-2,2
1993	92	97,0	+1,2
1994	102		
1995	106		

$$m_{\Delta k} = \frac{S_{\Delta k}}{\sqrt{6}}$$

$$\overline{\Delta}_1 = \frac{3{,}8 + 0{,}4 - 3{,}6 - 3{,}0 - 1{,}8 - 2{,}2}{6} = -1{,}07$$

$$S_1 = \sqrt{\frac{37{,}82}{6-1}} = 2{,}75$$

$$m_1 = S_1 : \sqrt{6} = 1{,}123$$

$$\overline{\Delta}_2 = \frac{-6{,}6}{6} = -1{,}1$$

$$S_2 = \sqrt{\frac{17{,}50}{6-1}} = 1{,}87$$

$$m_2 = S_2 : \sqrt{6} = 0{,}763$$

Средняя случайная ошибка разностей двух выборочных средних оценок определяется как корень квадратный из суммы квадратов ошибок каждой из средних, т. е.

$$m_{d(I-II)} = \sqrt{m_1^2 + m_2^2} = \sqrt{1{,}123^2 + 0{,}763^2} = 1{,}356$$

Критерий Стьюдента для существенности различия двух среднегодовых приростов (изменений) составит:

$$t = \frac{\overline{\Delta}_1 - \overline{\Delta}_2}{m_{d(I-II)}} = \frac{-1{,}07 - (-1{,}10)}{1{,}356} = 0{,}022$$

Критическое значение критерия при уровне значимости 0,05 и при (6-1) + (6-1) = 10 степенях свободы равно 2,23. Фактическое значение много меньше. Следовательно вероятность того, что различие среднегодовых приростов в разные подпериоды случайно, превышает 0,05 и гипотеза о равенстве приростов не отклоняется. А значит, тенденцию динамики на всем протяжении ряда можно считать линейной.

4. Методика измерения параметров тренда

После установления типа тренда вычисляются оптимальные значения параметров тренда исходя из фактических уровней. Для этого обычно используют метод наименьших квадратов (МНК). В данном случае оптимизация состоит в минимизации суммы квадратов отклонений фактических уровней ряда от тренда. Для каждого типа тренда МНК дает систему нормальных уравнений, решая которую вычисляют параметры тренда.

Для *линейного тренда* нормальные уравнения МНК имеют вид:

$$\begin{cases} na + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} y_i t_i \end{cases}$$

где:

у, - уровни исходного ряда динамики;

I - номера периодов или моментов времени;

n- число уровней ряда.

Систему можно упростить, перенеся начало отсчета времени t_i в середину ряда. Тогда $\sum t_i$ (а также суммы всех нечетных степеней t_i) будет равна нулю и система приобретет вид:

$$\begin{cases} na = \sum_{i=1}^{n} y_i \\ b\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} y_i t_i \end{cases}$$

откуда:

$$a = \bar{y};$$

$$b = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2}$$

Значения t, расположенные выше средины, имеют отрицательные значения, а расположенные ниже средины – положительные. При нечетном числе членов ряда изменения t даются с интервалом 1, при четном – с интервалом 2, например, -5, -3, -1, 1, 3, 5.

Значения $\sum t^2$ определяется по следующим формулам:

для нечетного числа членов ряда:

$$\sum t^2 = \frac{n(n^2-1)}{12}$$

для четного числа членов ряда:

$$\sum t^2 = \frac{n(n^2-1)}{3}$$

Порядок расчета параметров линейной функции рассмотрим на примере динамического ряда добычи нефти в регионе (табл. 8.8).

Таблица 8.8.

Динамика добычи нефти

Год	y_t, млн. т	t	y_tt	Y_t=74,5+3,8t

1990	63,5	-3	-190,5	63,1
1991	66,8	-2	-133,6	66,9
1992	71,0	-1	-71,0	70,7
1993	74,3	0	0	74,5
1994	76,9	+1	76,9	78,3
1195	82,2	+2	164,4	82,1
1196	86,8	+3	260,4	85,9
Всего	521,5	0	106,6	521,5

По данным табл. 8.8. имеем:

$$a = \overline{y} = \frac{521{,}5}{7} = 74{,}5;$$

$$b = \frac{\sum\limits_{i=1}^{n} y_i t_i}{\sum t_i^2} = \frac{106{,}6}{28} = 3{,}8$$

Уравнение линейного тренда выглядит следующим образом:

$$\widehat{y} = 74{,}5 + 3{,}8t$$

Нормальные уравнения МНК для параболы 2-го порядка имеют вид:

$$\begin{cases} na + b \cdot \sum\limits_{i=1}^{k} t_i + c \cdot \sum\limits_{i=1}^{k} t_i^2 = \sum\limits_{i=1}^{k} y_i \\ a\sum\limits_{i=1}^{k} t_i + b \cdot \sum\limits_{i=1}^{k} t_i^2 + c \cdot \sum\limits_{i=1}^{k} t_i^3 = \sum\limits_{i=1}^{k} y_i t_i \\ a\sum\limits_{i=1}^{k} t_i^2 + b \cdot \sum\limits_{i=1}^{k} t_i^3 + c \cdot \sum\limits_{i=1}^{k} t_i^4 = \sum\limits_{i=1}^{k} y_i t_i^2 \end{cases}$$

После переноса начала отсчета t в середину ряда имеем:

$$\begin{cases} na + c \cdot \sum\limits_{t=1}^{n} t_i^2 = \sum\limits_{i=1}^{n} y_i \\ b \cdot \sum\limits_{i=1}^{n} t_i^2 = \sum\limits_{i=1}^{n} y_i t_i \\ a\sum\limits_{i=1}^{n} t_i^2 + c \cdot \sum\limits_{t=1}^{n} t_i^4 = \sum\limits_{i=1}^{n} y_i t_i^2 \end{cases}$$

Отсюда:

$$b = \frac{\sum\limits_{i=1}^{n} y_i t_i}{\sum\limits_{i=1}^{n} t_i^2};$$

$$\begin{cases} na + c \cdot \sum\limits_{t=1}^{n} t_i^2 = \sum\limits_{i=1}^{n} y_i \\ a \cdot \sum\limits_{i=1}^{n} t_i^2 + c \cdot \sum\limits_{t=1}^{n} t_i^4 = \sum\limits_{i=1}^{n} y_i t_i^2 \end{cases}$$

Разделим первое уравнение на n

$$\begin{cases} a + \dfrac{c}{n} \cdot \sum\limits_{t=1}^{n} t_i^2 = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} y_i \\ a \cdot \sum\limits_{i=1}^{n} t_i^2 + c \cdot \sum\limits_{t=1}^{n} t_i^4 = \sum\limits_{i=1}^{n} y_i t_i^2 \end{cases}$$

Умножим первое уравнение на $\sum\limits_{i=1}^{n} t_i^2$

$$\begin{cases} a \cdot \sum\limits_{i=1}^{n} t_i^2 + \dfrac{c}{n} \cdot \left(\sum\limits_{i=1}^{n} t_i^2 \right)^2 = \dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} t_i^2}{n} \\ a \cdot \sum\limits_{i=1}^{n} t_i^2 + c \cdot \sum\limits_{t=1}^{n} t_i^4 = \sum\limits_{i=1}^{n} y_i t_i^2 \end{cases}$$

Вычтем из второго уравнения первое:

$$\frac{c}{n} \cdot \left(\sum_{i=1}^{n} t_i^2 \right)^2 - c \cdot \sum_{t=1}^{n} t_i^4 = \frac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} t_i^2}{n} - \sum_{i=1}^{n} y_i t_i^2$$

$$c \cdot \left(\frac{\left(\sum\limits_{i=1}^{n} t_i^2 \right)^2}{n} - \sum_{t=1}^{n} t_i^4 \right) = \frac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} t_i^2}{n} - \sum_{i=1}^{n} y_i t_i^2$$

$$c = \frac{\dfrac{\sum\limits_{i=1}^{n} y_i \cdot \sum\limits_{i=1}^{n} t_i^2}{n} - \sum\limits_{i=1}^{n} y_i t_i^2}{\dfrac{\left(\sum\limits_{i=1}^{n} t_i^2 \right)^2}{n} - \sum\limits_{t=1}^{n} t_i^4}$$

$$a = \frac{\sum\limits_{i=1}^{n} y_i - c \cdot \sum\limits_{i=1}^{n} t_i^2}{n}$$

Суммы включают значения t от –(n-1) до +(n-2):2, при этом сумма биквадратов может быть вычислена по формуле:

$$\sum t^4 = \frac{3n^5 - 10n^3 + 7n}{240}$$

Нормальные уравнения МНК для *экспоненты* имеют следующий вид:

$$\begin{cases} n \ln a + \ln k \cdot \sum\limits_{t=1}^{k} t_i = \sum\limits_{i=1}^{k} y_i \\ \ln a \cdot \sum\limits_{i=1}^{k} t_i + \ln k \cdot \sum\limits_{t=1}^{k} t_i^2 = \sum\limits_{i=1}^{k} \ln y_i t_i \end{cases}$$

После переноса начала отсчета t, в середину ряда получим:

$$n\ln a = \sum \ln y\text{, откуда } \ln a = \overline{\ln y}$$

$$\ln k \sum t_i^2 = \sum (\ln y)_i t_i\text{, откуда } \ln k = \frac{\sum (\ln y_i) t_i}{\sum t_i^2}$$

Уравнение логарифмической параболы имеет вид:

$$y_t = ab^t c^{t^2}$$

Для определения параметров этого уравнения поместим t=1 в начало ряда и будем его изменять по уровням ряда с шагом равным 1.

Параметры a, b и c уравнения определяются путем потенцирования соответствующих значений, определенных по формулам:

$$\alpha = \frac{\sum \alpha_t \log y_t}{M};\ \beta = \frac{\sum \beta_t \log y_t}{M};\ \gamma = \frac{\sum \gamma_t \log y_t}{M}$$

Параметры α_t, β_t, γ_t и M определяются по специальным таблицам в зависимости от числа уровней исходного динамического ряда [см. Четыркин Е.М. Статистические методы прогнозирования. М.: «Статистика». - 1975. - С. 178-179].

Произведем расчет трендового уравнения по данным о производстве электроэнергии в Украине в 2002 г. [исходные данные приводятся на основании: Динаміка виробництва найважливіших видів промислової продукції // Економіст. – 2003. - №9. – С. 48-55]. Результаты расчетов сведены в табл. 8.9.

Таблица 8.9

Расчет параметров трендового уравнения логарифмической параболы по данным о производстве электроэнергии в Украине в 2002 г. (млрд. квт. час.)

Месяц	y_t	t	Табличные значения			$\alpha_t \log y_t$	$\gamma_t \log y_t$	$\beta_t \log y_t$	Трендовые значения
			α_t	β_t	γ_t				
январь	17419	1	3003	-869	55	12735,8	-3685,4	233,3	17738
февраль	14693	2	1911	-451	25	7963,3	-1879,4	104,2	15584
март	15404	3	1001	-111	1	4191,8	-464,8	4,2	14061
апрель	13741	4	273	151	-17	1129,7	624,8	-70,3	13030
май	12195	5	-273	335	-29	-1115,5	1368,9	-118,5	12401
июнь	11393	6	-637	441	-35	-2584,1	1789,0	-142,0	12120
июль	12574	7	-819	469	-35	-3357,5	1922,7	-143,5	12166
август	12382	8	-819	419	-29	-3352,0	1714,9	-118,7	12541
сентябрь	12663	9	-637	291	-17	-2613,3	1193,8	-69,7	13277
октябрь	15074	10	-273	85	1	-1140,7	355,1	4,2	14436
ноябрь	15680	11	273	-199	25	1145,3	-834,9	104,9	16120
декабрь	18914	12	1001	-561	55	4281,1	-2399,3	235,2	18486
Итого						17284,0	-294,6	23,2	

Табличное значение M=4004. Следовательно $\alpha = \frac{17284}{4004} = 4{,}317$;

$\beta = \frac{-294{,}6}{4004} = -0{,}074$; $\gamma = \frac{23{,}2}{4004} = 0{,}006$. Потенцирование дало следующие результаты: a=20733,867; b=0,844; c=1,013. Искомое уравнение логарифмической параболы имеет вид:

$$y = 20733{,}867 * 0{,}844^{t} * 1{,}013^{t^2};$$

Соответствующие трендовые значения, рассчитанные с помощью полученного уравнения приведены в последней колонке табл. 8.9.

5. Методика изучения и показатели колеблемости

Выделяют три основных типа колебаний статистических показателей: ***пилообразную*** или ***маятниковую*** колеблемость, ***циклическую долгопериодическую*** и ***случайно распределенную*** во времени колеблемость. Колеблемостью называются отклонения уровней динамического ряда от тренда. свойства и отличия друг от друга хорошо видны при графическом изображении рис. 8.1-8.3.

Пилообразная или ***маятниковая колеблемость*** состоит в попеременных отклонениях уровней от тренда в одну и в другую сторону (рис. 8.1). Для циклической долгопериодической колеблемости свойственна редкая смена знаков отклонений от тренда и кумулятивный (накапливающийся) эффект отклонений одного знака, который может тяжело отражаться на экономике (рис. 8.2). Случайно распределенная во времени колеблемость - нерегулярная, хаотическая. Она может возникать при наложении множества колебаний с разными по длительности циклами либо хаотической колеблемости главной причины существования колебаний (рис. 8.3).

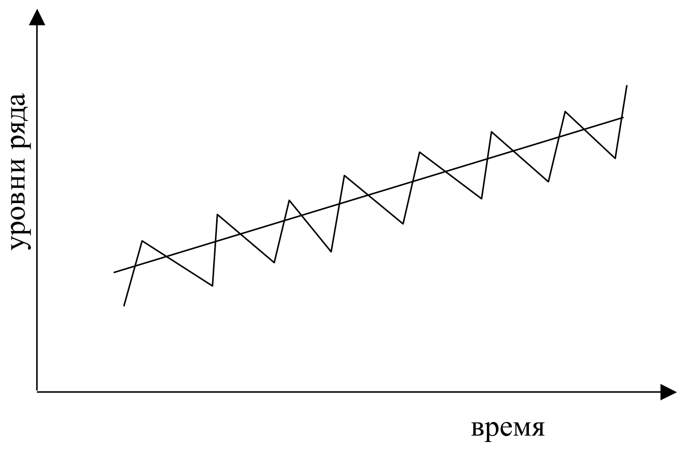


Рис. 5.1. Пилообразная или маятниковая колеблемость

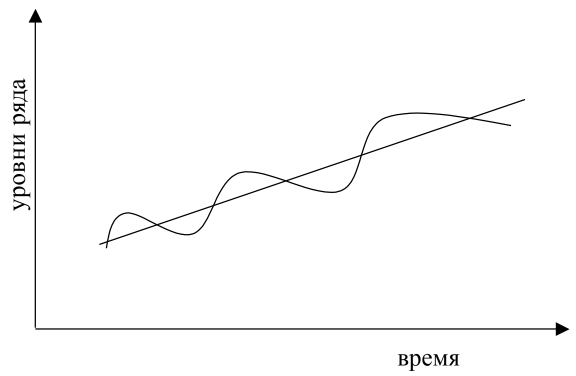


Рис. 5.2. Долгопериодическая циклическая колеблемость

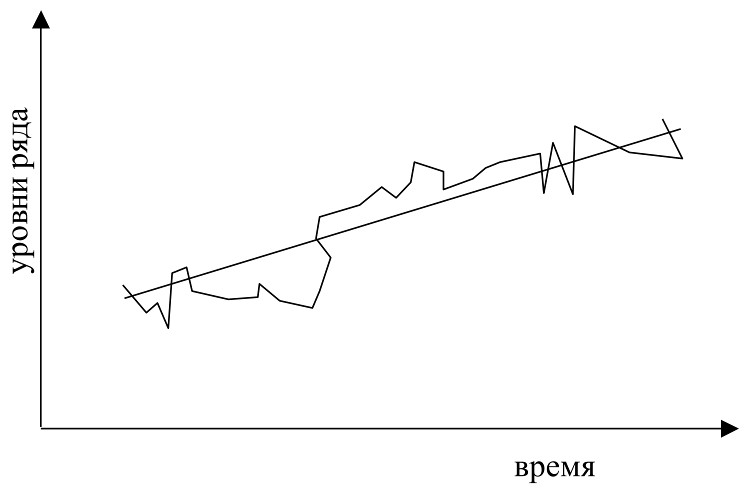


Рис. 5.3. Случайно распределенная во времени колеблемость

Показателями силы колебаний уровней являются: амплитуда отклонений уровней отдельных периодов или моментов от тренда (по модулю), среднее абсолютное отклонение уровней от тренда (по модулю), среднее квадратическое отклонение уровней от тренда. К относительным показателям колеблемости относятся коэффициент колеблемости (аналог коэффициента вариации) и относительное линейное отклонение от тренда.

Для определения типа колебаний используют следующие методы: ***графический, метод «поворотных точек» М. Кендалла, вычисление коэффициентов автокорреляции отклонений от тренда***.

Вариация в пространстве и колеблемость во времени принципиально различны. Эти различия сводятся к следующим:

- вызываются разными факторами, из которых одни вызывают влияние на тренд, а другие на колеблемость;
- значения варьирующего признака в пространственной совокупности можно считать в основном не зависимыми друг от друга, напротив, уровни динамического ряда, как правило, являются зависимыми: это показатели разви-

вающегося процесса, каждая стадия которого связана с предыдущими состояниями;

- вариация в пространственной совокупности измеряется отклонениями индивидуальных значений признака от среднего значения, а колеблемость уровней динамического ряда измеряется не их отличиями от среднего уровня (эти отличия включают и тренд, и колебания), а отклонениями уровней от тренда.

Для выявления типов колебаний применяют следующие методы:

- графический;
- поворотных точек Кендалла;
- автокорреляции отклонений от тренда.

Графический метод состоит в визуальном распознавании типа колеблемости согласно построенного графика.

Метод поворотных точек Кендалла состоит в подсчете отклонений от тренда u_i. Отклонение, либо большее по алгебраической величине, либо меньшее двух соседних, отмечается точкой. При маятниковой колеблемости все отклонения, кроме двух крайних, будут «поворотными», следовательно, их число составит n-2. При долгопериодических циклах на цикл приходятся один минимум и один максимум, а общее число точек составит *2(n:l),* где *l* - длительность цикла. При случайно распределенной во времени колеблемости число поворотных точек в среднем составит: 2/3 *(n*-2).

Автокорреляция - это корреляция между уровнями ряда или отклонениями от тренда, взятыми со сдвигом во времени: на 1 период (год), на 2, на 3 и т.д., поэтому говорят о коэффициентах автокорреляции разных **порядков:** первого, второго и т. д.

Одна из основных формул для расчета коэффициента автокорреляции отклонений от тренда имеет вид:

$$r_1^{ul} = \frac{\sum_{i=1}^{n-1} u_i u_{i+1}}{\frac{u_1^2}{2} + \sum_{i=2}^{n-1} u_i^2 + \frac{u_n^2}{2}}$$

В данной формуле u – отклонения от тренда, т.е. $u_i = y_i - \hat{y}_i$

При маятниковой колеблемости все произведения в числителе будут отрицательными величинами, и коэффициент автокорреляции первого порядка будет близок к -1. При долгопериодических циклах будут преобладать положительные произведения соседних отклонений, а смена знака происходит лишь дважды за цикл. Чем длиннее цикл, тем больше перевес положительных произведений в числителе, и коэффициент автокорреляции первого порядка ближе к +1. При случайно распределенной во времени колеблемости знаки отклонений чередуются хаотически, а число положительных произведений близко к числу отрицательных, ввиду чего коэффициент автокорреляции близок к нулю.

6. *Измерение устойчивости в динамике*

По отношению к статистическому изучению динамики рассматривается два аспекта понятия «устойчивость»: устойчивость как категория, противоположная колеблемости; устойчивость направленности изменений, т.е. устойчивость тенденции.

В первом понимании показатель устойчивости, который может быть только относительным, должен изменяться от нуля до единицы (100%). Это разность между единицей и относительным показателем колеблемости. Если коэффициент колеблемости составил 9,0%, то коэффициент устойчивости равен 100% - 9,0% = 91,0%. Этот показатель характеризует близость фактических уровней к тренду и не зависит от характера последнего. Слабая колеблемость и высокая устойчивость уровней в данном смысле могут существовать даже при полном застое в развитии, когда тренд выражен горизонтальной прямой.

Устойчивость во втором смысле характеризует процесс направленного изменения уровней. С этой точки зрения полной устойчивостью направленного изменения уровней динамического ряда следует считать такое изменение, в процессе которого каждый следующий уровень либо выше всех предшествующих (устойчивый рост), либо ниже всех предшествующих (устойчивое снижение). Всякое нарушение строго ранжированной последовательности уровней свидетельствует о неполной устойчивости изменений.

Из определения понятия устойчивости тенденции вытекает и метод построения ее показателя. В качестве показателя устойчивости можно использовать *коэффициент корреляции рангов* Ч. Спирмэна:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\Delta_i^2}{n^3 - n}$$

n - число уровней;

Δ_i – разность рангов уровней и номеров периодов времени.

При полном совпадении рангов уровней, начиная с наименьшего, и номеров периодов (моментов) времени по их хронологическому порядку коэффициент корреляции рангов равен +1. Это значение соответствует случаю полной устойчивости возрастания уровней. При полной противоположности рангов уровней рангам лет коэффициент Спирмэна равен -1, что означает полную устойчивость процесса сокращения уровней. При хаотическом чередовании рангов уровней коэффициент близок к нулю, это означает неустойчивость какой-либо тенденции. Приведем расчет коэффициента корреляции Спирмэна по данным о динамике индекса цен (табл. 8.7) в табл. 8.10.

Таблица 8.10

Расчет коэффициентов корреляции рангов Спирмена

Годы	Уровни, y_i	Ранг лет, Px	Ранг уровней, Py	*Px-Py*	*(Px-Py)*^{2}
1979	105	1	8	7	49
1980	111	2	13	11	121
1981	110	3	12	9	81

1982	106	4	9,5	5,5	30,26
1983	118	5	16	11	121
1984	124	6	17	11	11
1985	113	7	14,5	7,5	56,26
1986	92	8	3,5	4,5	20,26
1987	91	9	1,5	7,5	56,26
1988	109	10	11	1	1
1989	113	11	14,5	3,5	12,26
1990	100	12	6	6	36
1991	94	13	5	8	64
1992	91	14	1,5	12,5	156,26
1993	92	15	3,5	11,5	132,26
1994	102	16	7	9	81
1995	106	17	9,5	7,5	56,26
$\sum$	1777	-	-	-	1141

Ввиду наличия трех пар «связанных рангов» применяем формулу:

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(P_x - P_y)^2 - A}{\sqrt{(n^3 - n)(n^3 - n - A)}} = 1 - \frac{6*1141 - 2}{\sqrt{(17^3 - 17)(17^3 - 17 - 2)}}$$

Отрицательное значение коэффициента корреляции рангов Спирмена указывает на наличие тенденции снижения уровней, причем устойчивость этой тенденции ниже средней. При этом следует иметь в виду, что даже при 100%-ной устойчивости тенденции в ряду динамики может быть колеблемость уровней, и коэффициент их устойчивости будет ниже 100%. При слабой колеблемости, но еще более слабой тенденции, напротив, возможен высокий коэффициент устойчивости уровней, но близкий к нулю коэффициент устойчивости тренда. В целом же оба показателя связаны, конечно, прямой зависимостью: чаще всего большая устойчивость уровней наблюдается одновременно с большей устойчивостью тренда.

7. Сезонные колебания

Сезонными называют периодические колебания, возникающие под влиянием смены времени года. Поскольку интервальные уровни зависят от длительности интервалов времени, а длина месяцев не равная, точнее проводить анализ сезонных колебаний не по фактическим месячным уровням, а по уровням, пересчитанным на равную (30-дневную) длительность всех месяцев или среднесуточным. Если изучаются сезонные колебания за отдельный год, то обычно тренд не принимается во внимание, и отклонения месячных (30-дневных) уровней, исчисляются от среднемесячного уровня за год. Кроме рассмотренных показателей колеблемости для характеристики сезонных колебаний важное значение имеет форма сезонной «волны», изучаемая с помощью относительных показателей - отношений месячных уровней к среднемесячному (так называемый «индекс сезонности»). Изучать сезонные колебания следует за несколько лет, чтобы сгладить случайные колебания и точнее измерить сезонные.

Рассмотрим сезонность потребления воды населением, соответствующие расчеты которой приведены в табл. 8.11.

Таблица 8.11.

Расчет уровня сезонности потребления воды населением

Месяц	Среднесуточное потребление воды, тыс. куб. м.				В % к среднемесячному показателю	Отклонения от средней	Квадрат отклонения от средней
	1999	2000	2001	в среднем за 3 года			
Январь	80,3	89,5	56,3	75,4	109,24	6,37	40,61
Февраль	85,3	96,4	59,4	80,3	116,43	11,34	128,54
Март	77,3	82,1	54,6	71,3	103,38	2,33	5,45
Апрель	75,3	86,4	56,0	72,6	105,17	3,57	12,73
Май	72,7	66,2	57,9	65,6	95,08	-3,40	11,54
Июнь	72,6	74,1	57,4	68,0	98,63	-0,95	0,90
Июль	73,5	60,8	54,1	62,8	91,04	-6,18	38,18
Август	73,1	61,6	54,2	63,0	91,25	-6,04	36,43
Сентябрь	82,0	66,5	55,4	68,0	98,52	-1,02	1,04
Октябрь	82,9	59,2	53,2	65,1	94,36	-3,89	15,15
Ноябрь	92,8	59,9	54,9	69,2	100,30	0,21	0,04
Декабрь	90,4	56,5	53,1	66,6	96,60	-2,35	5,51
Средняя	79,9	71,6	55,5	69,0	100,00	-	296,14

По данным табл. 8.11 потребление воды возрастает в зимние месяцы. Среднее квадратичное отклонение составляет 4,97. Коэффициент колеблемости равен 7,2% (4,97/69,0), что позволяет сделать выводы о достаточно слабой сезонной колеблемости потребления воды населением.

8. *Прогнозирование на основе тренда и колеблемости*

Методика статистического прогноза по тренду и колеблемости основана на том предположении, что основные факторы, влияющие на формирование того или иного тренда и колеблемости, остаются постоянными на протяжении как самого динамического ряда, так и прогнозируемого периода. В этом случае говорят об экстраполяции, под которой понимают получение представлений о будущем на основе информации, относящейся к прошлому и настоящему. Период экстраполяции не должен превышать одной трети исходного динамического ряда. Прогнозные значения динамического ряда, рассчитанные на основе некоторого уравнения тренда, называются теоретическими значениями ряда (в отличие от фактических значений исходного ряда). Период времени, на который делается прогноз, принято называть периодом упреждения.

Существует два принципиально различных класса моделей, ориентированных на использование тренда и колеблемости. Модели первого класса предполагают, что тренд и колеблемость между собой никаким образом не связаны, то есть изменения тренда и изменения колеблемости происходят независимо друг от друга. Модели второго класса предполагают взаимное влияние тренда и колеблемости.

При прогнозировании в данном случае рассматриваются модели первого класса, а потому каждое трендовое значение представляет собой сумму трендового значения по экспоненциальному тренду и трендового значения по колеблемости. Для этого необходимо рассчитать параметры уравнения колеблемости, чего вычисляются отклонения фактических значений динамического ряда от трендовых значений.

Если выявлены долгопериодические циклические колебания, то аналитическое сглаживание ряда отклонений фактический значений от тренда производят с помощью ряда Фурье. В этом случае уравнение колеблемости выглядит следующим образом:

$$\widehat{Y}_t = a_0 + \sum_{n=1}^{m} (a_n \cos kt + b_n \sin kt)$$

где: k – номер гармоники;

t – период времени, выраженный в радианах или градусах и представляющий собой последовательные промежутки времени, для которых построен динамический ряд. Начальное значение t принимается равным нулю. Шаг, с которым изменяется t рассчитывается по формуле:

$$\Delta t = \frac{2\pi}{l},$$

где *l*- период колебаний, определенный ранее. К примеру, если обнаружены долгопериодические циклические колебания с периодом, равным одному году (12 месяцев), то шаг, с которым изменяется t будет равен $\Delta t = \frac{2\pi}{12} = \frac{\pi}{6}$.

Синусоиды, описываемые рядом Фурье, представляют собой гармоники различного порядка. Их наложение и позволяет составить суммарное уравнение колеблемости. При этом используется различное количество гармоник, однако практический опыт показывает, что их количество редко превышает 10 гармоники. Выбор количества гармоник проводят на основе анализа полученных данных. Для одной гармоники уравнение Фурье принимает вид:

$$\widehat{Y}_t = a_0 + a_1 \cos t + b_1 \sin t$$

Для двух гармоник:

$$\widehat{Y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t$$

Для трех гармоник:

$$\widehat{Y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t + a_3 \cos 3t + b_3 \sin 3t$$

и т.д.

Параметры уравнения Фурье вычисляются по следующим формулам:

$$a_0 = \frac{\sum y}{n}$$

$$a_k = \frac{2\sum y \cdot \cos kt}{n}$$

$$b_k = \frac{2\sum y \cdot \sin kt}{n}$$

Параметры уравнения колеблемости вычисляют для первых 10 гармоник. Оптимальное число гармоник находят по наименьшему значению квадрата отклонений тренда от фактических значений. Именно для этого количества гармоник и определяют параметры уравнения колеблемости.

Точность модели прогноза может быть оценена с помощью показателя относительной ошибки аппроксимации:

$$\bar{\delta} = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} \cdot 100\%$$

Данный показатель используется для сравнения точности прогнозов по различным моделям. Точность прогноза принято считать высокой при значении показателя меньше 10%, если его значения от 10% до 20%, то точность считается хорошей, 20-50% - удовлетворительной.

На заключительном этапе прогнозирования необходимо определить ошибку полученного прогноза.

Для получения достаточно надежных границ прогноза положения тренда, скажем, с вероятностью 0,95 того, что ошибка будет не больше указанной, следует среднюю ошибку умножить на величину критерия Стьюдента при указанной вероятности (или значимости 1-0,95 = 0,05) и при числе степеней свободы, равном *n-2*. Предельная ошибка с заданной вероятность определяется как:

$$\alpha = m_{\hat{y}_k} \cdot t_{Стьюдента}$$

где: $m_{\hat{y}_k}$ - средняя ошибка прогноза.

В этом случае, с вероятностью 0,95 можно ожидать, что тренд для момента времени t_k пройдет между значениями $\hat{Y}_{t_k} + \alpha$ и $\hat{Y}_{t_k} - \alpha$.

Стандартная ошибка прогноза определяется по формуле:

$$\sigma_p = \sigma_{ост} \cdot \sqrt{\frac{n+1}{n} + \frac{3(n+2v-1)^2}{n(n^2-1)}}$$

где:

$\sigma_{ост}$ - оценка остаточной дисперсии;

v- период упреждения.

$$\sigma_{ост} = \sqrt{\frac{1}{n-1}\sum_{1}^{n}(y_i - \hat{Y}_i)^2}$$

Средняя ошибка прогноза:

$$m_{\tilde{y}_k} = \frac{\sigma_p}{\sqrt{n}}$$

9. Смыкание рядов

Рассмотрим пример: в АО «Меркурий» до 1996 года входило 6 филиалов, а в 1996 году влилось еще 2 филиала из другой фирмы. Реализация продукции АО «Меркурий» выразилась следующим рядом динамики (табл. 8.10):

Таблица 8.10

Реализация продукции фирмы «Меркурий», млн. грн.

Год	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Продукция 6 филиалов	120	123	148	162	
Продукция 8 филиалов				210	306
Сомкнутый ряд	155,6	159,5	192,0	210,0	306,0

Для получения сомкнутого ряда определяем коэффициент пересчета по 1996 году как $K = \frac{210}{162} = 1{,}297$, а затем производим пересчет данных о реализации продукции за 1993,1994 и 1995 годы. Реализация продукции в 1993 г. = 1,297·120=155,6 млн. грн. и т.д.

Контрольные вопросы к теме 8

1. Какие вы знаете виды рядов динамики?
2. Какие показатели применяются для характеристики изменений уровней ряда динамики?
3. Какие приемы применяются для преобразования несопоставимых рядов в сопоставимые? Приведите примеры.
4. От чего зависит способ расчета по формуле средней хронологической?
5. Как рассчитать средний темп роста и темп прироста уровней ряда динамики?
6. Назовите виды колебаний уровней временного ряда.
7. Какие важнейшие приемы обработки и анализа динамических рядов вы знаете?
8. Для чего и каким образом приводят динамические ряды к одному основанию и как это делается?

Тема 9. Индексный метод в экономическом анализе

1. *Понятие и виды индексов*
2. *Среднеарифметическая и среднегармоническая формы построения индексов*
3. *Индексный метод анализа динамики среднего уровня*

1. *Понятие и виды индексов*

Индекс – это показатель сравнения двух состояний одного и того же явления (простого или сложного, состоящего из соизмеримых, или несоизмеримых элементов). Это относительная величина, показывающая изменения сложных явлений во времени и в пространстве или в сравнении с нормативными значениями.

С помощью индексов можно измерить изменение сложных явлений, выявить роль и место отдельных факторов в развитии явления.

По степени охвата единиц совокупности индексы делятся на индивидуальные и сводные, которые в свою очередь делятся на групповые и общие. Общие относятся ко всей совокупности, в то время как групповые охватывают какую-то ее часть.

Индивидуальные индексы (i) выражают изменение отдельных элементов совокупности.

Период, уровень которого сравнивается, называется текущим или отчетным периодом и обозначается подстрочной цифрой «1». Период, с уровнем которого производится сравнение, называется базисным периодом и обозначается подстрочной цифрой «0».

Приняты следующие обозначения показателей: количество обозначается буквой «q», цена – «p», себестоимость – «z», трудоемкость – «t», выработка продукции на 1 работника – «v», товарооборот – «pq», затраты на производство – «zq», затраты труда – «tq».

Индивидуальный индекс цены рассчитывается по формуле:

$$i_p = \frac{p_1}{p_0},$$

где:

p_1 и p_0 – цены на продукт в отчетном и базисном периодах.

Индекс количества продукции рассчитывается по формуле:

$$i_q = \frac{q_1}{q_0},$$

где:

q_1 и q_0 – количество произведенной продукции в отчетном и базисном периодах.

Общие индексы показывают соотношение совокупности явлений, состоящей из разнородных, непосредственно несоизмеримых элементов.

По базе сравнения различают цепные и базисные индексы.

По характеру использования показателя веса различают индексы с постоянными и переменными весами. Покажем их взаимосвязь на примере индекса физического объема продукции:

Базисные индексы с постоянными весами:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0};\ I_q = \frac{\sum q_2 p_0}{\sum q_0 p_0};\ I_q = \frac{\sum q_3 p_0}{\sum q_0 p_0} \text{ и т.д.}$$

Базисные индексы с переменными весами:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0};\ I_q = \frac{\sum q_2 p_1}{\sum q_0 p_1};\ I_q = \frac{\sum q_3 p_2}{\sum q_0 p_2} \text{ и т.д.}$$

Цепные индексы с постоянными весами:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0};\ I_q = \frac{\sum q_2 p_0}{\sum q_1 p_0};\ I_q = \frac{\sum q_3 p_0}{\sum q_2 p_0} \text{ и т.д.}$$

Цепные индексы с переменными весами:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0};\ I_q = \frac{\sum q_2 p_1}{\sum q_1 p_1};\ I_q = \frac{\sum q_3 p_2}{\sum q_2 p_2} \text{ и т.д.}$$

Между цепными и базисными индексами существует следующая взаимосвязь: для индексов с постоянными весами произведение цепных индексов равно базисному индексу крайних периодов:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} \bullet \frac{\sum q_2 p_0}{\sum q_1 p_0} \bullet \frac{\sum q_3 p_0}{\sum q_2 p_0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}$$

По содержанию явления различают индексы постоянного и переменного состава. Индексы, в которых изменяется одна величина, являются индексами постоянного состава, две и более – переменного состава.

Индексы переменного состава можно представить как произведение соответствующих индексов постоянного состава. Например, индекс товарооборота:

$$I_{pq} = I_p I_q$$

$$\frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \bullet \frac{\sum q_1 p_1}{\sum q_1 p_0}$$

По форме построения различают агрегатные, среднеарифметические и среднегармонические индексы. Рассмотренные выше формы индексов являются агрегатными, т.к. и числитель и знаменатель этих индексов представляет собой объединение разнородных элементов.

Агрегатные индексы считаются основной формой индексов. Они выполняют две функции: **синтетическую и аналитическую.** Первая функция обеспечивается тем, что в одном индексе обобщаются (синтезируются) непосредственно несоизмеримые явления. Например, цены на разные товары или разные товары, абсолютно не сопоставимые между собой в натуральном выражении. Благодаря использованию ценового соизмерителя можно агрегировать данные по различным товарам.

Аналитическая функция следует из взаимосвязи индексов. Дело в том, что практически каждый индекс можно рассматривать как составляющую некоей системы индексов, в которой его роль сводится к измерению одного из факторов общего изменения сложного явления и вклада этого фактора в совокупное изменение.

Рассмотрим построение общих индексов на примере исчисления индекса товарооборота (табл. 9.1).

Таблица 9.1

Количество и цены по поставке продукции

Вид продукции	Количество		Цена за единицу, ден. ед.		Стоимость товаров, ден. ед.	
	квартал		квартал		квартал	
	3 q_0	4 q_1	3 p_0	4 p_1	3 p_0q_0	4 p_1q_1
Продукт А	4	6	5	6	20	36
Продукт Б	10	18	9	10	90	180
Итого					110	216

Общее изменение стоимости товаров можно определить, сопоставив общую стоимость продукции в отчетном периоде (4 кв.) с общей стоимостью продукции в базисном периоде (3 кв.):

$$i_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{216}{110} = 1{,}8 \text{, или } 196\%$$

Таким образом, стоимость продукции увеличилась в отчетном периоде по сравнению с базисным на 96%, или в абсолютном выражении на (216 – 110) = 106 ден. ед.

Формула индекса стоимости продукции является ***агрегатной***. Она показывает относительное и абсолютное изменения сложных экономических явлений.

Общее изменение цены определяется по агрегатной форме

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{6 \cdot 6 + 10 \cdot 18}{5 \cdot 6 + 9 \cdot 18} = \frac{216}{192} = 1{,}125,$$

или 112,5%, т.е. цены в среднем увеличились на 12,5%.

При построении агрегатных индексов удобно пользоваться такими понятиями, как «индексируемый признак» и «признак-вес». **Индексируемый** - это признак, изменение которого характеризует данный индекс. Значение индексируемого признака изменяется: отчетное значение сопоставляется с базисным.

Признак-вес выполняет функцию веса по отношению к индексируемому признаку; его значение в данном индексе принимается неизменным, так как он не должен искажать оценку изменения индексируемого признака. Индексируемый признак можно назвать **фактором** изменения общего результата, а признак-вес - **характеристикой условий,** в которых оценивается это изменение.

Рассматривая построение индекса цены можно сказать, что в числителе и в знаменателе стоит произведение сомножителей:

$$i_p = \frac{\sum p_1}{\sum p_0} \quad \text{X} \quad \frac{q_1}{q_1}$$

индексируемый показатель — вес

Первый сомножитель – индексируемый показатель (качественный), второй сомножитель берется в числителе и в знаменателе количественный показатель на уровне отчетного периода. Такое построение индекса, когда в качестве весов используется показатель отчетного периода, принято называть построением Пааше.

Изменение физического объема товарооборота можно определить по формуле агрегатного индекса:

$$i_q = \frac{\sum q_1}{\sum q_0} \quad \text{X} \quad \frac{p_0}{p_0}$$

индексируемый показатель — соизмеритель

Где первый сомножитель – индексируемая величина, второй сомножитель берется качественный показатель на уровне базисного периода и в числителе, и в знаменателе. Такое построение индекса рассчитано по формуле Ласпейреса. Применительно к нашему примеру: $I_q = \frac{6\cdot5+18\cdot9}{4\cdot5+10\cdot9} = 1{,}745$ или 174,5%, т.е. физический объем товаров в среднем увеличился на 74,5%.

Если индексы рассматриваются в системе, то должна обеспечиваться взаимосвязь между ними, т.е. должно выполняться равенство:

$$I_q I_p = I_w$$

Это равенство не обеспечивается, если индексы построены только с отчетными или только с базисными весами. Только когда взаимосвязанные индексы строятся с весами разных периодов, увязка их в систему выполняется:

$$\frac{\sum_j q_1 p_1}{\sum_j q_0 p_1} * \frac{\sum_j q_0 p_1}{\sum_j q_0 p_0} = \frac{\sum_j q_1 p_1}{\sum_j q_0 p_0}$$

или

$$\frac{\sum_j q_1 p_0}{\sum_j q_0 p_0} * \frac{\sum_j q_1 p_1}{\sum_j q_1 p_0} = \frac{\sum_j q_1 p_1}{\sum_j q_0 p_0}$$

Из этих двух вариантов отечественная статистика долгое время отдавала предпочтение второму. Существовало правило определения периода весов: индексы первичных признаков строятся на весах базисного периода, вторичных - на весах отчетного периода. Это правило признавало неравное значение признаков в системе: первичный признак выступает как основа формирования нового (отчетного) значения результативного признака. Этим объясняется то, что индекс первичного признака оценивает изменение этого признака при сохранении базисных условий, тогда как изменение вторичного признака оценивается уже в изменившихся условиях, когда первичный признак принял значение отчетного периода.

2. Среднеарифметическая и среднегармоническая формы построения индексов

Кроме основной формы построения индексов есть среднеарифметическая и среднегармоническая формы построения индексов.

Среднеарифметическая форма построения тождественна агрегатной форме. Она получается путем замены индексируемой величины в числителе агрегатного индекса на произведение индивидуального индекса на индексируемую величину противоположного периода.

Так, индивидуальный индекс физического объема продукции равен $i_q = \frac{q_1}{q_0}$, откуда $q_1 = i_q \cdot q_0$, следовательно,

$$i_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum i_q \cdot q_0 p_0}{\sum q_0 p_0}.$$

Рассмотрим метод определения среднего арифметического индекса физического объема продукции на следующем примере. Имеющиеся данные представлены в табл. 9.2:

Таблица 9.2

Данные для расчета

Товарные группы	Продано товаров в базисном периоде, млн. грн. p_0q_0	Изменение количества проданных товаров в отчетном периоде по сравнению с базисным, % i_q
1	360,0	-1,5
2	240,0	+10,0
3	280,0	+4,0
Итого	880,0	-

Индивидуальные индексы:

$$i_{q_1} = 100\% - 1{,}5\% = 98{,}5\% \text{ или } 0{,}985$$

$$i_{q_2} = 100\% + 10\% = 110\% \text{ или } 1{,}1$$

$$i_{q_3} = 100\% + 4\% = 104\% \text{ или } 1{,}04.$$

Вычислим средний арифметический индекс физического объема продукции по трем товарным группам

$$i_q = \frac{0{,}985 \cdot 360 + 1{,}1 \cdot 240 + 1{,}04 \cdot 280}{360 + 240 + 280} = \frac{354{,}6 + 264{,}0 + 291{,}2}{880{,}0} = \frac{909{,}8}{880{,}0} = 1{,}034$$

или 103,4%. Физический объем продукции трех товарных групп увеличился на 3,4% (103,4%-100%).

Среднегармонический индекс тождественен агрегатному индексу. Он получается путем замены индексируемой величины, стоящей в знаменателе агрегатного индекса, на частное от деления индексируемой величины противоположного периода на индивидуальный индекс. Так, индивидуальный индекс себестоимости равен $i_z = \frac{z_1}{z_0}$, откуда $z_0 = \frac{z_1}{i_z}$. Тогда преобразование агрегатного индекса себестоимости в средний гармонический имеет вид:

$$i_z = \frac{\sum z_1 q_1}{\sum z_0 q_1} = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}.$$

Покажем расчет среднего гармонического индекса себестоимости на следующем пример (табл. 9.3):

Таблица 9.3

Данные для расчета среднего гармонического индекса себестоимости

Изделие	Затраты на производство во 2 квартале, млн. грн. $z_1 q_1$	Индексы себестоимости i_z
А	7,2	0,97
Б	10,7	1,21
В	15,2	1,18

Исчислим средний гармонический индекс себестоимости:

$$i_z = \frac{7{,}2 + 10{,}7 + 15{,}2}{\frac{7{,}2}{0{,}97} + \frac{10{,}7}{1{,}21} + \frac{15{,}2}{1{,}18}} = \frac{33{,}1}{7{,}42 + 8{,}84 + 12{,}88} = \frac{33{,}1}{29{,}14} = 1{,}136$$ или 113,6%.

Себестоимость продукции увеличилась на 13,6% а удорожание продукции составило 3,96 млн. грн. (33,1-29,14).

3. Индексный метод анализа динамики среднего уровня

В статистике приходится иметь дело с показателями, которые связаны между собой. Эту связь можно выразить через индексный метод. Система взаимосвязанных индексов позволяет изучать влияние отдельных факторов на изменение сложного явления.

Так, индекс товарооборота зависит от изменения цены и от изменения физического объема товарооборота. Эту связь можно выразить через индексный метод – $I_{pq} = I_p \cdot I_q$.

Индекс издержек производства равен произведению индекса себестоимости на индекс физического объема – $I_{zq} = I_z \cdot I_q$. Индекс затрат труда равен произведению индекса трудоемкости на индекс физического объема продукции, т.е. $I_{tq} = I_t \cdot I_q$. Необходимость в применении индексов переменного и постоянного состава возникает в том случае, когда динамические уровни общественных явлений выражаются средними величинами.

Индекс постоянного состава определяется по обычной агрегатной форме индекса. Общая формула для всех индексов следующая: $I_x = \frac{\sum x_1 f_1}{\sum x_0 f_1}$. Индекс постоянного состава показывает влияние только одного фактора – самой индексируемой величины.

Индекс переменного состава определяется как отношение двух средних арифметических взвешенных с переменными весами. Его изменение зависит от влияния двух факторов: от изменения самой индексируемой величины и от влияния структурных сдвигов.

Формула индекса переменного состава в общем виде:

$$I_{\bar{x}} = \bar{x}_1 : \bar{x}_0 = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}.$$

Чистое влияние структурных сдвигов определяют отношением индекса переменного состава к индексу постоянного состава, т.е. $I_{стр.сдвиг.} = I_{\bar{x}} / I_x$. Рассмотрим расчет показателей на следующем примере (табл. 9.4)

Таблица 9.4

Динамика таможенного тарифа на экспортируемую продукцию фирмой «Конус»

Вид продукции	1 квартал		2 квартал	
	Тарифная ставка, %	Сумма экспорта, дол.	Тарифная ставка, %	Сумма экспорта, дол.
А	0,90	8240,0	0,96	8927,0
Б	0,07	11,2	0,10	14,3
В	22,00	51296,1	24,50	7456,3

Требуется определить:

1. индекс среднего тарифа постоянного состава;
2. индекс среднего тарифа переменного состава;
3. индекс структурных сдвигов.

Вычислим индекс постоянного состава

$$I_{тар} = \frac{0{,}96 \cdot 8927 + 0{,}1 \cdot 14{,}3 + 24{,}5 \cdot 7456{,}3}{0{,}9 \cdot 8927 + 0{,}07 \cdot 14{,}3 + 22 \cdot 7456{,}3} = \frac{8569{,}9 + 1{,}43 + 182679{,}4}{8034{,}3 + 1{,}001 + 164038{,}6} = \frac{191250{,}73}{172073{,}9} = 1{,}112$$

или 111,2%.

Увеличение в среднем таможенного тарифа на 11,2% (111,2%–100%) зависит от изменения тарифной ставки во 2 квартале по сравнению с 1 кварталом.

Определим изменения среднего таможенного тарифа за 1 и 2 кварталы:

$$\overline{T}_1 = \frac{0{,}9 \cdot 8240 + 0{,}07 \cdot 11{,}2 + 22 \cdot 51296{,}1}{8240 + 11{,}2 + 51296{,}1} = \frac{7416{,}0 + 0{,}78 + 1128514{,}2}{59547{,}3} =$$

$$= \frac{1135930{,}9}{59547{,}3} = 19{,}08\%$$

$$\overline{T}_2 = \frac{0{,}96 \cdot 8927 + 0{,}1 \cdot 14{,}3 + 24{,}5 \cdot 7456{,}3}{8927 + 14{,}3 + 7456{,}3} = \frac{8569{,}9 + 1{,}43 + 182679{,}35}{16397{,}6} =$$

$$= \frac{191250{,}68}{16397{,}6} = 11{,}66\%$$

Вычислим индекс среднего таможенного тарифа переменного состава:

$i_{\overline{T}} = \overline{T}_1 : \overline{T}_0 = 0{,}1166 : 0{,}1908 = 0{,}612$ или 61,2%

Следовательно, за счет влияния тарифной ставки и сдвигов в структуре экспорта средний таможенный тариф снизился на 38,8% (100%–61,2%). Индекс структурных сдвигов составил:

$$I_{стр.сдвиг.} = I_{перем.сост.} : I_{пост.сост} = 0{,}612 : 1{,}112 = 0{,}551$$

или 55,1%.

Изменение структуры во 2 квартале повлекло снижение таможенного тарифа дополнительно на 44,9% (100%–55,1%).

Пример 2.

Имеются следующие данные о численности работников в бригадах и их должностном окладе:

	Базисный период		Отчетный период	
	$Ч_0$	$З_0$	$Ч_1$	$З_1$
№1	20	400	15	380
№2	10	300	15	330

Определить абсолютное и относительное изменение фонда заработной платы в целом по двум бригадам и в том числе под действием различных факторов.

Индекс фонда заработной платы переменного состава

$$\frac{15*380+15*330}{20*400+10*300}=\frac{5700+4950}{8000+3000}=\frac{10650}{11000}=0{,}968$$

Абсолютное изменение:

10650-11000=-350

Индекс фонда заработной платы постоянного состава

$$\frac{20*380+10*330}{20*400+10*300}=\frac{7600+3300}{11000}=\frac{10900}{11000}=0{,}991$$

Абсолютное изменение

10900-11000=-100

Индекс фонда заработной платы структурных сдвигов

$$\frac{15*380+15*330}{20*380+10*330}=\frac{10650}{10900}=0{,}977$$

Абсолютное изменение

10650-10900=-250

Проверка:

-350=-100+(-250)=-350

Контрольные вопросы к теме 9

1. Какова роль индексного метода анализа в экономических исследованиях?
2. В чем сущность общих индексов и как они строятся?
3. В чем состоит различие агрегатных индексов Ласпейреса и Пааше и какие факторы оказывают влияние на изменения в этих индексах?
4. Что представляют собой индексы с постоянными и переменными весами?
5. Что характеризует общие базисные индексы с постоянными и переменными весами?
6. Что отражают общие цепные индексы с постоянными и переменными весами?
7. Возможен ли переход от цепных индексов к базисным и наоборот?
8. Когда возникает необходимость преобразования агрегатного индекса в среднеарифметический и среднегармонический?
9. Что представляет собой система взаимосвязанных индексов и что характеризуют индексы постоянного и переменного состава?
10. Что характеризует индекс структурных сдвигов и как он исчисляется?
11. Какая взаимосвязь существует между индексами переменного, фиксированного состава и индексом структурных сдвигов?