

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«ГРОДНЕНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ
УНИВЕРСИТЕТ»

*Кафедра финансов и анализа в АПК
Кафедра бухгалтерского учета и контроля в АПК*

СТАТИСТИКА

ЧАСТЬ I

Курс лекций
для студентов сельскохозяйственных высших учебных
заведений экономических специальностей

Гродно
ГГАУ
2012

УДК 311: 330.101.541 (075.8)

ББК 65. 051 Я73

С 78

Авторы: доцент, кандидат экономических наук Л.Н. Кривенкова-Леванова,
старший преподаватель С.В. Грудько.

Рецензенты: кандидат экономических наук, доцент В.И. Сильванович;
кандидат экономических наук, доцент М.В. Пестис

Статистика. Курс лекций. Ч.I. / Л.Н. Кривенкова-Леванова,
С 78 С.В. Грудько. – Гродно : ГГАУ, 2012 – 156 с.

Курс лекций предназначен для подготовки студентов бухгалтерских и экономических специальностей к предмету «Статистика».

Рассмотрены основные понятия, задачи и методы статистики: наблюдение, средние величины и показатели вариации, группировка, корреляционно-регрессионный анализ, ряды динамики и индексный метод.

УДК 311: 330.101.541 (075.8)

ББК 65. 051 Я73

Рекомендовано межфакультетской учебно-методической комиссией экономического факультета и факультета бухгалтерского учета УО «ГГАУ» (Протокол № 6 от 13.02.2009 г.).

© Л.Н. Кривенкова-Леванова, С.В. Грудько, 2012

© УО «ГГАУ», 2012

ОГЛАВЛЕНИЕ

Тема 1. Предмет статистики, ее задачи и организация.....	4
Тема 2. Статистическое наблюдение.....	10
Тема 3 Средние величины и показатели вариации.....	20
Тема 4. Группировка.....	35
Тема 5. Выборочное наблюдение.....	55
Тема 6. Корреляционно-регрессионный анализ.....	70
Тема 7.Ряды динамики.....	100
Тема 8. Индексный метод	126

Тема 1. Предмет статистики, ее задачи и организация

1. Предмет и метод статистики.
2. Основные принципы статистики.
3. Задачи статистики. Связь статистики с общественно-экономическими науками.
4. Организация статистической службы в РБ.
5. Закон РБ о государственной статистике.
6. Основные понятия статистики.

ЛИТЕРАТУРА

1. Громыко Г.А. *Общая теория статистики: Практикум.* – М.: ИНФРА – М, 1999. – 139 с.
2. Елисеева И.И., Юзбашев М.М. *Общая теория статистики:* М.: Финансы и статистика, 1998.
3. Ефимова М.Р., Петрова В.В. *Общая теория статистики. Учебник.* – 2-е издание, – М.: ИНФРА – М. 2002, – 413 с.
4. Едророва В.С. *Общая теория статистики: Учебник / – М.: Юристъ, 2001, – 511 с.*
5. *Статистика: Учебник / Ред. И.И. Елисеева.* – М.: Проспект, 2002. – 448 с.

В 1. Предмет и метод статистики.

Статистика начинает свою историю с 4-го века до н.э. с Аристотеля, который сделал первые описательные сообщения о государствах (Спарта, Греция, Рим), об их экономической мощи.

12-17 вв. – период Возрождения – привел к появлению числовой информации о рождаемости, смертности, бракосочетаниях, налогах (уже тогда появились акцизные налоги). Информация бралась из податных книг. Были установлены первые статистические закономерности: на 100 родившихся девочек рождалось 105-106 мальчиков; на 10 человек умерших приходилось 13 рожденных.

Statistica (термин был введен Готфридом Ахенвалем (1719-1772)) – «сумма знаний, нужных купцам, политикам, военным, всем культурным людям».

Зарождение статистики связано с зарождением учета, в частности бухгалтерского учета. Следует отличать понятие

statistica от понятия statista. В 17 веке статистами называли людей, искусенных в политике.

Существует около 50 определений статистики, например, одно из них приведено ниже.

Статистика – это наука о количественных и качественных характеристиках жизнедеятельности человека, их изменениях, взаимосвязях и методах анализа.

Предметом статистики является любой вид деятельности человека, который требует числового учета, учета качественных показателей, их систематизации и анализа, с последующим прогнозом на перспективу.

Методы статистики:

1. Диалектический метод. Рассматривает явление в пространстве, времени и связях.
2. Монографический (описательный). Как правило, применяется для описания отдельно взятого или “уникального” явления.
3. Метод массового наблюдения.
4. Группировки и сводки.
5. Выборочный метод.
6. Метод дисперсионного анализа.
7. Корреляционно-регрессионный метод.
8. Индексный анализ и т.д.

Методы очень широко комбинируются, сочетаются друг с другом.

Как наука статистика может рассматриваться двумя своими составляющими:

- 1) как общественная наука;
- 2) как метод исследований и анализа.

В 2. Основные принципы статистики.

К основным принципам статистики как науки относятся:

- 1) неразрывность количественного и качественного анализа изучаемых явлений;
- 2) описание объекта с помощью обобщающих величин;
- 3) аналитичность изучения явлений, т.е. раскрытие типов совокупностей, их динамики и взаимодействия.

В соответствии с законом РБ “О государственной статистике”, основными принципами организации государственной статистики являются:

- 1) объективность и достоверность статистической информации;
- 2) стабильность и сопоставимость статистических данных;
- 3) доступность и открытость статистической информации в пределах, установленных законодательством РБ.

В 3. Задачи статистики. Связь статистики с общественно-экономическими науками.

Задачи статистики можно рассматривать под несколькими ракурсами:

- контрольная деятельность;
- аналитическая;
- хозяйственно-финансовая;
- информационно-пропагандистская.

Задачи:

1. Статистика должна давать фундамент фактов, служащих основанием для выводов, теоретических построений, базирующихся на закономерностях развития.
2. Проверка правильности научных положений, конкретизация научных выводов.
3. Определение того нового, что появилось в жизни, но не нашло своего отображения в соответствующих теоретических построениях.
4. Совершенствование организации и методов учета.
5. Проверка правильности, достоверности учетно-статистических данных.
6. Разработка и утверждение форм статистической отчетности.
7. Сбор и обобщение тех статистических данных, которые не охватываются организациями (перепись населения, социологические исследования).
8. Организация работы по механизации и автоматизации обработки статистической информации.
9. Гласность статистической информации:

Отрасли статистики наглядно представлены на рисунке 1.

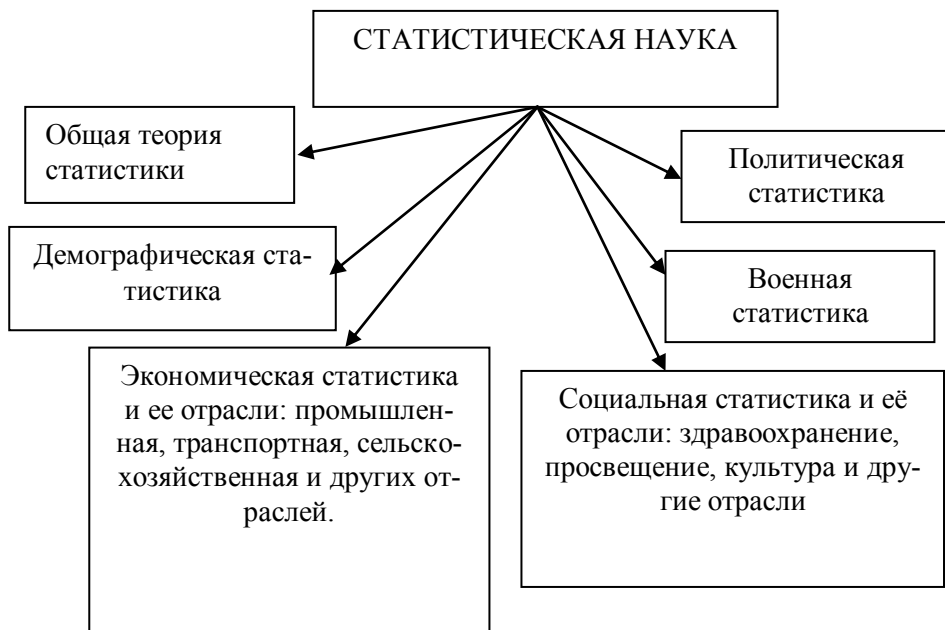


Рисунок 1. Отрасли статистики

Статистика связана с такими науками, как:

- 1) бухгалтерский учет;
- 2) теория вероятностей;
- 3) математика;
- 4) социология, демография;
- 5) экономика;
- 6) биология, естественные науки;
- 7) планирование и прогнозирование.

В 4. Организация статистической службы в РБ.

Существует два подхода к организации статистической службы:

- централизованный;
- децентрализованный.

При централизованном способе организации статистики вся сфера деятельности статистических служб регламентируется

государством. В РБ деятельность статистических органов и служб подчиняется закону "О государственной статистике в РБ", "Положением о коммерческой тайне" и двум статьям закона "О предприятии".

При децентрализованной системе организации существуют вне государственные статистические службы как при предприятиях, так и самостоятельные, которые выполняют частные заказы по сбору, обработке и анализу информации государства и различных предприятий.

Проблемы в организации статистической работы:

1. Подготовка квалифицированных кадров.
2. Обеспечение автоматизации и механизации сбора, обработки и передачи информации снизу вверх и сверху вниз.
3. Проблема технического обеспечения статистических служб (персональные компьютеры, включенные в единую глобальную сеть, программное обеспечение, защита от вирусов и др.).
4. Унификация статистической информации с учетом общепринятой мировой системы национального счетоводства.

В 5. Закон РБ о государственной статистике.

Закон РБ о государственной статистике принят Палатой представителей 29 октября 2004 года с изменениями и дополнениями от 4 января 2008 года.

Закон состоит из 6 глав:

1. Общие положения. В данной главе рассматриваются основные термины и их определения; отношения, регулируемые настоящим законом; задачи и принципы статистики; законодательство о государственной статистике и международное сотрудничество в области государственной статистики.
2. Субъекты правоотношений в области государственной статистики, их права и обязанности. В данной главе приводится перечень органов государственной статистики; их права и обязанности; права и обязанности уполномоченных государственных организаций; права и обязанности респондентов.
3. Осуществление государственной статистической деятельности. Рассматривается ведение, планирование, финансовое обеспечение статистической деятельности, а также государственные статистические наблюдения.

4. Представление, распространение и использование статистических данных (информации). В данной главе рассматривается: кому представляются статистические данные, в каком виде, как используются, кем и каким образом формируется и ведется статистический регистр.

5. Ответственность за нарушение законодательства Республики Беларусь в области государственной статистики. В главе предусмотрена ответственность за нарушение порядка представления и искажение данных государственной статистической отчетности, а также за нарушения, допущенные при осуществлении государственной статистической деятельности.

6. Заключительные положения. В данной главе предусмотрены сроки вступления в силу данного нормативного акта, документы, утратившие силу.

В 6. Основные понятия статистики.

Статистическая совокупность – это массовое явление, взятое в определенных пространственных и временных границах. Например, совокупность СПК Гродненской области, совокупность торговых точек г. Гродно, совокупность студентов экономического факультета УО «ГГАУ».

Признаки статистических совокупностей – это свойства единиц объектов. В наших примерах единицами объектов могут служить каждое СПК, торговая точка, студент.

Статистический показатель – это обобщающий сводный показатель признака.

Система показателей – это совокупность показателей, связанных между собой. Показатели бывают: натуральные (штуки, килограммы, тонны, гектары и др.), условно-натуральные (лошадиные силы, эталонные гектары, условные банки, погонные метры), стоимостные (в денежных единицах), относительные.

Статистическая совокупность количественно характеризуется:

- 1) индивидуальными значениями признака для единиц совокупности;
- 2) средней величиной показателя (признака) для всей совокупности или ее части;
- 3) величиной (силой) вариации (изменчивости).

Тема 2. Статистическое наблюдение

1. Этапы статистического исследования.
2. Этапы статистического наблюдения.
3. Виды статистического наблюдения.
4. Способы статистического наблюдения.
5. Объект и единица наблюдения.
6. Методика и организация проведения наблюдения.
7. Достоверность наблюдений и её своевременность.
8. Классификация ошибок наблюдения. Выявление и исключение ошибок наблюдений.

ЛИТЕРАТУРА:

1. Громыко Г.А. *Общая теория статистики: Практикум.* – М.: ИНФРА- М, 1999. –139 с.
2. Дружинин Е.К. *Выборочное наблюдение и эксперимент.* М.: Статистика, 1997.
3. Елисеева И.И., Юзбашев М.М. *Общая теория статистики:* М.: Финансы и статистика, 1998.
4. Ефимова М.Р., Петрова В.В. *Общая теория статистики. Учебник/– 2- е издание., – М.: ИНФРА- М, 2002, – 413 с.*
5. Едренова В.С. *Общая теория статистики: Учебник / – М.: Юристъ, 2001, – 511 с.*
6. *Статистика: Учебник/ Ред. И.И. Елисеева.* – М.: Проспект, 2002. – 448 с.
7. *Статистика “Выборочный метод наблюдения”: задачи для экономических специальностей.*– Гродно, ГГСХИ, 2000.
8. *Статистика: показатели и методы анализа: справ. пособие / Н.Н. Бондаренко, Н.С. Бузыгина, Л.И. Василевская и др.; под ред. М.М. Новикова.* – Мн.: «Современная школа», 2005.

В 1. Этапы статистического исследования.

Статистическое исследование – это сложный и продолжительный процесс. Как правило, статистическое исследование сопровождает любое технико-экономическое исследование и должно соответствовать его цели. Например, необходимо оценить уровень развития племенного дела в животноводстве в хозяйствах Гродненской области или, например, выявить финансовое состояние сельскохозяйственных предприятий по итогам

финансового года и решить вопросы дополнительного государственного финансирования для развития отраслей в конкретных хозяйствах. Чтобы провести такое исследование, необходимо знать основные положения об исследовании и наблюдении.

Статистические исследования состоят из **трёх этапов**:

- проведение наблюдения
- сводка собранной информации
- анализ полученной информации статистическими методами.

Основное требование к статистическим исследованиям – его достоверность

Статистическое наблюдение представляет собой научно организованный, планомерный и систематический процесс сбора массовых данных о различных экономических, технологических и социальных явлениях и фактах.

Сводка – это процесс обобщения данных статистического наблюдения, обеспечивающее получение статистических показателей как сводных признаков. Чаще всего представляется в виде наблюдений (правила заполнения таблицы).

Анализ собранной информации статистическими методами: необходимо дать оценку проведённого наблюдения, выявить динамику явления, установить связи факторов, их взаимодействия, дать прогноз дальнейшего развития явления или события.

Так, в нашем примере об анализе финансового состояния сельскохозяйственных предприятий необходимо регулярно отслеживать результаты их финансовой деятельности, собирать и систематизировать данные о выручке от реализации продукции растениеводства и животноводства, определять все виды рентабельности сельскохозяйственных подотраслей и выявлять убыточные в целом хозяйства и отрасли в регионе, методами статистики устанавливать положительные и негативные причинно-следственные связи факторов для принятия управленческих решений на уровне области и государства.

В 2. Этапы статистического наблюдения.

Мы уже знаем определение понятия статистического наблюдения. В современных условиях при наличии государственной информационной базы данных «БУХСТАТ» и электронно-вычислительной техники можно выбрать любой объем стати-

стических данных по хозяйствам Гродненской области (и по другим областям РБ), перенести их в электронные таблицы и проводить обработку числовой информации методами статистики, используя возможности ПЭВМ. Однако опыт показывает, что в указанной базе данных могут быть ошибки, которые снижают достоверность информации и могут привести к дальнейшим искажениям в исследовании. Поэтому работать с любой базой данных необходимо очень внимательно, соблюдая все правила статистического наблюдения.

Этапы наблюдения:

1. Проектирование наблюдения по его цели.
2. Подготовка проведения наблюдения.
3. Производство собственного наблюдения.
4. Контроль полученных материалов на достоверность.

На первом этапе необходимо установить цель исследования и продумать, какая система данных нам понадобится. Часто по ходу обработки собранной информации приходится добавлять необходимые показатели, пропущенные при первом отборе.

На втором этапе необходимо определить, какие формы статистической отчетности нам понадобятся, в каком виде и в какой последовательности будет представлена исходная информация, сколько и какие будут составлены таблицы с базой данных и т.д.

На третьем этапе необходимо произвести собственно сбор цифровой информации, представить ее в электронном и печатном виде.

На последнем этапе следует произвести контроль информации на ее достоверность. Как уже говорилось, любая исходная база данных может иметь ошибки. Кроме того, мы сами можем их допускать, если неправильно выберем или выпишем вручную необходимую информацию.

В 3. Виды статистического наблюдения

Статистическое наблюдение классифицируют по видам.

1. Первый вид оценивает совокупность по количеству исследуемых единиц, т.е. по объему. Поскольку статистическая совокупность всегда достаточно большая по числу своих единиц, то исследователям необходимо решить вопрос о том,

сколько единиц необходимо отобрать из совокупности (все или только какую-то часть) для того, чтобы получить достоверные результаты своего исследования.

Сплошное наблюдение охватывает все единицы изучаемой совокупности и обеспечивает всю полноту информации.

Несплошное наблюдение охватывает определённую часть изучаемой совокупности. Число единиц, отобранных для исследования, может быть различным:

1.1) **Выборочным** называется наблюдение части единиц исследуемой совокупности, выделенной методом случайного отбора. Подробно этот способ наблюдения и исследования рассматривается в отдельной теме.

1.2) **Наблюдение основного массива** охватывает обследование определённых, наиболее существенных по значимости изучаемых единиц совокупности.

Например, среди сельскохозяйственных предприятий Гродненской области выбраны только те, которые занимаются выращиванием и реализацией рапса в достаточно больших объёмах.

Целью указанных наблюдений является оценка всей совокупности по анализируемой системе показателей с определённой степенью точности.

1.3) **Монографическое наблюдение** глубоко и всесторонне характеризует отдельные единицы совокупности. При этом не ставится цель охарактеризовать всю совокупность.

Например, в Гродненском районе есть только два предприятия, занимающихся пушным звероводством (СПК «Прогресс-Вертелишки» и зверохозяйство «Озеры»). Цель этого вида наблюдения – подробно и всесторонне охарактеризовать интересующие нас единицы наблюдения.

2. Второй вид наблюдения связан со временем проведения наблюдения и интервалом между двумя соседними наблюдениями.

Непрерывное (текущее) наблюдение осуществляется путём непрерывной регистрации фактов по мере их возникновения.

Например, оприходование приплода по мере отела или опороса маточного поголовья КРС и свиней.

Прерывное наблюдение производится либо регулярно через равные промежутки времени, либо по мере надобности.

Например, обследование частных подворий в сельской местности с целью установления объемов производимой и потребляемой в домашних хозяйствах сельскохозяйственной продукции.

3. Третий вид наблюдения связан с источниками информации.

Непосредственное наблюдение осуществляется путём фиксации фактов, лично устанавливаемых регистраторами в результате осмотра, измерения, подсчёта признака, изучаемого явления.

Такие наблюдения в сельском хозяйстве подробно рассматриваются в курсе «Статистика предприятий АПК», когда требуется провести оперативное наблюдение за развитием растений на полях и видах на урожай, в животноводстве – за физиологическим состоянием животных и т.д.

Документальное наблюдение основано на использовании в качестве источника информации различных документов.

Такое наблюдение широко используется при проведении ревизии или аудита в бухгалтерском учете, при оценке финансового состояния предприятий со стороны банков и т.д.

Опрос базируется на получении данных в виде ответов опрашиваемых лиц с целью получения общественного мнения.

Такой вид наблюдения практикуется во всех социологических исследованиях. Следует помнить, что не всякий опрос является статистическим наблюдением, а только тот, который касается массового исследования.

В 4. Способы статистического наблюдения.

Каждый вид наблюдения должен заканчиваться одним из способов предоставления информации. В статистической литературе их называют способами статистического наблюдения. Различают следующие способы:

- 1) отчётный,
- 2) саморегистрации,
- 3) экспедиционный,
- 4) анкетный.

Отчётный способ заключается в представлении всеми предприятиями отчётов в установленные сроки по определённой форме.

Формы и сроки статистической отчетности устанавливаются министерствами и ведомствами. Одной из таких форм отчетности является годовой отчет предприятий и его первая часть – бухгалтерский баланс.

Экспедиционный способ заключается в том, что к каждой единице наблюдения посылают «счётчик».

«Счетчиком» в данном случае называют человека, который фиксирует снимаемые с приборов показатели или информацию по производству с.-х. продукции домашними хозяйствами.

Этот способ используется при переписи населения, при обследовании домашних хозяйств в сельской местности, при контроле за правильностью оплаты по газо-, водо- и электро-счетчикам.

При **саморегистрации** юридические и физические лица, которые обязаны предоставить информацию в статистические органы, снабжаются бланками, инструкциями к ним, самостоятельно с ними работают, т.е. заполняют и предоставляют в статистические органы. Лица, заполняющие бланки статистической отчетности, несут административную ответственность за правильность предоставляемой информации. Эта форма отчетности и статистического наблюдения имеет минимальное применение в РБ, т.к. фермеры или индивидуальные предприниматели, как правило, не умеют качественно оформлять бланочную документацию.

Анкетное наблюдение проводится добровольно и анонимно. Основным требованием при составлении анкеты должно быть условие анонимности, т.е. в анкете не должно содержаться вопросов, при ответе на которые можно определить личность анкетированного. При этом способе наблюдения снижается достоверность информации, так как заполнение анкеты весьма субъективно. Выбор ответов на вопросы часто зависит от настроения анкетированного, его состояния здоровья, степени усталости после работы и т.д.

В 5. Объект и единица наблюдения.

Объектом наблюдения называется совокупность общественных явлений или процессов, подлежащих исследованию (совокупность хозяйств, предприятий, совокупность студентов УО «ГГАУ», совокупность торговых точек региона и так далее).

Единицей наблюдения является составной элемент объекта наблюдения, представляющий собой источник информации (предприятие, студенческая группа, студент, торговая точка и так далее).

Единицей совокупности называют составной элемент объекта наблюдения, который служит основой счёта и обладает признаками, подлежащими исследованию.

Например, сельскохозяйственные предприятия (как единицы совокупности) обладают такими признаками, как уровень специализации, уровень рентабельности затрат и выручки при реализации с.-х. продукции, показателями обеспеченности ресурсами и эффективности из использования и т.д.

В результате статистического наблюдения мы должны собрать информацию о каждой единице совокупности по анализируемым экономико-технологическим признакам. Причем в результате сбора информации может выясниться, что не все единицы совокупности обладают требуемыми признаками. Например, если нам необходимо оценить уровень развития свиноводства в хозяйствах Гродненской области, то мы увидим, что не все хозяйства имеют эту подотрасль животноводства.

В 6. Методика и организация проведения наблюдения.

Когда определена цель статистико-экономического исследования, необходимо четко и грамотно организовать сбор информации. Для этого надо соблюдать все этапы проведения наблюдения. Кроме того, надо установить место, время и продолжительность проведения наблюдения. Если сбор информации связан с общением с лицами, занятыми своей производственной деятельностью, то необходимо заручиться их согласием, предоставить вам нужную для исследования информацию. А это не всегда возможно (см. Закон о государственной статистике). До проведения наблюдения необходимо подготовить материальную базу (дискеты, диски, бумагу, ручки, измерительные приборы и

др.), позаботиться о том, чтобы было обеспечение ПЭВМ, подготовить «счетчиков» для грамотного сбора информации.

Само наблюдение должно проводиться в кратчайшие сроки с учетом объема собираемой информации и ее сложности.

После завершения сбора информации необходимо обеспечить проверку ее на достоверность и наличие ошибок в источниках информации или допущенных при сборе информации из первоисточников.

В 7. Достоверность наблюдений и её своевременность.

Под достоверностью понимается степень объективного отображения статистическими данными, сущности явлений и процессов.

Для обеспечения достоверности статистической отчетности в законодательстве и нормативных актах предусмотрены меры наказания за искажение информации. В обязанности статистических органов входит обеспечение выполнения контролирующей функции статистики. Поэтому статистические управления проводят как плановые, так и внеплановые проверки правильности составления статистической отчетности. Если же статистическое исследование проводится вне рамок информации вносимой в формы статистической отчетности, то за достоверность такой информации несут ответственность лица, проводившие наблюдение и исследование.

Своевременность характеризует поступление данных в сроки, соответствующие целям проводимого наблюдения и исследования.

В 8. Классификация ошибок наблюдения. Выявление и исключение ошибок наблюдений.

В силу того, что исследование и наблюдение проводят люди со своим типом нервной системы, биоритмами, физическим состоянием и настроением, степенью усталости, то они могут допускать и допускают в своей работе ошибки. Если при проведении наблюдения используются какие-то приборы или техника, то их техническое состояние тоже может влиять на точность информации и результаты ее исследования. В силу этого ошибки наблюдения возникают как по объективным, так и по субъективным причинам. Поэтому исследователям необходимо

знать классификацию ошибок, уметь их избегать, выявлять и исправлять или минимизировать.

Рассмотрим виды ошибок (таблица 2.1).

Если при статистическом наблюдении допускаются регулярно ошибки с разными знаками ($\pm$), то они чаще всего бывают «случайными». Например, при раздаче комбикормов коровам вручную, количество комбикорма каждой корове (при их одинаковой продуктивности и физиологическом состоянии) может незначительно отличаться.

Систематической будет ошибка, если ее знак (+ или -) сохраняется постоянно. Например, продавец пользуется неисправными весам и не знает об их неисправности. Здесь ошибка будет постоянной до момента обнаружения неисправности прибора.

Таблица 2.1. Ошибки наблюдения

Классификация ошибок	Виды ошибок
Характер ошибок	– Случайные – Систематические
Стадия возникновения	– Ошибки регистрации – Ошибки при подготовке данных к машинной обработке – Ошибки машинной обработки
Причины возникновения	– Ошибки измерения – Ошибки репрезентативности – Преднамеренные ошибки – Непреднамеренные ошибки

Ошибки регистрации чаще всего возникают из-за невнимательности «счетчика». Например, уставший или больной работник может неправильно фиксировать числовую информацию, или во время сбора числовой информации работники-счетчики разговаривают, отвлекаются от работы и т.д.

При подготовке числовой информации к обработке на ЭВМ ошибки у оператора могут возникать по тем же причинам, что говорилось выше. Кроме того, ошибки могут возникать при западении клавиш на клавиатуре и нечеткой записи информации на бумажном носителе или по какой-либо другой причине.

Ошибки машинной обработки, как правило, связаны с двумя причинами – это наличие вируса в ПЭВМ или плохо от-тестированная программа.

Ошибки измерения возникают из-за точности используемых приборов. Так «ошибка» сантиметровой линейки равна 1 мм, а ошибка весов со шкалой до 1 кг – 1 г и т.д.

Ошибки репрезентативности зависят от объема (количества) отобранных единиц совокупности и силы изменчивости значений анализируемых показателей (характеристик или признаков) в выборочной совокупности. Подробно вопросы ошибок репрезентативности рассмотрим в теме «Выборочное наблюдение». Избежать ошибок репрезентативности невозможно, т.к. они возникают по объективным причинам и не зависят от воли человека. Их (ошибки) можно только минимизировать.

Если «счетчик» допускает ошибки умышленно, то они являются **преднамеренными**. Если преднамеренные ошибки в статистической отчетности приводят к серьезным отрицательным экономическим последствиям для государственного предприятия или бюджета, то это чревато уголовным расследованием и судом.

Непреднамеренные ошибки, как правило, тоже приводят к отрицательным результатам (неправильным выводам по результатам анализа и принятием ошибочного управленческого решения). Однако здесь последствия для работников не такие жесткие, как в предыдущем случае (можно получить административное взыскание).

При проведении статистического исследования и наблюдения все виды ошибок (кроме ошибок репрезентативности) необходимо выявлять и устранять.

Для минимизации ошибок проводится контроль за качеством статистических наблюдений на всех стадиях. Виды и содержание контроля можно представить в виде таблицы 2.2.

Из всего вышеизложенного можно сделать вывод о том, что качество и достоверность статистической информации, используемой при различных исследованиях, имеет очень важное народно-хозяйственное значение.

Таблица 2.2. Виды и содержание контроля

Способ контроля	Что проверяется
Статистический (вручную)	Структура документа, полнота документов (наличие реквизитов) полнота заполнения строк
Логический (на ЭВМ)	Соответствие кодов и наименование признаков, наличие отклонений от заданных значений и единиц измерений, наличие логических связей между показателями
Арифметический (на ЭВМ)	Соответствие построчных контрольных сумм документа и контрольных сумм ЭВМ, соответствие пографных контрольных сумм документов и контрольных сумм ЭВМ (балансовый метод), логические связи между показателями (выполнение равенств в расчетных формулах)

Тема 3. Средние величины и показатели вариации

1. Понятие средней величины.
2. Формы средних и их применение.
3. Свойства средней арифметической.
4. Понятие вариации.
5. Абсолютные показатели вариации.
6. Относительные показатели вариации.
7. Дисперсия как показатель вариации.

ЛИТЕРАТУРА:

1. Громыко Г.А. *Общая теория статистики: Практикум.* – М.: ИНФРА – М, 1999. –139 с.
2. Елисеева И.И., Юзбашев М.М. *Общая теория статистики:* М.: Финансы и статистика, 1998.
3. Ефимова М.Р., Петрова В.В. *Общая теория статистики. Учебник/ – 2-е издание., – М.: ИНФРА- М, 2002, – 413 с.*
4. Едронова В.С. *Общая теория статистики: Учебник / – М.: Юристъ, 2001, - 511 с.*

5. *Статистика: Учебник / Ред. И.И. Елисеева. – М.: Проспект, 2002. – 448 с.*
6. *Статистика: показатели и методы анализа: справ.пособие / Н.Н. Бондаренко, Н.С. Бузыгина, Л.И. Василевская и др.; под ред. М.М. Новикова. – Мн.: «Современная школа», 2005.*
7. *Экономическая статистика. 2-е изд., доп.: Учебник / Под ред. Ю.Н. Иванова. – М.: ИНФРА – М, 2001.*

В 1. Понятие средней величины.

Из истории статистики известно, что в 17-18 веках политические статистики (англ. Чарльз Давенант (1656-11714), англичанин Вилиям Пети (1623-1687), англичанин Грегори Кинг (1648-1712)) при проведении анализа результатов переписи населения во Франции и Англии использовали средние величины: средний душевой доход, средняя продолжительность жизни, средний размер семьи, среднее дневное пропитание одного взрослого работника и др.

Понятие средней величины встречается в работах Адольфа Кетле (1796-1874), бельгийца по национальности и математика по образованию. В своей работе он отмечал: «Понятие о средней величине существует вне науки, которая придает ему определенность и точность» (1866 г.)

Теория средних дальнейшее развитие получила в работах ученых 19 в. (Карл Пирсон (1857-1936), Макс Отто Лоренц (1876-1944), Родольфо Беннини (1862-1956) и др.).

Большой вклад в развитие теории средних величин внесли советские ученые: А.Я. Боярский, В.Н. Старовский, В.И. Хотимский, Б.С. Ястремский и др. В 1930 г. в Москве был опубликован учебник этих авторов «Теория математической статистики».

В результате сбора математической информации получаем базу данных, которую в последствии необходимо оценить. Если мы рассматриваем результаты деятельности сельскохозяйственных организаций по районам области, то возникает необходимость оценить значение анализируемых показателей в целом по области и в каждом районе. Если объектом исследования являются поля или фермы конкретного сельскохозяйственного предприятия, то необходимо оценить работу в полеводческих бригадах или на фермах и т.д.

В таких случаях, когда оцениваются показатели в целом по нескольким объектам, прибегают к расчету средних значений анализируемых показателей. Однако выбор формул для расчета средних не так прост, как кажется на первый взгляд. Например, как рассчитать среднюю урожайность по культуре, если она занимает несколько полей с различной площадью?

В экономических и технологических исследованиях возникает необходимость дать характеристику всей совокупности в целом, не рассматривая подробно состояние отдельных единиц этой совокупности. Например, есть необходимость оценить уровень развития молочного скотоводства по районам Гродненской области для принятия решения об инвестировании в строительство новых молочно-товарных ферм, где обеспечено содержание коров по современным интенсивным технологиям. Для чего и необходимо провести комплексное исследование совокупности молочно-товарных ферм в каждом районе.

Однако такой анализ требует особых знаний и умений для его проведения. Поэтому мы рассмотрим более простые задачи.

Разберем понятие средней величины на конкретном примере.

Задача. Группа студентов из 18 человек сдала экзамен по истории Беларуси при десятибалльной системе оценок знаний. Необходимо вычислить средний балл по этой дисциплине в группе, если получены следующие оценки: 9, 9, 9, 8, 6, 6, 9, 9, 6, 9, 8, 8, 8, 6, 6, 9, 8, 6. Средний балл можно вычислить двумя способами:

$$1) \quad \bar{x} = \frac{(9+9+9+8+6+\dots+9+8+6)}{18} = 7$$

2) Подсчитаем, сколько каких оценок получили студенты.

Тогда

"девяток"	– 7;
"восьмерок"	– 5;
<u>"шестерок"</u>	<u>– 6;</u>
всего	18.

Здесь средний балл можно вычислить иначе:

$$\bar{x} = \frac{9 \cdot 7 + 8 \cdot 5 + 6 \cdot 6}{7 + 5 + 6} = 7$$

Как видно, в данном случае результат (средний балл) не зависит от способа расчета. Кроме того, второй способ быстрее дает результат, который отражает успеваемость группы, а не конкретного студента. Однако так бывает не всегда. Очень часто выбор формулы для расчета среднего значения анализируемого признака существенно влияет на качество исследования. Некоторые формы средних используются для расчетов в конкретных статистических методах (например, рядах динамики).

При анализе статистической совокупности возникают два вопроса:

- 1) что отражает средний показатель;
- 2) как его вычислить.

Средней величиной называется обобщающий показатель, характеризующий типичный уровень варьирующего признака в качественно однородной совокупности.

Значит, средняя величина отражает или характеризует всю совокупность, а не отдельные ее единицы. По точности расчета (числу знаков после запятой) средняя величина может совпадать с индивидуальными значениями признака, а может не совпадать (например, средняя оценка в группе).

Поскольку признак варьирует (изменяется), то совокупность должна быть однородной, то есть изменчивость показателя не должна быть сильной.

Средние, характеризующие систему, совокупность в целом, называются **системными или общими средними**.

Средние, характеризующие подсистемы, типические группы или подгруппы, **называются групповыми** (см. дальше тему "Статистические группировки").

В.2 Формы средних и их применение.

Введем обозначения:

X_i – варианты;

f_i – частота (повторяемость) варианты;

W_i – объем явления ($X_i \cdot f_i = W_i$).

1. Средняя арифметическая:

- **простая**

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n} \quad (3.1)$$

- **взвешенная**

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum x_i \cdot f_i}{\sum f_i} \quad (3.2)$$

Средняя арифметическая простая применяется, когда варианта X_i – встречается один раз или одинаковое число раз.

Средняя арифметическая взвешенная используется в тех случаях, когда варианты встречаются не одинаковое число раз или когда определяют среднюю из средних.

Обе формы средних были использованы нами при расчете средней оценки в группе по истории Беларуси.

2. Средняя гармоническая:

- **простая**

$$\bar{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \left(\frac{1}{x_i} \right)} \quad (3.3)$$

-

-

-

- **взвешенная**

$$\bar{x} = \frac{w_1 + w_2 + \dots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_n}{x_n}} = \frac{\sum w_i}{\sum \left(\frac{w_i}{x_i} \right)} \quad (3.4)$$

Понятие частоты и веса не равнозначны: в качестве веса может служить показатель доли или структуры явления, варианта может представлять собой частное. Например, трудоемкость продукции равна дроби:

$$\frac{\text{общие затраты времени}}{\text{объем продукции}}$$

В этих случаях формулы средней арифметической не годятся, и поэтому используется средняя гармоническая. Рассмотрим следующие примеры на применение формулы средней гармонической.

Задача 1. Найти среднюю цену реализации продукции, используя данные, приведенные в таблице 3.1.

Таблица 3.1. Расчет средней цены реализации 1ц продукции

Сортность продукции	Цена 1 ц, ден. ед.	Удельный вес в выручке, %	Отношение удельного веса выручки к цене
	X_i	W_i	W_i/X_i
1) Первый сорт	262	79,0	0,302
2) Второй сорт	245	16,0	0,065
3) Нестандартное	228	5,0	0,022
ИТОГО	x	100	0,389

Тогда средняя цена реализации 1 центнера определяется по формуле:

$$\bar{x} = \frac{w_1 + w_2 + \dots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_n}{x_n}} = \frac{100}{0,389} = 257 \text{ ден.ед. за 1 ц}$$

Задача 2. По трем бригадам хозяйства получены следующие показатели.

Таблица 3.2. Урожайность и валовой сбор зерновых по бригадам

Бригада	Урожайность, ц/га	Валовой сбор, ц
Обозначения	X_i	W_i
Первая	32	52648
Вторая	51	42986
Третья	39	69154
Итого и в среднем	?	164788

Найти среднюю урожайность зерновых в хозяйстве.
Среднюю урожайность найдем по формуле:

$$\bar{x} = \frac{52648 + 42986 + 6915}{\frac{52648}{32} + \frac{42986}{51} + \frac{69154}{39}} =$$

$$= \frac{52648 + 42986 + 69154}{1645,2 + 842,9 + 1773,2} = 38,67 \text{ ц/га}$$

В этом примере каждая дробь в знаменателе дает значение посевной площади под зерновыми культурами в каждой бригаде.

3. Средняя хронологическая:

$$\bar{x} = \frac{0,5 \cdot x_1 + x_2 + \dots + 0,5 \cdot x_n}{n - 1} \quad (3.5)$$

Эта формула используется в моментных рядах динамики (см. дальше тему "Ряды динамики") для расчета среднего уровня ряда.

Например, дана стоимость основных производственных фондов на первое число каждого месяца.

Таблица 3.3. Стоимость основных производственных фондов на первое число каждого месяца

Дата	Стоимость основных производственных фондов, млн. руб.
01.01	11602
01.02	11589
01.03	11664
01.04	11840

Найти среднемесячную стоимость основных производственных фондов за первый квартал можно по формуле (3.5).

$$\bar{x} = \frac{0,5 \cdot 11602 + 11589 + 11664 + 0,5 \cdot 11840}{4 - 1} = 11658 \text{ млн. руб.}$$

Графически это представлено на рисунке 3.1.

Следовательно, средняя месячная стоимость ОПФ будет составлять 11658 млн. руб.

4. Средняя геометрическая:

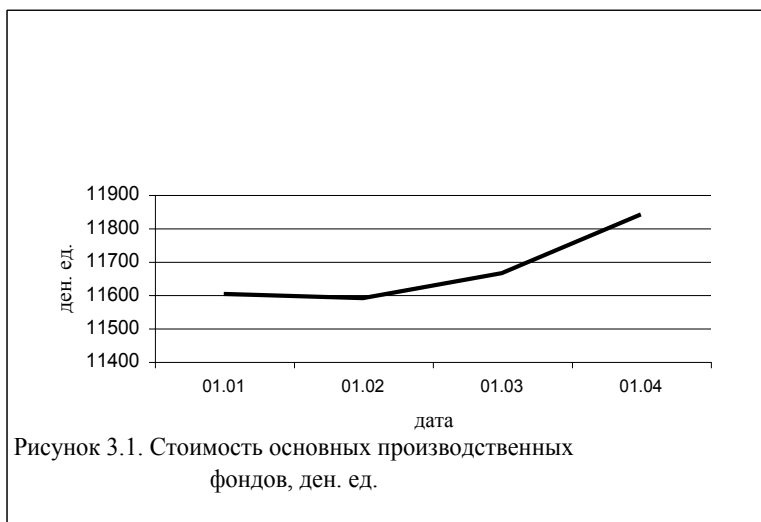
$$K = \sqrt[n]{K_1 \cdot K_2 \cdot \dots \cdot K_n} \quad (3.6)$$

Где

$$K_i = \frac{y_i}{y_{i-1}} \quad \text{– цепной темп роста, или цепной}$$

коэффициент роста,

y_i, y_{i-1} – уровни ряда динамики.



Эта формула используется в рядах динамики для определения средней скорости изменения явления во времени.

По этой формуле можно вычислить средний темп уровня инфляции.

5. Средняя квадратическая:

- простая

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}} \quad (3.7)$$

- **взвешенная**

$$\bar{x} = \sqrt{\frac{\sum x_i^2 \cdot f_i}{\sum f_i}} \quad (3.8)$$

Эта форма средней применяется при расчете среднего квадратического отклонения, где вариантой служит величина отклонения показателя от своей средней.

Примечание: выбор формулы для расчета средней величины зависит от цели исследования, вида анализируемой совокупности и наличия частоты (повторяемости) у анализируемого признака (показателя).

6. К описательным средним относятся мода и медиана, иногда их называют **структурными средними**.

Модой (Мо) называется величина признака, которая чаще всего встречается в данной совокупности.

Медианой (Ме) называется варианта, расположенная в середине ранжированного ряда. Если ряд содержит четное число элементов, то медиана равна полусумме срединных вариантов.

Примечание: Подробнее формулы моды и медианы рассматривают в курсе теории вероятностей и математической статистики.

Свойство мажорантности средних. Между средними существует следующее соотношение:

$$x_{\text{гарм}} < \bar{x}_{\text{геометр}} < \bar{x}_{\text{арифм}} < \bar{x}_{\text{квадр}} < \bar{x}_{\text{кубич}} < \dots \quad (3.9)$$

В 3. Свойства средней арифметической.

Свойства средней арифметической базируются на свойствах математического ожидания (смотрите теорию вероятностей).

1. Средняя от постоянной величины равна самой себе:

$$A = A \quad (3.10)$$

2. Произведение средней на сумму частот равно сумме произведений вариант на частоты:

$$\bar{x} \cdot \sum f_i = \sum (x_i \cdot f_i) \quad (3.11)$$

3. Изменение каждой варианты на одну и ту же величину изменяет среднюю на ту же величину:

$$\frac{\sum (x_i \pm A) \cdot f_i}{\sum f_i} = \bar{x} \pm A \quad (3.12)$$

4. Изменение каждой варианты в одно и то же число раз изменяет среднюю во столько же раз.

$$\frac{\sum (A \cdot x_i) \cdot f_i}{\sum f_i} = A \cdot \bar{x} \quad (3.13)$$

5. Изменение каждого из весов (частот) в одно и то же количество раз не изменяет средней величины.

$$\frac{\sum x_i \cdot (A \cdot f_i)}{\sum A \cdot f_i} = \bar{x} \quad (3.14)$$

6. Алгебраическая сумма отклонений всех вариантов от средней равна нулю.

$$\sum (x_i - \bar{x}) = 0 \quad (3.15)$$

7. Средняя суммы равна сумме средних.

$$\overline{X + Y} = \bar{X} + \bar{Y} \quad (3.16)$$

8. Сумма квадратов отклонений вариантов от средней арифметической меньше, чем от любой другой величины:

$$\sum (x_i - \bar{x})^2 = \min \quad (3.17)$$

$$\sum (x_i - \bar{x})^2 \cdot f_i = \min \quad (3.18)$$

Свойства 3...5 используют для упрощения расчетов средней величины на микрокалькуляторе.

В 4. Понятие вариации.

Колеблемость, многообразие, изменяемость значений величины признака совокупности называется **вариацией**.

Развитие статистической науки шло от анализа средних величин в 19в. к исследованию изменчивости анализируемых показателей в начале 20в. Р.А.Фишер писал: «Вариация, взятая сама по себе, не была предметом изучения, и на нее смотрели разве только как на досадное обстоятельство, приводящее к снижению точности средней величины. С современной точки зрения изучение причин изменчивости любой переменной величины, начиная с урожайности пшеницы и кончая интеллектом человека, должно производиться на основе измерения и анализа вариации, которая и сама по себе имеет большое значение».

При экономическом анализе нас может интересовать не только и не столько обобщенный средний показатель, как поведение индивидуальных значений признака.

Например, в двух соседних районах средняя урожайность зерновых составила 47 и 47,5 ц/га, в одном из районов урожайность в отдельных хозяйствах достигает 70 ц/га и может быть очень низкой – 33 ц/га, в другом районе урожайность по хозяйствам незначительно отличается от средней (около 5 ц/га). Встает вопрос о том, где более стабильное производство, выдерживается технология, обеспечено качественное управление отраслью.

Понятие вариации признака включает в себя оценку индивидуальных значений признака по сравнению со средними. В некоторых учебниках прибегают к оценке силы вариации при попарном сравнении индивидуальных признаков.

Применение показателей вариации весьма разнообразно:

- 1) для оценки однородности совокупностей (совокупности);
- 2) для оценки концентрации и специализации производства;
- 3) для выявления аномальностей в совокупностях;
- 4) для сравнения вариации в различных совокупностях и их оценки.

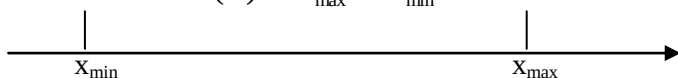
В 5. Абсолютные показатели вариации.

Абсолютные показатели вариации показывают, на сколько отличаются индивидуальные значения признака от среднего или друг от друга. Значения анализируемого показателя можно изобразить на шкале, расположив значения показателя от его минимума до максимального значения. Поэтому кроме расчетных

формулы рассмотрим и графическую интерпретацию показателей. К абсолютным показателям вариации относятся:

1. Размах вариации или амплитуда:

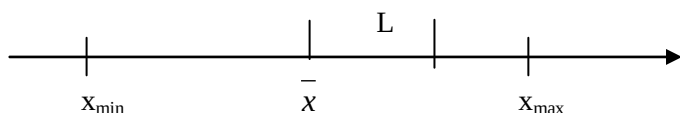
$$A(R) = X_{\max} - X_{\min} \quad (3.19)$$



2. Среднее линейное отклонение:

$$L = \frac{\sum |x_i - \bar{x}|}{n} \quad (3.20)$$

$$L = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i} \quad (3.21)$$

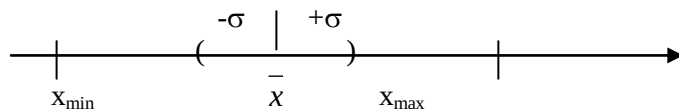


3. Среднее квадратическое (среднеквадратическое) отклонение (σ - сигма):

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (3.22)$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}} \quad (3.23)$$

Этот показатель отражает средний разброс индивидуальных значений вокруг среднего значения анализируемого признака.



Следует отметить, что формулы для расчета среднего линейного отклонения и среднеквадратического отклонения выбирают в зависимости от того, по какой формуле (простой или взвешенной) рассчитывается среднее значение анализируемого показателя.

Рассмотрим расчет абсолютных показателей вариации для нашего примера с оценками по истории Беларуси.

Таблица 3.4. Расчет абсолютных показателей вариации

Оцен- ка	Пов- торяе- мость оценки	Расчетные показатели				
		$X_i * f_i$	$ X_i - \bar{X} $	$(X_i - \bar{X})^2$	$ X_i - \bar{X} * f_i$	$(X_i - \bar{X})^2 * f_i$
9	7	63	2	4	14	28
8	5	40	1	1	5	5
6	6	36	1	1	6	6
Ито- го	18	139	x	x	25	39

Произведем расчеты абсолютных показателей вариации:

$$L = \frac{25}{18} = 1,3(8)$$

$$\bar{x} = \frac{9 \cdot 7 + 8 \cdot 5 + 6 \cdot 6}{18} = 7$$

$$\sigma = \sqrt{\frac{39}{18}} = 1,472$$

$$R = 9 - 6 = 3$$

Абсолютные показатели вариации измеряются в тех же единицах, что и варианта.

В 6. Относительные показатели вариации.

Если сравнивать абсолютное отклонение со средней величиной, приняв ее за 100%, то мы получим относительные показатели вариации.

К ним относятся:

1. Относительный размах

$$\rho = \frac{R}{x} \cdot 100\% \quad (3.24)$$

2. Относительное линейное отклонение:

$$q = \frac{L}{x} \cdot 100\% \quad (3.25)$$

3. Коэффициент вариации:

$$V = \frac{\sigma}{x} \cdot 100\% \quad (3.26)$$

Для коэффициента вариации существует шкала оценки силы вариации:

- если $V < 10\%$, то вариация показателя слабая;
- если $10\% \leq V < 20\%$, то вариация показателя средняя;
- если $V \geq 20\%$, то вариация показателя сильная.

Часто используют "полутона" силы вариации.

Например,

$V=3,1\%$ – вариация слабая, практически отсутствует;

$V=9,3\%$ – вариация слабая, ближе к средней;

$V=11,4\%$ – вариация средняя, ближе к слабой;

$V=18,5\%$ – вариация средняя, ближе к сильной;

$V=21-22\%$ – вариация сильная, ближе к средней.

В нашем примере относительные показатели будут:

$$\rho = \frac{3}{7} \cdot 100\% = 42,9\% \quad q = \frac{1,3(8)}{7} \cdot 100\% = 19,86\%$$

$$V = \frac{1,472}{7} \cdot 100\% = 21,03\%$$

Таким образом, вариация успеваемости студентов сильная, ближе к средней. Это говорит о том, что не все студенты добросовестно относятся к изучению данной дисциплины.

Рассмотрим использование коэффициентов вариации для оценки уровня концентрации отрасли и специализации районов. Используя первичную информацию по годовым отчетам хозяйств в двух районах, были рассчитаны значения показателей вариации по поголовью свиноматок, т.к. в первую очередь их число, плодовитость и многоплодие определяют объемы выращивания и откорма свиней в целом.

Оценить уровень специализации и концентрации производства в двух районах, если есть следующие данные по поголовью свиноматок.

Таблица 3.5. Анализ поголовья свиноматок.

Район	Среднее поголовье свиноматок, гол.	Число хозяйств в районе	Значения показателей вариации			
			A (R)	σ	q	V
Первый	266	12	339,1	471,5	1,500	1,772
Второй	81	20	139,6	272,2	1,698	3,321

В обоих районах свиноматок содержат только в трех хозяйствах. Поголовье свиноматок на одно хозяйство в первом районе много больше, чем во втором.

Выводы: концентрация и специализация свиноводства выше в первом районе, чем во втором.

В 7. Дисперсия как показатель вариации.

Среди показателей вариации особое место занимает дисперсия. Она позволяет из общей вариации (или общей дисперсии) выделить долю изменчивости результативного показателя за счет изменчивости факторов, на него влияющих. Рассматривается два вида факторов – первый отражает влияние факторов на результат (например, дозы органических удобрений и балльность пашни влияют на урожайность), второй – «случайные» факторы, которые невозможно предусмотреть в опыте и их исключить (например, количество выпавших осадков на конкретном поле).

Дисперсию в статистике часто называют объемом вариации (и обозначают буквой W) или суммой квадратов (и обозначают буквой D).

Основные формулы для расчета дисперсии:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \overline{x^2} - (\bar{x})^2 \quad (3.27)$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} \quad (3.28)$$

В математической статистике рассматривается и доказывается следующая формула, связывающая дисперсии в равенство:

$$\sigma_{общая}^2 = \sigma_{факторная}^2 + \sigma_{остаточная}^2 \quad (3.29)$$

Это равенство будет нами использоваться в очень многих статистических методах. Расчет дисперсий и их использование мы рассмотрим в следующей теме (Статистические группировки).

Тема 4. Группировка

1. Сводка как один из этапов статистического наблюдения и исследования.
2. Группировка как составная часть сводки.
3. Виды группировок:
 - типологическая,
 - структурная,
 - аналитическая.
4. Виды интервалов группировок и их определение.
5. Равночастотные группировки и их использование.
6. Система группировок. Регруппировка.
7. Использование дисперсионного анализа в группировках.
 - 7.1 Анализ достоверности связей факторов.
 - 7.2 Анализ тесноты связи факторов.
8. Правила оформления таблиц.

ЛИТЕРАТУРА:

1. Ефимова М.Р., Петрова В.В. *Общая теория статистики. Учебник* / – 2-е издание., – М.: ИНФРА-М, 2002, – 413 с.
2. Едророва В.С. *Общая теория статистики: Учебник* / – М.: Юристъ, 2001, – 511 с.
3. *Статистика: показатели и методы анализа: справ. пособие* / Н.Н. Бондаренко, Н.С. Бузыгина, Л.И. Василевская и др.; под ред. М.М. Новикова. – Мн.: «Современная школа», 2005.
4. *Экономическая статистика. 2-е изд., доп.: Учебник* / Под ред. Ю.Н. Иванова. – М.: ИНФРА – М, 2001.

В 1. Сводка как один из этапов статистического наблюдения и исследования.

Статистическая информация фиксируется по мере ее поступления и сбора. Поэтому в первичной информации отсутствует какая-либо закономерность. Исходная информация после проверки ее на достоверность подвергается систематизации. При рассмотрении понятия статистического исследования и его этапов, было определено понятие сводки как второго этапа исследования. Например, из годовых отчетов выписана информация из формы № 9-АПК "Производство и себестоимость продукции растениеводства" по сахарной свекле в хозяйствах Гродненского района, которая фиксировалась без системы. Необходимо установить связи факторов и закономерности в этой информации. Для этого следует как-то систематизировать выписанную информацию, т.е. произвести ее сводку.

Сводка представляет собой комплекс последовательных действий по обобщению конкретных единичных данных, образующих совокупность в целях выявления типичных черт и закономерностей, присущих изучаемому явлению в целом.

Задача сводки – охарактеризовать исследуемый предмет с помощью систем статистических показателей, выявить и измерить существенные черты и особенности изучаемой совокупности (приложение 1-3).

Эта задача решается на трёх этапах:

1. систематизация данных, их группировка,
2. уточнение системы показателей,
3. исчисление системы показателей и их обобщение.

Сводки различают по глубине обработки статистической информации:

- **простой сводкой** является итоговый результат по совокупности наблюдаемых единиц;
- **сложной является сводка** не только по всему объекту, но и по частям (группам) объекта, в т.ч. промежуточным результатам.

В 2. Группировка как составная часть сводки.

Впервые необходимость упорядочения большого количества информации возникла в 19 веке в России. Земскими статистиками за 15 лет к концу 1894 года были собраны, обработаны и опубликованы данные подворных переписей крестьянских дворов по 172 уездам (около 4 миллионов дворов). Большой объем информации необходимо было систематизировать для анализа, что и явилось толчком для разработки самого распространенного сегодня статистического метода "Статистические группировки".

Статистическая группировка – это метод, позволяющий выделить из массы первичной статистической информации однородные группы, обладающие общим сходством в количественном и качественном отношении и позволяющий выделить социально-экономические типы, изучить структуру совокупности и проанализировать связи между отдельными признаками.

Среди разработчиков этого метода можно назвать Орлова Василия Ивановича (1848-1885, Москва), Червинского Петра Петровича (1849-1931, Чернигов) и др.

Группировка как метод сложилась к концу 19 и началу 20 веков. Этот метод в своих работах широко использовал В.И. Ленин. В полном собрании сочинений можно найти много цифрового материала, отражающего экономическую или политическую ситуацию в различных государствах того времени. Примером может служить группировка фермерских хозяйств США по размеру земельной площади и стоимости производимого продукта (В.И. Ленин, полн. собр. соч., т 27, с 190).

Таблица 4.1. Влияние размеров ферм на уровень интенсивности земледелия (США, 1900 г.)

Группы ферм по размеру земельной площади, акр	Приходится на 1 акр всей земли в долларах			
	Расходы на наемный труд	Расходы на удобрения	Стоимости всего скота	Стоимости орудий и машин
До 3 акров	40,30	2,36	456,76	27,57
От 3 до 10 акров	2,95	0,60	16,32	6,71
10-20	1,12	0,33	8,30	2,95
20-50	0,55	0,20	5,21	1,65
50-100	0,46	0,12	4,51	1,47
100-175	0,45	0,07	4,09	1,14
175-260	0,52	0,07	3,96	1,00
260-500	0,48	0,04	3,61	0,77
500-1000	0,47	0,03	3,16	0,57
1000 и более	0,25	0,02	2,15	0,29

В данной таблице видно, что с увеличением площади землевладения, уменьшаются затраты из расчета на 1 акр на наемный труд, на удобрения и т.д. На сколько интенсивно осуществляется ведение сельского хозяйства можно увидеть из следующей таблицы.

Таблица 4.2. Влияние размеров ферм на уровень интенсивности земледелия (США, 1900 г.).

Группы ферм по стоимости продукта, долл.	Приходится на 1 акр всей земли, долларов			
	расход на наемный труд	расход на удобрения	стоимости всего скота	стоимости орудий и машин
1	2	3	4	5
0	0,08	0,01	2,97	0,19
1-50	0,06	0,01	1,78	0,38
50-100	0,08	0,03	2,01	0,48
100-250	0,11	0,05	2,46	0,62
250-500	0,19	0,07	3,00	0,82
500-1000	0,36	0,07	3,75	1,07
1000-2500	0,67	0,08	4,63	1,21
2500 и более	0,72	0,06	3,98	0,72

Эта группировочная таблица позволяют сделать определенные выводы о том, насколько крупные фермерские хозяйства эффективно работают.

В 3. Виды группировок

Как метод статистического анализа группировка позволяет решить следующие задачи:

- 1) выделить социально-экономические типы, единиц совокупности;
- 2) изучить структуру однотипной совокупности;
- 3) обнаружить и изучить существующие связи и зависимости между признаками исследуемого явления.

Исходя из этого, определяется три вида или **типа группировок**:

- 1) типологическая,
- 2) структурная,
- 3) аналитическая.

Признак (статистический показатель), который положен в основу проведения группировки, называют *группировочным признаком*. Как правило, это наиболее важный технологический или экономический показатель, соответствующий целям исследования. От степени сложности изучаемого массового явления и от задач анализа группировки могут быть простыми, то есть по одному группировочному признаку и сложными или комбинационными (по двум или более группировочным признакам). Группировки, приведенные в таблицах 4.1 и 4.2, являются простыми, так как в каждой из таблиц один группировочный признак (это показатель, отраженный в первой графе).

Рассмотрим все виды группировок.

Типологической принято считать группировку совокупности на классы, социально-экономические типы, однородные группы.

В сельском хозяйстве типологической модно считать группировку сельскохозяйственных угодий по виду их использования (пашня, сенокосы, пастбища, многолетние насаждения) или группировку сельскохозяйственных животных по половозрастным группам.

Рассмотрим типологическую группировку студентов по полу и анализ успеваемости по различным дисциплинам (Таблица 4.3).

В данном примере группировка студентов проведена по признаку, который не имеет числового значения, причем пол студента рассматривается как факторный признак, а успеваемость – как результативный. По результатам группировки можно сказать, что в данной группе студентов девушки более добросовестно относятся к учебе, но юноши превосходят их в точных науках (высшей математике, информатике и программировании).

Таблица 4.3. Успеваемость студентов экономического факультета.

Пол	Количество, чел.	Средний балл успеваемости по предметам				
		Высшая математика	Фило-софия	Микро-экономика	Экономическая теория	Информатика и программирование
Девушки	96	7,5	7,5	7,1	6,9	6,9
Юноши	48	8,1	7,0	6,3	6,8	7,8

Но выбор типа не так прост, как кажется на первый взгляд. Например, как произвести группировку продукции по сортам? Здесь надо учитывать много показателей и государственные стандарты на их величину. Поэтому в типологических группировках группировочные признаки делят на атрибутивные и количественные.

Если группировочный признак *атрибутивный* (сорт, порода, форма собственности, и др.), то единица совокупности или обладает этим признаком или нет.

Если группировочный признак *числовой*, то возникают проблемы:

- 1) как определить число групп (интервалов),
 - 2) какие должны быть границы группировочного признака.
- Поэтому придерживаются следующих правил:

- 1) Число интервалов при типологической группировке должно соответствовать числу предварительно намеченных частных совокупностей;

2) Границы интервалов подбирают и устанавливают таким образом, чтобы переход через них означал переход к иному типу или в новое качество.

В таких группировках группировочный признак может быть факторным, а показатели, характеризующие группы единиц, могут быть результативными, т.е. можно установить некоторые зависимости между показателями. Например, группировка коров по их классности (элита-рекорд, элита, 1 класс и т.д.) позволяет установить зависимость продуктивности коров и других важных характеристик дойного стада от классности животных.

Также примером такой группировки может служить группировка продукции по ее качеству или сортам. Государственными стандартами на сельскохозяйственную продукцию предусмотрены параметры по очень многим показателям, отражающим чистоту продукции от химических элементов, которые могут в последствии принести вред здоровью человека. Так, качество молока оценивается по его жирности, по содержанию белка, по микробиологической чистоте и др.

Следующий вид группировки – это **структурная группировка**.

Группировка, расчленяющая однородную совокупность единиц по долям признака, выраженного в процентах, называется **структурной** или **вариационной**.

Примером может служить группировка работников по стажу работы (таблица 4.4).

Таблица 4.4. Группировка рабочих по стажу работы

Интервалы по стажу работы, лет	Удельный вес численности работников, %	
	Базисный год	Текущий год
0 – 4	21,7	23,2
5 – 9	15,5	9,1
10 – 14	13,8	9,1
15 – 19	20,4	12,1
20 – 24	22,7	22,7
25 – 29	4,7	19,2
30 – 34	1,5	4,6
Итого	100%	100%

Для построения такой группировки необходимо знать численность рабочих каждого интервала по стажу работы. Анализ таблицы позволяет сделать вывод о том, что на предприятии увеличились доли работников с небольшим стажем и людей предпенсионного и пенсионного возраста, что для предприятия не может считаться положительной динамикой, так как производительность труда этих групп работников всегда минимальна. Эта группировка является неполной, так как здесь отсутствуют другие показатели, отражающие качество работы, производительность труда, оплату труда и т.д.

В аналитической группировке группировочный признак X_i может принимать любое числовое значение на каком-то промежутке от $X_{\min}$ до $X_{\max}$. Группировочный признак выбирается с учетом цели и задач исследования.

Аналитическая группировка в качестве группировочного признака всегда использует числовой показатель и позволяет установить связи и зависимости между явлениями и признаками.

В аналитической группировке, также как и в других видах группировок, выделяют два вида признаков:

- 1) факторные,
- 2) результативные.

Факторным является тот признак, который оказывает влияние на изменение других показателей.

Результативные признаки сами изменяются под влиянием факторных.

Чаще всего в аналитической группировке в качестве группировочного показателя выбирают факторный признак. Однако это необязательно. В аналитической группировке устанавливается:

- 1) наличие связей между показателями,
- 2) достоверность установленных связей,
- 3) направление связи (прямая или обратная),
- 4) теснота связи.

Примечание.

Независимо от вида группировки в каждой из них должны быть указаны:

- 1) количество единиц совокупности в каждой группе,
- 2) группировочный признак, и если он числовой, то его интервал и среднее значение для каждой группы,

3) система технолого-экономических показателей, характеризующих каждую группу совокупности и отражающих связи признаков,

4) итоговые строки и (или) графы.

Группировка должна быть оформлена в виде таблицы с учетом всех требований ГОСТа (см. вопрос 8).

В 4. Виды интервалов группировок и их определение.

Проблема интервала группировочного признака и количества групп в группировке решается в зависимости от типа (вида) группировки и числа единиц совокупности.

В типологической группировке по атрибутивному показателю число групп соответствует количеству видов признака, и каждая единица совокупности проверяется на наличие признака (например, период лактации у коров или наличие сортов растениеводческой культуры). Если группировочный признак числовой, то, как правило, он принимает 1-4 значения, установленных нормативными документами (например, базисная жирность молока).

В структурной или вариационной группировке, как правило, весь объем явления, определяющего группировочный признак, принимается за 100% или 1, а затем находятся его доли, которые и определяют число групп.

В аналитической группировке по количественному признаку число групп и величины интервалов зависят от цели исследования, объема совокупности и колеблемости группировочного признака и могут быть определены несколькими способами:

1) экспериментатором–исследователем, используя формулу

$$i = \frac{x_{\max} - x_{\min}}{n} \quad (4.1)$$

2) с использованием формулы Г.С. Стерджесса (америк.)

Пусть N – число единиц анализируемой совокупности, тогда число групп найдем как $n = 1 + 3,322 \lg N$, а величина интервала тогда определится по формуле:

$$i = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg N} \quad (4.2)$$

Эта формула даёт хорошие результаты, когда N – достаточно велико, а распределение группировочного показателя подчиняется нормальному закону.

3) с произвольными интервалами разбивается совокупность, если она мала по объему или группировочный признак имеет сильную колеблемость и неравномерность.

Интервалы группировок бывают:

Равные и неравные; открытые и закрытые, например:

0–10	0–10	до 10
10–20	10–15	10–15
20–30	15–27	15–27
80–90	27–40	свыше 27

Примечания:

1) Числа, повторяющие границу, входят в интервал, где они являются нижней границей.

2) Группировка не допускает групп, в которых число единиц совокупности равно 1 или 0. Такая ситуация возникает, если распределение группировочного признака крайне неравномерно (см. вопрос по использованию равночастотных группировок).

При проведении группировки по количественному признаку удобно использовать ранжировку группировочного признака, то есть упорядочение единиц совокупности в порядке возрастания или убывания группировочного показателя.

В 5. Равночастотные группировки и их использование.

Иногда приходится проводить исследование в совокупности, где распределение группировочного признака крайне неравномерно или совокупность является небольшой по объему.

Так, на кафедре учета и анализа в АПК была проведена равночастотная сводка и группировка районов Гродненской области по системе показателей, отражающих наличие ресурсов, их соотношение с площадью сельскохозяйственных угодий и эффективность использования указанных ресурсов (приложения 1, 2 и 3). Анализ таблиц позволяет установить факт большего наличия качественного, и интенсивного использования ресурсов в первой группе районов. Третья группа районов сложилась из тех районов, где самая низкая обеспеченность ресурсами и эффективность их использования. Эта информация может служить

поводом для принятия управленческих решений как на уровне областного управления, так и на районном уровне.

Вторым примером может служить группировка европейских компаний по объему капитала. В российском журнале «Финансовые известия» № 5, 1994, дана информация по пятистам крупнейшим компаниям Европы по размеру капитала, в млрд. \$. Эти компании могут быть сгруппированы следующим образом с использованием формулы Стерджесса:

Таблица 4.5. Распределение компаний по размеру капитала

№ п/п	Группа по размеру капитала, млрд. \$	Число компаний	Доля в % к итогу
1	До 9	439	87.8
2	9-18	41	8.2
3	18-27	12	2.4
4	27-36	6	1.2
5	36-45	1	0.2
6-9	45-81	-	-
10	81 и более	1	0.2
	итого	500	100

Компании распределились по группам неравномерно, а четыре группы являются пустым множеством. Однако рассматривать совокупность без последней компании нельзя, так как на рынке производства и сбыта продукции эта компания может быть "законодателем моды". Поэтому в анализе таких совокупностей также прибегают к равночастотным группировкам, т.е. таким, когда в каждой группе одинаковое или почти одинаковое число единиц. Следовательно, в рассматриваемом примере компании можно распределить иначе, включив в каждую группу одинаковое число компаний, предварительно ранжировав их по объему капитала.

При наличии такой группировки можно проводить более глубокий анализ. Видно, что даже среди крупнейших компаний капитал распределён крайне не равномерно. В первой группе всего 2,12% всего капитала, в десятой – 46,27%. Для более полного анализа необходимо иметь дополнительную информацию по финансовым показателям этих компаний за анализируемый финансовый период, как это принято в странах Западной Европы.

Таблица 4.6. Равночастотная группировка компаний.

№ п/п	Число компаний	Интервал капитала, млрд. \$,	Сумма их капитала в млрд. \$	Доля к итогу капитала 500 компаний в %
1	50	0.93–1.08	50	2.12
2	50	1.08–1.33	60	2.54
3	50	1.33–1.55	72	3.05
4	50	1.55–1.95	86	3.51
5	50	1.95–2.24	103	4.33
6	50	2.24–2.83	126	5.34
7	50	2.83–3.85	167	7.08
8	50	3.88–5.73	229	9.70
9	50	5.78–10.50	379	16.06
10	50	10.50–80.78	1092	46.27
Всего	500	–	2364	100

В 6. Система группировок. Перегруппировка.

Чаще всего в экономическом анализе одну и ту же совокупность нужно систематизировать (группировать) по различным существенным показателям, т.е. образовать систему группировок.

Система группировок – это ряд взаимосвязанных статистических группировок по наиболее характерным, существенным признакам, всесторонне отражающим важнейшие стороны изучаемых процессов и явлений.

Например, совокупность сельхозпредприятий Гродненской области можно группировать по площади сельскохозяйственных угодий, по баллу пашни, по поголовью животных и т.д., т.е. получить систему группировок, оценивающих наличие ресурсов и эффективность их использования.

При сравнении различных совокупностей или одной и той же совокупности, но за разные периоды времени может возникнуть проблема: сравнение различных группировок с разными интервалами. Здесь возникает несопоставимость показателей, которую необходимо устранять. Например, распределение рабочих двух сельскохозяйственных предприятий по уровню средней заработной платы. Если рассматривать напрямую оплату труда в двух хозяйствах, то непонятно, в каком из хозяйств все же опла-

та труда выше, а, следовательно, и выше заинтересованность работников в результатах своей работы.

По имеющимся результатам сложно произвести сравнение и оценить, где оплата труда выше. Следует привести информацию к сопоставимому виду. Если имеется вся исходная информация и группировка проводится на персональном компьютере, то перегруппировку можно произвести для обеих совокупностей и с любым одинаковым интервалом. В нашем случае это невозможно.

Таблица 4.7. Сравнительный анализ по оплате труда

СПК "Заря"		СПК "Победа"	
Группы по заработной плате	Удельный вес, %	Группы по заработной плате	Удельный вес, %
До 100	1	До 100	1
100 - 120	1	100 - 150	27
120 - 140	10	150 - 200	36
140 - 160	14	200 - 250	24
160 - 180 *	10	250 - 300	8
180 - 200	18	Свыше 300	4
200 - 220	13		
220 - 240	12		
240 - 260 **	6		
260 - 280	4		
280 - 300	5		
Свыше 300	6		
Итого	100	Итого	100

Получение новых групп на основании имеющихся возможно двумя способами:

1) перегруппировка по величине интервалов первичной группировки,

2) перегруппировка по удельному весу отдельных групп в общем их итоге.

В нашем случае есть возможность произвести перегруппировку информации, укрупнив интервалы в первой совокупности (СПК «Заря»).

В таблице 4.7. видно, что первые группы совпадают, а далее трансформируем интервалы первого хозяйства в интервалы второго предприятия. Там, где границы совпадают, оставляем их без

изменения. Если интервал необходимо разделить на две части, то делим его пополам. В результате получим вторичную группировку с укрупненными интервалами по СПК "Заря".

По результатам вторичной группировки (табл 4.8) можно сделать вывод, что оплата труда выше в СПК "Заря", так как там больше доля работников с более высокой оплатой труда. Однако необходимо помнить, что в данном случае группировочная таблица не является полной, т.к. в ней нет показателей по производительности труда, по оплате труда по категориям работников и т.д.

Таблица 4.8. Вторичная группировка по средней заработной плате

Группы по оплате труда	СПК "Заря", %	СПК "Победа", %
До 100	1	1
100–150	$1+10+7=18$	27
150–200	$7+10+18=35$	36
200–250	$13+12+3=28$	24
250–300	$3+4+5=12$	8
Свыше 300	6	4
Итого	100	100

В 7. Использование дисперсионного анализа в группировках.

В результате проведения анализа показателей, включенных в группировку, можно установить наличие связей между ними. Однако наши выводы о наличии связей могут оказаться ошибочными. Поэтому установленные связи должны быть подтверждены с помощью критериев достоверности или другими статистическими методами. Если группировочный признак является факторным, а остальные показатели, включенные в анализ – результативные признаки, то достоверность зависимости результатов от фактора можно доказать с помощью критерия Фишера.

В экономическом анализе принято предполагаемое утверждение называть **рабочей гипотезой**. В противовес рабочей гипотезе о наличии связей между признаками может быть установлена **"ноль-гипотеза"**, утверждающая отсутствие достоверных связей между анализируемыми показателями.

Немецкий статистик и экономист Вильгельм Лексис (1837–1914) вывел широко известное правило сложения дисперсий, которое было центральным местом в теории устойчивости (теории дисперсий).

В предыдущих темах мы уже рассмотрели показатели вариации, особое место среди которых занимает дисперсия. Но вариация результативного показателя обычно обусловлена многими факторами, которые могут быть разделены на две большие группы: систематические и случайные. И очень важно уметь из общего объёма влияния выделить вариацию за счёт постоянно действующих факторов.

Из темы "Показатели вариации" мы знаем, как определить общую дисперсию признака, отражающую объем вариации показателя независимо от факторов на него влияющих. Используем ту же систему обозначений, что и в теме "Средние величины и показатели вариации".

$$\sigma_{общ}^2 = \frac{\sum (x_i - \bar{x}_{общ})^2 \cdot f_i}{\sum f_i} \quad (4.3)$$

где X_i – значение анализируемого показателя у каждой единицы объекта,

$\bar{x}_{общ}$ – среднее значение показателя по всей совокупности,

f – частота значений анализируемого показателя.

Дисперсия, отражающая вариацию результата под воздействием конкретного фактора, называется межгрупповой или факторной. Ее вычисляют по формуле:

$$\sigma_{м.г.}^2 = \frac{\sum (x_i - \bar{x}_i)^2 \cdot f_i}{\sum f_i} \quad (4.4)$$

где $\bar{x}_i$ – среднее значение показателя в каждой группе.

Однако на результативный показатель влияет не только группировочный фактор, рассматриваемый в анализе, но и другие факторы. Кроме того, всегда присутствует случайная колеблемость анализируемого показателя. Сила случайной (остаточной) колеблемости оценивается по величине остаточной дисперсии, которую можно вычислить двумя способами.

При первом – используем правило сложения дисперсий, которое рассматривается в курсе математической статистики:

$$\sigma_{\text{общ}}^2 = \sigma_{\text{м.г.}}^2 + \sigma_{\text{случ. (остаточное)}}^2 \quad (4.5)$$

Из нее получаем формулу для расчета остаточной дисперсии:

$$\sigma_{\text{остаточное}}^2 = \sigma_{\text{общ}}^2 - \sigma_{\text{м.г.}}^2 \quad (4.6)$$

Однако такая последовательность расчетов дисперсий не позволяет выявить ошибки расчетов. Поэтому лучше остаточную дисперсию вычислить по формуле средней взвешенной из внутригрупповых дисперсий, а затем проверить правильность расчетов по формуле сложения дисперсий.

$$\sigma_{\text{остат. (случайн.)}}^2 = \frac{\sum \sigma_i^2 f_i}{\sum f_i} \quad (4.7)$$

На базе полученных дисперсий вычисляют значение F-критерия (критерия Фишера) по одной из формул:

$$F_{\text{расч.}} = \frac{\sigma_{\text{м.г.}}^2}{\sigma_{\text{ост.}}^2} \cdot \frac{\nu_{\text{ост.}}}{\nu_{\text{м.г.}}} \quad (4.8)$$

Символом "ν" обозначается "число степеней свободы".

В группировке число степеней свободы определим по формулам:

$$\nu_{\text{м.г.}} = m - 1, \quad (4.9)$$

где m – число групп

$$\nu_{\text{ост.}} = n - m, \quad (4.10)$$

где n – объём изучаемой совокупности

В учебниках по теории статистики для обозначения числа степеней свободы часто используют следующие символы:

k_1 , или ν_1 , или $\nu_{\text{м.г.}}$

k_2 , или ν_2 , или $\nu_{\text{ост. (случ)}}$

Отношение $\frac{\sigma_i^2}{\nu_i}$ называется средним квадратом и часто обозначается как S_i^2 , тогда

$$F_{\text{расч.}} = \frac{S_{\text{факторное}}^2}{S_{\text{остаточное}}^2} \quad (4.11)$$

F-критерий можно рассчитать, используя значение такого показателя тесноты связи, как индекс детерминации (см. ниже по тексту):

$$F_{\text{расч.}} = \frac{\eta^2}{1 - \eta^2} \cdot \frac{\nu_{\text{ост.}}}{\nu_{\text{м.г.}}} \quad (4.12)$$

Расчетное значение F-критерия необходимо сравнить с табличным значением критерия Фишера. Табличное значение F-критерия находим по таблицам Фишера, определив столбец по значению $\nu_{\text{м.г.}}$ и строку по $\nu_{\text{ост.}}$.

Если $F_{\text{расчетное}} > F_{\text{табл.}}$, то подтверждается рабочая гипотеза о наличии связи между показателями и опровергается "ноль-гипотеза".

Если $F_{\text{табл.}} \geq F_{\text{расч.}}$, то подтверждается "ноль-гипотеза" и опровергается рабочая гипотеза.

Если установлена достоверная связь показателей, то нужно установить тесноту связи, используя эмпирические коэффициенты корреляции и детерминации.

Эмпирический коэффициент корреляции η отражает силу связи между результативным и факторным показателем.

Если $0 < \eta \leq 0,3$ – связь слабая

$0,3 < \eta \leq 0,7$ – связь средняя

$0,7 < \eta \leq 1$ – связь тесная

Эмпирический коэффициент корреляции вычисляют по формуле:

$$\eta = \sqrt{\frac{\sigma_{\text{м.г.}}^2}{\sigma_{\text{общ.}}^2}} \quad (4.13)$$

Эмпирический коэффициент детерминации η^2 показывает, какая часть вариации результативного показателя зависит от вариации факторного признака. Он может быть выражен в процентах.

Рассмотрим числовой пример. Среди молодых механизаторов сельскохозяйственных предприятий есть прошедшие обучение в СПТУ (колледжах) и есть прошедшие обучение непосредственно в хозяйстве. Был проведен конкурс молодых механизаторов. В результате проведенного конкурса известна производительность труда всех молодых механизаторов и необходимо установить, зависит ли их производительность труда от формы подготовки. Для анализа составим сложную типологическую группировку по формам обучения и показателям производительности труда в каждой группе (таблица 4.9).

Расчеты по группировке.

Средние значения производительности труда в группах и в целом по совокупности найдем по формуле средней арифметической взвешенной:

$$\bar{x}_1 = \frac{409}{40} = 10,225 \qquad \bar{x}_2 = \frac{666}{60} = 11,100$$

$$\bar{x}_{общ} = \frac{1073}{100} = 10,730$$

Вычислим все необходимые дисперсии. Причем остаточную дисперсию рассчитаем двумя способами и проверим выполнение правила сложения дисперсий.

$$\sigma_{общ}^2 = \frac{400,79}{100} = 4,0079 \qquad \sigma_{факт}^2 = \frac{18,415}{100} = 0,1842$$

$$\sigma_{ост}^2 = 4,0079 - 0,1842 = 3,824$$

или

$$\sigma_1^2 = \frac{106,975}{40} = 2,674 \quad \sigma_2^2 = \frac{275,4}{60} = 4,590$$

$$\sigma_{ост}^2 = \frac{2,674 \cdot 40 + 4,59 \cdot 60}{40 + 60} = 3,824$$

Таблица 4.9. Анализ производительности труда

Способ подготовки механизаторов	Обозначения	i	Произведено продукции, усл. ед.	x_i	Число механизаторов, чел.-век	f_i	Общий объем выполненных работ, усл. ед.	$x_i f_i$	Расчетные показатели			
									$x_i - \bar{x}_{общ}$	$(x_i - \bar{x}_{общ})^2 \cdot f_i$	$(x_i - \bar{x}_{общ})^3 \cdot f_i$	$(x_i - \bar{x}_j)^3 \cdot f_i$
СПУ (коллек.)	1		8	8	64				-2,73	59,623	X	39,605
	2		9	11	99				-1,73	32,922		16,507
	3		11	6	66				+0,27	0,437		3,604
	4		12	15	180				+1,27	24,194		47,259
Итого и в среднем			10,225	40	409				X	X	10,201	106,975
Ученичество	5		9	17	153				-1,73	50,879	X	74,970
	6		10	18	180				-0,73	9,592		21,780
	7		12	14	168				+1,27	22,581		11,340
	8		15	11	165				+4,27	200,562		167,310
Итого и в среднем			11,100	60	666				X	X	8,214	275,400
Всего и в среднем			10,730	100	1075				X	400,8	18,415	382,375

Установим достоверность зависимости производительности труда молодых механизаторов от формы подготовки, используя F-критерий, рассчитав его по формуле (4.12) с использованием имперического коэффициента детерминации.

$$F_{расч} = \frac{0,04596}{1 - 0,04596} \cdot \frac{98}{1} = \frac{4,50408}{0,95404} = 4,72$$

$$F_{табл} = 3,98 \quad F_{расч} > F_{табл}$$

Поскольку $F_{расч} > F_{табл}$, можно утверждать, что производительность труда молодых механизаторов зависит от формы подготовки, т.е. "ноль-гипотеза" является несостоятельной.

Эмпирические коэффициенты корреляции и детерминации показывают, что связь между анализируемыми показателями слабая ($\eta = 0,214$) и вариация производительности труда только на 4,6% зависит от формы подготовки механизаторов.

$$\eta^2 = \frac{0,182}{4,0049} \cdot 100\% = 4,595\%; \quad \eta = \sqrt{0,045959} = 0,214$$

В 8. Правила оформления таблиц.

Статистическая таблица является наиболее компактной формой представления информации. Кроме того, в статистических исследованиях используются промежуточные расчетные таблицы, которые позволяют упростить математические вычисления или использовать электронные таблицы ПЭВМ. Виды таблиц весьма разнообразны, но ко всем предъявляются некоторые стандартные требования.

1. Если таблиц много, то они должны быть пронумерованы.
2. Каждая таблица должна иметь заголовок, отражающий ее суть. Заголовок может содержать указание на объект, время и место исследования.
3. Заголовки граф и строк должны быть краткими и не допускают сокращений, кроме общепринятых.
4. Графы и строки должны содержать единицы измерения, проставленные через запятую (например: Численность работников, чел.)
5. Округление чисел в пределах одной и той же строки или графы следует проводить с одинаковой степенью точности.

6. Отсутствие числового значения в зависимости от причины может быть отмечено следующими символами:

а) "X" – если позиция в таблице не подлежит заполнению,

в) "... " или "нет свед.", или "н.св." – если отсутствуют сведения,

с) "—" при отсутствии явления.

7. Если числовое значение показателя очень мало по сравнению с другими значениями, то используют обозначения "0,0" или "0,00".

8. Таблицы должны содержать итоговые строки и графы.

9. Желательно, чтобы таблица целиком размещалась на странице.

10. Допускается расположение таблицы поперек страницы.

11. Если таблица не помещается на одной странице, то допускается перенос ее части на другую страницу с указанием на то, что это продолжение таблицы с номерами граф.

12. В тексте кроме ссылки на таблицу должен быть дан ее анализ.

Тема 5. Выборочное наблюдение

1. Выборочное наблюдение как основной вид сплошного наблюдения.

2. Обозначения показателей при выборочном наблюдении, ошибки выборки.

3. Способы отбора единиц; виды выборки.

4. Математические основы выборочного наблюдения (закон больших чисел).

5. Выбор формул для оценки результатов выборочного наблюдения при повторном и бесповторном способах отбора.

6. Особенности малой выборки.

ЛИТЕРАТУРА:

1. *Елисеева И.И., Юзбашев М.М. Общая теория статистики: М.: Финансы и статистика, 1998.*

2. *Ефимова М.Р., Петрова В.В. Общая теория статистики. Учебник/ – 2-е издание., – М.: ИНФРА- М, 2002, – 413 с.*

3. Едронова В.С. *Общая теория статистики: Учебник* / – М.: Юристъ, 2001, – 511 с.
4. Плошко Б.Г., Елисеева И.И. *История статистики: Учеб.пособие* / – М.: Финансы и статистика, 1990, – 295с.
5. *Статистика “Выборочный метод наблюдения”: задачи для экономических специальностей.* – Гродно, ГГСХИ, 2000.
6. *Статистика: Учебник/ Ред. И.И. Елисеева.* – М.: Проспект, 2002. – 448 с.
7. *Статистика: показатели и методы анализа: справ. пособие* / Н.Н. Бондаренко, Н.С. Бузыгина, Л.И. Василевская и др.; под ред. М.М. Новикова. – Мн.: «Современная школа», 2005.

В 1. Выборочное наблюдение как основной вид несплошного наблюдения.

Первое выборочное наблюдение связано с именем Пьера Лапласа (1749–1827). В 1802 г. выборочным методом была проведена перепись населения Франции. "Выборка охватила почти 7% общего числа жителей; затем на основе отношения числа жителей к числу рождений, приходящихся на переписанные общины, и данных об общем числе рождений в стране (по метрическим книгам) было исчислено население всей Франции с ошибкой, меньше полумиллиона".

В 19 в. несплошные наблюдения (опросы населения) проводились в США, Германии, Бельгии, Англии, Норвегии. И уже в 1903 г. на Берлинской сессии Международного статистического института была принята резолюция о том, что выборка может дать точные результаты, если соблюдаются условия отбора наблюдаемых единиц.

В современном экономико-технологическом исследовании выборочный метод применяется по двум основным причинам:

- 1) когда экономически не выгодно сплошное наблюдение;
- 2) когда наблюдение связано с уничтожением единиц наблюдения (анализ качественных характеристик продукции).

Первая причина возникает при исследовании больших совокупностей, когда сплошное наблюдение связано с большими материальными затратами и затратами времени. Например, сплошной опрос населения по каким-либо политическим, демографическим или другим вопросам.

Вторая причина, как правило, связана с оценкой качества продукции при проведении лабораторных исследований. Например, оценка качества молока, зерна и другой с.-х. продукции при реализации на предприятия переработки.

Выборочное наблюдение проводится по следующей схеме:

- 1) определяют, какая часть совокупности подлежит выборочному наблюдению;
- 2) устанавливают, каким способом произвести отбор части анализируемой совокупности в выборочную совокупность;
- 3) производят отбор единиц совокупности;
- 4) определяют, как получить необходимые характеристики всей совокупности по результатам выборки;
- 5) производят расчеты для оценки точности результатов проведенного выборочного наблюдения, т.е. оценивают величину ошибки выборки.

Цель выборочного наблюдения – по результатам выборки оценить всю совокупность.

Например, когда берут пробы молока или зерна при реализации продукции на предприятия переработки, то в соответствии с качеством сделанных проб устанавливают цену для всей реализованной продукции, считая, что вся продукция обладает установленным качеством, а не только взятые пробы. При этом исследователи понимают, что при выборочном исследовании допускается ошибка репрезентативности, величину которой необходимо установить.

В 2. Обозначения показателей при выборочном наблюдении, ошибки выборки.

Вся совокупность, из которой производят отбор, называется **генеральной совокупностью**.

N – численность генеральной совокупности,

$\tilde{x}_{ген.}$ – среднее значение анализируемого показателя в генеральной совокупности.

Часть генеральной совокупности, попавшая в выборку, называется **выборочной совокупностью или выборкой**.

n – численность выборочной совокупности,

$\bar{x}$ – среднее значение анализируемого показателя в выборочной совокупности,

$\tilde{\sigma}^2$ – дисперсия генеральной совокупности,

σ^2 – дисперсия выборочной совокупности,

$\tilde{p}$ – доля генеральной совокупности,

ω – доля выборочной совокупности

Величина доли вычисляется по формулам:

$$\tilde{p} = \frac{M}{N}; \quad \omega = \frac{m}{n} \quad (5.1)$$

Долей называется та часть генеральной или выборочной совокупности, которая обладает определенным признаком.

M и **m** – число единиц соответственно генеральной и выборочной совокупности, обладающих анализируемым признаком.

Например, из 125 коров 48 дают 9000 кг молока в год. Следовательно, доля высокопродуктивных коров будет равна $48/124=0,384$, или 38,4%.

Отклонение значения характеристики выборочного наблюдения от характеристики генеральной совокупности называется ее **ошибкой**.

$\Delta x = \bar{x} - \tilde{x}_{ген.}$ – *ошибка репрезентативности* выборочной средней,

$\Delta \omega = \omega - p$ – ошибка частности (доли выборки),

$\Delta \sigma^2 = \sigma^2 - \tilde{\sigma}^2$ – ошибка дисперсии.

Чем меньше эти ошибки, тем лучше (точнее) выборочная совокупность характеризует генеральную.

Силу изменчивости ошибки относительно среднего значения можно оценить по величине относительной ошибки выборки. Чем меньше этот показатель, тем качественнее проведено выборочное наблюдение при условии выполнения всех требований к выборочной совокупности.

Относительной ошибкой выборки считается доля абсолютной ошибки относительно значения средней выборочной величины анализируемого показателя или аналогично по доле выборки.

$$\Delta_{\%} = \frac{\Delta}{\bar{x}} \cdot 100\%; \quad \Delta_{w\%} = \frac{\Delta_w}{w} \cdot 100\% \quad (5.2)$$

При проведении выборочного наблюдения возникает ряд вопросов, которые должен решить исследователь:

Сколько может быть выборочных средних? Сколькими способами можно отобрать, например, 10 студентов из 40 присутствующих для оценки успеваемости? И каждый раз будем получать новую среднюю успеваемость. Какова должна быть величина выборки? Как нужно отбирать единицы в выборочную совокупность? и т.д. Со всеми этими проблемами нам помогает разобраться теория выборочного наблюдения.

В 3. Способы отбора единиц; виды выборки.

Под способом отбора понимают порядок отбора единиц из генеральной совокупности в выборочную.

Различают

1) повторный и

2) бесповторный способы отбора.

При повторном способе отбора отобранные и обследованные единицы совокупности возвращаются в генеральную совокупность и снова могут участвовать в выборочном исследовании. Например, оценивают скорость молокоотдачи при дойке коров выборочным методом. По сути, здесь коровы, попавшие в обследование, находятся в стаде (в генеральной совокупности и не изымаются из нее).

При бесповторном способе отбора исследуемые отобранные единицы не возвращаются в генеральную совокупность или не могут быть возвращены и не могут участвовать в других выборочных исследованиях. Например, в г. Минске на Комаровском рынке в воскресенье проводили социологический опрос покупателей по ценам на товары и по объемам сделанных покупок по продуктам питания. Вероятность того, что эти же покупатели будут участвовать в следующем выборочном обследовании, равна нулю, поэтому такое наблюдение является бесповторным.

Виды выборки:

1) Собственно-случайная (жеребьевка) – отбор в этом случае производится либо по жребию, либо по таблицам случайных чисел.

2) Механический – в выборку попадает каждая пятая, десятая и т.д. единица.

3) Типическая (районированная) – предварительно генеральная совокупность по признаку типизации разбивается на частные группы (типы, районы), а затем в пределах этих групп производится выборка.

4) Серийно-гнездовая – генеральная совокупность делится на одинаковые по объему группы-серии и производится отбор не единиц совокупности, а серий.

5) Комбинированная – берется комбинация различных видов выборки.

Все виды выборки могут использоваться в сельском хозяйстве.

Способы совмещаются с видами.

Принципы отбора:

1) Обеспечение случайности, т.е. при отборе все единицы генеральной совокупности имеют равную возможность (вероятность) попасть в выборку.

2) Обеспечение достаточного числа отобранных единиц, т.е. репрезентативности (представительности) выборки.

Выборка является репрезентативной, если она обладает всеми качественными характеристиками генеральной совокупности.

В 4. Математические основы выборочного наблюдения (закон больших чисел).

Современные теоретические основы выборочного наблюдения сформировались в трудах петербургской математической школы – П.Ч. Чебышева (1821–1894), А.М. Ляпунова (1857–1918), А.А. Маркова (1856–1922) и др. В 1901 г. русский экономист и статистик В.И. Борткевич (1863–1931) указал, что теоретической основой выборочного метода должно служить исчисление вероятностей.

Результатами выборочного наблюдения можно пользоваться, если решены три вероятностные задачи:

1. Если определены возможные пределы ошибки репрезентативности с заданной вероятностью. Если эти пределы больше заданных, то такими результатами пользоваться нельзя.

2. Если определена вероятность "Р" того, что возможные пределы ошибки репрезентативности не превосходят заданных величин. При недостаточной вероятности тоже нельзя пользоваться результатами выборочного наблюдения.

3. Необходимо решить вопрос о том, каким должен быть объем выборки "n" для того, чтобы получить результаты с заранее заданной точностью, и чтобы можно было гарантировать эту точность с заранее заданной вероятностью.

Решение всех трех задач обеспечивается **теоремой П.Л. Чебышева**:

$$P\{|\bar{x} - \tilde{x}_{ген.}| < \varepsilon\} \geq 1 - h, \quad (5.3)$$

где ε и h – как угодно малые величины, а объем выборки "n" – достаточно большое число.

Но поскольку при проведении выборочного наблюдения возможно получить много различных значений средней величины $\bar{x}_{выбор.}$, то ответ на вопрос, когда можно пользоваться выборочным наблюдением, дает **теорема А.М. Ляпунова**:

Случайная переменная величина, состоящая из большого числа взаимно независимых слагаемых, среди которых нет ни одного, резко выдающегося своей колеблемостью, имеет нормальное распределение.

На практике это позволяет принимать равенство дисперсий в выборочной и генеральной совокупностях:

$$\sigma_{ген.}^2 = \sigma_{выборочн}^2 \quad (5.4)$$

Здесь не важно, какое распределение в генеральной совокупности. Главное, чтобы ошибка выборки была распределена нормально.

Ошибка выборки, не превышающая величины среднеквадратического отклонения, называется **средней ошибкой**.

При этом вероятность Р того, что ошибка выборочной средней (Δ) не превысит среднюю ее ошибку,

$\pm\sigma$ - равна $P=0,68269$

$\pm 2\sigma$ – $P=0,95450$

$\pm 3\sigma$ – $P=0,99730$

Коэффициент при среднеквадратическом отклонении (σ), показывающий кратность средней ошибки выборки, обозначается t и называется коэффициентом доверия. Связь между P и t выражается интегралом:

$$F(t) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_0^t e^{-\frac{t^2}{2}} dt \quad (5.5)$$

Величина $\Delta = \sigma$ называется средней или стандартной ошибкой выборки.

Произведение $t\sigma$ называется предельной ошибкой выборки.

Можно доказать, что величина предельной ошибки может быть вычислена по формуле

$$\Delta_x = t \cdot \sqrt{\frac{\sigma^2}{n}} \quad (5.6),$$

т.е. предельная ошибка (Δ_x) зависит от трех факторов: σ^2 , n и t (дисперсии, объема выборки, величины коэффициента доверия или вероятности) следующим образом:

1) величина дисперсии (σ^2) отражает силу вариации, поэтому чем выше сила вариации анализируемого показателя, тем больше ошибка выборки;

2) чем больше объем выборки n , тем меньше ошибка исследования;

3) чем больше вероятность P , тем меньше ошибка.

Значительно величина ошибки (Δ_x) зависит не только от объема выборки n , но и от доли выборки относительно генеральной совокупности.

Величина значения коэффициента доверия

$$t = \frac{\bar{x} - \tilde{x}_{ген.}}{\sigma_{ген.}} \quad (5.7)$$

зависит только от того, с какой вероятностью необходимо гарантировать пределы ошибки выборки.

Если известны величина дисперсии σ^2 и коэффициент доверия t , то из формулы (5.6) можно получить формулу для расчета объема выборки

$$n = \frac{t^2 \cdot \sigma^2}{\Delta_x^2} \quad (5.8)$$

На основании формул (5.6) и (5.8) решаются все три поставленные задачи.

Для доли выборки эти формулы трансформируются:

$$\Delta_\omega = t \cdot \sqrt{\frac{\omega \cdot (1 - \omega)}{n}}; \quad (5.9)$$

$$t = \frac{\Delta_\omega}{\sqrt{\frac{\omega(1 - \omega)}{n}}}; \quad (5.10)$$

$$n = \frac{t^2 \cdot \omega \cdot (1 - \omega)}{\Delta_\omega^2} \quad (5.11)$$

В 5. Выбор формул для оценки результатов выборочного наблюдения при повторном и бесповторном способах отбора.

Система формул (5.6-5.11) является основой для построения формул для оценки предельной ошибки выборки при различных способах и видах выборки, что отражено в следующей таблице 5.1. Необходимо помнить, что выражение, стоящее под радикалом, дает значение средней ошибки выборки.

Из таблицы 5.1 видно, что при бесповторном способе отбора в формулах добавляется разность между 1 и долей выборки относительно объема генеральной совокупности $(1 - n/N)$.

В таблице 5.2 отражены формулы для расчета объема выборки. Все представленные в этой таблице формулы получают из формул таблицы 5.1 путем алгебраических преобразований.

Таблица 5.1. Предельная ошибка выборки для некоторых способов формирования выборочной совокупности

Метод отбора Вид выборки	Повторный		Бесповторный	
	для средней	для доли	для средней	для доли
1. Собственно-случайная и механическая	$t \cdot \sqrt{\frac{\sigma^2}{n}}$	$t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}}$	$t \cdot \sqrt{\frac{\sigma^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$	$t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)}$
2. Типическая (при пропорциональном отборе групп)	$t \cdot \sqrt{\frac{\sigma_i^2}{n}}$	$t \cdot \sqrt{\frac{\omega_i \cdot (1-\omega_i)}{n}}$	$t \cdot \sqrt{\frac{\sigma_i^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$	$t \cdot \sqrt{\frac{\omega_i \cdot (1-\omega_i)}{n} \cdot \left(1 - \frac{n}{N}\right)}$
3. Серийная (гнездовая)	$t \cdot \sqrt{\frac{\delta_x^2}{r}}$	$t \cdot \sqrt{\frac{\delta_\omega^2}{r}}$	$t \cdot \sqrt{\frac{\delta_x^2}{r} \cdot \frac{R-r}{R-l}}$	$t \cdot \sqrt{\frac{\delta_\omega^2}{r} \cdot \frac{R-r}{R-l}}$

Их используют тогда, когда уже были ранее проведены аналогичные исследования и рассчитаны дисперсии и ошибки выборки. Тогда расчет объема выборки позволяет не делать лишнюю ненужную работу, при этом сохраняя достаточно высокую точность и надежность выборочного исследования.

Таблица 5.2. Необходимый объем выборки для некоторых способов формирования выборочной совокупности

Виды выборочного наблюдения	Повторный отбор	Бесповторный отбор
1	2	3
Собственно-случайная выборка: а) при определении среднего размера признака	$n = \frac{t^2 \cdot \sigma^2}{\Delta_x^2}$	$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \sigma^2}$
б) при определении доли признака	$n = \frac{t^2 \cdot W(1-W)}{\Delta_W^2}$	$n = \frac{t^2 \cdot W(1-W) \cdot N}{\Delta_W^2 \cdot N + t^2 \cdot W(1-W)}$

Продолжение таблицы 5.2

1	2	3
---	---	---

Механическая выборка	то же	то же
Типическая выборка: а) при определении среднего размера признака	$n = \frac{t^2 \cdot \sigma^2}{\Delta_x^2}$	$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \sigma^2}$
б) при определении доли признака	$n = \frac{t^2 \cdot W(1-W)}{\Delta_W^2}$	$n = \frac{t^2 \cdot W(1-W) \cdot N}{\Delta_W^2 \cdot N + t^2 \cdot W(1-W)}$
Серийная выборка: а) при определении среднего размера признака	$r = \frac{t^2 \cdot \delta^2}{\Delta_x^2}$	$n = \frac{t^2 \cdot \delta^2 \cdot R}{\Delta_x^2 \cdot R + t^2 \cdot \delta^2}$
б) при определении доли признака	$r = \frac{t^2 \cdot W_r(1-W_r)}{\Delta_W^2}$	$r = \frac{t^2 \cdot W_r(1-W_r) \cdot R}{\Delta_W^2 \cdot R + t^2 \cdot W_r(1-W_r)}$

В 6. Особенности малой выборки.

В перерабатывающей промышленности и в сельском хозяйстве, где необходимо установить качество продукции используют метод оценки проб продукции в лабораторных условиях. Как правило, количество проб небольшое. В каждом конкретном случае экспериментальным способом отрабатывались методики взятия проб и оценки качества продукции.

Малой выборкой считают выборку, объем которой находится в пределах от 5 до 30 единиц, т.е. $5 < n \leq 30$.

При малой выборке нельзя принимать равенство дисперсий $\sigma_{ген.}^2 = \sigma_{выб.}^2$, как это делается для большой выборки. Особенность малой выборки в том, что ее случайные ошибки не подчиняются закону нормального распределения.

Закон распределения случайных ошибок малой выборки был найден английским ученым Вильямом Госсетом (1876–1937) и опубликован в 1908 г. под псевдонимом «Стьюдент».

Для малой выборки расчет предельной ошибки выборки найдем по формуле:

$$\Delta_{x_g} = t \cdot \mu_{м.б.}, \quad (5.12)$$

где $t = \frac{\bar{X}_{в.} - \bar{X}_{ген.}}{\mu_{м.в.}}; \mu_{м.в.} = \frac{\sigma_{выбор.}}{\sqrt{n-1}}$ – стандартная ошибка малой

выборки.

Разность $(n-1)$ называется *числом степеней свободы* и обозначается $\nu = n-1$. В некоторых учебниках эта разность обозначается буквой k .

Стьюдентом разработаны таблицы, которыми все пользуются для определения предельного значения t , при различных значениях вероятности (P).

Современные таблицы Стьюдента, которые имеются в каждом учебнике статистики и теории вероятностей, более подробные.

Объем выборки (n)	P=0,95	P=0,99
3	3,18	5,84
4	2,78	4,60
5	2,57	4,03
...		

Рассмотрим несколько задач по выборочному наблюдению.

Задача 1.

Для определения средней урожайности картофеля в области проведена 20%-ная серийная бесповторная выборка, в которую вошло 5 районов из 25.

Таблица 5.3. Результаты обследования по районам.

№ района	Средняя урожайность, ц с 1 га	Посевная площадь, га	Валовой сбор, тыс.ц.
	$\bar{X}_i$	f_i	$\bar{X}_i \cdot f_i$
1	250	800	200
2	260	1000	260
3	275	1200	330
4	280	1200	336
5	300	2800	1406
Итого, в среднем	280	7000	1406

Определить с вероятностью 0,954 пределы, в которых будет находиться средняя урожайность картофеля по области.

Решение.

Из условия задачи выпишем значения имеющихся показателей:

$n=5$; $N=25$; $P=0,954$; по таблицам найдем значение $t=2,00$. Надо определить предельную ошибку выборки Δx - ?

Чтобы произвести необходимые расчеты ошибки выборки, надо вычислить среднее значение урожайности картофеля и величину дисперсии.

$$\bar{x}_{общ} = \frac{\sum \bar{x}_i \cdot f_i}{\sum f_i} = 280 \text{ ц/га.}$$

Таблица 5.4. Расчет дисперсии

п/п	$(\bar{x}_i - \bar{x}_{общ})^2$	$(\bar{x}_i - \bar{x}_{общ})^2 \cdot f_i$
1	900	720000
2	400	400000
3	25	30000
4	0	0
5	400	1120000
Итого	X	2270000

$$\sigma_{сер.}^2 = \frac{\sum (\bar{x}_i - \bar{x}_{общ})^2 \cdot f_i}{\sum f_i} = 337 \text{ ц/га.}$$

Теперь можно вычислить величину предельной ошибки урожайности:

$$\Delta_x = t \cdot \sqrt{\frac{\sigma_{сер.}^2}{n_c} \left(1 - \frac{n_c}{N}\right)} = 2 \cdot \sqrt{\frac{337}{5} \cdot \left(1 - \frac{5}{25}\right)} = \pm 7.34 \text{ ц/га}$$

Средняя урожайность картофеля в области (в генеральной совокупности) составит

$$\tilde{x}_{ген.} = \bar{x}_{общ} \pm \Delta_x = 280 \pm 7.34 \text{ ц/га}$$

или результат можно представить в интервальном виде

$$272,66 \text{ ц/га} < \tilde{x}_{ген} < 287,34 \text{ ц/га}$$

Задача 2.

Рассмотрим постановку и решение задач на использование механического бесповторного отбора. Используя исходную информацию по группировке покупателей можно решить несколько задач. В нашем случае они указаны по пунктам 1-3.

Условие:

При опросе покупателей на рынке отбирается каждый десятый покупатель со стоимостью покупки до 35 тыс. руб. и каждый десятый со стоимостью выше 35 тыс. руб. Покупатели распределились по выборке следующим образом:

Таблица 5.5. Распределение покупателей на рынке.

Стоимость покупок, тыс. руб.	Количество покупок (n_i)	Средняя стоимость покупок, тыс. руб., ($\bar{X}_i$)	Внутригрупповая дисперсия, тыс. руб., (σ_i^2)
До 36 тыс. руб.	150	26	49
36 тыс. руб. и выше	250	42	81
Итого	400	36	69

Определить:

1) Каковы возможные пределы ошибки выборочной средней (т.е. средней цены покупки), если ее надо гарантировать с вероятностью $P=0,954$.

2) Какова вероятность того, что предельная ошибка выборочной средней стоимости покупок на 1 покупателя не превысит 1,6 тыс. руб., а предельная ошибка частоты покупателей, приобретших товаров на сумму более 60 тыс. руб., не превысит 0,04 при объеме выборки $n=400$, если по предыдущим обследованиям известно, что доля таких покупок равна $\omega=0,2$.

3) Каким должен быть объем выборки для того, чтобы можно было гарантировать с вероятностью $p=0.954$, что пределы возможной средней ошибки выборки (средней цены покупки) не превзойдут 1,6 тыс. руб., а возможной ошибки доли (частоты) не превзойдут 0,04, если по предыдущим обследованиям известно, что $\omega=0,2$.

Решение.

Для решения всех трех задач необходимо найти среднюю общую стоимость одной покупки и общую дисперсию:

$$\bar{x}_{общ} = \frac{\sum x_i \cdot n_i}{\sum n_i} = \frac{26 \cdot 150 + 42 \cdot 250}{150 + 250} = \frac{1440}{400} = 36 \text{ тыс. руб.}$$

Средняя из частных внутригрупповых дисперсий

$$\bar{\sigma}_i^2 = \frac{49 \cdot 150 + 81 \cdot 250}{150 + 250} = \frac{27600}{400} = 69$$

1) Для расчета величины ошибки используем формулу предельной ошибки средней выборочной и подставим в нее все необходимые значения.

$$\Delta_x = t \cdot \sqrt{\frac{\bar{\sigma}_i^2}{n} \cdot \left(1 - \frac{n}{N}\right)} = 2 \cdot \sqrt{\frac{69}{400} \cdot \left(1 - \frac{400}{4000}\right)} = 0.78 \text{ тыс. руб.}$$

t=2 при P=0,954 (из таблицы)

N=4000, т.к. осуществлен десятипроцентный отбор и n=400

Вывод: средняя цена покупки в генеральной совокупности будет

$$X_{ген} = 36 \pm 0,78 \text{ тыс. руб.}$$

2) В пункте 2 необходимо произвести расчеты по величине вероятности для цены покупки и для доли покупателей, поэтому из формул ошибок средней и доли необходимо вычислить величину коэффициента доверия t, а затем по таблицам найти значение вероятностей.

$$\text{а) } t = \frac{\Delta_x}{\sqrt{\frac{\bar{\sigma}_i^2}{n} \cdot \left(1 - \frac{n}{N}\right)}} = \frac{1.6}{\sqrt{\frac{69}{400} \cdot \left(1 - \frac{400}{4000}\right)}} = 4.06 ;$$

$$P(4,06) = 0.99; \Delta_x = 1.6$$

$$\text{б) } t = \frac{\Delta_\omega}{\sqrt{\frac{\omega \cdot (1 - \omega)}{n} \cdot \left(1 - \frac{n}{N}\right)}} = \frac{0.04}{\sqrt{\frac{0.02 \cdot (1 - 0.02)}{400} \cdot \left(1 - \frac{400}{4000}\right)}} = 0,006$$

$$P(0,0066) = 0,008$$

Вывод: Вероятность средней цены покупки велика (0,99), а вероятность доли покупателей практически равна нулю.

3) В третьем пункте предлагается рассчитать объемы выборки для очередного выборочного наблюдения. Для этого мы выбираем формулу из таблицы 5.2. и подставляем в нее все необходимые числовые значения.

$$a) n = \frac{t^2 \cdot \bar{\sigma}_i^2 \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \bar{\sigma}_i^2} = \frac{2^2 \cdot 69 \cdot 4000}{1.6^2 \cdot 4000 + 2^2 \cdot 69} = 105$$

$t=2$ (из таблиц), $\Delta_x=1,6$

б)

$$n = \frac{t^2 \cdot \omega \cdot (1-\omega) \cdot N}{\Delta_\omega^2 \cdot N + t^2 \cdot \omega \cdot (1-\omega)} = \frac{2^2 \cdot 0.2 \cdot 0.8 \cdot 4000}{0.04^2 \cdot 4000 + 2^2 \cdot 0.2 \cdot 0.8} = 200$$

$\Delta_\omega=0,04$, $\omega=0,2$

Вывод: Для оценки средней цены покупки надо опросить 105 человек, а для оценки доли покупателей необходимо побеседовать с 200 покупателями.

Тема 6. Корреляционно-регрессионный анализ

1. Условия применения корреляционно-регрессионного анализа.
2. Этапы корреляционно-регрессионного анализа.
3. Виды парной корреляционно-регрессионной связи.
4. Парная линейная корреляционно-регрессионная модель.
5. Показатели тесноты связей линейной корреляционно-регрессионной модели.
6. Анализ достоверности парной линейной корреляционно-регрессионной модели.
7. Расчет параметров криволинейных уравнений.
8. Оценка тесноты связи при криволинейных моделях.
9. История развития многофакторного корреляционно-регрессионного анализа.
10. Расчет уравнения многофакторной корреляционно-регрессионной модели.
11. Оценка тесноты связи признаков и ранжир факторов по силе их влияния на результат в множественном корреляционно-регрессионном анализе.

12. Оценка достоверности результатов произведенного корреляционно-регрессионного анализа.
13. Прогнозирование по корреляционно-регрессионной модели.
14. Стандартизированный вид уравнения связи.
15. Рекомендации по составлению различных корреляционно-регрессионных моделей.

ЛИТЕРАТУРА:

1. Елисеева И.И., Юзбашев М.М. *Общая теория статистики*: М.: Финансы и статистика, 1998.
2. Ефимова М.Р., Рябцева В.М. *Общая теория статистики*. М.: Финансы и статистика, 1999.
3. Плошко Б.Г., Елисеева И.И. *История статистики: Учеб.пособие* / – М.: Финансы и статистика, 1990, – 295с.
4. *Статистика сельского хозяйства. Под ред. Замосковского О.П.*, М.: Финансы и статистика, 1990.
5. *Сельское хозяйство РБ: статистический сборник / ред.кол.: В.С. Метекс; В.Н. Синкевич; Л.Л. Выбчик и др.; Министерство статистики и анализа РБ.* – Минск, 2006. – 223 с.

В 1. Условия применения корреляционно-регрессионного анализа.

В курсе математики рассматриваются функциональные связи, где однозначно определяется значение одной переменной, если известно значение другой (или других). Многие экономические показатели связаны в жестко детерминированную систему, например, расчет уровня рентабельности затрат на производство и реализацию продукции. Однако в экономических и технологических явлениях часто значение одного показателя никогда не определяет однозначно значение другого показателя, хотя связь может наблюдаться. Например, уровень молочной продуктивности коров зависит от уровня кормления, аналогично уровень урожайности сельскохозяйственных культур зависит от доз органических и минеральных удобрений. В экономических исследованиях хорошо видно, что уровень себестоимости единицы сельскохозяйственной продукции зависит от уровня механизации производственного процесса, от уровня продуктивности

биологического объекта (животных или земельных угодий) и т.д. Такая связь показателей называется стохастической или вероятностной и может быть установлена только при большом числе наблюдений.

«Corelation» – английское слово, введенное в употребление биологом и статистиком Френсисом Гальтоном (1822–1911) в конце XIX века. И переводится оно как «соответствие», т.е. нежесткая математическая зависимость.

Связь называется корреляционно-регрессионной, если значению результативного показателя соответствует несколько значений факторного признака, и, наоборот, при одном и том же значении факторного показателя можно достичь разных значений результата.

За результативный показатель в каждом конкретном анализе выбирается более важный признак, отражающий результаты деятельности. Например, объем валовой продукции, объем прибыли, уровень рентабельности, уровень себестоимости или цены единицы продукции. Для этих результативных показателей факторными могут быть: наличие основных производственных фондов, фондообеспеченность, фондовооруженность, производительность труда или трудоемкость, урожайность отдельных культур, продуктивность животных и т.д.

В другом примере результативным показателем может быть производительность труда, которая зависит от:

- уровня механизации производственного процесса;
- квалификации работников и т.д.

Корреляционно–регрессионная связь – это связь вариаций (изменений) различных показателей, когда изменение одного признака влияет на изменение другого признака.

Корреляционно-регрессионный анализ заключается в построении и анализе экономико-математической модели в виде уравнения регрессии и оценке тесноты связи между анализируемыми показателями. Поэтому принято отличать понятия «регрессия» и «корреляция». Очень часто в статистической литературе **под регрессией понимают нахождение уравнения связи показателей, под корреляцией – определение тесноты связи изучаемых признаков.**

В общем виде уравнение регрессии записывается в следующем виде:

$$\tilde{Y}_{x_1; x_2; \dots x_n} = f(x_1; x_2; \dots x_n) \quad (6.1)$$

где n – число факторов.

Условия применения корреляционно-регрессионного анализа:

1. Для построения корреляционно-регрессионной модели надо иметь достаточно большое количество значений каждого показателя, т.е. объем статистической совокупности должен содержать не менее 50 единиц.

2. Распределение значений анализируемых признаков у единиц совокупности должно быть близким к нормальному, т.е. сила вариации показателей должна быть минимальной.

Из второго условия вытекает требование отсутствия в анализируемой совокупности «аномальных» точек, т.е. таких значений признаков, которые сильно отличаются от большинства значений по всей совокупности. Аномальные точки должны быть удалены из анализируемой совокупности.

В 2. Этапы корреляционно-регрессионного анализа:

1. Предварительный (априорный) анализ. Он проводится достаточно квалифицированным исследователем и заключается в предварительном отборе (установлений) результативных и факторных показателей, которыми обладают единицы анализируемой статистической совокупности в соответствии с целью проводимого исследования.

2. Сбор информации и ее первичная обработка. Здесь выявляются ошибки наблюдения, информация проверяется на нормальность распределения показателей и отсутствие «аномальных» точек, иногда проводят группировку для предварительного установления связей.

3. Выбор вида и построение модели (уравнения регрессии) с использованием математического аппарата. Оценка уравнения регрессии на достоверность по одному из критериев достоверности.

4. Оценка тесноты связей признаков с использованием показателей, отражающих тесноту и направление связи (если его можно установить).

5. Прогнозирование развития явления по модели на перспективу.

На первом этапе формулируется задача исследования, определяется методика измерения показателей или сбора информации, исключаются дублирующие факторы или факторы, связанные в жесткую систему.

На втором этапе совокупность должна быть достаточно большой по числу единиц и наблюдений: $N \gg 50$, число факторов, влияющих на результат, должно соответствовать N . Данные должны быть количественно и качественно однородны.

На третьем этапе определяется внешний вид аналитической функции (уравнения регрессии) и находятся ее параметры.

На четвертом этапе оценивается достоверность всех характеристик корреляционно-регрессионной модели с использованием критериев достоверности (Стьюдента или Фишера).

На пятом этапе осуществляется прогноз усредненных значений показателей, наилучших значений и наихудших. По модели возможен ранжир единиц совокупности по влиянию факторного показателя на результативный.

В 3. Виды парной корреляционно-регрессионной связи.

К самым простым корреляционно-регрессионным связям относят парные или однофакторные связи. Среди парных выделяют: линейные и криволинейные связи. Результативный показатель обозначается Y , факторный признак обозначается X . Исходную базу данных (цифровую информацию) можно представить в виде горизонтальной или вертикальной таблицы. Например,

Таблица 6.1. База данных для исследования связи показателей

№ по порядку	1	2	3					N
Результативный показатель	Y_1	Y_2	Y_3					Y_n
Факторный признак	X_1	X_2	X_3					X_n

В исходной информации вместо «№ по порядку» могут быть указаны наименования хозяйств (точек совокупности). Если задана информация в таком виде, то каждой паре чисел на декартовой системе координат в соответствие может быть поставлена точка. На рисунках 6.1 и 6.2 изображены некоторые виды таких графиков, которые называются точечными диаграммами.

Виды парной корреляционно-регрессионной связи:

1. Линейное уравнение регрессии:

$$\tilde{Y}_x = a_0 + a_1 x$$

2. Степенная связь факторов:

$$\tilde{Y}_x = a \cdot x^b \quad \text{или} \quad \tilde{Y} = a_0 \cdot x^{a_1}$$

Это уравнение может быть приведено к линейному уравнению путем логарифмирования:

$$\log Y = \log a + b \log x$$

3. Показательная связь факторов:

$$Y_x = a \cdot b^x$$

Уравнение приводится к линейному виду через логарифмирование:

$$\log y = \log a + (\log b) x$$

4. Гиперболическая зависимость результата от фактора:

$$\tilde{Y}_x = a + \frac{b}{x}$$

Это уравнение преобразуется в линейное уравнение подстановкой величины, обратной X, т.е. $Z = \frac{1}{X}$

тогда $\tilde{Y}_z = a + bz$.

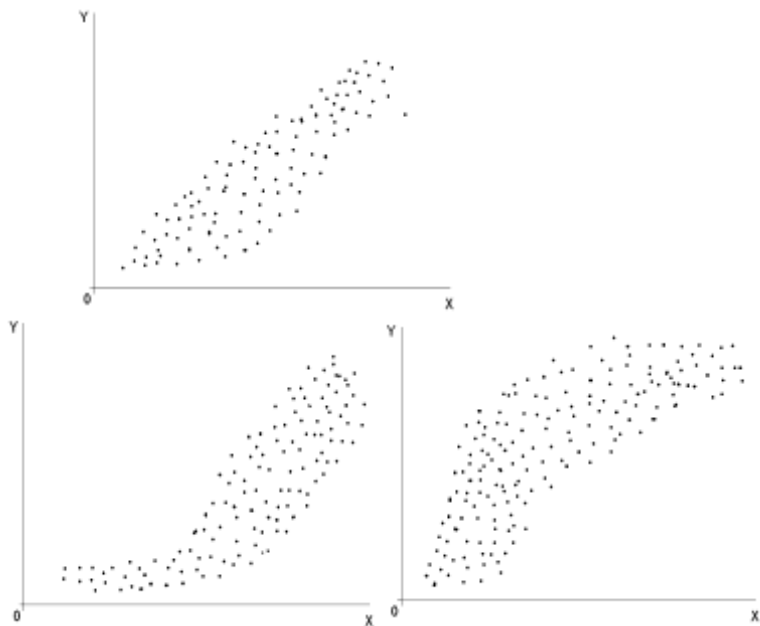


Рисунок 6.1. – Прямая (положительная) регрессия

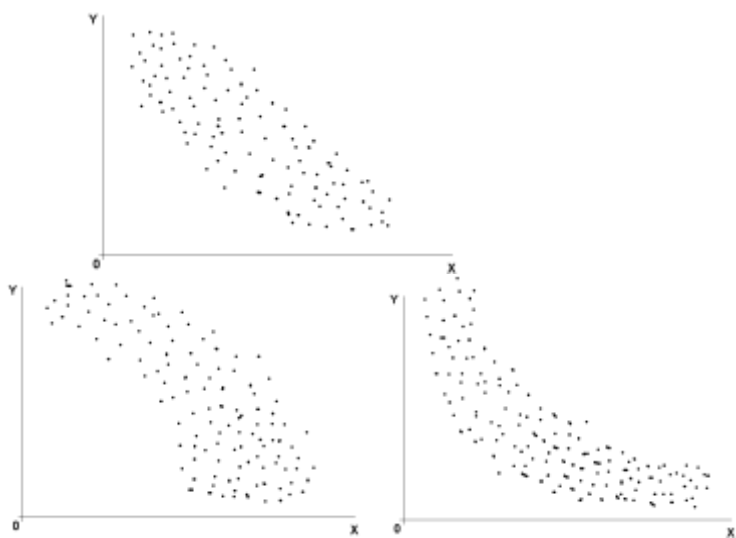


Рисунок 6.2. – Обратная (отрицательная) регрессия

5. Параболическая связь:

$$\tilde{Y}_x = a_0 + a_1x + a_2x^2$$

В 4. Парная линейная корреляционно-регрессионная модель.

Процесс построения корреляционно-регрессионной модели сводится к осреднению значения результата и фактора.

Пусть исходные данные X и Y сведены в таблицу и дополнительно рассчитаны X^2 , XU , Y_x .

Таблица 6.2. Информация для построения уравнения регрессии

№ п/п	X_i	Y_i	X^2	XU	$\tilde{Y}_x$
1	X_1	Y_1	X^2_1	X_1Y_1	$\tilde{Y}_{x_1}$
2	X_2	Y_2	X^2_2	X_2Y_2	$\tilde{Y}_{2n}$
...	...	...	...	...	...
N	X_n	Y_n	X^2_n	X_nY_n	$\tilde{Y}_{xn}$
Σ	ΣX_i	ΣY_i	ΣX^2	ΣXU	$\Sigma \tilde{Y}_x$

В дальнейшем, после расчета уравнения регрессии и нахождения по нему теоретических значений результативного показателя, необходимо будет проверить выполнение равенства:

$$\sum y_i = \sum \tilde{Y}_x$$

Для парной связи следует построить график (точечную диаграмму), и выявить наличие ошибок информации и (или) аномальные точки:

Уравнение регрессии должно быть таким, чтобы обеспечить минимум суммы квадратов отклонений эмпирических значений результативного показателя от теоретических значений, полученных по модели: $\tilde{y}_x$

$$\sum (y_i - \tilde{Y}_x)^2 = \min \quad (6.2)$$

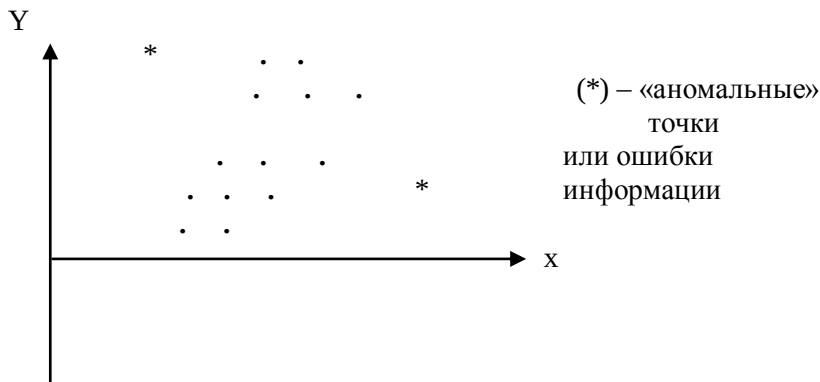


Рисунок 6.3. Точечная диаграмма связи показателей

Это достигается при использовании метода наименьших квадратов (МНК), разработанного К.Ф.Гауссом (1777–1855).

Для прямой линии $\tilde{Y}_x = a_0 + a_1 x$ составим линейную систему нормальных уравнений (два уравнения с двумя неизвестными).

Получим:

$$\begin{cases} N \cdot a_0 + a_1 \cdot \sum x = \sum y \\ a_0 \cdot \sum x + a_1 \cdot \sum x^2 = \sum xy \end{cases} \quad (6.3)$$

В учебниках по общей теории статистики, как правило, даются формулы для расчета параметров уравнения регрессии a_0 и a_1 :

$$a_0 = \frac{\sum y \cdot \sum x^2 - \sum xy \cdot \sum x}{N \cdot \sum x^2 - \sum x \cdot \sum x} \quad (6.4)$$

$$a_1 = \frac{N \cdot \sum xy - \sum y \cdot \sum x}{N \cdot \sum x^2 - \sum x \cdot \sum x} \quad (6.5)$$

Эти формулы получены при использовании правила определителей второго порядка для решения систем двух линейных уравнений с двумя неизвестными, где в знаменателе стоит значение главного определителя:

$$\Delta = \begin{vmatrix} N & \sum x \\ \sum x & \sum x^2 \end{vmatrix} = N \cdot \sum x^2 - \sum x \cdot \sum x \quad (6.6)$$

где Δ – главный определитель системы.

В числителях формул 6.4 и 6.5 стоят определители, где в главном определителе заменен один из столбцов на столбец свободных членов.

В полученном уравнении регрессии параметры носят следующие названия:

a_0 – свободный член;

a_1 – коэффициент регрессии.

В уравнении свободный член может иметь экономико-технологический смысл, а может не иметь. Например, если уравнение отражает уровень продуктивности животных в зависимости от уровня кормления, то свободный член должен показывать уровень продуктивности животных при «нулевом» кормлении, что является абсурдным. С точки зрения математики свободный член отражает значение точки пересечения прямой линии с осью ординат. Корреляционно-регрессионная связь существует только в определенной области (в области размаха вариации для X и Y)

Парный коэффициент регрессии всегда интерпретируем.

Парный коэффициент регрессии показывает, на сколько единиц своего измерения в среднем изменится результат (Y), если факторный показатель (X) изменится в среднем на единицу своего измерения.

Например, получено уравнение зависимости уровня рентабельности от уровня механизации производственного процесса:

$\hat{Y}_x = 10 + 0,301X$, где значения обоих параметров уравнения имеют смысл:

$a_0=10\%$ будет отражать уровень рентабельности при полном отсутствии механизации труда;

$a_1=0,301\%$ показывает, что уровень рентабельности увеличится на 0,3%, если уровень механизации вырастет на 1%.

Имея значения параметров уравнения регрессии, вычисляют теоретические значения результативного показателя, подставляя в уравнение фактические значения X_i (см. таблицу 6.2), для дальнейшей работы с моделью.

В 5. Показатели тесноты связей линейной корреляционно-регрессионной модели.

Насколько близко фактические точки разбросаны вокруг теоретической линии, оценивают по показателям тесноты связей, к которым относятся

1.) Парный коэффициент корреляции:

$$r = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \quad (6.7)$$

где в знаменателе стоит произведение среднеквадратических отклонений факторов

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (6.8)$$

или

$$\sigma_x = \sqrt{(\overline{x^2}) - (\bar{x})^2} \quad (6.9)$$

Аналогично рассчитывается среднеквадратическое отклонение Y .

Величина $\overline{x \cdot y} - \bar{x} \cdot \bar{y}$ называется ковариацией и обозначается

$\overline{x \cdot y} - \bar{x} \cdot \bar{y} = COV_{xy}$ (это показатель величины совместной вариации X и Y).

Вторая формула для расчета парного коэффициента корреляции:

$$r = \frac{\sum (y_i - \bar{y}) \cdot (x_i - \bar{x})}{n \cdot \sigma_x \cdot \sigma_y} \quad (6.10)$$

Третья формула для расчета парного коэффициента корреляции

$$r = \frac{n \cdot \sum y \cdot x - \sum x \cdot \sum y}{\sqrt{[n \cdot \sum y^2 - (\sum y)^2] \cdot [n \cdot \sum x^2 - (\sum x)^2]}} \quad (6.11)$$

Парный коэффициент корреляции интерпретируется в зависимости от его величины и знака, т.к. он указывает не только на тесноту связи, но и направление связи. Знак парного коэффициента корреляции всегда совпадает со знаком парного коэффициента регрессии.

Парный коэффициент корреляции всегда принадлежит одному из интервалов $0 < r < 1$ или $-1 < r < 0$.

Если $0 < r < 1$, то связь между факторами прямая.

Если $-1 < r < 0$, то связь факторов обратная.

Если $|r| \leq 0,3$, то связь признаков слабая;

если $0,3 < |r| < 0,7$, то связь факторов средняя;

если $|r| \geq 0,7$, то связь между показателями сильная (или тесная).

2.) Парный коэффициент детерминации

$$r^2 \cdot 100\%$$

Он показывает, на сколько процентов **вариация результата** зависит от **вариации фактора**.

3.) Коэффициент эластичности (Э). В курсе микроэкономики рассматривается очень много парных зависимостей показателей, относительно которых говорят об эластичности графика (например, эластичность спроса или предложения какого-либо товара на рынке сбыта). В курсе статистики оценивают величину или силу эластичности, используя значение вполне конкретного показателя.

Коэффициент эластичности находят как первую производную от уравнения регрессии:

$$\Xi = a_1 \cdot \frac{\bar{x}}{\bar{y}} \quad (6.12)$$

Определение: Коэффициент эластичности результативного признака относительно факторного показателя показывает, на сколько процентов изменится в среднем результат (Y) при изменении фактора (X) на 1% своего среднего значения.

Коэффициент эластичности может быть рассчитан не только для всей совокупности, но и для отдельной точки совокупности. Тогда в формуле средние значения показателей меняют на значения показателей в данной точке.

В 6. Анализ достоверности парной линейной корреляционно-регрессионной модели.

Поскольку изучаемая совокупность по объему всегда ограничена, то значения параметров уравнения регрессии и коэффициентов корреляции и детерминации могут искажаться действием случайных факторов. Поэтому нужно с помощью какого-либо критерия достоверности оценить все показатели, т.е. подтвердить или опровергнуть ноль-гипотезу.

Для линейной связи обычно используют t-критерий (критерий Стьюдента). Определяют t расчетные для параметров уравнения и силы связи факторов a_0 , a_1 и r . Сравнивают полученные значения с t табличными. Если расчетные t больше t табличных, то ноль-гипотеза о недостоверности уравнения отвергается. Таким достоверным уравнением можно пользоваться для анализа и прогноза показателей.

Значения t-критерия для параметров вычисляют по формулам:

$$t_{a0} = \left| a_0 \frac{\sqrt{n-2}}{\sigma_{ocm}} \right| \quad (6.13)$$

$$t_{a1} = \left| a_1 \cdot \frac{\sqrt{n-2}}{\sigma_{ocm}} \cdot \sigma_x \right| \quad (6.14)$$

$$t_r = \left| r \cdot \frac{\sqrt{n-2}}{1-r^2} \right| = \left| r \cdot \sqrt{\frac{n-2}{\sigma_{ocm}^2}} \cdot \sigma_y \right| \quad (6.15)$$

$\nu = n - 2$ – число степеней свободы.

$$\sigma_{ocm}^2 = \sigma_{\text{м.з.}}^2 - \sigma_{ocm}^2 \quad (6.16)$$

$$\sigma = \sqrt{\sigma_{ocm}^2} \quad (6.17)$$

t табличное находим по значению ν преимущественно при вероятности $p=0.90$; $p=0.95$; $p=0.99$.

В статистике принято для каждого параметра определять доверительные интервалы. Это можно сделать, используя коэффициенты Стьюдента.

$$a_i \pm t \cdot \Delta a_i \quad (6.18)$$

$$\Delta a_0 = \frac{\sigma_{ocm}}{\sqrt{n-2}} \quad (6.19)$$

$$\Delta a_i = \frac{\sigma_{ocm}}{\sigma_x \cdot \sqrt{n-2}} \quad (6.20)$$

В 7. Расчет параметров криволинейных уравнений.

Гипербола:

$$\tilde{Y}_x = a_0 + \frac{a_1}{X} \quad (6.21)$$

$$a_0 = \frac{\sum Y \sum \frac{1}{X} - \sum \frac{Y}{X} - \sum \frac{1}{X}}{n \sum \frac{1}{X^2} - \sum \frac{1}{X} - \sum \frac{1}{X}} \quad (6.22)$$

$$a_1 = \frac{n \sum \frac{Y}{X} - \sum Y \sum \frac{1}{X}}{n \frac{1}{X^2} - \sum \frac{1}{X} - \sum \frac{1}{X}} \quad (6.23)$$

Степенная функция:

$$\tilde{Y}_x = a_0 X^{a_1} \quad (6.24)$$

$$\lg \tilde{Y}_x = \lg a_0 + a_1 \lg X \quad (6.25)$$

$$\lg a_0 = \frac{\sum \lg Y \sum (\lg Y)^2 - \sum (\lg Y \lg X) \sum \lg X}{n \sum (\lg X)^2 - \sum \lg X \sum \lg X} \quad (6.26)$$

$$a_1 = \frac{n \sum (\lg Y \lg X) - \sum \lg Y \sum \lg X}{n \sum (\lg X)^2 - \sum \lg X \sum \lg Y} \quad (6.27)$$

Парабола:

$$\tilde{Y}_x = a_0 + a_1 X + a_2 X^2 \quad (6.28)$$

$$\begin{cases} a_0 n + a_1 \sum X + a_2 \sum X^2 = \sum Y \\ a_0 \sum X + a_1 \sum X^2 + a_2 \sum X^3 = \sum YX \\ a_0 \sum X^2 + a_1 \sum X^3 + a_2 \sum X^4 = \sum YX^2 \end{cases} \quad (6.29)$$

В 8. Оценка тесноты связи при криволинейных моделях.

В отличие от линейной корреляционно-регрессионной зависимости, при криволинейных связях тесноту связи признаков определяют по индексам корреляции и детерминации. Смысл индексов тот же, что и коэффициентов корреляции и детерминации.

Схема нахождения индекса детерминации:

1. Определяют общую дисперсию

$$\sigma^2 = \frac{\sum (Y_i - \bar{Y})^2}{n} \quad (6.30)$$

2. Определяют меру колеблемости расчетных значений признака около среднего:

$$\sigma^2_{\tilde{y}_x} = \frac{\sum (\tilde{Y}_x - \bar{Y})^2}{n} \quad (6.31)$$

3. Определяют меру колеблемости результативных показателей вокруг теоретических значений, вычисленных по уравнению регрессии.

$$\sigma^2_{остат} = \frac{\sum (Y_i - \tilde{Y}_x)^2}{n} \quad (6.32)$$

4. В математической статистике доказывают, что общая дисперсия равна факторной дисперсии плюс остаточная дисперсия:

$$\sigma^2(Y) = \sigma^2_{\tilde{y}_x} + \sigma^2_{остат} \quad (6.33)$$

Индексом детерминации i^2 называется отношение факторной дисперсии к общей дисперсии, т.е.

$$i^2 = \frac{\sigma^2_{\tilde{y}_x}}{\sigma^2_y} \quad (6.34)$$

Этот индекс объясняет, какая часть вариации результативного признака (**Y**) объясняется вариацией факторного признака (**X**). Индекс детерминации может быть выражен в процентах.

5. Индекс корреляции равен корню квадратному из индекса детерминации взятого в долях (а не в процентах):

$$i = \sqrt{i^2} \quad (6.35)$$

Индекс корреляции всегда положителен и не может указывать направление связи. Например: при параболической связи

признаков одна ветвь показывает возрастание, другая убывание факторов и общего направления быть не может.

6. Достоверность индексов корреляции и детерминации при криволинейных связях оценивают по критерию Фишера :

$$F_{рас} = \frac{\sigma_{\tilde{y}x}^2}{\sigma_{остат}^2} \times \frac{n-m}{m-1} = \frac{i^2}{1-i^2} \times \frac{n-m}{m-1} \quad (6.36)$$

где : **n** — число единиц в совокупности,

m — число параметров уравнения.

F расчетное сравниваем с **F** табличным. Строку в таблицах находим по значению

$$\nu_1 = n - m ; \quad (6.37)$$

столбец по значению

$$\nu_2 = m - 1 \quad (6.38)$$

Если **F** расчетное больше **F** табличного, то нуль-гипотеза отвергается и принимается утверждение, что криволинейная связь в виде полученного уравнения регрессии, индекс корреляции и детерминации достоверны. Если же соотношение $F_{расч.} < F_{табл.}$, то наша модель недостоверна.

В 9. История развития многофакторного корреляционно-регрессионного анализа.

Начало корреляционного и регрессионного анализа относится ко второй половине XIX века и связано с именем двоюродного брата Ч. Дарвина – Френсисом Гальтоном (1822–1911). Он ввел понятие «закона регрессии», связав его со средним снижением роста сыновей по сравнению с ростом отцов (1899 г). Ему же принадлежит введение числовой меры, оценивающей силу связи показателей (корреляцию). Поэтому началом разработки корреляционно-регрессионного анализа ученые-статистики считают статью Ф. Гальтона «Регрессия, наследственность и панмиксия» (1896 г), в которой автор «дал определение корреляции, построил теоретическую модель совместного измерения двух переменных, ввел понятие линии регрессии и корреляционного индекса «г».

Экономистами и математиками разработаны различные модели, оценивающие влияние нескольких факторов на результат. Например, в США в 1929 г. при анализе развития обрабатывающей промышленности за 1899–1922 гг. была построена мультипликативная производственная функция, отражающая зависимость выхода продукции от затрат живого труда и наличия капитала, которая получила название функции Кобба-Дугласа (там же, с.176):

$$Y = a L^{\alpha} K^{\beta}, \quad (6.39)$$

где Y – объем выпускаемой продукции, a – коэффициент размерности, L – объем затрат живого труда или численность работников, K – объем капитала (основного или совокупного), α и β – коэффициенты эластичности производства продукции по труду и капиталу.

Для сельского хозяйства данную модель можно расширить, включив в нее еще один важный ресурс – это площадь сельскохозяйственных угодий (S) с соответствующим коэффициентом эластичности. Тогда данная функция будет иметь вид:

$$Y = a L^{\alpha} K^{\beta} S^{\gamma}, \quad (6.40)$$

Существуют и другие нелинейные модели, отражающие различные связи факторов, некоторые из которых рассматриваются в курсе «Планирование и прогнозирование». Однако большинство из них не могут быть интерпретированы системой экономических параметров и сложны в экономическом обосновании.

Например, что может отражать экономический логарифм производительности труда или производительность труда в степени n и так далее? То есть модели могут быть использованы для прогнозов без экономической интерпретации параметров. Наилучшим образом экономически интерпретируется многофакторные линейные модели вида:

$$\approx Y_{x_1 x_2 x_3 \dots x_n} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (6.41)$$

где: n – отражает число факторов.

При составлении модели встает вопрос отбора факторов, которые могут быть включены в многофакторную модель. Как правило, имеется таблица с базой данных, где указаны числовые значения факторов интересующих исследователя. Общий вид такой таблицы с информацией о значениях технико-экономических показателей следующий:

Таблица 6.3. База данных для исследования связей факторов

№ п/п	Y	X ₁	X ₂		X _n
1					
2					
...					
N					

При исследовании необходимо решить целый ряд проблем, одна из которых заключается в отборе факторов для их включения в модель (уравнение регрессии). Здесь существенную роль играют знания исследователя об экономических закономерностях развития процессов и явлений, знания экономики конкретной отрасли (сельского хозяйства легкой и тяжелой промышленности, транспорта и т.д.). После сбора информации и проведения ее априорного анализа можно провести расчеты, позволяющие достаточно качественно отобрать факторы для проведения корреляционно-регрессионного анализа.

Для отбора факторов для модели часто используют матрицу парных коэффициентов корреляции, расчет и интерпретация которых рассматривалась в вопросах 5 и 6 темы.

Таблица 6.4. Матрица парных коэффициентов корреляции

	Y	X ₁	X ₂		X _n
Y	1	r_{YX1}	r_{YX2}		r_{YXn}
X ₁		1	r_{X1X2}		r_{X1Xn}
X ₂			1		r_{X2Xn}
....					
X _n					1

Матрицу используют следующим образом:

По строке $\mathbf{Y}$ анализируют значения парных коэффициентов корреляции r_{ij} и отбирают в модель те факторы, для которых $r_{iy} > 0,2$.

Используя остальные строки матрицы, устанавливают наличие или отсутствие мультиколлинеарности факторов. Если выявляется наличие таких пар факторов, то в модель включают только один из них.

Факторы являются **мультиколлинеарными**, если связь между ними близка к функциональной или функциональна.

Например, производительность труда и трудоемкость – показатели обратные друг другу, и для них парный коэффициент корреляции равен единице. В модель можно включить только один из этих факторов.

Можно считать, что фактор является незначимым, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не изменяя суммы квадратов остатков, то есть:

$$\sum (Y_i - \bar{Y})^2 \approx \text{const} \quad (6.42)$$

Если при включении в модель факторного признака увеличивается величина множественного коэффициента корреляции и детерминации, а коэффициенты регрессии меняются незначительно, то данный признак существен, и его включение в уравнение регрессии обязательно.

В 10. Нахождение параметров уравнения многофакторной корреляционно-регрессионной модели.

Параметры уравнения регрессии (6.41), то есть коэффициенты регрессии a_i и свободный член a_0 находят из системы (6.43) нормальных уравнений по методу наименьших квадратов.

$$R(I) = \sqrt{\frac{\sigma_{\Sigma}^2}{\sigma_{Y_i}^2}} = \sqrt{R^2} \quad (6.45)$$

Совокупный коэффициент детерминации R^2 может быть выражен в процентах. Он показывает, какая часть вариации результативного показателя объясняется вариацией факторов, включенных в модель.

Совокупный коэффициент корреляции всегда $0 < R < 1$. Он отражает только тесноту связи и не может отражать направление связи (как парный коэффициент корреляции). Чем ближе значение R к 1, тем влияние факторов на результат сильнее, чем ближе к 0 – тем влияние слабее.

Расчет совокупного коэффициента детерминации можно произвести, используя связь его с парными коэффициентами корреляции r_{ij} и коэффициентами регрессии в стандартизированном виде, т.е. β –коэффициенты (см. вопрос 14)

$$R^2 = \beta_1 \cdot r_{yx_1} + \beta_2 \cdot r_{yx_2} + \dots + \beta_n \cdot r_{yx_n} \quad (6.46)$$

Если рассматривается зависимость результата от двух факторов, то расчет совокупных коэффициентов корреляции и детерминации можно упростить, используя значения парных коэффициентов корреляции и детерминации.

$$R^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_2x_1}^2} \quad (6.47)$$

При множественной корреляционно-регрессионной связи необходимо выделить тесноту связи результативного показателя индивидуально с каждым фактором, для чего вычисляют **коэффициенты раздельной корреляции и детерминации**.

Коэффициентом раздельной детерминации называется произведение парного коэффициента корреляции фактора X_i на его β –коэффициент

$$d_i^2 = r_{yx_i} \cdot \beta_i \quad (6.48)$$

Последняя формула отражает тоже равенство, что и формула (6.46). Корень квадратный из коэффициента раздельной детерминации даст коэффициент раздельной корреляции.

$$\sum d_i^2 = R^2 \quad (6.49)$$

При построении уравнения регрессии важным моментом является последовательность включения факторов в уравнение регрессии. И здесь большую роль играет системная связь между каждой парой факторов, включенных в модель, и их группами. Поэтому важным представляется выделение дополнительной доли вариации результивного показателя (Y) после включения в модель дополнительно фактора X_k . Такая вариация объясняется **частными коэффициентами корреляции и детерминации**.

В общем виде частный индекс или коэффициент детерминации находят по формуле:

$$I_{Yn}^2 = \frac{\sigma_{0,1,2,\dots(n-1)}^2 - \sigma_{Y1/2/\dots(n-1)}^2}{\sigma_{Y1/2/\dots(n-1)}^2} \quad (6.50)$$

Как правило, **частные коэффициенты** корреляции и детерминации **меньше парных коэффициентов** корреляции и детерминации.

В случае анализа модели Y по двум факторам X_1 и X_2 для расчета частных коэффициентов корреляции можно использовать следующие формулы:

$$r_{y x_1 (x_2)} = \frac{r_{y x_1} - r_{x_1 x_2} \cdot r_{y x_2}}{\sqrt{(1 - r_{y x_2}^2) \cdot (1 - r_{x_1 x_2}^2)}} \quad (6.51)$$

$$r_{y x_2 (x_1)} = \frac{r_{y x_2} - r_{x_1 x_2} \cdot r_{y x_1}}{\sqrt{(1 - r_{y x_1}^2) \cdot (1 - r_{x_1 x_2}^2)}} \quad (6.52)$$

В формуле 6.51 отражена связь между Y и X_1 при условии неизменности X_2 , в формуле 6.52 – связь между Y и X_2 при условии постоянства X_1 .

Частные коэффициенты детерминации найдем, возведя в квадрат частные коэффициенты корреляции. Их сумма близка к значению совокупного коэффициента детерминации.

Однако не следует упрощать смысл анализируемых показателей связи, т.к. вопросы анализа силы влияния факторов на результативный показатель можно рассматриваться в зависимости от последовательности включения факторов в модель от их «системного» влияния и т.д. Многие проблемные вопросы оценки силы влияния факторов на результативный показатель рассматриваются в современных учебниках статистики российских авторов.

Следующая группа показателей, отражающих связи факторов, включенных в модель, – это **коэффициенты эластичности** и β – **коэффициенты**.

Коэффициенты эластичности вычисляются на базе первых частных производных от функции связи.

Коэффициент эластичности показывает, на сколько процентов в среднем изменится результат (Y) при изменении фактора X_i в среднем на 1% при условии неизменности остальных факторов, входящих в модель.

$$\varepsilon_i = \frac{\partial \bar{y}}{\partial x_i} = a_i \frac{\bar{X}_i}{\bar{Y}} \quad (6.53)$$

$$\beta_i = a_i \frac{\sigma_{x_i}}{\sigma_{y_i}} \quad (6.54)$$

β – **коэффициент** показывает, на сколько среднеквадратических отклонений изменяется результат (Y) при изменении фактора X_i на одно свое среднеквадратическое отклонение, при неизменности остальных факторов входящих в уравнение.

Примечание:

Для парной линейной регрессии выполняется равенство $\beta = r$. Поэтому в парном корреляционно-регрессионном анализе β – коэффициент не рассматривался.

В 12. Оценка достоверности результатов произведенного корреляционно-регрессионного анализа.

Одним из требований при построении многофакторных моделей является требование к объему анализируемой совокупности, т.е. выборка должна быть репрезентативной (представительной). Однако это требование не всегда выполняется. Поэтому рассмотрим вопросы оценки достоверности полученных параметров уравнения и тесноты связи как для достаточно большой совокупности, так и для малой выборки.

а) Для репрезентативной выборки:

Оценка происходит по той же схеме, что и при парной линейной зависимости, для чего могут быть использованы критерий Стьюдента и критерий Фишера.

Расчет параметров t-критерия и F-критерия. Для каждого частного коэффициента регрессии рассчитывается значение t-критерия по формуле:

$$t_i = \frac{|a_i|}{\sqrt{\sigma_{a_i}^2}} \quad (6.55)$$

где в знаменателе стоит дисперсия частного коэффициента регрессии

$$\sigma_{a_i} \approx \frac{\sigma_y \sqrt{1 - R^2}}{\sigma_{x_i} \sqrt{n} \sqrt{1 - R_i}} \quad (6.56)$$

где: R_i — величина множественного коэффициента корреляции по фактору X_i с остальными факторами.

Однако проще использовать F-критерий, т.к. с его помощью можно оценить достоверность всех полученных показателей (параметров и числовых характеристик).

$$F_{pac} = \frac{\frac{1}{n+1} \sum (\tilde{Y})^2}{\frac{1}{N-n-1} \cdot \sum (Y_i - \tilde{Y})^2} \quad (6.57)$$

где: n — число факторов в модели,
 N — объем совокупности.

Табличное значение F-критерия найдем по таблицам Фишера, определив столбец по значению числа степеней свободы $v_1=n+1$, строку – по $v_2=N-n-1$.

Если $F_{\text{расч.}} \geq F_{\text{табл.}}$, то нуль-гипотеза отвергается и подтверждается достоверность произведенного корреляционно-регрессионного анализа.

Если при построении модели используется только два фактора (X_1 и X_2), то можно использовать упрощенную формулу

$$F_{\text{расч.}} = \frac{\frac{1}{2} R^2}{\frac{1}{N-3} (1-R^2)} \quad (6.58)$$

Табличное значение F-критерия находим по значениям $v_1=2$, $v_2=N-3$.

v_1 – определяет графу,

v_2 – строку таблицы.

б) Для малой выборки:

При небольшом числе наблюдений (а это часто бывает при исследовании небольшой совокупности, например, только по хозяйствам одного-двух районов), величина множественного коэффициента корреляции и детерминации завышается. Поэтому чтобы оценить реальную тесноту связи и ее достоверность, необходимо произвести следующие расчеты. Сначала проверим выполнение соотношения

$$\frac{N-n}{n} \geq 20$$

Если это соотношение выполняется, то все дальнейшие расчеты выполняем по пункту а), если не выполняется, то необходимо скорректировать значение множественного коэффициента корреляции и оценить его достоверность.

Рассчитаем скорректированный совокупный коэффициент корреляции

$$R_{\text{скор}}^2 = R^2 \frac{N-1}{N-n}$$

$$R_{\text{скор}} = \sqrt{R^2 \frac{N-1}{N-n}} \quad (6.59)$$

Произведем оценку достоверности скорректированного множественного коэффициента корреляции, используя формулы (6.57) или (6.58) и соответствующий алгоритм использования критерия Фишера.

В 13. Прогнозирование по корреляционно-регрессионной модели.

Поскольку корреляционно-регрессионные модели статичны, то есть привязаны к месту и времени, то они обеспечивают прогноз на небольшом временном интервале при стабильных условиях экономической ситуации.

Различают прогнозы:

1. Наилучшие.
2. Наихудшие.
3. Усредненные.

Наилучшего значения результат **Y** будет достигать, если в модель подставить наилучшее значение факторов X_i .

Наихудшее значение результат **Y** будет достигать, если в уравнение связи подставить самые “неблагоприятные” значения факторов X_i .

Усредненное значение результат **Y** будет достигать, если в модель подставить среднее значение факторов.

Такие модели и прогнозы по ним хорошо работают для конкретных предприятий, т.к. можно рассчитать возможные значения результативного показателя при возможных значениях факторных признаков. Составляют, как правило, оптимистические и пессимистические прогнозы.

В 14. Стандартизированный вид уравнения связи.

В экономическом анализе часто возникает необходимость сопоставить эффективность работы предприятий по многим факторам. При этом приходится учитывать, что предприятия находятся в различных экономических условиях, обладают различными ресурсами и т.д. В этих случаях исходную информацию необходимо привести к какому-либо сопоставимому виду. Таким «видом» является стандартизированный вид. А далее ис-

пользуются методы многомерного сопоставления или кластерный анализ.

В курсе математической статистики в разделе «Выборочный метод и законы распределения случайной величины» устанавливается зависимость между индивидуальными значениями показателя и его средним квадратичным отклонением. Введем обозначения новых **стандартизированных переменных, которые отражают долю отклонения индивидуального значения переменной от среднего значения относительно среднеквадратического отклонения:**

$$Z = \frac{Y_i - \bar{Y}_i}{\sigma_y} \quad t_i = \frac{X_i - \bar{X}_i}{\sigma_x} \quad (6.60)$$

где Z_i – стандартизированная результативная переменная.

Значения t_i являются новыми факторными переменными в нашей многофакторной модели, которая называется стандартизированной моделью и выглядит следующим образом:

$$\tilde{Z}_{t_i} = \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_n t_n \quad (6.61)$$

где стандартизированными коэффициентами регрессии являются β – коэффициенты. В стандартизированном уравнении нет свободного члена.

Стандартизированное уравнение регрессии может быть найдено (построено) несколькими способами:

1) Первый способ заключается в том, чтобы, не находя значений стандартизированных переменных, рассчитать значения β -коэффициентов по известной нам формуле (6.54)

$$\beta_i = a_i \frac{\sigma_x}{\sigma_y}. \quad \text{И затем записать уравнение вида (6.61)}$$

2) Второй способ заключается в том, чтобы, используя исходную базу данных значений всех переменных (результативно-го и факторных признаков), по формулам (6.60) построить таблицу стандартизированных переменных

Таблица 6.5. Стандартизированные переменные

№ п/п	Z_i	t_{1i}	t_{2i}	...	t_{ni}
1					
2					
N					

Используя значения стандартизированных переменных, можно построить и решить следующую систему линейных уравнений относительно переменных t_i :

$$\begin{cases} \beta_1 \sum t_1^2 + \beta_2 \sum t_1 t_2 + \beta_3 \sum t_1 t_3 + \dots \beta_n \sum t_1 t_n = \sum t_1 Z \\ \beta_1 \sum t_1 t_2 + \beta_2 \sum t_2^2 + \beta_3 \sum t_2 t_3 + \dots \beta_n \sum t_2 t_n = \sum t_2 Z \\ \dots \\ \beta_1 \sum t_n t_1 + \beta_2 \sum t_n t_2 + \beta_3 \sum t_n t_3 + \dots \beta_n \sum t_n^2 = \sum t_n Z \end{cases} \quad (6.62)$$

Эта система на один параметр в уравнении меньше, чем система (6.43).

Значения переменных таблицы 6.5 могут быть использованы для многомерного или кластерного анализа, которые в данном курсе не рассматриваются.

3) Третий способ заключается в использовании матрицы парных коэффициентов корреляции для нахождения значений β -коэффициентов.

В математической статистике доказывается, что $r_{ij} = r_{ji}$ и $r_{ij} = (1/N) \sum t_i t_j$. Кроме того, мы уже отмечали, что парный коэффициент корреляции равен β -коэффициенту. Поэтому оказывается возможным использовать для расчетов при построении стандартизированного уравнения матрицу парных коэффициентов корреляции.

$$\begin{cases} \beta_1 r_{11} + \beta_2 r_{12} + \dots + \beta_n r_{1n} = r_{x_1 y} \\ \beta_1 r_{21} + \beta_2 r_{22} + \dots + \beta_n r_{2n} = r_{x_2 y} \\ \dots \\ \beta_1 r_{n1} + \beta_2 r_{n2} + \dots + \beta_n r_{nn} = r_{x_n y} \end{cases} \quad (6.63)$$

β -коэффициенты находим по формулам Крамера или методом Гаусса или другим способом.

$$\beta_i = \frac{\Delta_i}{\Delta} \quad (6.64)$$

где Δ_i – частный определитель,
 Δ – главный определитель.

В 15. Рекомендации по составлению различных корреляционно-регрессионных моделей.

Все виды моделей можно классифицировать следующим образом:

1) Объемные или количественные модели. Это такие модели, где и результативный признак и факторы, на него влияющие, являются объемными или количественными показателями. К таким моделям можно отнести зависимость выхода продукции от наличия (объема) различных ресурсов.

2) Интенсивные модели. В них все показатели являются качественными или интенсивными. Например, зависимость уровня себестоимости единицы продукции растениеводства от удельных затрат труда на 1 га, от урожайности, от стоимости 1 тонны удобрений НРК действующего вещества и т.д.

3) Третий вид моделей – смешанный, где результативным показателем выступает объемный признак, а среди факторов могут быть как объемные, так и интенсивные или качественные признаки. Здесь главное выдержать наличие технико-экономической связи показателей, включенных в модель.

Однако существуют рекомендации, чего нельзя допускать в многофакторных корреляционно-регрессионных моделях.

1) Нельзя допускать включение в модель факторов, которые сами зависят от результативного показателя. Например, в

модель себестоимости продукции нельзя включать уровень рентабельности, т.к. он сам зависит от себестоимости.

2) Признаки-факторы не должны быть составными частями результативного показателя. Например, нельзя строить модель зависимости уровня себестоимости от составляющих ее статей или элементов затрат.

3) Как уже говорилось в вопросе 1, нельзя включать в модель факторы, которые относительно друг друга являются мультиколлинеарными.

4) Нельзя в модель интенсивного типа включать объемные факторы. Например, размер посевной площади никак не может влиять на урожайность.

Тема 7. Ряды динамики

1. Понятие о динамических рядах. Виды рядов динамики.
2. Показатели ряда динамики.
3. Приемы выравнивания рядов динамики.
4. Тренд. Аналитическое выравнивание ряда. Прогнозы по тренду.
5. Сезонные колебания в рядах динамики.
6. Модели рядов с учетом тенденции и сезонности.
7. Корреляция рядов динамики. Регрессия рядов динамики.
8. Понятия интерполяции и экстраполяции в рядах динамики.

ЛИТЕРАТУРА:

1. Елисеева И.И., Юзбашев М.М. *Общая теория статистики*: М.: Финансы и статистика, 1998.
2. Ефимова М.Р., Петрова В.В. *Общая теория статистики. Учебник* / – 2-е издание., – М.: ИНФРА-М, 2002, – 413 с.
3. Едронова В. С. *Общая теория статистики: Учебник* / – М.: Юристъ, 2001, – 511 с
4. Кривенкова Л.Н. *Статистические методы анализа и моделирования свиноводства (на уровнях отдельного предприятия и региона). Диссертация на соискание ученой степени к.э.н.* С.Петербург, 1992, – 149с.
5. Плошко Б.Г., Елисеева И.И. *История статистики: Учеб.пособие* / – М.: Финансы и статистика, 1990, – 295с.

6. *Статистика: показатели и методы анализа: справ. пособие/ Под ред. М.М. Новикова. – Мн.: “Современная школа”, 2005. – 268 с.*
7. *Сельское хозяйство РБ: статистический сборник / ред.кол.: В.С. Метекс; В.Н. Синкевич; Л.Л. Выбчик и др.; Министерство статистики и анализа РБ. – Минск, 2006. – 223 с.*
8. *Теория статистики: Учебник / Под ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 199. – 464 с.*

В 1. Понятие о динамических рядах. Виды рядов динамики.

Согласно основному диалектическому методу анализа в статистике показатели рассматриваются в динамике, т.е. во времени. Государственными органами статистики выпускаются сборники, отражающие обобщенную информацию по годам.

К концу XIX века статистический учет всех важнейших областей общественной жизни начали проводить систематически на основе периодически повторяемых и текущих работ. У истоков исследований рядов динамики стоит бельгиец по национальности, математик по образованию Адольф Кетле (1796–1874). Особое внимание он уделял уголовной статистике Франции и выделял среди преступлений такие, число которых во времени почти не меняется и для которых орудия убийства остаются постоянными.

Теория устойчивости возникла как способ, позволяющий объяснить повторяемость данных в динамике, выявленную А. Кетле. Автор этой теории – немецкий статистик и экономист Вильгельм Лексис (1837–1914). Он ввел деление рядов динамики на типы:

- эволюторный (главное проявление основных тенденций),
- ундуляторный (волнообразное изменение во времени),
- периодический (правильное повторение волн),
- осцилляторный (беспорядочные колебания уровней).

Современная статистика рассматривает как динамику отдельных технолого-экономических показателей, так и группы показателей, связанных друг с другом во времени, используя при этом современные приемы исследования.

Ряды статистических величин, характеризующие изменение явлений во времени, называются **динамическими или хронологическими, или временными рядами**.

Ряды динамики состоят из двух элементов: Y – уровень ряда и t – время, к которому эти уровни относятся.

Чаще всего информация задается в виде вертикальных (таблица 7.1) или горизонтальных (таблица 7.2) таблиц:

Таблица 7.1. База данных ряда динамики

№ п/п	Время (t)	Уровень ряда (Y_i)
1	2000	13
2	2001	18
3	2002	14
...	...	...
n	2010	21

Таблица 7.2. База данных ряда динамики

№ п/п	1	2	3	...	n
Время (t)	2000г.	2001г.	2002г.	...	2008г.
Уровень ряда, (Y_i)	347	384	315	...	397

Ряд динамики может быть нанесен на декартову систему координат, где время отражается по оси абсцисс, а значения анализируемого показателя – по оси ординат. Точки ряда динамики соединяют ломаной линией (рисунок 7.1).

Например, при исследовании развития сельскохозяйственного предприятия можно рассматривать динамику ресурсов и их соотношения (площадь с.-х. угодий, стоимость ОПФ, численность работающих, поголовье животных, фондовооруженность и фондообеспеченность и т.д.), показатели урожайности с.-х. культур и продуктивности животных, финансовые результаты производственной деятельности, показатели финансового состояния и др. Поэтому уровни ряда могут быть любой статистической величиной, т.е. объемными показателями, интенсивными (качественными) или относительными.

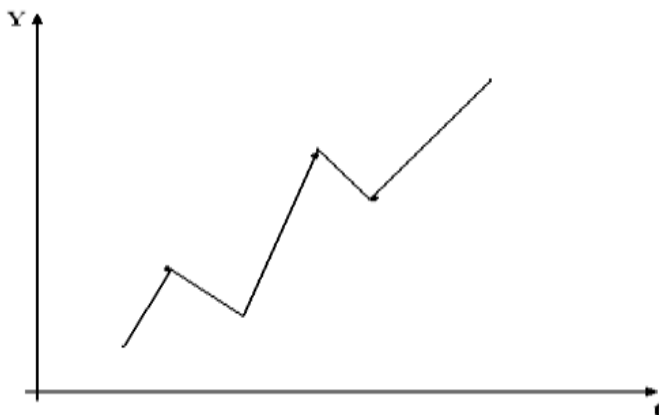


Рисунок 7.1. Ломаная ряда динамики

Параметр t может отражать временной интервал или определенный момент времени, к которому относится значение анализируемого показателя: **в первом случае ряд называется интервальным, во втором – моментным.** Это один из способов классифицировать ряды динамики (но существуют и другие классификации: по способу выражения уровней ряда, в зависимости от расстояния между уровнями ряда, при наличии или отсутствии основной тенденции). Например:

1) интервальный ряд

Таблица 7.3. Средний объем реализации продуктов питания на 1 человека в месяц, млн. руб.

Y_i	3,1	3,2	2,9	...
t	январь	февраль	март	...

В интервальных рядах величина t может быть различной: неделя, декада, месяц, год, пятилетка и т.д., в зависимости от цели и задач исследования.

2) моментный ряд

Таблица 7.4. Температура тела человека ежедневно в 18.00

Y_i	36,7	36,0	36,9	...
t	Понедельник	Вторник	Среда	...

В моментных рядах параметр времени точно указывает время, к которому относится каждое значение уровня ряда. К моментным рядам динамики в сельском хозяйстве можно отнести учет выходного поголовья в животноводстве на первое число каждого месяца или списочную численность работников на начало месяца.

При проведении исследований рядов динамики, чтобы не искажались результаты и выводы, необходимо соблюдать определенные требования.

Условия правильного построения рядов:

1. Сопоставимость рядов по территории. Несопоставимость может возникнуть при административном пересмотре границ сельскохозяйственных предприятий, бригад, районов, областей и т.д.

2. Сопоставимость по единицам измерения или единицам счета. Несопоставимость данного вида возникает при различном подходе к расчетным показателям. Например, производительность труда может быть вычислена как выход продукции на 1 работника или на 1 рабочего.

3. Сопоставимость по продолжительности анализируемых периодов, к которым относятся уровни ряда. Такая несопоставимость возникает, если проводить помесечный учет, когда “длина” одного месяца колеблется и различие в продолжительности в 1-3 дня может привести к искажению информации.

Существуют методики приведения рядов к сопоставимому виду.

В 2. Показатели ряда динамики.

При изучении динамики явлений статистика решает ряд задач:

1. Оценка скорости изменений явления во времени и выявление особенностей ряда.

2. Определение основной тенденции или тенденций на определенных интервалах времени.

3. Выявление факторов, обуславливающие изменение тенденции.

4. Расчет прогнозов развития явления (показателей) на перспективу.

Первая задача решается путем расчета следующих показателей ряда динамики:

Δ – абсолютный прирост;

K – темп роста (или коэффициент роста);

T – темп прироста;

A – абсолютное значение 1% прироста.

Первые 3 показателя вычисляются двумя способами: цепным и базисным, последний – преимущественно цепным.

При цепном способе расчета сравниваются два соседних уровня ряда. При базисном способе все показатели ряда сопоставляются с одним определенным уровнем, взятым за базу сравнения. Чаще всего за базисный уровень принимается начальный уровень ряда. Кроме того, вычисляются средние показатели ряда:

- 1) средний уровень ряда;
- 2) средний абсолютный прирост;
- 3) средний темп роста;
- 4) средний темп прироста.

Формулы для расчета показателей ряда динамики.

Подстрочные символы “ц” и “б” отражают цепной и базисный способы расчета показателей ряда динамики.

Абсолютный прирост показывает, на сколько больше или меньше значение уровня ряда с тем значением, с которым проводится сравнение (соседним или выбранным за базу сравнения).

$$\Delta_{\text{ц}} = Y_i - Y_{i-1} \quad \text{или} \quad \Delta_{\text{б}} = Y_i - Y_0 \quad (7.1)$$

Коэффициент роста (темп роста) показывает, во сколько раз больше или меньше значение уровня ряда с тем значением, с которым проводится сравнение (соседним или выбранным за базу сравнения). Коэффициент роста может быть выражен в процентах.

$$K_{\text{ц}} = \frac{Y_i}{Y_{i-1}} \cdot [100\%] \quad \text{или} \quad K_{\text{б}} = \frac{Y_i}{Y_0} \cdot [100\%] \quad (7.2)$$

Темп прироста характеризует относительную скорость изменения уровня ряда в единицу времени. Темп прироста показывает, на какую долю (или процент) уровень ряда рассматри-

ваемого периода больше или меньше базового (выбранного для сравнения).

$$T = \frac{\Delta_i}{Y_i} \cdot 100\% \text{ или } T_{ц,б} = K_{ц,б} - 1 \text{ или } T_{ц,б} = K_{ц,б} \% - 100\%; \quad (7.3)$$

Предыдущий показатель не отражает «вес» каждого процента прироста. Например, в двух бригадах урожайность зерновых выросла по сравнению с предыдущим годом на 1,2%. Но это не значит, что в натуральном выражении это будет составлять одинаковое количество центнеров. Оценить каждый процент изменения позволяет следующий показатель.

Абсолютное значение 1% прироста показывает натуральную величину показателя, приходящегося на 1% изменения относительно базисного (взятого за основу сравнения) значения.

$$A = \frac{\Delta_u}{T_u} \quad (7.4)$$

Формулы для расчета средних значений показателей ряда динамики:

1) Средний уровень ряда может быть вычислен по одной из формул (7.5 – для интервальных рядов; 7.6 – для моментных рядов)

$$\bar{Y} = \frac{\sum Y_i}{n} \text{ или } \bar{Y} = \frac{\sum Y_i \cdot t}{\sum t} \quad (7.5)$$

$$\bar{Y} = \frac{0,5Y_1 + Y_2 + \dots + Y_{n-1} + 0,5Y_n}{n-1} \quad (7.6)$$

2) Средний абсолютный прирост:

$$\overline{\Delta_u} = \frac{\sum \Delta_i}{n-1} \text{ или } \overline{\Delta_o} = \frac{Y_n - Y_0}{n-1} \quad (7.7)$$

где (n – 1) – число периодов.

3) Средний темп роста:

$$\overline{K}_y = \sqrt[n]{K_1 \cdot K_2 \cdot K \cdot \dots \cdot K_n}, \quad (7.8)$$

где n – количество цепных темпов роста уровня ряда

$$\overline{K}_o = \sqrt[n-1]{\frac{Y_n}{Y_0}}, \quad (7.9)$$

где n – число уровней ряда

4) Средний темп прироста:

$$\overline{T}_{y,o} = \overline{K} - 1 \quad \text{или} \quad \overline{T}_{\%} = \overline{K} \% - 100\% \quad (7.10)$$

Средний уровень ряда рассчитаем по формуле средней арифметической простой $Y=1145,057$

Средний темп роста вычислим по любой из формул 8.8 или 8.9 и получим один и тот же результат $K=0,9913=99,13\%$. Это значение отражает среднее ежегодное снижение численности населения по сравнению с предыдущим периодом.

Средний темп прироста (снижения) $T=-0,0087=-0,87\%$, т.е. ежегодно на 0,87% уменьшалась численность населения РБ.

Абсолютное среднее ежегодное уменьшение численности населения будут равно $\Delta=-9,97$ тыс. чел.

Все эти значения показывают ухудшающуюся демографическую ситуацию в РБ.

Например:

Таблица 7.5. Численность сельского населения РБ на начало года, тыс. чел.

Г о д ы	Y	Δ		K		T		Δ
		ц	б	ц	б	ц	б	ц
1	1173,9	-	-	-	-	-	-	-
2	1166,2	-7,7	-7,7	0,993441	0,993441	-0,006559	-0,006559	1173,9
3	1156,5	-9,7	-17,4	0,991682	0,985178	-0,008318	-0,014822	1166,2
4	1146,1	-10,4	-27,8	0,991007	0,976318	-0,008993	-0,023682	1156,5
5	1135,1	-1,1	2309	0,990402	0,966948	-0,009598	-0,033052	1146,1
6	1123,5	-11,6	-50,4	0,989781	0,957066	-0,010219	-0,042934	1135,1
7	1114,1	-9,4	-59,8	0,991633	0,949059	-0,008367	-0,050941	1123,5

В 3. Приемы выравнивания рядов динамики.

Часто в рядах динамики сложно определить общую тенденцию развития. Второй задачей рядов динамики является выявление таких периодов развития, которые достаточно однородны по своим условиям и взаимодействиям связей между показателями.

Под тенденцией понимается общее направление к росту, снижению или стабилизации уровня явления с течением времени.

Тенденция по периодам может не совпадать с общей тенденцией. Иногда вообще сложно определить общую тенденцию. В этом случае прибегают к различным методам выравнивания ряда для определения его тенденции:

- 1) способу укрупнения интервалов;
- 2) сглаживанию ряда с помощью скользящей средней;
- 3) производят аналитическое выравнивание и получают уравнение тренда.

Первый способ заключается в переходе от меньших временных интервалов к большим и расчете усредненных уровней ряда за укрупненный интервал. Временной промежуток для укрупнения интервала производят с учетом общей длины анализируемого ряда и величины исходных интервалов. Так, например, если исходный ряд содержит информацию за каждый месяц, то можно перейти к укрупненному интервалу величиной в квартал. Если представлена информация по годам, то «укрупнение» можно произвести за 2,3,4,5 лет в зависимости от длины исходного ряда.

График этого ряда изображен на рисунке 7.2, где видно, что имеются периоды спада производства продукции и периоды подъема. Если рассмотреть только часть графика, начиная с 1998 года по 2005, то сложно установить общую тенденцию производства продукции данного вида. Поэтому более четко общая тенденция будет просматриваться, если произвести выравнивание ряда одним из методов. Так, использование метода укрупнения интервалов дает результаты, представленные в таблице 7.6 и 7.7.

Например:

Таблица 7.6. Производство цельномолочной продукции в пересчете на молоко, тыс. тонн.

Годы	Цельномолочная продукция в пересчете на молоко	Годы	Цельномолочная продукция в пересчете на молоко
1995	801	2001	963
1996	744	2002	906
1997	813	2003	927
1998	952	2004	992
1999	1032	2005	1122
2000	954	2006	1284

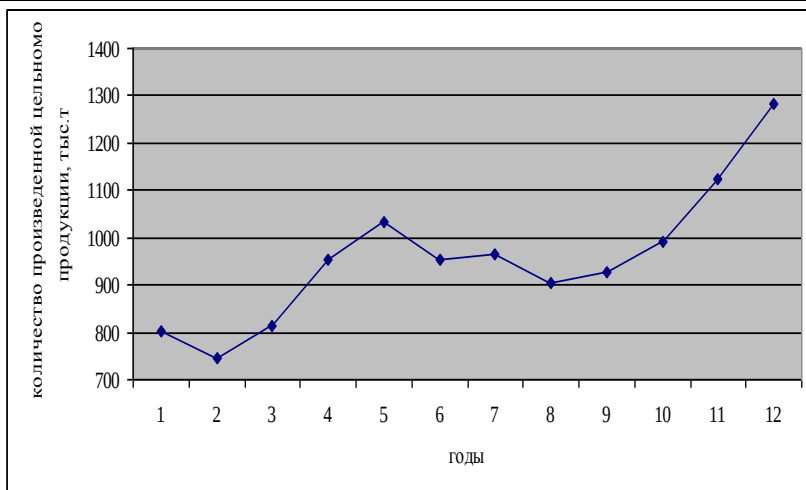


Рисунок 7.2. График динамики производства цельномолочной продукции в РБ

Таблица 7.7. Расчет производства цельномолочной продукции по укрупненным периодам (трехлетиям).

Укрупненные периоды	Сумма цельномолочной продукции в пересчете на молоко	Среднегодовое производство цельномолочной продукции в пересчете на молоко
1995-1997	2358	786,0
1998-2000	2938	979,3
2001-2003	2796	932,0
2004-2006	3398	1132,7

Алгоритм расчета:

$$801+744+813 = 2358; 2358/3=786,0$$

$$952+1032+954=2938; 2938/3=979,3 \text{ и т.д.}$$

Из последнего столбика таблицы 7.7 и рисунка 7.3 более четко видна тенденция к росту производства цельномолочной продукции.

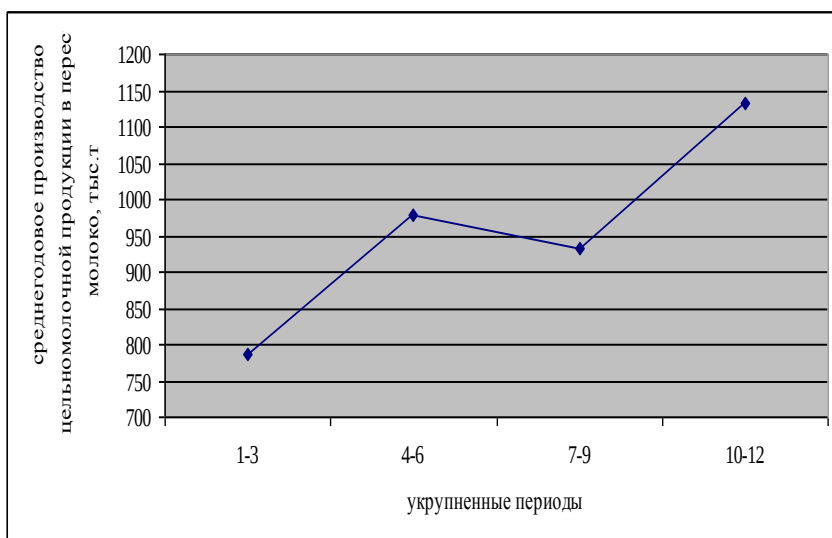


Рисунок 7.3. График ряда, выровненного по укрупненным интервалам

Второй способ заключается в последовательном расчете средних уровней за выбранный лаг времени, передвигая этот лаг

на 1 интервал по времени t . Величину лага L выбирают произвольно в зависимости от длины анализируемого ряда: $L = 2, 3, 4, \dots, n$. Например, мы выберем трехлетний лаг. И тогда выровненный ряд можно представить в следующей таблице (7.8).

Алгоритм расчета:

$$801+744+813=2358; 2358/3=786,0$$

$$744+813+952=2509; 2509/3=836,3$$

$$813+952+1032=2797; 2797/3=932,3 \text{ и т.д.}$$

В настоящее время в основном используют третий метод, который рассмотрим в следующем вопросе.

Таблица 7.8 – Расчет трехлетней скользящей средней.

Годы	Цельномолочная продукция в пересчете на молоко, тыс. т	Сумма цельномолочной продукции в пересчете на молоко по трехлетиям, тыс. т	Трехлетняя скользящая средняя, тыс. т
1995	801	-	—
1996	744	-	—
1997	813	2358	786,0
1998	952	2509	836,3
1999	1032	2797	932,3
2000	954	2938	979,3
2001	963	2949	983,0
2002	906	2823	941,0
2003	927	2796	932,0
2004	992	2825	941,7
2005	1122	3041	1013,7
2006	1284	3398	1132,7

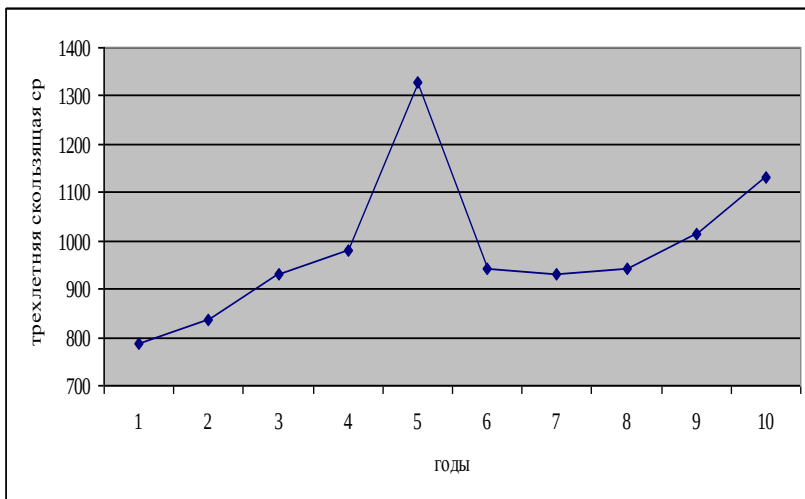


Рисунок 7.4. График ряда, выровненного по скользящей средней

В 4. Тренд. Аналитическое выравнивание ряда. Прогнозы по тренду.

Самые точные результаты выравнивания ряда динамики дает аналитическое выравнивание, с помощью которого строят математическое уравнение, наилучшим образом описывающее тенденцию изменения показателя.

Понятие об уравнении тенденции динамики ввел в 1902 году английский ученый Р.Гукер. Он предложил называть такое уравнение трендом (the trend).

Под **трендом** понимают уравнение линии во времени, вдоль которой расположена ломаная ряда динамики.

Этапы построения тренда:

1. Выявляют этапы развития явления (спады, подъемы, однородные участки и т.д.).
2. Анализируют показатели ряда динамики на этих этапах и выбирают вид уравнения (вид тренда).
3. Вычисляют параметры тренда, используя метод наименьших квадратов.

4. Оценивают адекватность уравнения развитию анализируемого показателя, оценивая колеблемость фактических уровней ряда вокруг теоретических, т.е. вычисленных по тренду.

Рассмотрим 3-ий этап на простейшем примере, когда трендом является уравнение прямой линии. Будем искать уравнение в виде:

$$\hat{Y}_t = a_0 + a_1 t \quad (7.11)$$

где t – параметр времени; Y_t – соответствующие времени уровни ряда, $\hat{Y}_t$ – выровненные уровни ряда, т.е. вычисленные по тренду.

Исходную информацию и расчетные показатели представим в виде таблицы 7.9.

Используя метод наименьших квадратов, построим систему уравнений:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum t = \sum Y \\ a_0 \cdot \sum t + a_1 \cdot \sum t^2 = \sum Yt \end{cases} \quad (7.12)$$

Таблица 7.9. Схема таблицы для расчета параметров тренда

t	Y_i	t^2	$Y \cdot t$	$\hat{Y}$
1	Y_1			$\hat{Y}$
2	Y_2			
3	Y_3			
...	...			
n	Y_n			$\hat{Y}$
$\sum t$	$\sum Y$	$\sum t^2$	$\sum(Y \cdot t)$	$\sum \hat{Y}$

Решение этой системы позволит найти значения параметров уравнения. Если уравнение построено качественно, то $\sum Y_i = \sum \hat{Y}_t$.

Эти расчеты можно упростить, если иначе производить нумерацию параметра t . Ее нужно произвести следующим образом: начало отсчета должно находиться в середине анализируемого ряда. Если количество точек нечетное, то в середине ряда t ставим 0; если количество точек четное, то нуль выбрасывается.

К началу ряда счет идет со знаком «-», к концу – со знаком «+», тогда $\sum t$ становится равной 0 (т.е. $\sum t=0$).

Таблица 7.10. Схема номерации уровней ряда от середины ряда

t для нечетного количества уровней ряда	t для четного количества уровней ряда
-к	-к
...	...
-2	...
-1	-2
0	-1
+1	+1
+2	+2
...	...
+к	+к
$\sum k_i=0$	$\sum k_i=0$

Тогда в системе нормальных уравнений (7.12) исчезнут слагаемые, в которые входит сумма t ($\sum t=0$), т.е. мы получим упрощенную систему

$$\begin{cases} na_0 = \sum Y \\ a_1 \sum t^2 = \sum Yt \end{cases} \quad (7.13)$$

где n – количество точек анализируемого ряда.

Отсюда:
$$a_0 = \frac{\sum Y}{n} \quad (7.14)$$

$$a_1 = \frac{\sum Yt}{\sum t^2} \quad (7.15)$$

Параметр a_1 соответствует абсолютному приросту за единицу периода времени.

Тенденции или тренды могут быть выражены в виде кривых. Это могут быть: парабола, ветка гиперболы, экспонента или показательная кривая, логарифмическая линия и т.д. Виды трендов в виде кривых и их построение рассмотрим ниже.

Прогноз по тренду. Используя уравнение можно построить точечный прогноз на последующие периоды времени, подставляя в уравнение тренда (7.11) номера t , следующие за по-

следним фактическим номером, используемым при построении тренда. Например:

Таблица 7.11. Схема расчета прогноза при различной нумерации показателей ряда динамики

Номерация t от начала ряда	Номерация t от середины ряда
1	-6
2	...
...	-1
11	+1
прогноз 12	+2
прогноз 13	...
прогноз 14	+6
прогноз 15	прогноз 7
и т.д.	прогноз 8

Параметры a_0 и a_1 в уравнении тренда, построенных различным способом нумерации t , будут отличаться, а прогнозы будут совпадать.

Оценка точности тренда. В экономическом анализе необходимо оценить силу разброса фактических точек вокруг расчетных (трендовых). Это оценивается по величине остаточного среднеквадратического отклонения и коэффициенту вариации:

$$\sigma_{ост.} = \sqrt{\frac{\sum (Y_i - \hat{Y}_t)^2}{n - m}} \quad (7.16)$$

где n – число уровней ряда; m – число параметров в уравнении тренда (например, для прямой $m=2$, для параболы $m=3$).

$$V = \frac{\sigma_{ост.}}{\bar{Y}_i} \cdot 100\% \quad (7.17)$$

где $\bar{Y}_i$ – средний уровень ряда.

Чем меньше значения $\sigma_{ост.}$ и V , тем лучше тренд отражает тенденцию изменения показателя.

Имея оценку качества тренда, производят оценку качества прогноза на перспективу.

Для точечного прогноза по тренду, также как и для выборочного наблюдения, необходимо оценить среднюю ошибку прогноза.

Для линейного тренда средняя ошибка прогноза рассчитывается:

$$m_y = \sigma_{ост.} \cdot \sqrt{1 + \frac{1}{n} + \frac{t_{прогноза}^2}{\sum t_{факт.}^2}} \quad (7.18)$$

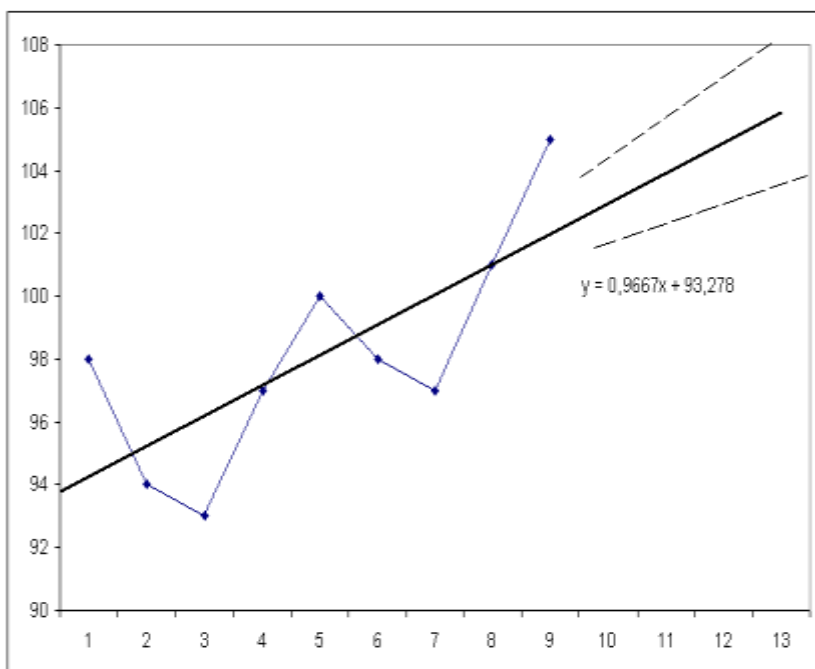


Рисунок 7.5. Точечное и интервальное прогнозирование

Доверительный интервал или предельная ошибка прогноза равна средней ошибке умноженной на коэффициент доверия t-Стьюдента, при доверительных вероятностях $p=0,90; 0,95; 0,99$. Строчку в таблицах находим по числу степеней свободы тренда $n - m$.

В 5. Сезонные колебания в рядах динамики

Если ряд динамики насчитывает достаточное количество уровней, например 100, то можно обнаружить колебания, которые повторяются, т.е. наличие регулярных спадов и подъёмов. Такие макроэкономические колебания называются **циклическими**. Такого вида колебания рассматриваются в курсе макроэкономики, когда оценивают динамику экономического развития отдельных государств.

Внутри годовичные колебания, имеющие регулярный характер, т.е. повторяющиеся из года в год, называются **сезонными**.

Сезонные подъёмы и спады в производстве снижают равномерность, устойчивость технологического процесса и реализации продукции и должны учитываться при планировании производства.

Сезонные колебания показателя могут накладываться на тенденцию роста или спада фактора, а могут наблюдаться при постоянной случайной колеблемости. В зависимости от этого графики могут иметь следующий вид:

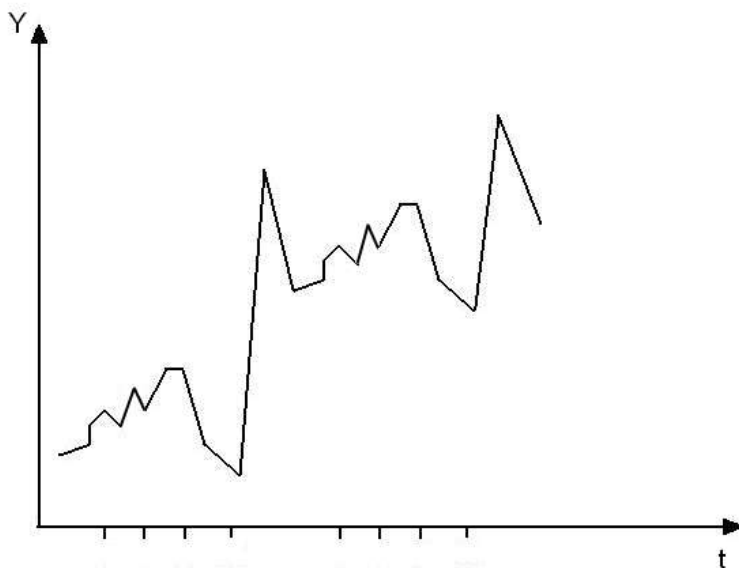


Рисунок 7.6. Наличие тенденции, случайных и сезонных колебаний

На рисунке хорошо видны спады в четвертом квартале и подъемы в первом. В остальные периоды наблюдается случайная колеблемость при наличии тенденции к росту показателя.

На рисунке 7.7. видно, что изменение показателя не имеет тенденции ни к росту, ни к снижению. Однако видна четко выраженная сезонная колеблемость показателя, т.е. есть периоды «впадин» и «подъемов».

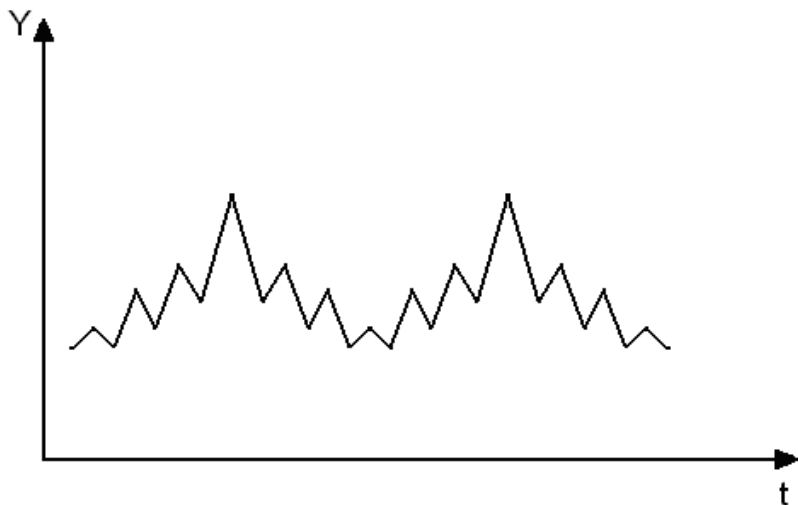


Рисунок 7.7. Наличие сезонных и случайных колебаний

Чаще всего степень сезонных колебаний определяют по величине индекса сезонности. Различают индивидуальные индексы сезонности и средние, с учётом тренда и без его наличия.

Если тренд отсутствует, то индивидуальный индекс сезонности может быть вычислен как отношение величины месячного (квартального) уровня к среднегодовому:

$$I_{сез} = \frac{Y_{мес(кварт)}}{Y_{год}} \cdot [100\%] \quad (7.19)$$

Средний индекс сезонности равен сумме индивидуальных индексов сезонности, делённой на количество сезонов n . Так как анализ может проводиться за несколько полных лет и один (последний) неполный год, то число n будет различным для той

части времени, где учитываются только полные годы и для той, где есть дополнительно значения ряда за неполный год.

$$\overline{I_{сез}} = \frac{\sum I_{i сез}}{n} \quad (7.20)$$

В сельском хозяйстве анализ сезонных колебаний на основе данных поквартального учёта несколько ухудшает показатель сезонности, потому что времена года (сезоны) не совпадают с календарными периодами. Поэтому исследование наличия сезонных колебаний **лучше проводить на основании данных месячного учета.**

Если анализируемый показатель имеет не только сезонные колебания, но и достоверную тенденцию (тренд), то в знаменателе берётся показатель ряда, взятый по тренду:

$$I_{сез} = \frac{Y_{i факт}}{\hat{Y}_{i тренд}} \cdot [100\%] \quad (7.21)$$

Оценка индекса сезонности осуществляется по среднему линейному или среднеквадратическому отклонению для каждого года. Если коэффициенты сезонности (линейный и среднеквадратический) увеличиваются, то это свидетельствует об усилении сезонных колебаний; если уменьшаются, то наоборот.

Среднее линейное отклонение коэффициента сезонности рассчитывают по формуле:

$$(L)K_{сез} = \frac{\sum |I_{сез} - 100\%|}{n} = \frac{\sum |I_{сез} - 1|}{n} \quad (7.22)$$

Среднеквадратическое отклонение коэффициента сезонности рассчитывают по формуле:

$$(\sigma)K_{сез} = \sqrt{\frac{\sum (I_{сез} - 100\%)^2}{n}} = \sqrt{\frac{\sum (I_{сез} - 1)^2}{n}} \quad (7.23)$$

где n – число сезонов.

Чем меньше значения этих показателей, тем меньше сезонная колеблемость и выше стабильность показателя.

В 6. Модели рядов с учетом тенденции и сезонности.

Тенденцию ряда и наличие сезонности можно представить в виде двух различных моделей:

1) Ряд динамики в виде произведения уравнения тренда на средние индексы сезонных колебаний

$$\hat{Y}_t = (a_0 + a_1 t) \overline{I}_{сез} \quad (7.24)$$

Такого вида модель можно использовать, если есть информация помесечных или поквартальных данных за несколько лет. Последний год может быть полный, а может быть незавершенный. Тогда средние индексы сезонности должны быть рассчитаны с учетом имеющихся данных. В нашем примере рассматриваются квартальные данные за три полных года (таблица 7.12).

Используя данные столбцов t и Y и метод наименьших квадратов в системе (7.12) найдем параметры уравнения тренда $\hat{Y}=12,18-0,18t$.

Таблица 7.12. Расчеты модели по данным квартальных отчетов.

Год	Квартал	t	Y	Тренд	I	Модель
1	I	1	12,7	11,99808	1,058503	11,48332
1	II	2	11,7	11,81934	0,989903	11,29328
1	III	3	11,4	11,64059	0,979331	11,65556
1	IV	4	12,3	11,46185	1,073125	12,44982
2	I	5	9,03	11,28311	0,800311	10,79903
2	II	6	10,2	11,10437	0,914055	10,61014
2	III	7	11,7	10,92563	1,074538	10,93967
2	IV	8	12,6	10,74689	1,16871	11,67323
3	I	9	10,7	10,56815	1,012476	10,11474
3	II	10	10	10,38941	0,962519	9,926997
3	III	11	9,7	10,21066	0,949987	10,22379
3	IV	12	10,2	10,03192	1,016754	10,89664
Прогноз				по тренду	средние индексы	по модели
4	I	13	–	9,853182	0,957097	9,430449
4	II	14	–	9,674441	0,955492	9,243853
4	III	15	–	9,495699	1,001285	9,507905
4	IV	16	–	9,316958	1,086197	10,12005

Подставив в это уравнение порядковые номера ряда, рассчитаем теоретические значения показателя по тренду и занесем их в следующий столбец. Поквартальные индексы сезонности найдем как отношение фактических значений уровня ряда к уровням ряда, рассчитанным по тренду. Тем самым мы заполним следующий столбец таблицы.

Чтобы рассчитать значения показателя на перспективу (**прогноз**), необходимо прогноз по тренду умножить на средние значения индекса сезонности. Все проведенные расчеты и исходная информация отражены на рисунке 7.8.

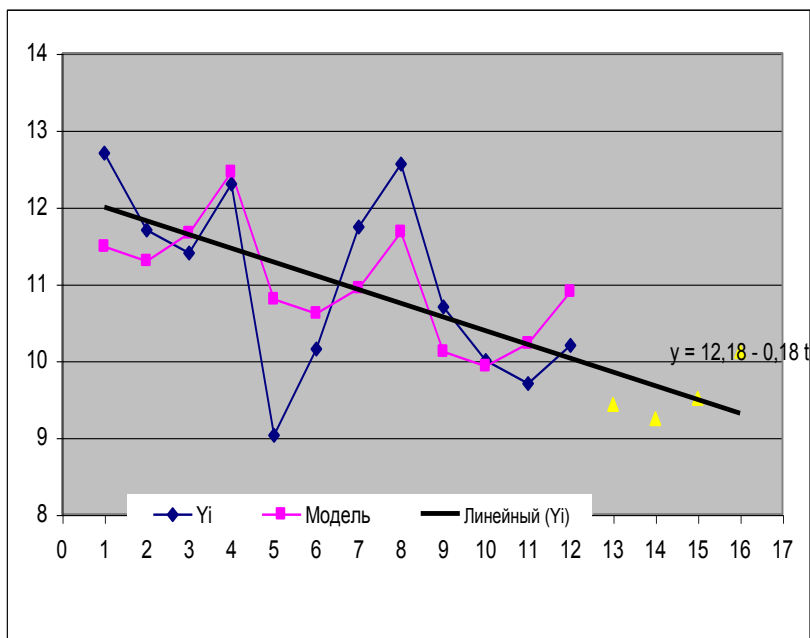


Рисунок 7.8. График результатов анализа динамики показателя с учетом тенденции и сезонных колебаний.

Из рисунка видно, что четко выраженная тенденция к уменьшению показателя накладывается на сезонные подъемы в четвертом квартале каждого года и спады в другие кварталы.

2) Уравнение в виде «гармоник ряда Фурье»:

$$\hat{Y}_t = a_0 + (ak \cos kt + bk \sin kt) \quad (7.25)$$

где k определяет номер гармоник с точностью до четырёх знаков.

Под гармоникой понимают полную волну синусоиды (гармонику Фурье), где

$$a_0 = \frac{\sum Y}{n} ; ak = \frac{2}{n} \sum Y_i \cos kt ; bk = \frac{2}{n} \sum Y_i \sin kt \quad (7.26)$$

Этот метод анализа сезонных колебаний дает хорошие результаты, когда нет необходимости в четком построении уравнения тренда.

В 7. Корреляция рядов динамики. Регрессия рядов динамики.

Статистика в анализе рядов динамики ставит перед собой задачу совместного анализа рядов динамики, уровни которых технологически или экономически связаны друг с другом. Например, ряд уровня рентабельности связан с рядом уровня себестоимости продукции, с рядом уровня трудоёмкости и т.д. При этом изменение одного показателя вызывает изменение другого. Поэтому перед статистическим исследованием “связанных” друг с другом рядов стоит проблема:

1) оценить тесноту связи между значениями уровней различных рядов;

2) построить уравнение регрессии, связывающее результативный показатель, факторный показатель и временной параметр t .

Проблема оценки тесноты связи осложняется возможным наличием автокорреляции в рядах динамики. Поскольку технологический процесс производства и реализации сельскохозяйственной продукции растянут во времени, то может оказаться, что каждый последующий уровень ряда зависит с определённой величиной лага L от предыдущих значений уровней ряда. Это явление называется **автокорреляцией**. Наличие автокорреляции в рядах динамики искажает результаты исследования. Поэтому разные авторы в учебной литературе предлагают несколько раз-

личных методик оценки тесноты связи и построения уравнения регрессии в рядах динамики и позволяющих исключить влияние автокорреляции.

Пусть есть два ряда с трендами:

$$\hat{Y}_t = a_0 + a_1 t \quad \hat{X}_t = b_0 + b_1 t \quad (7.27)$$

Чтобы избежать автокорреляции или влияния автокорреляции на результаты исследований, некоторые авторы предлагают проводить анализ не уровней ряда, а их отклонений от теоретических значений по тренду, если тренд существует, или от средних значений, если отсутствует тренд и колебания показателей случайны:

$$\Delta y = Y_i - \bar{Y} \quad \text{или} \quad \Delta y = Y_i - \bar{Y} \quad \Delta x = X_i - \bar{X} \quad \text{или} \quad \Delta x = X_i - \bar{X} \quad (7.28)$$

Составляется таблица для расчета парного коэффициента корреляции на основании значений абсолютных отклонений:

Таблица 7.13. Схема расчетов для оценки тесноты связи факторов

Период	Δx	Δy	$\Delta x \Delta y$	$\Delta^2 x$	$\Delta^2 y$
...	...	...	...	...	...
...	...	...	...	...	...
Итого	$\sum \Delta x$	$\sum \Delta y$	$\sum \Delta x \Delta y$	$\sum \Delta^2 x$	$\sum \Delta^2 y$

Парный коэффициент корреляции факторов X и Y будет равен:

$$r_{\Delta x \Delta y} = \frac{\sum \Delta x \Delta y}{\sqrt{\sum \Delta^2 x \sum \Delta^2 y}} \quad (7.29)$$

Если отсутствует автокорреляция, то парный коэффициент корреляции можно находить по обычной формуле r_{xy} .

Уравнение связи (регрессии) отклонений имеет вид:

$$\sum \Delta y = a_1 \sum \Delta x \quad (7.30)$$

В этом уравнении a_0 всегда равно 0, коэффициент регрессии a_1 вычисляют по формуле:

$$a_1 = \frac{\sum \Delta x \Delta y}{\sum \Delta_x^2} \quad (7.31)$$

Ряд авторов предлагает совместить показатели рядов в едином уравнении, т.е. построить многофакторное уравнение, отражающее зависимость результата (Y) от фактора (X) и его изменение во времени (t)

$$\tilde{Y}_{xt} = a_0 + a_1 x + a_2 t \quad (7.32)$$

Это уравнение отражает зависимость изменений значений одного ряда показателей (Y) от изменений значений другого ряда (X) во времени t. Время t задаётся как натуральный ряд чисел $t=1,2,3,\dots, n$ или нумерацию производят от середины ряда, как это показано в вопросе 4. Параметры уравнения регрессии можно найти, используя метод наименьших квадратов, если составить и решить следующую систему трех уравнений с тремя неизвестными

$$\begin{cases} na_0 + a_1 \sum x + a_2 \sum t = \sum y, \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum xt = \sum yx \\ a_0 \sum t + a_1 \sum xt + a_2 \sum t^2 = \sum yt \end{cases} \quad (7.33)$$

Система упростится, если нумерацию уровней ряда проводить от его середины так, чтобы $\sum t=0$. Тогда система примет несколько упрощенный вид:

$$\begin{cases} na_0 + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum xt = \sum yx \\ a_1 \sum xt + a_2 \sum t^2 = \sum yt \end{cases} \quad (7.34)$$

Затем исследование проводится также, как в многофакторном корреляционно-регрессионном анализе.

В учебниках по общей теории статистики не отражена проблема построения уравнения регрессии при наличии тренда и при наличии сезонных колебаний.

В 8. Понятия интерполяции и экстраполяции в рядах динамики.

В рядах динамики часто теряется информация при сборе данных для них.

Слово «интерполяция» латинского происхождения (interpolatio) – означает изменение. Современное использование этого слова в экономических науках имеет несколько другой, более широкий смысл.

Интерполяция – это способ определения промежуточных значений динамического ряда:

а) на основе известных соседних значений (как среднее арифметическое);

б) на основе взаимосвязей с другими рядами, количественные выражения которых известны;

в) на основе средних значений показателей ряда динамики (см. вопрос 2).

При проведении интерполяции предполагается, что выявленная тенденция и ее характеристики не претерпели существенных изменений в тот период времени, уровни которого отсутствуют.

При прогнозировании развития анализируемых показателей на перспективу используется метод экстраполяции, при проведении которого предполагается, что условия развития явления в будущем не изменятся. Такая экстраполяция называется перспективной. Чем короче срок экстраполяции, тем точнее и надежнее ее результаты. За короткий срок чаще всего условия развития резко не меняются. Однако история знает и другие примеры, такие как развал Советского Союза и образование целого ряда независимых государств, в т.ч. и Республики Беларусь, финансовый и экономический кризис 2008–2009 гг.

Термин «экстраполяция» произошел от соединения двух латинских слов «extra» – сверх и «polire» – делать гладким. Соединение этих двух терминов в одном слове означает нахождение по ряду известных значений величины данной таблицы или графика других ее значений, находящихся вне ее ряда.

Экстраполяция – это метод определения количественных характеристик для совокупностей, не подвергающихся изучению

непосредственно, а на основе изучения предыдущего их развития или развития аналогичных совокупностей.

Чем дальше прогнозирование от реального фактического времени, тем больше интервальная ошибка результатов экстраполяции. Поэтому с течением времени необходимо с учетом свершившегося факта оценивать динамику ряда и осуществлять корректировку прогноза.

По сути дела методика экстраполяции и расчет значений уровня ряда на перспективу рассмотрен нами в вопросе 4 данной темы.

Оба эти метода возможны только в условиях стабильного развития технологического-экономических показателей.

Тема 8. Индексный метод

1. История индексного метода. Общее понятие об индексах.
2. Система обозначений в индексном анализе.
3. Классификация индексов.
4. Синтетическая концепция индексов.
5. Аналитическая концепция индексов.
6. Агрегатная форма построения индексов. Классификация агрегатных индексов по составу.
7. Классификация агрегатных индексов по содержанию.
8. Классификация агрегатных индексов по базе сравнения.
9. Средние индексы.
10. Индексы средних величин (индексы структуры).
11. Другие формы построения индексов
 - 11.1) геометрический индекс
 - 11.2) индекс общей численности товара (продукции)
 - 11.3) индекс ассортиментных сдвигов товаров и его экономический смысл.
12. Индексный анализ мультипликативных, кратных и комбинированных моделей.
13. Метод абсолютных разностей в индексном анализе.

ЛИТЕРАТУРА:

1. *Адамов В.Е. Факторный индексный анализ. М.: Статистика, 1997.*

2. Едронова В.С. *Общая теория статистики: Учебник* / – М.: Юристъ, 2001, – 511 с.
3. Елисеева И.И., Юзбашев М.М. *Общая теория статистики: М.: Финансы и статистика, 1999, – 408 с.*
4. Ефимова М.Р., Петрова В.В. *Общая теория статистики. Учебник/ – 2-е издание., – М.: ИНФРА- М, 2002, – 413 с.*
5. Ефимова М.Р., Рябцева В.М. *Общая теория статистики. М.: Финансы и статистика, 1999.*
6. Громыко Г.А. *Общая теория статистики: Практикум. – М.: ИНФРА- М, 199, –139 с.*
7. *Теория статистики: Учебник / Под ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 199,., – 464 с.*
8. *Теория статистики: Учебник / Под ред. Проф. Г.Л. Громыко. – М.:ИНФРА – М, 2000, – 414 с.*

В 1. История индексного метода. Общее понятие об индексах.

История индексного анализа насчитывает более четырёх веков. С развитием торговли купцов всегда интересовал вопрос различия цен на одни и те же товары в разных краях света. Однако аналитический подход к исследованию цен начал закладываться только в 17 веке. Термин имеет латинское происхождение «index» — указатель, показатель.

Первый индекс был предложен в 1738 г. французским экономистом Шарлем Дюто для оценки динамики цены товара. Он сравнивал динамику цены как сумму цен текущего периода (отчетного) с суммой цен предыдущего периода (базисного):

$$I_p = \frac{\sum p_1}{\sum p_0} \quad (8.1)$$

где p_0 и p_1 — цены товара в базисный и отчетный периоды (обозначение «р» взято от латинского слова «pretium» — цена).

Итальянец Джаномо Ринальдо Карли в 1751 г. предложил анализировать изменение цен по другой формуле, где учитывается индивидуальное изменение цены по каждому товару:

$$I_p = \frac{\sum i_p}{n} = \frac{\sum \frac{p_1}{p_0}}{n} \quad (8.2)$$

где i_p – индивидуальный индекс цены i -го товара, отражающий динамику цены каждого вида товара, n – количество различных видов товара. Данный подход был шагом вперед. Однако и он не учитывал динамику количества проданного товара.

Следующий шаг в развитии теории индексного анализа связан с именами немецких ученых экономистов Этьеном Ласпейресом и Германом Пааше. Поскольку изменению подвергается не только цена товара, но количество и доля в выручке за различный товар, то этими авторами был предложен и обоснован такой подход к исследованию, где учитываются оба показателя – цена товара и объем его реализации (подробнее эти индексы рассмотрены в вопросе 3).

Развитие теории индексного анализа связано с именами английских экономистов Стенли Джевонсона и Альфреда Маршалла. Свой вклад в теорию индексного анализа внесли С.Г. Митчелл, И. Фишер, В.И. Борткевич и другие экономисты.

В экономической науке существует около 30 определений индекса. Рассмотрим два из них:

ОПРЕДЕЛЕНИЕ 1: Экономические индексы – это относительные величины, которые характеризуют изменения экономических явлений во времени, в пространстве или по сравнению с любым эталоном (плановым, нормативным или лучшим по совокупности).

ОПРЕДЕЛЕНИЕ 2: Статистический индекс – это сложный относительный показатель, характеризующий среднее изменение непосредственно несоизмеримых или соизмеримых показателей.

Оба определения отражают суть одного и того же процесса или явления – сопоставляют (сравнивают) значения какого-то показателя друг с другом. Кроме того, коль это относительный показатель, то он может быть представлен в виде коэффициента или в процентах.

Рассмотрим некоторые понятия, которые упростят нам понимание данного метода исследования и которые содержатся во втором определении индекса.

Продукция (товар) является *соизмеримой (сопоставимой)*, если она может подвергаться непосредственному суммированию в натуральных единицах измерения. Например, гектары, штуки, килограммы, литры и т. д.

Если продукция в своём натуральном выражении непосредственно не суммируется, то она *несоизмерима (несопоставима)*.

Например, в сельском хозяйстве единицей измерения продукции являются центнеры и тонны, но напрямую объемы сельскохозяйственной продукции суммировать нельзя, кроме некоторых видов (зерновые культуры, овощные культуры). Следовательно, сельскохозяйственная продукция является не соизмеримой (кроме групп зерновых и овощных культур).

В первом определении указано, что сравнение может быть проведено во времени, т.е. более поздние значения показателей можно сравнивать с более ранними. Если рассматриваются разные объекты в один и тот же период времени, то их характеристики также могут быть сравнимы, если показатели одного из объектов взять за базу сравнения.

Например, во времени можно сравнивать показатели одного и того же хозяйства, т.е. оценивать динамику (развитие) предприятия. Сравнивая результаты производственно-финансовой деятельности предприятия с другими сельскохозяйственными предприятиями можно оценить свое «место» в регионе (районе, области)

Задачи индексного анализа:

1. Определить динамику изменения технологического-экономических показателей производства и реализации продукции;
2. Выявить влияние на результат факторов, определяющих изменение результата.

Решить обе эти задачи позволяют синтетическая и аналитическая концепции индексов (см. вопросы 4 и 5).

Рассмотрим **простейший пример решения задачи** с помощью индексов.

Сельскохозяйственное предприятие произвело в прошлом году 100 тыс. тонн продукции (q_0), в текущем году – 120 тыс. тонн (q_1). Продукция является сопоставимой (соизмеримой). Цена за каждую тонну продукции в прошлом году составила 20 условных денежных единиц (p_0), в текущем – 18 усл. ден. ед. (p_1). Общая стоимость продукции в прошлом году составила 2000 тыс. усл. ден. ед. (S_0), в текущем году – 2160 тыс. усл. ден. ед. (S_1).

Динамику показателей можно оценить системой относительных и соответствующих им абсолютных изменений. Индивидуальные индексы обозначим буквой i , абсолютное изменение – символом Δ (дельта). Подстрочная сноска будет отражать обозначение анализируемого показателя.

Динамика количества продукции:

$$i_q = \frac{q_1}{q_0} = \frac{120}{100} = 1,200 = 120,0\%,$$

$$\Delta_q = q_1 - q_0 = 120 - 100 = +20 \text{ тыс. т}$$

Этот расчет показывает, что объем производства в текущем году по сравнению с предыдущим составил 120,0%, т.е. прирост составляет 20% или +20 тыс. тонн.

Динамику цен проанализируем аналогично:

$$i_p = \frac{p_1}{p_0} = \frac{18}{20} = 0,900 = 90,0\%$$

$$\Delta_p = p_1 - p_0 = 18 - 20 = -2 \text{ усл. ден. ед.}$$

Расчеты показывают, что цена 1 тонны в текущем периоде по сравнению с предыдущим составляет 90,0%, т.е. цена снизилась на 10% или на 2 условные денежные единицы.

Динамика стоимости продукции будет следующая:

$$i_s = \frac{S_1}{S_0} = \frac{2160}{2000} = 1,080 = 108,0\%$$

$$\Delta_s = S_1 - S_0 = 2160 - 2000 = +160 \text{ тыс. ден. ед.}$$

Анализ проделанных расчетов показывает, что стоимость произведенной продукции в отчетный период относительно пре-

дыдущего времени составила 108,0%, т.е. прирост составил 8,0%, или 160 тыс.ден.ед.

Следует отметить, что данная задача с учетом второй цели индексного анализа может иметь более развернутое решение, принципы которого будут рассмотрены ниже.

В 2. Система обозначений в индексном анализе.

Мы уже рассмотрели некоторые обозначения, используемые в индексном анализе. Однако принятая система обозначений несколько шире. В отечественной экономической литературе в индексном анализе принята следующая система обозначений, которой рекомендуется придерживаться.

i – индивидуальные (частные) индексы;

I – общие индексы;

q (Q) – количество (объём) продукции;

t (T) – затраты рабочего времени (затраты труда);

z – себестоимость единицы продукции;

p – цена реализации единицы продукции;

m — материалоёмкость единицы продукции;

$_0$ – подстрочный символ, указывающий на базисный показатель или период, то есть тот, с которым сравнивают;

$_1$ – подстрочный символ, указывающий на то, что показатель подлежит сравнению.

При проведении индексного анализа каких-либо технологических показателей по сельскому хозяйству для построения формул принято использовать первую букву русского слова. Например, P – поголовье животных или посевная площадь и т.д.

В 3. Классификация индексов.

При построении индексов между экономистами возникало много дискуссий по поводу правил и принципов их построения, по поводу их экономического смысла. В основу современной методологии построения легли тесты американского статистика И. Фишера (1867–1947). Эти тесты основаны на логике построения экономико-статистических показателей и позволяют всегда проверить правильность проведенных расчетов.

В примере, который мы рассмотрели в первом вопросе, были рассчитаны индивидуальные индексы показателей и соответствующие им абсолютные показатели. Там мы не ставили

своей задачей проанализировать влияние факторов (цены и количества продукции) на результат (полную стоимость всей продукции). Такая задача может быть решена при использовании более сложных **общих индексов**.

Существует несколько классификаций индексов. Однако большинство авторов выделяют индексы, отвечающие следующим понятиям, в смысле которых нам предстоит разобраться. При классификации конкретного индекса его соотносят с характеристикой из каждой группы так, как это показано в примерах.

Классификация индексов

I. По содержанию:

- 1.1. объёмные;
- 1.2. качественные;

II. По форме образования:

- 2.1. агрегатные;
- 2.2. средние;
- 2.3 средних величин (индексы структуры)

III. По базе сравнения:

- 3.1. базисные;
- 3.2. цепные;

IV. По виду весов (по составу):

- 4.1. постоянного состава;
- 4.2. переменного состава.

Например, в соответствии с этой классификацией:

$$1) I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0} \quad - \text{этот индекс в соответствии с классификацией будет объёмный, агрегатный, базисный, постоянного состава.}$$

Он отражает изменение затрат в зависимости от изменения объемов производства различных видов продукции при постоянных затратах на единицу произведенной продукции.

$$2) I_{\bar{y}} = \frac{\bar{Y}_1}{\bar{Y}_0} = \frac{\sum Y_1 \Pi_1}{\sum \Pi_1} \div \frac{\sum Y_0 \Pi_0}{\sum \Pi_0},$$

где Π – посевная площадь сельскохозяйственных культур, га; Y – урожайность, ц с 1га. Этот индекс – качественный, средней величины, базисный, переменного состава. Он может отражать изменение средней урожайности по группе однотипных культур (зерновых или овощных) и одной культуры в

зависимости от сортов, производственных бригад или участков и т.д.

Общие индексы строят разными способами в зависимости от поставленной задачи и наличия исходных данных.

Основным и обязательным условием построения индексов является соответствие индексов абсолютным изменениям.

Первому индексу соответствует абсолютное изменение, показывающее на сколько денежных единиц изменились затраты за счет изменения объемов производства.

$$\Delta_{qz(q)} = \sum q_1 z_0 - \sum q_0 z_0 = \sum (q_1 - q_0) z_0 \quad (8.3)$$

Для второго индекса соответствует два абсолютно разных изменения:

$$\begin{aligned} 2.1.) \Delta_y^I &= \bar{Y}_1 - \bar{Y}_0; \\ 2.2.) \Delta_y^{II} &= (\bar{Y}_1 - \bar{Y}_0) \times \sum \Pi_0. \end{aligned} \quad (8.4)$$

Первое из них (2.1) отражает изменение собственно средней урожайности, а второе (2.2) – изменение валового сбора за счет изменения средней урожайности.

Как видим из примеров индексов, надо четко себе представлять, как соотносить конкретный индекс с каждой позицией из классификации.

В 4. Синтетическая концепция индексов.

В конце XIX века немецкие статистики-экономисты Герман Пааше (в 1874г) и Этьен Ласпейрес (в 1864г) предложили свою трактовку индекса цен с учетом объема реализованной продукции.

$$\text{Формула Пааше} \quad I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \quad (8.5)$$

$$\text{Формула Ласпейреса} \quad I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} \quad (8.6)$$

Оба экономиста в своих расчетах учитывали динамику цен и влияние её на объем товарооборота, не учитывая **динамики объемов** продукции (q). Причём Г. Пааше использовал итоговые

объёмы реализации (текущие), а Э. Ласпейрес предлагал расчёт при базовых (прошлых) объёмах реализации. **В этих формулах шагом вперёд было объединение двух разноимённых показателей цены и количества товара и получение третьего нового показателя – стоимости реализованной продукции или выручки (товарооборота).**

Эти экономисты впервые предали индексу новый экономический смысл, заменив простое суммирование цен различных товаров подсчётом стоимости определённой массы товаров. В результате получили сложный экономический показатель, в котором отдельные части (элементы) непосредственно несоизмеримы, т.е. они ввели первый соизмеритель объема продукции – его денежное выражение. А деньги можно суммировать.

Соединение в один показатель разнотипных технико-экономических факторов, дающих новый показатель, подлежащий анализу, и является основой синтетической концепции индексного анализа. Синтез (произведение) цены и объёма продукции дал новый показатель – объём товарооборота или объём реализации продукции.

Современная синтетическая концепция индексов позволяет проводить анализ достаточно сложных технико-экономических связей.

Например, в структуре себестоимости продукции важное место занимают затраты материалов (кормов для с.-х. животных, минеральных удобрений и средств защиты растений от вредителей и болезней и т.д.). Тогда стоимость потраченных материалов в соответствии с синтетической концепцией можно представить в виде произведения показателей **(мультипликативной связи факторов)**

$$M = \sum q \cdot m \cdot p. \quad (8.7)$$

Если факторы только суммируются, то такая их связь называется аддитивной. Аддитивную модель можно построить при анализе начисления заработной платы (Зп), которая состоит из оклада (Ок), надбавки за стаж (Нс), надбавки за квалификацию (Нк), премии (Пр), отчислений – подоходного налога (Нп), в пенсионный фонд (Пф) и др. (Од). Тогда модель заработной платы работника можно представить:

$$Зп=Ок+Нс+Нк+Пр-Нп-Пф-Од. \quad (8.8)$$

Простейшую формулу рентабельности (R) можно представить в виде **кратной модели**, если разделить прибыль (Пр) на затраты (З).

$$R = \frac{Пр}{З} \quad (8.9)$$

Сложные или комбинированные системы связей факторов содержат различные арифметические действия над показателями, входящими в модель. Так, модель рентабельности может быть представлена следующей формулой:

$$R = \frac{\sum q(p-z)}{\sum qz} \quad (8.10)$$

В этой формуле использованы принятые в статистике обозначения количества продукции, цены и себестоимости единицы продукции.

В 5. Аналитическая концепция индексов.

В аналитической концепции индексного анализа реализуется современная цель экономического исследования, проводимого с помощью индексов (см. вопрос 1). При проведении исследования очень важно установить не только общее изменение результативного показателя, но и какая доля или часть изменения приходится на влияние динамики факторов, входящих в модель.

Аналитическая концепция индексов позволяет установить:

- 1.) динамику каждого индивидуального показателя,
- 2.) в какой мере общее изменение результативного показателя зависит от изменения каждого фактора, его составляющего.

Чтобы грамотно использовать аналитическую концепцию индексов, необходимо факторы в модели располагать в следующей последовательности:

– модель (формула) должна быть составлена в строгом соответствии с фактическими связями технико-экономических факторов;

– на первое место в построенной модели выносятся количественные (объемные) показатели, такие как объем произведенной продукции, количество животных или площадь пашни и т.д.;

– на втором месте указывается структурный показатель, например, доля каждого вида продукции в общем объеме реализации;

– на третьем месте в модели стоят качественные (интенсивные) показатели, такие как цена или себестоимость продукции;

Рассмотрим использование аналитической концепции индекса на примере общепринятое разложение индекса товарооборота:

Анализ требует установить, как изменится объем товарооборота под влиянием:

1) динамики количества товаров;

2) динамики цен;

3) изменения структуры различных видов товаров.

Простейшее разложение этого индекса будет иметь вид

$$I_{qp} = I_q * I_p \quad (8.11)$$

или в развернутом виде:

$$I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (8.12)$$



Индекс объёма
товарооборота



Индекс
количества
и структуры
товаров



Индекс цены
товаров

Рассмотрим составляющие этой формулы. Исходным в анализе является индекс товарооборота или выручки.

$$5.1) I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \quad (8.13)$$

– этот индекс отражает общую динамику стоимости товара без анализа факторов, влияющих на это изменение.

$$5.2) I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \quad (8.14)$$

– этот индекс называется индексом количества и структуры товара и отражает динамику стоимости товара за счёт динамики объёма реализации;

$$5.3) I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (8.15)$$

– этот индекс носит название индекса цен и анализирует динамику стоимости товара за счёт изменения цен.

Например, если $I_{qp} = 1,2658 = 126,58\%$, т.е. общее изменение стоимости товара составило 126,6% или возросло на 26%, то составляющие индекса выручки (или товарооборота) покажут, какая доля влияния приходится на динамику объёма товара, какая – на динамику цены. Так, например, $I_q = 1,1543$ и $I_p = 1,0966$. Значения этих индексов показывают, что выручка (товарооборот) выросла на 15,43% за счет роста объемов реализации продукции и улучшения структуры товара и на 9,66% – за счет роста цен.

Сторонники такого анализа считают, что с помощью индексов решаются **три главные задачи**:

- 1) измеряются факторы в общей динамике показателей;
- 2) обособляется влияние структуры явлений от изменения индексируемого признака при анализе динамики вторичных признаков;
- 3) измеряются результаты изменения признаков с несоизмеримыми элементами.

Эти три задачи отражают следующее:

- 1) индекс – есть относительная величина, вследствие чего мы абстрагируемся от абсолютного размера явления;
- 2) индекс выражает изменение одного показателя во взаимосвязи с другими факторами, от изменений которых мы абстрагируемся, предполагая поочередно их величину неизменной, т.е. в индексе всегда есть элемент условности.

В 6. Агрегатная форма построения индексов. Классификация агрегатных индексов по составу.

Поскольку виды продукции разнородны, и непосредственно их суммировать нельзя, экономистами найдены два основных показателя, позволяющих объединять любые виды продукции в единую систему: это денежное (стоимостное) выражение про-

дукции или выражение её (продукции) в затратах труда, т.е. в виде произведений qr , qz , qt .

Произведения вида qr , qz , qt – называются **агрегатами**, а индексы, построенные на их основе – **агрегатными**, т. е. в агрегатном индексе и числитель и знаменатель представлены в виде суммы агрегатов.

По сути Г. Пааше и Э. Ласпейрес заложили основу для построения агрегатных индексов. Поэтому все индексы, входящие в формулу (8.9), являются агрегатными. Рассмотрим подробнее систему показателей, входящих в агрегатный индекс.

Показатели z , p , t , m являются интенсивными или качественными, а показатель q – объёмным или количественным.

В агрегатном индексе могут меняться (**индексироваться**) или оба показателя, или один из них.

Показатель, который в индексе меняется во времени, называется **индексирваемым**, который не меняется — **весами** или **соизмерителем**.

В зависимости от того, сколько показателей изменяется, различают индексы **переменного и постоянного состава**.

Например, рассмотрим систему агрегатных индексов, отражающих затраты на производство продукции.

$$6.1) I_{qz} = \frac{\sum q_1 z_1}{\sum q_0 z_0} \quad (8.16)$$

агрегатный индекс, в котором индексируются (изменяются) оба показателя. Следовательно, этот индекс будет индексом **переменного состава или с переменными весами**. В экономическом исследовании он будет показывать, во сколько раз изменились затраты во времени или по сравнению с эталоном (планом и т.д.).

Данный индекс может быть представлен в виде произведения двух следующих индексов постоянного состава. В соответствии с порядком индексирования факторов, входящих в агрегат (модель), сначала необходимо проиндексировать количественный показатель **q**, при неизменном значении качественного фактора **z**.

$$6.2) \text{ Индекс } I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0} \quad (8.17)$$

– является агрегатным индексом **постоянного состава**, т.к. здесь индексируемая величина количество продукции **q**, а **z** (себестоимость единицы продукции) является соизмерителем, т.к. суммировать можно только денежное выражение несопоставимой продукции. При построении данного индекса используется еще одно **правило: соизмеритель (z) берется за базисный период**. Данный индекс показывает, во сколько раз изменились затраты на производство продукции за счет изменения объема производства и его структуры.

Затем будем индексировать качественный фактор, входящий в модель (агрегат) – **z** (себестоимость единицы продукции).

$$6.3) \text{ Индекс } I_z = \frac{\sum q_1 z_1}{\sum q_1 z_0} \quad (8.18)$$

– также является агрегатным, **постоянного состава**. Однако индексируемая величина **z** здесь качественный показатель (уровень себестоимости единицы продукции), а **q** (количество продукции) – веса (частота или повторяемость) себестоимости. Поскольку влияние на затраты объема продукции **q** было отражено в предыдущем индексе (8.17), то **его значение берется за отчетный период**.

Этот индекс отражает изменение затрат на производство, связанное с изменением себестоимости единицы продукции.

Примечания:

– Поскольку в формулах появляются показатели, где один фактор взят за отчетный период, а второй – за базисный (например, $\sum q_1 p_0$), то такое произведение называют условным показателем за условный период.

– Если в экономической модели больше двух показателей, то в индексе переменного состава могут индексироваться (изменяться) все показатели.

– Общий индекс переменного состава разлагается на столько индексов постоянного состава, сколько факторов включено в модель.

– В индексе 8.17 соизмерителем является качественный показатель (z), его принято брать за базисный период; в индексе с 8.18 весами является количественный показатель (q), его берут за отчётный период. При ином распределении “времени” весов и соизмерителя может нарушаться соответствие между индексами и абсолютными показателями, т.е. нарушается равенство и аналитическая концепция индексов.

Итак, в этом вопросе мы выяснили не только понятие агрегатных индексов, но и деление их на индексы по виду весов (по составу) – переменного и постоянного состава в соответствии с классификацией (см. вопрос 3).

В 7. Классификация агрегатных индексов по содержанию.

Рассмотрим деление агрегатных индексов на качественные и количественные на примере индексов объёма товарооборота переменного состава и его разложения на индексы постоянного состава (формула 8.12).

$$7.1) I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \quad (8.19)$$

Это агрегатный индекс переменного состава. Кроме того, он будет объёмным (или количественным) в силу того, что здесь индексируется (изменяется) обобщающий объёмный показатель – товарооборот (выручка или стоимость товара).

Для индексов постоянного состава этот вопрос решается иначе. Отнесение индекса к количественному или качественно-му зависит от того, какой фактор (количественный или качественный) индексируется (изменяется).

$$7.2) I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \quad (8.20)$$

– Это агрегатный индекс постоянного состава. Он показывает, во сколько раз изменился товарооборот (выручка) за счет изменения количества и структуры товара. В индексе количества

и структуры товара отражена динамика объёмного показателя (q), поэтому этот индекс является объёмным или количественным.

$$7.3) I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (8.21)$$

– Данный агрегатный индекс цен постоянного состава отражает изменение выручки за счет изменения цен. В индексе цен определяющей является динамика качественного показателя – цены (p) и влияние её на товарооборот, поэтому этот индекс является качественным.

В 8. Классификация агрегатных индексов по базе сравнения.

Если в индексном анализе сравниваются показатели за два периода, то эти индексы носят название **базисных**. Если система индексов построена на показателях, взятых как минимум за 3 периода и более, причём каждый последующий показатель сравнивается как с начальным уровнем, взятым за базу сравнения, так и с предыдущим, то первые из них индексы являются **базисными**, а вторые – **цепными**. Кроме того, при построении системы необходимо учитывать состав индексов, в зависимости от цели исследования и базы сравнения показателей.

В формулах это может выглядеть следующим образом.

Базисные индексы выручки (стоимости товара) **переменного состава**:

$$\frac{\sum q_1 p_1}{\sum q_0 p_0}, \frac{\sum q_2 p_2}{\sum q_0 p_0}, \dots, \frac{\sum q_n p_n}{\sum q_0 p_0} \quad (8.22)$$

Цепные индексы выручки (стоимости товара) **переменного состава**:

$$\frac{\sum q_1 p_1}{\sum q_0 p_0}; \frac{\sum q_2 p_2}{\sum q_1 p_1}; \dots; \frac{\sum q_n p_n}{\sum q_{n-1} p_{n-1}} \quad (8.23)$$

Можно построить системы цепных и базисных индексов постоянного состава. Например,

базисные индексы физического объема **постоянного состава**:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_0}{\sum q_0 p_0}; \dots; \frac{\sum q_n p_0}{\sum q_0 p_0} \quad (8.24)$$

цепные индексы физического объема постоянного состава:

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_0}{\sum q_1 p_0} \dots; \frac{\sum q_n p_0}{\sum q_{n-1} p_0} \quad (8.25)$$

Кроме того, между цепными и базисными индексами установлены математические зависимости, которые могут быть использованы в конкретном экономическом анализе. Зная базисные индексы, можно построить цепные и наоборот.

В 9. Средние индексы.

В разной научной и учебной литературе предлагаются кроме агрегатных другие формы индексов. В технико-экономическом анализе часто решается задача исследования поведения индивидуальных значений показателей, входящих в модель и их влияния на результативные показатели. Например, как влияют на выручку или товарооборот индивидуальные изменения цены и объема каждого вида продукции. И здесь используется другая форма построения индексов – общие средние индексы из индивидуальных индексов. Такая форма индексов получила название **средних индексов: средние арифметические и средние гармонические**. Рассмотрим эти формы построения общих индексов.

1) Средний арифметический индекс физического объема продукции получим в том случае, если частотой (или повторяемостью) для индивидуального индекса i_q будет служить объем товарооборота $q_0 p_0$. Построенный индекс будет равен агрегатному индексу количества и структуры товара, т.е.

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \quad (8.26)$$

Это равенство среднего и агрегатного индексов следует и соотношения

$$i_q = \frac{q_1}{q_0} \Rightarrow q_1 = i_q \cdot q_0$$

При построении **среднеарифметического индекса цен** необходимо определиться с весами (в нашей формуле это веса Г. Пааше). Тогда построенный средний арифметический индекс цен будет равен агрегатному индексу цен Пааше.

$$I_p = \frac{\sum q_1 i_p p_0}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (8.27)$$

Данное равенство индексов обеспечивается за счет соотношения:

$$i_p = \frac{p_1}{p_0} \Rightarrow p_1 = i_p \cdot p_0$$

2) Средний гармонический индекс.

Средний гармонический индекс цен с весами Г.Пааше будет иметь вид:

$$I_p = \frac{\sum q_1 p_1}{\sum \frac{1}{i_p} q_1 p_1} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (8.28)$$

Равенство среднего гармонического индекса и агрегатного индекса обеспечивается соотношением:

$$i_p = \frac{p_1}{p_0} \Rightarrow p_0 = \frac{1}{i_p} \cdot p_1$$

Аналогично **средний гармонический индекс количества и структуры** товара будет построен следующим образом.

$$I_q = \frac{\sum q_1 p_0}{\sum \frac{1}{i_q} \cdot q_1 p_0} = \frac{\sum q_1 p_0}{q_0 p_0} \quad (8.29)$$

Данное равенство обеспечивается соотношением

$$i_q = \frac{q_1}{q_0} \Rightarrow q_0 = \frac{1}{i_q} q_1$$

В 10. Индексы средних величин (индексы структуры).

В экономическом анализе для характеристики совокупности в целом используются средние значения показателей. Поэтому представляет собой интерес анализ поведения средних значений признака таких, как, например, уровень средних цен, уровень себестоимости единицы продукции и т.д. (см. тему 3). Поэтому рассмотрим правила построения индексов средних величин и их экономический смысл.

10.1) Индекс средней цены переменного состава. Этот индекс получают как частное двух средних цен за отчетный и базисный период. В данном случае индекс отражает изменение средней цены за счет динамики как количества и структуры товара, так и за счет изменения цен на отдельные его виды.

$$I_{\bar{p} \text{ (nc)}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\sum q_1 p_1}{\sum q_1} \div \frac{\sum q_0 p_0}{\sum q_0} \quad (8.30)$$

Этому индексу соответствует два абсолютных изменения. Первое из них показывает, на сколько денежных единиц изменилась средняя цена реализации одной единицы продукции.

$$\Delta'_{\bar{p}} = \bar{p}_1 - \bar{p}_0 \quad (8.31)$$

Второе абсолютное изменение отражает изменение выручки за счет динамики средней цены.

$$\Delta''_{qp(\bar{p})} = (\bar{p}_1 - \bar{p}_0) \cdot \sum q_1 \quad (8.32)$$

10.2) Индекс средней цены с постоянным состава (фиксированного состава). Этот индекс отражает изменение средней цены при условии, что объем и структура произведенной или проданной продукции взята на уровне отчетного периода.

$$I_{\bar{p} \text{ (фс)}} = \frac{\bar{p}_1}{\bar{p}_{\text{усл}}} = \frac{\sum q_1 p_1}{\sum q_1} \div \frac{\sum q_1 p_0}{\sum q_1} \quad (8.33)$$

Этому индексу, как и в предыдущем случае, соответствует два абсолютных изменения, отражающих собственно изменение средней цены и выручки

$$\Delta'_{\bar{p}} = \bar{p}_1 - \bar{p}_{\text{усл.}} \quad (8.34)$$

$$\Delta''_{\text{qr}(\bar{p})} = (\bar{p}_1 - \bar{p}_{\text{усл.}}) \cdot \sum q_1 = \sum q_1 p_1 - \sum q_1 p_0 \quad (8.35)$$

10.3) Индекс структурных сдвигов. Это индекс **постоянного (фиксированного) состава**, так как здесь неизменным является цена продукции, взятая за базисный период. Он (индекс) отражает изменение средней цены продукции за счет изменения долей в реализации различной продукции. Структура может стать хуже, тогда средняя цена снижается. Если структура улучшается, то средняя цена реализации увеличивается.

$$I_{\text{структурных сдвигов}} = \frac{\bar{p}_{\text{усл.}}}{\bar{p}_0} = \frac{\sum q_1 p_0}{\sum q_1} \div \frac{\sum q_0 p_0}{\sum q_0} \quad (8.36)$$

или этот индекс можно представить как частное агрегатного индекса количества и структуры товара и индекса численности продукции

$$I_{\text{структурных сдвигов}} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \div \frac{\sum q_1}{\sum q_0}, \quad (8.37)$$

Этому индексу (как и предыдущим двум) соответствует два абсолютных изменения

$$\Delta' = \bar{p}_{\text{усл.}} - \bar{p}_0 \quad (8.38)$$

$$\Delta''_{\text{qr(структ.)}} = (\bar{p}_{\text{усл.}} - \bar{p}_0) \cdot \sum q_1 \quad (8.39)$$

Соотношение индексов:

$$I_{\text{структурных сдвигов}} = I_{\bar{p}(\text{нс})} \div I_{\bar{p}(\text{фс})} \quad (8.40)$$

В 11. Другие формы построения индексов.

В индексном анализе мы сталкиваемся с тем, что цель исследования может быть достигнута, если использовать различные алгебраические преобразования уже известных индексов.

11.1) Геометрические индексы (идеальный индекс Фишера):

Этот индекс был получен как корень квадратный из произведения индекса Пааше на индекс Ласпейреса.

$$11.1) I_p = \sqrt{\frac{\sum q_1 p_1}{\sum q_1 p_0} \cdot \frac{\sum q_0 p_1}{\sum q_0 p_0}} \quad (8.41)$$

По аналогии может быть построен геометрический индекс количества товара.

$$11.2) I_q = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1}} \quad (8.42)$$

Эти индексы сложны в экономической интерпретации и не имеют соответствующего абсолютного изменения.

11.2) В экономическом анализе используется общий индекс количества сопоставимой продукции, отражающий динамику только численности товара или продукции. По форме образования он больше похож на индивидуальный индекс и его нельзя отнести к какой-либо другой классификации. С использованием этого индекса мы столкнулись в пункте 10.3.

$$11.3) I'_q = \frac{\sum q_1}{\sum q_0} \quad (8.43)$$

Этот индекс получают как составную часть агрегатного индекса количества и структуры товара (см. формулу 8.36 и п.10.3), т.е. здесь имеет место равенство агрегатного индекса произведению индекса структуры и индекса численности товара

$$11.4) I_q = I'_q \cdot I_{\text{структ.}}$$

или

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum q_1}{\sum q_0} \cdot \left(\frac{\sum q_1 p_0}{\sum q_1} \div \frac{\sum q_0 p_0}{\sum q_0} \right) \quad (8.44)$$

11.3) Следующая комбинация индексов Пааше и Ласпейреса дает нам значение **индекса ассортиментных сдвигов товара**. Расширение ассортимента товара позволяет увеличивать приток покупателей, увеличивать объемы реализации продукции и массу выручки и незначительно повышать розничные или средние цены реализации, что, в конечном счете, позволяет увеличить прибыль предприятия.

Под **ассортиментом товара** понимаем перечень товаров, обязательных для производства или реализации.

Если произошло изменение ассортимента товара, т.е. от одного вида продукции отказались, а ввели другие виды продукции, то здесь возможны два варианта:

- 1) изменилось качество и объём товаров;
- 2) изменились наименования товаров.

На динамику цен оказывает влияние оба этих изменения, и индекс ассортиментных изменений равен:

$$11.4) I_{\text{ассорт. изменений}} = I_{\text{Пааше}} / I_{\text{Ласпейреса}} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \div \frac{\sum p_1 q_0}{\sum p_0 q_0} = 1,004009 = 100,4009\%.$$

$$\text{Например: } I_{\text{Пааше}} = 0,995716 = 99,57 \%$$

$I_{\text{Ласпейреса}} = 0,991740 = 99,17\%$ (если мы не изменяем объём реализации, то объём товарооборота при такой динамике цен уменьшится на 0,83%).

Вывод: цены выросли на 0,4% за счёт изменения ассортимента или изменения ассортимента обусловило рост цен на 0,4%. Нужно отметить, что эти два вывода равносильны.

В 12. Индексный анализ мультипликативных, кратных и комбинированных моделей.

В вопросе 9 мы рассматривали проблему использования синтетической концепции индексов для составления достаточно сложных факторных моделей. Поэтому есть необходимость хотя бы схематично рассмотреть алгоритмы построения индексов и соответствующих им абсолютных изменений.

12.1) Индексный анализ мультипликативных моделей.

$$Y = \sum a \cdot b \cdot c \cdot d$$

$$I_{y_{(обш)}} = \frac{\sum a_1 \cdot b_1 \cdot c_1 \cdot d_1}{\sum a_0 \cdot b_0 \cdot c_0 \cdot d_0} ;$$

$$\Delta_{y_{(обш)}} = \sum a_1 \cdot b_1 \cdot c_1 \cdot d_1 - \sum a_0 \cdot b_0 \cdot c_0 \cdot d_0$$

$$I_{Y(обш)} = I_{y(a)} \cdot I_{y(b)} \cdot I_{y(c)} \cdot I_{y(d)}$$

$$I_{y(a)} = \frac{\sum a_1 \cdot b_0 \cdot c_0 \cdot d_0}{\sum a_0 \cdot b_0 \cdot c_0 \cdot d_0} ;$$

$$\Delta_{y(a)} = \sum a_1 b_0 c_0 d_0 - \sum a_0 b_0 c_0 d_0 = \sum (a_1 - a_0) \cdot b_0 c_0 d_0$$

$$I_{y(b)} = \frac{\sum a_1 b_1 c_0 d_0}{\sum a_1 b_0 c_0 d_0} ;$$

$$\Delta_{y(b)} = \sum a_1 b_1 c_0 d_0 - \sum a_1 b_0 c_0 d_0 = \sum a_1 (b_1 - b_0) c_0 d_0$$

$$I_{y(c)} = \frac{\sum a_1 b_1 c_1 d_0}{\sum a_1 b_1 c_0 d_0} ;$$

$$\Delta_{y(c)} = \sum a_1 b_1 c_1 d_0 - \sum a_1 b_1 c_0 d_0 = \sum a_1 b_1 (c_1 - c_0) d_0$$

$$I_{y(d)} = \frac{\sum a_1 b_1 c_1 d_1}{\sum a_1 b_1 c_1 d_0} ;$$

$$\Delta_{y(d)} = \sum a_1 b_1 c_1 d_1 - \sum a_1 b_1 c_1 d_0 = \sum a_1 b_1 c_1 (d_1 - d_0)$$

12.2) Индексный анализ кратных моделей (этот анализ имеет ограничение, т.к. его нельзя использовать для знакопеременных факторов). В анализе трудоемкости продукции или производительности труда используются более сложные кратные модели, чем мы здесь рассматриваем.

$$y = \frac{a}{b}$$

$$I_{y(общ)} = \frac{y_1}{y_0} = \frac{a_1}{b_1} \div \frac{a_0}{b_0} \quad ; \quad \Delta_{y(общ)} = y_1 - y_0 = \frac{a_1}{b_1} - \frac{a_0}{b_0}$$

$$I_{y(a)} = \frac{y_{ysl}}{y_0} = \frac{a_1}{b_0} \div \frac{a_0}{b_0} \quad ; \quad \Delta_{y(a)} = y_{ysl} - y_0 = \frac{a_1}{b_0} - \frac{a_0}{b_0}$$

$$Y_{y(b)} = \frac{y_1}{y_{ysl}} = \frac{a_1}{b_1} \div \frac{a_1}{b_0} \quad ; \quad \Delta_{y(b)} = y_1 - y_{ysl} = \frac{a_1}{b_1} - \frac{a_1}{b_0}$$

12.3) Индексный анализ комбинированных моделей. К таким моделям можно отнести модели прибыли и рентабельности как отдельных видов продукции, так и в целом предприятия. Для этих экономических показателей можно рассчитать индексы только в том случае, если результативные показатели в отчетный и базисный период имеют одинаковый знак.

$$Y = \sum a \cdot (b - c)$$

$$I_{y(общ)} = \frac{\sum a_1(b_1 - c_1)}{\sum a_0(b_0 - c_0)} \quad ;$$

$$\Delta_{y(общ)} = \sum a_1(b_1 - c_1) - \sum a_0(b_0 - c_0)$$

$$I_{y(a)} = \frac{\sum a_1(b_0 - c_0)}{\sum a_0(b_0 - c_0)} \quad ;$$

$$\Delta_{y(a)} = \sum a_1(b_0 - c_0) - \sum a_0(b_0 - c_0)$$

$$I_{y(b)} = \frac{\sum a_1(b_1 - c_0)}{\sum a_1(b_0 - c_0)} \quad ;$$

$$\Delta_{y(b)} = \sum a_1(b_1 - c_0) - \sum a_1(b_0 - c_0)$$

$$I_{y(c)} = \frac{\sum a_1(b_1 - c_1)}{\sum a_1(b_1 - c_0)} ;$$

$$\Delta_{y(c)} = \sum a_1(b_1 - c_1) - \sum a_1(b_1 - c_0)$$

В 13. Метод абсолютных разностей в индексном анализе.

При построении индексов мы использовали основное и обязательное условие (см. в.3) соответствия индексам абсолютных изменений. В большинстве случаев эти абсолютные изменения находились как разность между числителем и знаменателем индекса, а затем можно было провести алгебраические преобразования – вынести за скобки одинаковые показатели (см. предыдущий вопрос). Это и есть не что иное, как метод абсолютных разностей для нахождения абсолютных изменений результата под влиянием индексируемого фактора. В формулах для мультипликативной системы связей факторов это имеет вид:

$$\Delta_{y(обш)} = \sum a_1 b_1 c_1 d_1 - \sum a_0 b_0 c_0 d_0 \quad (8.45)$$

$$\Delta_{y(a)} = \sum a_1 b_0 c_0 d_0 - \sum a_0 b_0 c_0 d_0$$

$$\Delta_{y(b)} = \sum a_1 b_1 c_0 d_0 - \sum a_1 b_0 c_0 d_0$$

$$\Delta_{y(c)} = \sum a_1 b_1 c_1 d_0 - \sum a_1 b_1 c_0 d_0$$

$$\Delta_{y(d)} = \sum a_1 b_1 c_1 d_1 - \sum a_1 b_1 c_1 d_0$$

Приложение 1. Наличие ресурсов с/х предприятий в районах Гродненской области

№ группы	Наименование района	Количество предприятий	Сельскохозяйственные угодья		Численность работников, тыс. чел.	Отработано в с/х, тыс. час.	Внеоборотные активы, млн. руб.	Оборотные активы, млн. руб.	Валовая продукция, млн. руб.
			Гектары, тыс. га	Баллы, тыс. га					
I	A	1	2	3	4	5	6	7	8
	Берестовицкий	11	42	1529	3	6376	17492	7	5534
	Волковысский	10	34	1289	3	6674	15848	5	4047
	Гродненский	16	95	3355	9	16440	61156	14	15833
	Мостовский	13	49	1662	4	8984	22563	9	5280
II	Щучинский	14	54	1698	5	9769	18424	15	5241
	Итого	64	274	9533	24	48243	135483	50	35935
	Дятловский	17	60	1599	4	9893	19160	11	3525
	Зельвенский	14	56	1962	5	9072	18464	10	4424

Продолжение приложения 1

		А	1	2	3	4	5	6	7	8
		Корелицкий	12	64	2527	5	9729	23766	15	4855
		Новогрудский	17	60	1748	4	9162	21596	4	4049
		Свяслотский	10	29	1014	2	4910	11392	4	1813
		Слонимский	13	45	1500	3	6454	17393	5	3379
		Итого	83	314	10350	23	49220	111771	49	22045
III		Вороновский	15	63	2034	4	7869	13046	14	3116
		Иньевский	20	53	1730	4	7578	13035	18	2832
		Лидский	13	53	1633	4	7264	18758	13	2753
		Островецкий	14	53	1555	3	6978	14018	11	2660
		Ошмянский	11	45	1187	3	5530	10584	7	2278
		Сморгонский	11	49	1616	3	6526	16054	13	2217
		Итого	84	316	9755	21	41745	85495	76	15856
		Всего	231	904	29638	68	139208	332749	175	73836

Приложение 2. Обеспеченность ресурсами с/х предприятий в районах

№ группы	Наименование района	Качество почвы, балл	Коэффициент распаханности	Обеспеченность с/х угодий ресурсами				Фондовооруженность
				Трудовыми		Активами		
				чел./га	тыс. час./га	внеоборотные	оборотные	
I	A	1	2					7
	Берестовицкий	36	0,734	0,071	0,152	416,9	0,162	5846,4
	Волковский	38	0,809	0,091	0,195	462,3	0,137	5095,9
	Гродненский	35	0,738	0,092	0,173	641,8	0,152	6991,6
	Мостовский	34	0,604	0,084	0,184	461,3	0,194	5459,3
II	Щучинский	31	0,624	0,096	0,181	341,7	0,27	3569,8
	В среднем	35	0,7	0,088	0,176	493,8	0,182	5611,7
	Дятловский	27	0,672	0,074	0,166	321	0,186	4328
	Зельвенский	35	0,701	0,083	0,163	331,7	0,178	4013,9

Продолжение приложения 2

	A	1	2	3	4	5	6	7
	Зельвенский	35	0,701	0,083	0,163	331,7	0,178	4013,9
	Кореличский	40	0,611	0,074	0,152	371,6	0,228	5033
	Новогрудский	29	0,582	0,072	0,152	358,6	0,241	4991,1
	Свислочский	36	0,72	0,077	0,172	398,8	0,134	5171
	Слонимский	33	0,739	0,069	0,144	386,9	0,119	5572,9
III	В среднем	33	0,661	0,075	0,157	357	0,189	4776,5
	Вороновский	32	0,617	0,064	0,125	206,7	0,225	3213,4
	Ивьевский	27	0,6	0,068	0,12	206,5	0,277	3030,7
	Лидский	31	0,573	0,068	0,136	352,3	0,237	5146,1
	Островский	29	0,675	0,055	0,131	263	0,216	4750,1
	Опшмянский	27	0,704	0,062	0,124	237,1	0,166	3821
	Сморгонский	33	0,555	0,058	0,134	330,7	0,261	5743,9
	В среднем	30	0,619	0,063	0,128	262,2	0,233	4166
	В среднем по совокупности	32	0,658	0,075	0,152	364,3	0,203	4888,7

Приложение 3. Эффективность использования ресурсов в районах

№ группы	Наименование района	Эффективность производства с/х продукции (выход валовой продукции из расчета на единицу ресурса)					
		тыс руб./га	тыс руб./бал. га	тыс руб./чел	руб./час	тыс руб./ тыс руб	тыс руб./ тыс руб
	A	1	2	3	4	5	6
I	Берестовицкий	131,89	3,62	1849,63	867,96	0,316	815,76
	Волковысский	118,06	3,14	1301,36	606,42	0,255	859,65
	Гродненский	166,16	4,72	1810,13	963,09	0,259	1092,17
	Мостовский	107,94	3,18	1277,44	587,67	0,234	555,93
	Щучинский	97,21	3,09	1015,58	536,53	0,284	359,76
	В среднем	130,98	3,77	1488,44	744,89	0,265	717,92
II	Дятловский	59,05	2,2	796,35	356,36	0,184	317,84
	Зельвенский	79,46	2,26	961,69	487,63	0,24	446,93
	Корелицкий	75,92	1,92	1028,14	499,01	0,204	332,3
	Новогрудский	67,23	2,32	935,75	441,93	0,187	278,7
	Свислоский	63,49	1,79	823,16	369,33	0,159	473,23
	Слонимский	75,16	2,25	1082,7	523,57	0,194	630,19
	В среднем	70,42	2,13	942,12	447,9	0,197	371,63
III	Вороновский	49,36	1,53	767,49	395,99	0,239	219,44
	Ивьевский	44,85	1,64	658,38	373,67	0,217	161,67

Продолжение приложения 3

	1	2	3	4	5	6
А						
Лидский	51,7	1,69	755,22	378,96	0,147	218,3
Островецкий	49,91	1,71	901,49	381,24	0,19	231,39
Опмянский	51,02	1,92	822,33	411,91	0,215	306,62
Сморгонский	45,66	1,37	793,11	339,68	0,138	175,15
В среднем	48,64	1,63	772,61	379,82	0,185	208,88
В среднем по совокупности	80,83	2,49	1084,79	530,4	0,222	398,5

Учебное издание

Кривенкова-Леванова Людмила Николаевна
Грудько Светлана Владимировна

СТАТИСТИКА
КУРС ЛЕКЦИЙ.
Часть I

Учебно-методическое пособие

Компьютерная верстка: С.В. Грудько, И.В. Середа, А.А. Островская

Подписано в печать 02.04.2012.
Формат 60х84/16. Бумага офсетная.
Печать Riso. Усл. печ. л. 9,88. Уч.-изд. л. 8,83.
Тираж 160 экз. Заказ 2738

Издатель и полиграфическое исполнение:
Учреждение образования
«Гродненский государственный аграрный университет»
ЛИ № 02330/0548516 от 16.06.2009.
Ул. Терешковой, 28, 230008, г. Гродно.