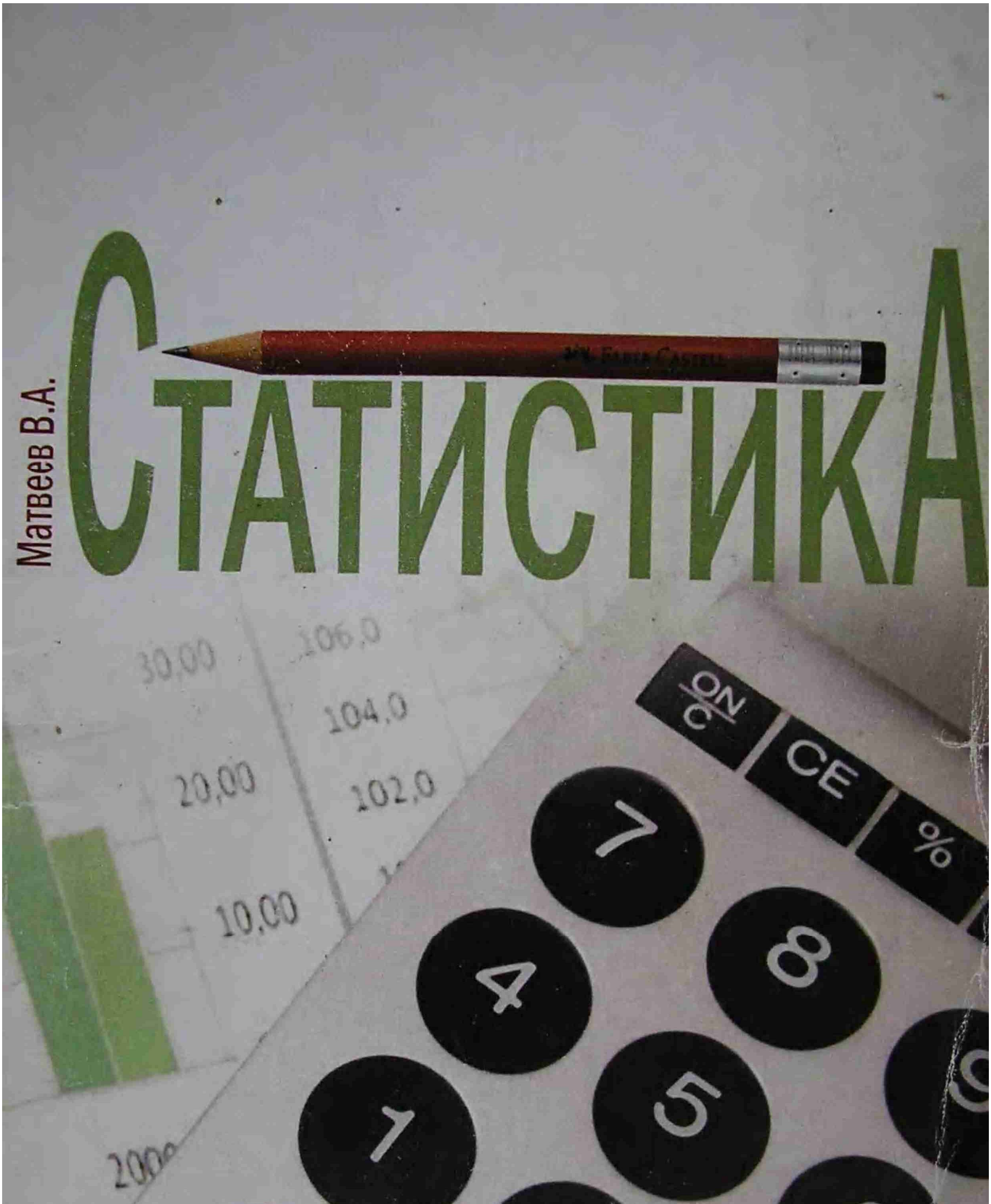
Матвеев В.А.

СтатистикА



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Нижегородский государственный университет им. Н.И. Лобачевского

Национальный исследовательский университет

Матвеев В.А.

СТАТИСТИКА

Учебно-методическое пособие

Нижний Новгород
2015

УДК 311(075.8)
ББК У051
М 33

М-33 Матвеев В.А. СТАТИСТИКА: Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2015.

Рецензенты: **проф. В.Н. Едронова**
проф. А.И. Козлов

Учебно-методическое пособие отражает основное содержание первого раздела общепрофессиональной учебной дисциплины «Статистика», являющейся базовой (общепрофессиональной) частью профессионального цикла ООП бакалавриата по направлениям подготовки «Экономика», «Менеджмент» и «Торговое дело».

УДК 311(075.8)
ББК У051

Предисловие

В современном обществе эффективность управления экономикой в значительной степени зависит от качества информационной базы, основным компонентом которой является статистика, осуществляющая сбор, научную обработку, обобщение и анализ информации и предоставляющая возможность выявления взаимосвязей в экономике, изучения динамики ее развития, проведения международных сопоставлений и т.д. Поэтому в системе экономического образования особое место отводится изучению статистики, которая является базовой научной дисциплиной, формирующей профессиональный уровень современного экономиста.

Цель настоящего учебно-методического пособия – дать представление о содержании статистики, познакомить с ее основными понятиями, методологией и методикой расчета важнейших статистических показателей, что является основой для применения статистики в конкретных исследованиях социально-экономических явлений и процессов.

Учебно-методическое пособие разработано в соответствии с учебной программой курса «Статистика», составленной на основе требований федерального государственного образовательного стандарта высшего профессионального образования.

Учебно-методическое пособие может оказаться полезным не только студентам и преподавателям, но и всем интересующимся проблемами анализа конкретных процессов в области экономики, финансов и управления, а также в иных ситуациях, связанных с анализом массовых статистических данных.

Содержание

Часть 1. Теория статистики

Глава 1. Предмет и методы статистики

1.1. История статистики

Слово «статистика» происходит от лат. status – состояние. В средние века оно означало политическое состояние государства. В науку этот термин был введен в XVIII в. немецким ученым Г. Ахенвалем.

Собственно как наука статистика возникла в конце XVII в., хотя статистический учет существовал уже в глубокой древности. Так, известно, что еще за 5 тыс. лет до н.э. проводились переписи населения в Китае, осуществлялось сравнение военного потенциала разных стран, велся учет имущества граждан в Древнем Риме и т.д.

У истоков статистической науки стояли две школы – немецкая описательная и английская школа политических арифметиков.

Представители описательной школы стремились систематизировать существующие способы описания государств, создать их теорию и разработать подробную схему, а также вести описание только в словесной форме, без цифр и лишь на момент наблюдения.

Политические арифметики видели основное назначение статистики в изучении массовых общественных явлений с помощью числовых характеристик. При этом они осознавали необходимость учета в статистических исследованиях требований закона больших чисел, поскольку закономерность может проявляться лишь при достаточно большом объеме изучаемой совокупности. Эта школа имела два направления: социальное и экономическое.

Как показала история, последнее слово в статистической науке осталось за школой политических арифметиков, благодаря которой статистика перешла от описания явлений и процессов к их количественному измерению и исследованию, а также выработке вероятных гипотез будущего развития.

Прогрессу статистической методологии в XIX и XX вв. способствовали работы английской математической школы, а в России – труды земских статистиков и представителей социологической и философско-математических школ. В XXI в. статистическая методология развивается на базе новых информационных технологий, сплошного статистического наблюдения и расширения сферы применения выборочных обследований.

1.2. Предмет и методы статистики

С развитием статистической науки и расширением сферы практической статистической работы изменилось и содержание понятия «статистика». В настоящее время данный термин употребляется в трех основных значениях:

1) под статистикой понимают отрасль практической деятельности по сбору, обработке, анализу и публикации массовых данных о самых различных явлениях общественной жизни (статистический учет);

2) статистикой называют цифровой материал, который служит для характеристики какой-либо области общественных явлений (статистические данные);

3) статистика – это отрасль знания, особая научная дисциплина, которая изучает с количественной стороны качественную сторону массовых общественных явлений или их содержание, а также количественное выражение закономерностей общественного развития в конкретных условиях места и времени.

Существует три группы методов статистики:

1) метод массовых наблюдений, который заключается в сборе первичного статистического материала;

2) метод группировок, суть которого заключается в том, чтобы подразделить изучаемые явления на однородные группы по каким-либо признакам;

3) метод обобщающих показателей, позволяющий характеризовать изучаемые явления и процессы при помощи абсолютных, относительных и средних величин с целью выявить их взаимосвязи, масштабы и закономерности развития, а также дать прогнозные оценки.

1.3. Задачи и организация статистики

В соответствии с Конституцией РФ руководство статистической деятельностью в стране осуществляет Федеральная служба государственной статистики (Росстат). Единая система государственной статистики включает Росстат (центральный аппарат и структурные подразделения), его территориальные органы в республиках, краях, областях, Москве и Санкт-Петербурге, в других районах и городах, а также подведомственные организации. Росстат определяет формы и методы сбора, обработки данных, методологию и методику расчета статистических показателей (статистические стандарты).

Кроме государственной статистики определенную часть статистической информации дает ведомственная статистика, статистика коммерческих и некоммерческих организаций.

Статистика решает несколько задач:

- предоставление органам государственной власти и местного самоуправления информации, необходимой при разработке экономической и социальной политики;
- обеспечение информацией о состоянии и развитии экономики и социальной сферы руководителей предприятий, менеджеров и предпринимателей, необходимой для принятия адекватных управленческих решений при организации и расширении производства, управлении производственным процессом, осуществлении инвестиций, сбыта продукции и т.д.;
- информирование об основных итогах и тенденциях социально-экономического развития общественности, научно-исследовательских

учреждений, международных организаций, а также всех заинтересованных лиц.

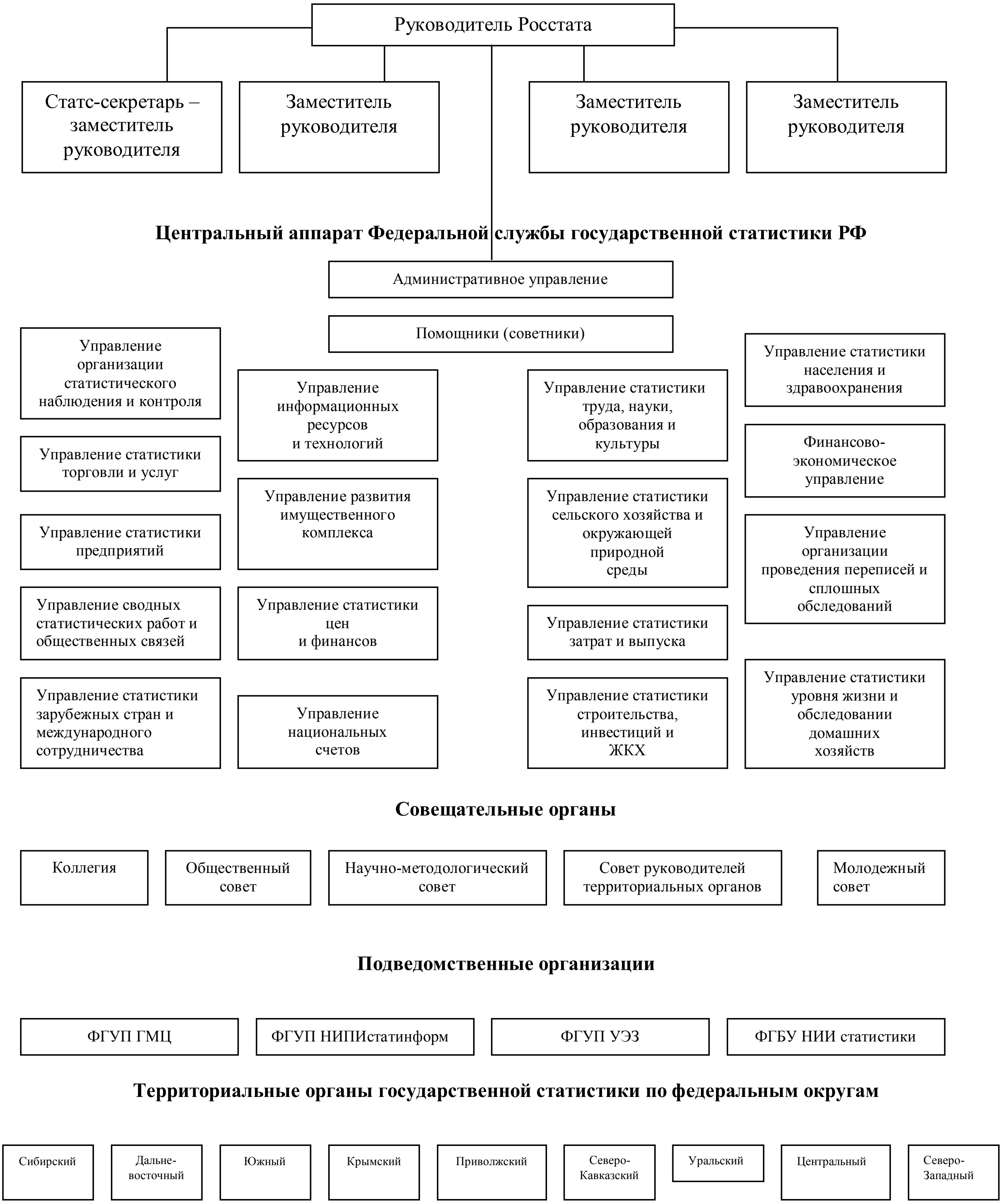


Рис. 1.1. Структура органов государственной статистики в РФ

1.4. Основные категории статистики

Статистика изучает свой предмет при помощи пяти основных категорий.

1. Статистическая совокупность – это множество социально-экономических объектов или явлений общественной жизни, объединенных качественной основой, общей связью, но отличающихся друг от друга отдельными признаками (например, совокупность домохозяйств, семей, предприятий и т.п.). Статистическая совокупность может быть однородной, для которой один или несколько существенных признаков являются общими, а в отношении остальных признаков она оказывается неоднородной, то есть состоит из объектов или явлений разного типа. Так, совокупность студентов может быть однородной по направлению подготовки, но неоднородной по успеваемости.

2. Единица совокупности – это первичный элемент статистической совокупности, который является носителем признаков, регистрируемых при статистическом наблюдении, а также основой счета.

3. Признак – качественная особенность единицы совокупности. По характеру отображения свойств изучаемого явления признаки подразделяются на две основные группы: количественные и неколичественные.

Количественные признаки выражаются числами (возраст человека, стаж работы, рост, вес и т.д.) и могут быть дискретные, значения которых известны точно, и интервальные, значения которых колеблются в определенных границах (от и до). Например, возраст человека по паспорту – дискретный признак, а возраст человека, определяемый на глаз, возраст нескольких человек (от и до) – интервальный признак.

Неколичественные признаки выражаются словами, и, если выражающих слов только два, то признаки называются альтернативными (например, пол человека может быть мужской или женский), а если выражающих слов больше двух, то признаки называются атрибутивными (например, профессия человека: учитель, врач, столяр, дворник и т.д.).

4. Статистический показатель – это категория, которая отображает количественные характеристики (размеры, соотношения признаков) социально-экономических явлений в условиях конкретного места и времени.

Статистические показатели могут быть объемные, связанные с измерением изучаемой совокупности (численность населения, работников предприятия, объем произведенной продукции и т.д.), и расчетные (средняя заработная плата, производительность труда, себестоимость продукции, рентабельность и т.д.), а также плановые, фактические и прогностические.

5. Система статистических показателей – это совокупность статистических показателей, отражающая взаимосвязи между изучаемыми явлениями и процессами как на макроуровне, так и на микроуровне.

Контрольные задания 1

Вариант 1

1. Термин «статистика» в науку был введен
а) Архимедом; б) Ахенвалем; в) Аристотелем; г) Петти.
2. Форма собственности предприятия – это
а) количественный признак; б) атрибутивный признак; в) альтернативный признак; г) дискретный признак.
3. Органами государственной статистики РФ в настоящее время является
а) Министерство финансов РФ; б) Росстат; в) Госкомстат; г) Министерство статистики.
4. Статистика решает задачу
а) предоставление экспертных оценок параметров социально-экономического развития; б) предоставления информации о состоянии и возможных направлениях развития природных явлений; в) предоставления информации об основных итогах и тенденциях социально-экономического развития общественности, научно-исследовательским учреждениям, международным организациям и т.д.; г) предоставления словесных описаний государства и муниципальных образований, необходимых при разработке экономической и социальной политики.
5. Совокупность – это
а) множество явлений общественной жизни или социально-экономических объектов, объединенных качественной основой, общей связью и отличающихся друг от друга отдельными признаками; б) несколько социально-экономических объектов или явлений общественной жизни, которые отличаются друг от друга отдельными признаками; в) ряд сходных по внешнему виду массовых общественных явлений; г) последовательность событий, условия существования каждого из которых определяются предыдущим событием.
6. Статистические показатели могут быть
а) прогностическими, плановыми и фактическими; б) детерминированными и стохастическими; в) атрибутивными и альтернативными; г) функциональными и корреляционными.
7. Атрибутивные признаки
а) принимают числовые значения в определенных границах (от и до); б) принимают точные числовые значения; в) выражаются тремя и более словами; г) имеют непосредственное количественное выражение.
8. Статистические совокупности могут быть
а) альтернативными и атрибутивными; б) однородными и неоднородными; в) прогностическими, плановыми и фактическими; г) абсолютными, относительными и средними.
9. К методам статистики относится
а) массовое наблюдение; б) экстраполяция; в) горизонтальный и вертикальный анализ; г) абстрагирование.
10. Под статистикой понимается
а) отрасль практической деятельности по сбору, обработке, анализу и публикации массовых данных о природных явлениях; б) отрасль практической деятельности по сбору, обработке, анализу и публикации данных о массовых общественных явлениях; в) отрасль знания, которая изучает с качественной стороны количественную сторону массовых общественных явлений; г) наука, которая изучает происхождение и закономерности развития человеческого общества, а также место в нём человека.

Вариант 2

1. Политические арифметики стремились
а) создать теорию и подробную схему словесного описания государств; б) измерить и исследовать массовые общественные явления с помощью числовых характеристик; в) построить математические модели социально-экономического развития; г) создать художественные изображения государств.
2. Стаж работников от 3 до 7 лет – это
а) интервальный признак; б) альтернативный признак; в) неколичественный признак; г) дискретный признак.
3. Органами государственной статистики РФ в настоящее время является
а) Федеральная миграционная служба РФ; б) Госкомстат; в) Федеральная служба государственной статистки; г) Статистический комитет РФ.
4. Статистика решает задачу
а) угадывания основных параметров социально-экономического развития; б) математического моделирования основных показателей состояния и развития экономики и социальной сферы; в) обеспечения информацией о состоянии и развитии экономики и социальной сферы организаторов производства, руководителей предприятий, менеджеров и предпринимателей, необходимой для принятия адекватных управленческих решений при организации и расширении производства, осуществлении инвестиций, сбыта продукции и т.д.
г) предоставления информации о стратегии и тактике управления социально-экономическими процессами, необходимых при разработке экономической и социальной политики.
5. Статистический показатель – это

а) количественная характеристика социально-экономических явлений в условиях конкретного места и времени; б) качественная характеристика количественно определенного социально-экономического явления; в) характеристика изменения во времени или распределения в пространстве конкретного социально-экономического явления; г) качественная особенность единицы совокупности.
6. Статистические показатели могут быть
а) дискретными и интервальными; б) альтернативными и альтернативными; в) объемными и расчетными; г) однородными и неоднородными.
7. Альтернативные признаки
а) имеют непосредственное количественное выражение; б) принимают точные числовые значения; в) принимают числовые значения в определенных границах (от и до); г) выражаются одним из двух слов.
8. Расчетные статистические показатели
а) регистрируются при статистическом наблюдении; б) определяются расчетным путем на основе первичных статистических данных; в) являются прогнозными оценками; в) связаны с измерением статистической совокупности.
9. К методам статистики относится
а) обобщение; б) сравнение; в) горизонтальный и вертикальный анализ; г) группировка.
10. Под статистикой понимается
а) наука, которая изучает с количественной стороны качественную сторону массовых общественных явлений, количественное выражение закономерностей общественного развития в условиях конкретного места и времени; б) наука о наиболее общих законах развития природы, человеческого общества и мышления; в) отрасль практической деятельности по сбору, регистрации, обработки, обобщению, хранению и предоставлению данных о хозяйственных операциях; г) наука, которая изучает величины, количественные отношения и пространственные формы.

Вариант 3

1. Статистическая наука возникла
а) в начале XX в.; б) во первой половине XIX в.; в) в конце XVII в.; г) в середине XIII в.
2. Стаж работника 5 лет – это
а) атрибутивный признак; б) альтернативный признак; в) словесный признак; г) дискретный признак.
3. Кроме органов государственной статистики статистическую информацию предоставляет
а) статистика коммерческих организаций; б) Росстат; в) Госкомстат; г) подведомственные организации.
4. Статистика решает задачу
а) перехвата информации об основных итогах и тенденциях социально-экономического развития; б) предоставления информации об основных итогах и тенденциях социально-экономического развития общественности, научно-исследовательских учреждений, международных организаций и т.д.; в) словесного описания итогов и основных тенденций социально-экономического развития; г) обеспечения слухами органов государственной власти и местного самоуправления, необходимых при разработке экономической и социальной политики.
5. Признак – это
а) количественная характеристика социально-экономических явлений в условиях конкретного места и времени; б) качественная особенность единицы совокупности; в) первичный элемент статистической совокупности; г) количественная особенность единицы совокупности.
6. Статистические показатели могут быть
а) абсолютными, относительными и средними; б) количественными и неколичественными; в) альтернативными и альтернативными; г) простыми и составными.
7. Дискретные признаки
а) не имеют непосредственного количественного выражения; б) принимают точные числовые значения; в) выражаются одним из двух слов; г) принимают числовые значения в определенных границах (от и до).
8. Объемные статистические показатели
а) определяются расчетным путем на основе первичных статистических данных; б) связаны с измерением статистической совокупности; в) являются прогнозными оценками; г) регистрируются при статистическом наблюдении.
9. К методам статистики относится
а) систематизация; б) обобщающие показатели; в) математическое моделирование; г) экспертные оценки.
10. Под статистикой понимается
а) наука, которая изучает происхождение и развитие человеческого общества, а также место в нём человека; б) цифровой материал, который служит для характеристики какой-либо области массовых общественных явлений; в) отрасль практической деятельности по сбору, регистрации, обработки, обобщению, хранению и предоставлению данных о хозяйственных операциях; г) отрасль практической деятельности по сбору, передаче, обработке, хранению и предоставлению информации.

Вариант 4

1. У истоков статистической науки стояла
а) русская социологическая школа; б) китайская философская школа; в) немецкая математическая школа; г) английская школа политических арифметиков.
2. Внешнеэкономическая операция (экспорт или импорт) – это
а) атрибутивный признак; б) альтернативный признак; в) интервальный признак; г) количественный признак.
3. Кроме органов государственной статистики статистическую информацию предоставляет
а) ведомственная статистика; б) подведомственные организации; в) Федеральная служба государственной статистки; г) Министерство статистики.
4. Статистика решает задачу
а) планирование и прогнозирование основных показателей социально-экономического развития; б) словесного описания состояния и развития экономики и социальной сферы; в) математического моделирования социально-экономического развития; г) обеспечения информацией органов государственной власти и местного самоуправления, необходимой при разработке экономической и социальной политики.
5. Единица совокупности – это
а) качественная характеристика социально-экономических явлений в условиях конкретного места и времени; б) качественная особенность единицы совокупности; в) первичный элемент статистической совокупности; г) конкретное числовое значение признака или показателя.
6. Статистические признаки могут быть
а) абсолютными, относительными и средними; б) количественными и неколичественными; в) объемными и расчетными; г) простыми и составными.
7. Интервальные признаки
а) принимают числовые значения в определенных границах (от и до); б) принимают целые числовые значения; в) выражаются одним или несколькими словами; г) не имеют непосредственного количественного выражения.
8. В неоднородной статистической совокупности
а) имеется один или несколько существенных признаков, которые являются общими; б) имеются объекты или явления разного типа; в) все числовые значения признаков являются одинаковыми; г) нет качественной основы или общей связи между объектами или явлениями.
9. К методам статистики относится
а) индукция; б) синтез; в) группировка; г) обобщение.
10. Под статистикой понимается
а) система данных, которая служит для характеристики имущественного и финансового положения экономического объекта и результатов ее хозяйственной деятельности; б) наука о наиболее общих законах развития природы, человеческого общества и мышления; в) наука, которая изучает величины, количественные отношения и пространственные формы; г) цифровой материал, который служит для характеристики какой-либо области массовых общественных явлений.

Вариант 5

1. У истоков статистической науки стояла
а) английская социологическая школа; б) немецкая описательная школа; в) немецкая математическая школа; г) русская философская школа.
2. Заработная плата работника – это
а) количественный признак; б) альтернативный признак; в) словесный признак; г) атрибутивный признак.
3. Система органов государственной статистики включает: а) статистику коммерческих организаций; б) центральный аппарат и структурные подразделения Федеральной службы государственной статистики; в) ведомственную статистику; г) статистику некоммерческих организаций.
4. Статистика решает задачу
а) предоставления информации о состоянии и развитии экономики и социальной сферы; б) качественного анализа состояния и развития экономики и социальной сферы; в) математического моделирования социально-экономического развития; г) определение причин и условий существования массовых общественных явлений.
5. Система статистических показателей – это
а) множество статистических показателей, которое отражает взаимосвязи на макроуровне и микроуровне между социально-экономическими явлениями и процессами; б) множество явлений общественной жизни или социально-экономических объектов, объединенных качественной основой, общей связью и отличающихся друг от друга отдельными признаками; в) множество статистических признаков, отображающих свойства социально-экономических явлений, а также взаимосвязи между ними на макроуровне и микроуровне; г) совокупность социально-экономических объектов или явлений, взаимосвязанных между собой, но отличающихся друг от друга отдельными признаками.
6. Статистические признаки могут быть

а) альтернативными и атрибутивными; б) функциональными и стохастическими; в) прогностическими, плановыми и фактическими; г) абсолютными, относительными и средними.
7. Количественные признаки
а) выражаются одним или несколькими словами; б) выражаются одним из двух слов; в) выражаются тремя и более словами; г) имеют непосредственное числовое выражение.
8. В однородной статистической совокупности
а) один или несколько существенных признаков являются общими; б) все числовые значения признаков являются одинаковыми; в) имеются объекты или явления разного типа; г) нет качественной основы или общей связи между объектами или явлениями.
9. К методам статистики относится
а) сравнительный анализ; б) массовые наблюдения; в) дедукция; г) абстрагирование.
10. Под статистикой понимается
а) отрасль практической деятельности по сбору, регистрации, обработки, обобщению, хранению и предоставлению данных о хозяйственных операциях; б) отрасль знания, которая изучает с количественной стороны качественную сторону массовых общественных явлений, количественное выражение закономерностей общественного развития в конкретных условиях места и времени; в) наука о наиболее общих законах развития природы, человеческого общества и мышления; г) отрасль практической деятельности по сбору, передаче, обработке, хранению и предоставлению информации.

Глава 2. Статистическое наблюдение и группировка

2.1. Статистическое наблюдение как первый этап статистического исследования

Статистическое наблюдение – это научно организованный по единой программе и заранее разработанному плану учет всех фактов, характеризующих изучаемые явления, а также сбор полученных на основе этого учета массовых данных. К полученным сведениям предъявляются требования полноты, достоверности, точности, единообразия и сопоставимости.

Любое статистическое исследование необходимо начинать с точной формулировки его цели и конкретных задач. Затем определяется объект (совокупность подлежащих исследованию социально-экономических явлений или точные границы, в пределах которых регистрируются сведения), единица наблюдения (составная часть объекта, которая служит основой счета и обладает признаками, которые подлежат регистрации), а также разрабатывается программа наблюдения. Последняя оформляется в виде бланка (анкеты, формуляра) и содержит перечень вопросов, по которым собираются первичные статистические данные.

После этого необходимо решить организационные вопросы, связанные с определением субъекта (органа, который осуществляет наблюдение), места и времени (периода, в течение которого проводится наблюдение, или момента, к которому относятся регистрируемые сведения), а также вида, формы и способа наблюдения.

2.2. Формы. виды и способы статистического наблюдения

В статистической практике используются три организационные формы наблюдения: 1) отчетность, когда единицы наблюдения самостоятельно предоставляют сведения о своей деятельности в виде бланков установленного образца; 2) специальное обследование (например, обследование бюджетов домашних хозяйств, переписи населения); 3) регистры – базы данных юридической, экономической и справочной информации, необходимые для получения априорной информации о численности и основных характеристиках изучаемой совокупности (например, Единый государственный регистр предприятий и организаций).

По времени регистрации фактов статистическое наблюдение может быть систематическим (постоянно охватывает факты по мере их возникновения), периодическим (охватывает факты регулярно через определенные отрезки времени) и единовременным (охватывает факты по мере необходимости).

По способу учета фактов статистическое наблюдение может быть непосредственное (путем пересчета, взвешивания, измерения, оценки и т.д.), документальное, основанное на систематических записях в первичных документах, подтверждающих тот или иной факт, а в ряде случаев прибегают к опросу населения (например, анкетирование, корреспондентский опрос и т.д.).

Рис. 2.1. Формы, виды и способы статистического наблюдения

С точки зрения полноты охвата фактов различают сплошное наблюдение, то есть полный учет всех единиц совокупности, и несплошное наблюдение, разновидностями которого являются: 1) выборочное наблюдение, при котором характеристику всей совокупности дают по некоторой ее части, отобранной в случайном порядке; 2) основной массив, когда отбирают наиболее крупные единицы, в которых сосредоточена значительная доля изучаемых фактов; 3) монографическое описание, применяемое для подробного изучения единичных, но типичных объектов (например, предприятий, домашних хозяйств, работников и т.д.).

2.3. Сводка и группировка статистических данных

Собранный в процессе статистического наблюдения первичный материал нуждается в определенной обработке, началом которой служит сводка. Сводка – это научно организованная (по единому плану и заранее разработанной программе) обработка материалов статистического наблюдения, включающая в себя контроль, систематизацию и группировку, составление таблиц, а также получение суммарных (итоговых) показателей как по группам, так и по совокупности в целом.

Группировка – это разбиение совокупности на группы, однородные по какому-либо признаку. Особым видом группировки является классификация, под которой понимается устойчивое разграничение объектов, каждый из которых может быть отнесен лишь к одной группе (например, классификации основных фондов, отраслей экономики, форм собственности). Классификация всегда основывается на существенных признаках, которые практически не меняются с течением времени и устанавливаются путем качественного анализа. Таким образом, классификация – это узаконенная, общепринятая, нормативная группировка.

Статистические группировки преследуют несколько целей:

- выделение качественно однородных совокупностей (типов);
- изучение структуры совокупности (строение типов);
- анализ связей между взаимодействующими признаками.

Каждой из этих целей соответствует определенный вид группировки: типологическая, структурная и аналитическая.

2.3. Виды статистических группировок

Статистические группировки можно классифицировать по нескольким критериям.

По числу группировочных признаков различают группировки простые (по одному признаку) и многомерные (по двум и более признакам).

По отношениям между признаками выделяют: иерархические группировки, которые производятся по двум и более признакам, при этом значения второго признака определяются областью значений первого (например, классификация отраслей промышленности по подотраслям); неиерархические группировки, когда строгой зависимости значений второго признака от первого не существует.

По очередности обработки информации группировки бывают первичные, составленные на основе первичных данных, и вторичные, являющиеся результатом перегруппировки ранее сгруппированного материала.

По времени различают статические группировки, дающие характеристику совокупности на определенный момент или за определенный период, и динамические, показывающие переходы единиц с течением времени из одних групп в другие.

2.4. Методологические вопросы построения группировок

При построении группировки по первичной статистической информации необходимо решить несколько вопросов.

1. Выбор группировочного признака X, по которому производится объединение отдельных единиц совокупности в однородные группы.

2. Определение числа групп и размаха интервалов. Интервалы очерчивают количественные границы групп и могут быть равные и неравные, а также закрытые, когда имеется верхняя и нижняя граница (от и до), и открытые, когда одна из границ отсутствует.

При этом не допускаются «пустые» группы, а если эта проблема возникает, строится группировка с неравными интервалами.

В случае равных интервалов для определения числа групп может использоваться формула Стерджесса:

$$n = 1 + 3{,}322 \cdot \lg N, \qquad (1.1)$$

где N – общее число единиц совокупности, n – число групп (интервалов).

Затем определяется размах интервала по формуле

$$h = \frac{X_{\max} - X_{\min}}{n}, \qquad (1.2)$$

где $X_{\max}$ и $X_{\min}$ – наибольшее и наименьшее значение группировочного признака.

3. Определение системы показателей для характеристики групп. Обязательным показателем является частота f, т.е. количество единиц в каждой группе, или частость d, т.е. доля каждой группы в общем объеме изучаемой соокупности:

$$d = \frac{f}{\sum f} \qquad (1.3)$$

Для наиболее наглядного и рационального представления статистических данных применяются статистические таблицы и графики.

2.5. Ряды распределения

Ряд распределения – это группировка, в которой для характеристики групп, расположенных упорядоченно по значению признака, применяется показатель численности группы. Ряды, построенные по качественному признаку, называются атрибутивными, а ряды, построенные по количественному признаку, называются вариационными и могут быть дискретными и интервальными.

Для характеристики заполненности групп в случае равных интервалов используют частоты или частости, а в случае неравных интервалов определяют плотность распределения как отношение частоты или частости к размаху интервала. Таким образом, в первом случае получается абсолютная, во втором – относительная плотность распределения. Определяются также накопленные частоты (частости) для каждого интервала путем суммирования его частоты (частости), а также частот (частостей) всех предшествующих интервалов.

Графическим изображением рядов распределения служат полигон и гистограмма. В первом случае строится замкнутый многоугольник, абсциссами вершин которого являются значения признака, а ординатами – соответствующие им частоты. Во втором случае по оси абсцисс откладывают границы интервалов, являющихся основаниями прямоугольников, площади которых равны или пропорциональны частотам или частостям в соответствующих интервалах.

При этом в практической деятельности важно фактический ряд распределения привести к известному теоретическому.

2.6. Критерий Пирсона

При построении статистических моделей широко применяется нормальное распределение. Функция плотности нормального распределения вероятностей для непрерывной случайной величины X имеет следующий вид:

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\bar{X})^2}{2\sigma^2}}, \tag{1.4}$$

где $\bar{X}$ – средняя величина; σ – среднее квадратическое отклонение.

Если подставить нормированное отклонение $t = \dfrac{X-\bar{X}}{\sigma}$, средняя величина для которого равна нулю, а среднее квадратическое отклонение равно единице, то функция принимает следующий вид:

$$p(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \tag{1.5}$$

Для проверки соответствия фактического распределения теоретическому определяется критерий согласия Пирсона по формуле

$$\chi = \sum \frac{(f-f')^2}{f'} \tag{1.6}$$

где f – фактические частоты; $f' = \dfrac{Nh}{\sigma} p(t)$ – теоретические частоты.

Табличные значения критерия Пирсона (см. Приложение 3) зависят от принятого уровня значимости α = 0,05 или 5% и числа степеней свободы v. Последнее можно определить путем вычитания числа связей теоретических и фактических частот, ограничивающих вариацию (средняя величина, среднеквадратическое отклонение и общее число единиц совокупности).

Таким образом, число степеней свободы будет равно

$$v = L - r - 1, \tag{1.7}$$

где L – число групп, r – число параметров распределения.

Если расчетное значение критерия меньше табличного, то считается, что фактическое распределение адекватно теоретическому, а расхождения между ними являются несущественными.

2.7. Методические указания

Пусть по нескольким предприятиям имеются следующие первичные статистические данные о выручке от продажи: 200, 350, 600, 800, 750, 680, 960, 150, 110, 120, 200, 600, 800, 450, 560, 130, 450, 600, 800, 450, 500, 800, 960, 750, 1000, 450, 500, 600, 120, 100 млн. р. Построим дискретную и интервальную группировку, полигон и гистограмму.

1. Определим общее число предприятий: N = 30.

2. Упорядочим предприятия по возрастанию выручки от продажи X, а также определим ее частоту f: 100, 110, 120, 120, 130, 150, 200, 200, 350, 450, 450, 450, 450, 500, 500, 560, 600, 600, 600, 600, 680, 750, 750, 800, 800, 800, 800, 960, 960, 1000. Например, выручка от продажи 100 млн. р. встречается (повторяется) 1 раз, 110 млн. р. – 1 раз, 120 млн. р. – 2 раза и т.д.

3. Построим дискретную группировку в виде статистической таблицы:

X	100	110	120	130	150	200	350	450	500	560	600	680	750	800	960	1000	Итого
f	1	1	2	1	1	2	1	4	2	1	4	1	2	4	2	1	

4. Для контроля подсчитаем итог: $\sum f = N = 30.$

5. Построим дискретную группировку в виде статистического графика (полигона):

Рис. 2.2. Полигон дискретного ряда распределения предприятий по выручке от продажи, млн. р.

6. Определим число групп или интервалов по формуле (1.1): $n = 1 + 3{,}322 \cdot \lg 30 = 5{,}907 \approx 6.$

7. Определим размах интервала по формуле (1.2): $h = \dfrac{1000 - 100}{6} = 150.$

8. Построим интервалы выручки от продажи по возрастанию *X*. Для этого примем нижнюю границу первого интервала равной 100, а его верхнюю границу найдем путем прибавления размаха интервала, т.е. 100 + 150 = 250. Таким образом, первый интервал имеет границы от 100 до 250 млн. р. Далее примем нижнюю границу второго интервала равной верхней границе предыдущего интервала, т.е. 250, а верхнюю границу найдем аналогично 250 + 150 = 400 и т.д. Таким образом, второй интервал имеет границы от 250 до 400 млн. р. Обычно первый и последний интервалы оставляют открытыми: до 250 млн. р. и 850 млн. р. и более.

9. Определим частоту каждого интервала *f*. Например, выручка от продажи от 100 до 250 млн. р. встречается у 8 предприятий, выручка от 250 до 400 млн. р. – только у 1 предприятия и т.д. При этом нижняя граница интервала включается в соответствующую группу, а верхняя – не включается. Например, от 100 млн. р. (включительно) до 250 млн. р. (не включительно), от 250 (включительно) до 400 млн. р. (не включительно) и т.д.

10. Построим интервальную группировку в виде статистической таблицы.

X	*f*
до 250	8
250-400	1
400-550	6
550-700	6
700-850	6
850 и более	3
Итого:	

11. Для контроля подсчитаем итог: $\sum f = N = 30.$

12. Построим дискретную группировку в виде статистического графика (гистограммы):

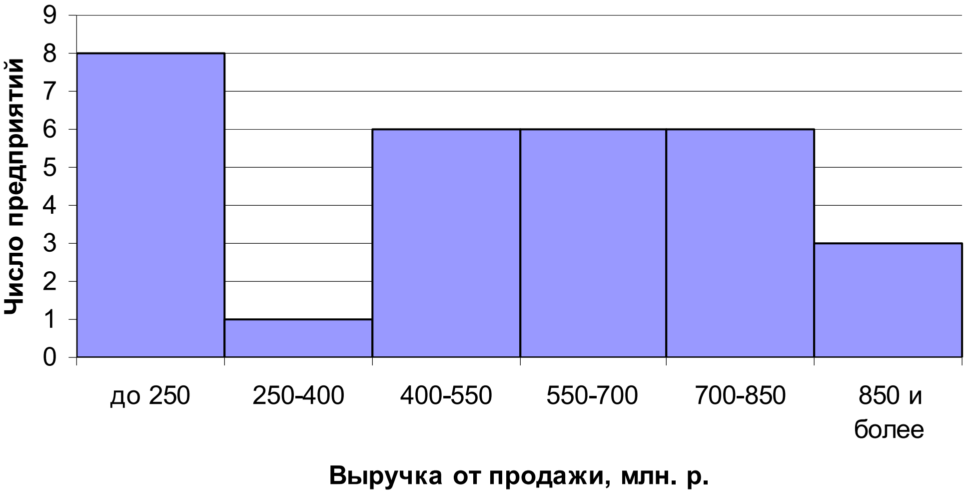


Рис. 2.3. Гистограмма интервального ряда распределения предприятий по выручке от продажи, млн. р.

Контрольные задания 2

Вариант 1

Имеются следующие первичные статистические данные о производственных затратах нескольких предприятий: 150, 1000, 400, 250, 100, 300, 200, 270, 400, 100, 800, 700, 150, 450, 700, 500, 800, 500, 800, 100, 600, 250, 300, 100, 700, 150, 450, 700 млн. р. Построить дискретную и интервальную группировку, полигон и гистограмму.

Вариант 2

Имеются следующие первичные статистические данные о запасах оборотных средств нескольких предприятий: 400, 350, 150, 200, 250, 600, 450, 500, 350, 200, 250, 400, 180, 650, 700, 150, 320, 380, 200, 250, 150, 350, 400, 220, 300, 350, 150, 200, 250 млн. р. Построить дискретную и интервальную группировку, полигон и гистограмму.

Вариант 3

Имеются следующие первичные статистические данные о чистой прибыли нескольких предприятий: 400, 600, 900, 400, 700, 600, 800, 120, 800, 900, 600, 500, 700, 700, 800, 100, 110, 500, 600, 300, 700, 800, 700, 500, 400, 400, 600, 900, 400, 700, 600 млн. р. Построить дискретную и интервальную группировку, полигон и гистограмму.

Вариант 4

Имеются следующие первичные статистические данные о наличии основных фондов нескольких предприятий: 250, 320, 410, 380, 450, 500, 540, 580, 650, 720, 830, 1000, 1100, 380, 450, 800, 650, 450, 800, 750, 450, 450, 800, 310, 450, 380, 450, 500 млн. р. Построить дискретную и интервальную группировку, полигон и гистограмму.

Вариант 5

Имеются следующие первичные статистические данные о собственном капитале нескольких предприятий: 160, 148, 190, 150, 174, 175, 174, 150, 148, 181, 190, 195, 165, 173, 174, 190, 158, 176, 168, 175, 174, 173, 180, 181, 200, 145, 165, 196, 187, 165 млн. р. Построить дискретную и интервальную группировку, полигон и гистограмму.

Глава 3. Абсолютные и относительные величины

3.1. Понятие и виды абсолютных величин

Результаты статистического наблюдения регистрируются в форме абсолютных величин, отражающих уровень развития изучаемого явления. Абсолютная величина по охвату совокупности может быть индивидуальной (охватывает часть совокупности) и суммарной (охватывает совокупность в целом). Например, численность группы студентов – индивидуальная абсолютная величины, а общая численность студентов ВУЗа – суммарная.

В статистике, в отличие от математики, абсолютная величина может быть как положительной, так и отрицательной (например, прибыль предприятия и убыток), а также всегда является именованной и измеряется в конкретных единицах, которые могут быть натуральными, условно-натуральными и стоимостными.

Кроме того необходимо различать моментные и периодные абсолютные величины. Первые показывают фактическое наличие изучаемого явления на определенный момент или дату (например, запасы оборотных средств на начало года и т.д.), а вторые относятся к итоговому накопленному результат за определенный отрезок времени (например, объем произведенной продукции за год и т.п.). При этом в отличие от моментных периодные абсолютные величины можно суммировать.

3.2. Единицы измерения абсолютных величин

Натуральные единицы применяются для исчисления качественно однородных объектов (штуки, тонны, литры, метры и т.д.) и могут быть простыми (например, масса груза в тоннах) и составными (например, грузооборот в тонно-километрах, выработка электроэнергии в кВт-часах, рабочее время в человеко-часах и т.д.).

Условно-натуральные единицы применяются для суммирования в натуральных единицах разновидностей объектов одного и того же типа. Например, разные виды топлива (нефть, газ, уголь и т.д.) переводятся в тонны условного топлива со стандартной теплотой сгорания 29 мДж/кг, алкогольные напитки – в литры условного 100%-го спирта, продукция консервной промышленности – в условные консервные банки со стандартной емкостью 0,33 л, разные виды школьных тетрадей – в условные школьные тетради со стандартным объемом 12 листов и т.д.

Например, на основании следующих статистических данных определим общее количество реализованных школьных тетрадей.

Наименование	Объем продаж, шт.	Объем продаж, у.ш.т.
1. Тетрадь школьная 12 листов	150	
2. Тетрадь школьная 18 листов	200	
3. Тетрадь общая 96 листов	50	
Итого:	-	

Если принять за стандартную школьную тетрадь объем 12 листов, тогда общее количество условных школьных тетрадей будет равно $150 \cdot (12/12) + 200 \cdot (18/12) + 50 \cdot (96/12) = 850$ у.ш.т.

Стоимостные единицы (рубли, доллары, евро и иная валюта) применяются для исчисления путем денежной оценки объектов разного типа (например, продукции или затрат на ее производство всех или нескольких видов), поскольку они не поддаются суммированию в натуральных единицах. Например, не имеет никакого смысла суммировать килограммы рыбы и мяса, поскольку в результате получится «ни рыба, ни мясо».

3.3. Понятие относительных величин

Абсолютная величина не дает полного представления об изучаемом явлении, поскольку не показывает его структуру, динамику, соотношение между отдельными частями, а также взаимосвязи. Эти функции выполняют относительные величины.

Относительная величина – это результат соотношения двух абсолютных величин и, если они являются одноименными, то относительная величина получается безразмерной (не имеет единицы измерения) и называется коэффициентом. Он показывает показывают, во сколько раз уровень изучаемого явления в данных условиях отличается от уровня того же явления в других условиях.

Часто применяется искусственная единица измерения путем умножения коэффициента на 100, 1000 или 10000, в результате чего получаются проценты, промилле или продецимилле. Наиболее часто используются проценты, однако искусственная единица измерения применяется, как правило, только в разговорной речи и в отчетах, а расчеты он затрудняет, поэтому существует «золотое» правило экономистов: «Говорим и учитываем процентом, а считаем коэффициентом».

Если относительная величина является результатом соотношения разноименных абсолютных величин, то она получает конкретное наименование, а также дробную единицу измерения. Например, если разделить выручку от продажи товара и на его объем в натуральных единицах, то получится такая относительная величина как цена в р./ед.

Относительные величины также называются индексами, которые по своему экономическому смыслу подразделяются на несколько видов.

3.4. Индекс динамики

Индекс динамики характеризует изменение во времени изучаемого явления и определяется по формуле

$$i_{\text{д}} = \frac{Y_1}{Y_0}, \tag{3.1}$$

где Y_1 – абсолютная величина (уровень) отчетного периода; Y_0 – уровень предыдущего (базисного) периода.

Критерием для него служит единица, сравнивая с которой можно сделать вывод о динамике уровня изучаемого явления. Если он больше единицы, то имеет место рост; равен единице – стабильность; меньше единицы – наблюдается спад. Путем вычитания из индекса динамики единицы и умножения на 100 можно получить темп динамики в процентах, критерием для которого служит нуль с такими же выводами о динамике: рост, стабильность или спад.

Например, если в марте произведено 138 тыс. тонн продукции, а в феврале – 108 тыс. тонн, то

$$i_{\text{д}} = \frac{138\text{ тыс. т}}{108\text{ тыс. т}} = 1,278 > 1,$$

т.е. имеет место рост производства продукции в 1,278 раза или на $T = (i_{\text{д}} - 1) \cdot 100 = 27,8\%$.

Если базисный и отчетный периоды не являются соседними, то можно определить средний индекс динамики по формуле

$$\overline{i_{\text{д}}} = \sqrt[t]{i_{\text{д}}}, \qquad (3.2)$$

где t – количество изучаемых периодов за исключением базисного.

Путем вычитанием из него единицы и умножения на 100 можно получить средний темп динамики в процентах.

3.5. Индексы планового задания и выполнения плана

Разновидностями индекса динамики являются индексы планового задания и выполнения плана, которые определяются по формулам:

$$i_{\text{пз}} = \frac{Y_1'}{Y_0}, \qquad (3.3)$$

$$i_{\text{вп}} = \frac{Y_1}{Y_1'}, \qquad (3.4)$$

где Y_1' – уровень, запланированная на отчетный период.

Критерием для них служит единица, а путем вычитания единицы и умножения на 100 можно получить соответственно запланированный процент роста (спада) и процент перевыполнения (недовыполнения) плана.

Если перемножить индексы планового задания и выполнения плана, то получится индекс динамики:

$$i_{\text{д}} = i_{\text{пз}} \cdot i_{\text{вп}}. \qquad (3.5)$$

Например, если выручка предприятия в прошлом году составила 20 млн. руб., в следующем году планировалось 28 млн. руб., а фактически получено 26 млн. руб., то $i_{\text{пз}} = \frac{28}{20} = 1,4,$ т.е. запланирован рост выручки на (1,4 – 1)·100 = 40%; $i_{\text{вп}} = \frac{26}{28} = 0,928,$ т.е. план недовыполнен на (0,928 – 1)·100 = –7,2%; $i_{\text{д}} = 1,4 \cdot 0,928 = 1,3,$ т.е. имеет место рост выручки на 30%.

3.6. Индексы структуры и координации

Индекс структуры, который характеризует удельный вес какой-либо части совокупности в ее общем итоге и определяется по формуле

$$i_c = \frac{Y}{\sum Y}, \qquad (3.6)$$

где Y – абсолютная величина какой-либо части совокупности (например, число лиц женского или мужского пола в группе студентов); ΣY – суммарная абсолютная величина (например, общая численность группы студентов).

Путем умножения индекса структуры на 100 можно получить долю какой-либо части совокупности в процентах.

Индекс координации характеризует отношение какой-либо части совокупности к другой ее части и определяется по формуле

$$i_к = \frac{Y}{Y_Б}, \qquad (3.7)$$

где $Y_Б$ – абсолютная величина части совокупности, принятой за базу (основу).

Например, если за основу (базу) принять число лиц мужского пола в группе студентов и поделить на нее число лиц женского пола в этой группе, то получится индекс координации лжп относительно лмп.

Путем умножения индекса структуры на 100 также можно получить процентное соотношение этих частей совокупности.

Например, если в составе российского ВВП произведено товаров на 980, оказано услуг на 740, собрано налогов на 393 млрд. руб., тогда общий итог равен 980 + 740 + 393 = 2113 млрд. долл.; $i_c^{т} = \frac{980}{2113} = 0,464$ или доля товаров от ВВП равна 46,4%; $i_c^{у} = \frac{740}{2113} = 0,350$ или доля услуг от ВВП 35,0%; $i_c^{н} = \frac{393}{2113} = 0,186$ или доля налогов от ВВП равна 18,6%. Если принять за основу (базу) производство товаров, то $i_к^{ут} = \frac{740}{980} = 0,755$ или процентное соотношение услуг и товаров равно 75,5%; $i_к^{нт} = \frac{393}{980} = 0,401$ или процентное соотношение налогов и товаров равно 40,1%.

3.7. Индексы сравнения и интенсивности

Индекс сравнения характеризует отношение одной и той же абсолютной величины, относящейся к одному и тому же периоду времени, но к разным объектам или территориям, и определяется по формуле

$$i_{ср} = \frac{Y_А}{Y_Б}. \qquad (3.8)$$

Критерием для этого индекса служит единица сравнивая с которой можно сделать вывод о том во сколько раз или на сколько процентов (путем вычитания единицы и умножения на 100) одна сравниваемая абсолютная величина больше (меньше) другой.

Например, если запасы воды в озере Байкал равны 23000 куб. км, а в Ладожском озере 911 куб. км, то они больше в $i_{ср}^{БЛ} = \frac{23000}{911} = 25,25$ раз. Наоборот, $i_{ср}^{ЛБ} = \frac{911}{23000} = 0,04 < 1$, т.е. запасы воды в Ладожском озере меньше по сравнению с Байкалом на (0,04-1)·100 = 96%.

Зная индексы динамики и начальные уровни $Y_А$ и $Y_Б$, можно найти условие их равенства в предстоящем периоде *t*:

$$Y_А \bar{i}_А^{\,t} = Y_Б \bar{i}_Б^{\,t}. \tag{3.9}$$

Отсюда

$$t = \frac{\lg(Y_А / Y_Б)}{\lg(\bar{i}_Б / \bar{i}_А)}. \tag{3.10}$$

Например, если объем добычи газа в России и США в базисном году равны 601 и 535 млн. т условного топлива, а среднегодовые темпы динамики соответственно –1,07% и 0,87%, то есть $\bar{i}_A = 0,0087 + 1 = 1,0087$ и $\bar{i}_Б = -0,0107 + 1 = 0,9893$, то объем добычи газа в США сравняется с российским через

$$t = \frac{\lg(Y_А / Y_Б)}{\lg(\bar{i}_Б / \bar{i}_А)} = \frac{\lg(601/535)}{\lg(0,9893/1,0087)} = 6 \text{ лет.}$$

Относительная величина интенсивности характеризует степень распределения или развития изучаемого явления в той или иной среде и определяется по формуле

$$i_{ин} = \frac{X}{Y}. \tag{3.11}$$

где X и Y – разноименные абсолютные величины, относящиеся к одному и тому же явлению и одинаковому периоду времени.

Таким образом, к относительным величинам интенсивности относится цена продукции в р./ед., производительность труда в руб./чел., плотность населения в чел./кв. км, скорость в км/час и т.д.

Например, если численность населения России в отчетном году равна 144,2 млн. чел., а ВВП 2113 млрд. долл., то можно получить такую относительную величину интенсивности, как среднедушевой ВВП $i_{ин} = \frac{2113000 \text{ млн. долл.}}{144,2 \text{ млн. чел.}} = 14653$ долл./чел.

3.8. Методические указания

Определим всевозможные индексы, используя следующие статистические данные (БП – базисный период; ОП – отчетный период; ПП – по плану):

Товар	Единица измерения	Торговое предприятие 1						Торговое предприятие 2					
		Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.			Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.		
		БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП
А	т	120	130	125	15	20	18	125	145	135	17	18	19
Б	м	240	220	245	25	35	28	280	195	215	24	38	31
В	шт.	350	290	325	18	24	21	235	275	340	19	25	22
Итого	-				-	-	-				-	-	-

1. Заполним итоги по выручке от продажи: 120 + 240 + 350 = 710, 130 + 220 +290 = 640, 125 + 245 + 325 = 695 и т.д. Итоги по объему продаж не заполняются, поскольку товары неоднородные.

2. Определим индексы динамики по первому торговому предприятию на примере выручки от продажи:

$i_д^A = \frac{130}{120} = 1,083 > 1$ – рост на (1,083 – 1)·100=8,3%; $i_д^Б = \frac{220}{240} = 0,917 < 1$ – спад на (0,917 – 1)·100 = -8,3%; $i_д^B = \frac{290}{305} = 0,829 < 1$ – спад на (0,829 – 1)·100 = -17,1%; $i_д^{\Sigma} = \frac{640}{710} = 0,901 < 1$ – спад на (0,901 – 1)·100 = -9,9%.

Аналогично можно определить индексы динамики по объему продаж, а также по второму торговому предприятию.

3. Определим индексы планового задания по первому торговому предприятию на примере выручки от продажи:

$i_{\text{пз}}^{A} = \frac{125}{120} = 1,042 > 1$ – рост на 4,2%; $i_{\text{пз}}^{Б} = \frac{245}{240} = 1,021 > 1$ – рост на 2,1%; $i_{\text{пз}}^{B} = \frac{325}{305} = 0,929 < 1$ – спад на 7,1%; $i_{\text{пз}}^{\Sigma} = \frac{695}{710} = 0,979 < 1$ – спад на 2,1%.

Аналогично можно определить индексы планового задания по объему продаж, а также по второму торговому предприятию.

4. Определим индексы выполнения плана по первому торговому предприятию на примере выручки от продажи:

$i_{\text{вп}}^{A} = \frac{130}{125} = 1,04 > 1$ – рост на 4%; $i_{\text{вп}}^{Б} = \frac{220}{245} = 0,898 < 1$ – спад на 10,2%; $i_{\text{вп}}^{B} = \frac{290}{325} = 0,892 < 1$ – спад на 10,8%; $i_{\text{вп}}^{\Sigma} = \frac{640}{695} = 0,921 < 1$ – спад на 7,9%.

Аналогично можно определить индексы выполнения плана по объему продаж, а также по второму торговому предприятию.

5. Определим индексы структуры по первому торговому предприятию на примере выручки от продажи в базисном периоде:

$i_{\text{с.БП}}^{A} = \frac{120}{710} = 0,169$ – доля товара А равна 16,9%; $i_{\text{с.БП}}^{Б} = \frac{240}{710} = 0,338$ – доля товара Б равна 16,9%; $i_{\text{с.БП}}^{B} = \frac{350}{710} = 0,493$ – доля товара В равна 49,3%.

Аналогично можно определить индексы структуры по выручке от продажи в отчетном периоде и по плану, а также по второму торговому предприятию.

6. Определим индексы координации по первому торговому предприятию на примере выручки от продажи в базисном периоде, принимая за базу (основу) товар В:

$i_{\text{к.БП}}^{AB} = \frac{120}{350} = 0,343$ – процентное соотношение товара А и В равно 34,3%; $i_{\text{к.БП}}^{БB} = \frac{240}{350} = 0,686$ – процентное соотношение товара Б и В равно 68,6%.

Аналогично можно определить индексы координации по выручке от продажи в отчетном периоде и по плану, а также по второму торговому предприятию.

7. Определим индексы сравнения первого торгового предприятия со вторым на примере выручки от продажи в базисном периоде:

$i_{\text{ср.БП}}^{A.1/2} = \frac{120}{125} = 0,96 < 1$ – меньше на 4%; $i_{\text{ср.БП}}^{Б.1/2} = \frac{240}{280} = 0,857 < 1$ – меньше на 14,3%; $i_{\text{ср.БП}}^{B.1/2} = \frac{350}{235} = 1,489 > 1$ – больше на 48,9%; $i_{\text{ср.БП}}^{\Sigma.1/2} = \frac{710}{640} = 1,109 > 1$ – больше на 10,9%.

Аналогично можно определить индексы сравнения первого торгового предприятия со вторым по выручке от продажи в отчетном периоде и по плану, а также по объему продаж.

8. Определим индексы интенсивности для первого торгового предприятия на примере базисного периода:

$i_{\text{ин.БП}}^{A} = \frac{120\ 000\ \text{тыс. р.}}{15\ \text{тыс. т}} = 8000$ р./т – цена товара А; $i_{\text{ин.БП}}^{Б} = \frac{240\ 000\ \text{тыс. р.}}{25\ \text{тыс. т}}$ =9600 р./м – цена товара Б; $i_{\text{ин.БП}}^{B} = \frac{350\ 000\ \text{тыс. р.}}{18\ \text{тыс. т}}$ = 19444 р./шт. – цена товара В.

Аналогично можно определить цены в отчетном периоде и по плану, а также по второму торговому предприятию.

Контрольные задания 3

Вариант 1

Определить всевозможные индексы, используя следующие статистические данные:

Товар	Единица измерения	Торговое предприятие 1						Торговое предприятие 2					
		Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.			Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.		
		БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП
А	шт.	210	330	275	18	19	21	225	245	200	18	17	25
Б	л	340	250	345	29	23	22	330	275	295	20	36	31
В	м2	250	270	235	18	22	19	230	300	310	18	23	22

Вариант 2

Определить всевозможные индексы, используя следующие статистические данные:

Товар	Единица измерения	Торговое предприятие 1						Торговое предприятие 2					
		Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.			Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.		
		БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП
Г	м3	340	330	250	17	16	22	250	245	210	25	22	21
Д	шт.	290	220	240	17	15	28	380	135	215	21	22	31
Е	т	150	190	125	26	24	11	370	335	340	15	26	24

Вариант 3

Определить всевозможные индексы, используя следующие статистические данные:

Товар	Единица измерения	Торговое предприятие 1						Торговое предприятие 2					
		Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.			Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.		
		БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП
Ж	ящ.	130	230	160	47	46	72	150	245	170	45	35	31
З	ц	140	240	240	57	55	48	180	145	195	53	42	51
И	м2	250	220	185	46	64	51	170	355	160	55	46	54

Вариант 4

Определить всевозможные индексы, используя следующие статистические данные:

Товар	Единица измерения	Торговое предприятие 1						Торговое предприятие 2					
		Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.			Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.		
		БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП
К	пог. м	430	530	460	37	26	32	470	445	470	35	45	46
Л	шт.	540	540	460	37	25	38	480	345	435	33	32	31
М	м3	650	620	485	26	24	21	370	255	560	35	36	44

Вариант 5

Определить всевозможные индексы, используя следующие статистические данные:

Товар	Единица измерения	Торговое предприятие 1						Торговое предприятие 2					
		Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.			Выручка от продажи, млн. руб.			Объем продаж, тыс. ед.		
		БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП	БП	ОП	ПП
Н	ящ.	910	830	880	63	56	62	870	745	980	45	55	66
О	т	920	940	980	67	55	58	880	845	785	53	42	61
П	шт.	930	920	985	66	54	61	970	855	670	65	66	44

Глава 4. Средние величины и показатели вариации

4.1. Необходимость и понятие средних величин

Необходимость расчета средней величины возникает при сравнении между собой нескольких совокупностей (предприятий, работников, студентов и т.д.) с различными числовыми значениями изучаемого признака (например, заработной платы работников, успеваемости студентов и т.д.).

Средняя величина – это обобщающий показатель совокупности, который характеризует типичный уровень изучаемого явления в конкретных условиях места и времени.

Средняя всегда погашает (сглаживает) индивидуальные различия единиц совокупности, обусловленные случайными обстоятельствами. Однако на основании средней величины нельзя делать однозначные выводы.

Так, если одна семья съедает две курицы в неделю, а другая – ни одной, то в среднем получается по одной курице на семью. Но из этого вовсе не следует, что все семьи питаются курятиной. Аналогично, если у одного работника предприятия заработная плата 10 тыс. р.., а у другого 50 тыс. р., то в среднем получается по 30 тыс. р. на работника, но из этого вовсе не следует, что эта заработная плата является типичной.

4.2. Общие принципы применения средних величин

В статистике должны соблюдаться несколько общих принципов применения средних величин.

1. Необходим обоснованный выбор единицы совокупности, для которой рассчитывается средняя (например, семья, предприятие, работник и т.д.).

2. При определении средней величины нужно учитывать экономический смысл осредняемого показателя (например, заработная плата, производительность труда и т.д.), а также имеющиеся исходные данные.

3. Средняя величина всегда должна рассчитываться по качественно однородной совокупности, которую позволяет получить метод группировок.

4. Общие средние должны подкрепляться групповыми средними (например, средняя заработная плата по стране должна подкрепляться средней заработной платой по регионам, отраслям экономики, профессиям и т.д.).

Средние величины делятся на два больших класса: степенные и структурные.

4.3. Понятие степенных средних величин

Степенные средние по форме представления исходных данных могут быть простыми и взвешенными. Простая средняя рассчитывается по несгруппированным данным и имеет следующую общую формулу:

$$\bar{X} = \sqrt[m]{\frac{\sum X_i^m}{N}}. \tag{4.1}$$

Взвешенная средняя рассчитывается по сгруппированным данным и имеет следующую общую формулу:

$$\bar{X} = \sqrt[m]{\frac{\sum X_i^m f_i}{\sum f_i}}, \tag{4.2}$$

где X_i – конкретное значение признака или середина группировочного интервала (от и до); f_i – частота, показывающая, сколько раз встречается (повторяется) конкретное значение признака; N – общее число таких значений (единиц совокупности); m – показатель степени, от которого зависят следующие виды степенных средних: при $m = -1$ – средняя гармоническая; при

$m=0$ – средняя геометрическая; при $m=1$ – средняя арифметическая; при $m=2$ – средняя квадратическая; при $m=3$ – средняя кубическая и т.д.

4.4. Виды степенных средних величин

Выведем формулу средней арифметической простой и взвешенной при $m=1$:

$$\bar{X}_{\text{ар}} = \frac{\sum X_i}{N}; \tag{4.3}$$

$$\bar{X}_{\text{ар}} = \frac{\sum X_i f_i}{\sum f_i}, \tag{4.4}$$

Выведем формулу средней гармонической простой и взвешенной при $m=-1$:

$$\bar{X}_{\text{гм}} = \frac{N}{\sum \frac{1}{X_i}}; \tag{4.5}$$

$$\bar{X}_{\text{гм}} = \frac{\sum f_i}{\sum \frac{f_i}{X_i}}. \tag{4.6}$$

Выведем формулу средней геометрической простой и взвешенной при $m=0$:

$$\bar{X}_{\text{г}} = \sqrt[N]{\prod X_i}; \tag{4.7}$$

$$\bar{X}_{\text{г}} = \sqrt[N]{\prod X_i^{f_i}}. \tag{4.8}$$

4.5. Правило применения средней арифметической и гармонической

В статистике наиболее часто используются средняя арифметическая и гармоническая для осреднения относительных величин интенсивности. При этом соблюдается следующее правило:

1. Если имеются данные, относящиеся к числителю осредняемого показателя, то применяется формула средней гармонической.

2. Если имеются данные, относящиеся к знаменателю осредняемого показателя, то применяется формула средней арифметической.

Остальные виды степенных средних используются в особых случаях. Так, формула средней геометрической применяется, если имеется последовательность индексов динамики, каждый из которых показывает изменение уровня изучаемого явления каждого последующего периода по сравнению с предыдущим.

Например, на основании следующих статистических данных по сельскохозяйственным предприятиям определим, в каком из них и на сколько больше средняя урожайность зерновых культур:

Таблица 4.1

Зерновая культура	Сельхозпредприятие 1		Сельхозпредприятие 2	
	Валовой сбор, ц	Урожайность, ц/га	Посевная площадь, га	Урожайность, ц/га
Пшеница	32500	25	1540	20
Рожь	1620	18	120	19
Ячмень	13640	22	460	18
Просо	1650	15	80	13
Итого	49410		2200	

Осредняемый показатель урожайности задан в центнерах на 1 га и может быть представлен как отношение валового сбора к посевной площади (взвешивающие показатели).

Поскольку по первому сельхозпредприятию имеются данные о валовом сборе, применяем формулу средней гармонической взвешенной:

$$\bar{X}_1 = \frac{32500+1620+13640+1650}{\frac{32500}{25}+\frac{1620}{18}+\frac{13640}{22}+\frac{13640}{15}} = \frac{ц}{\frac{ц}{ц/га}} = 23,31 \text{ ц/га}$$

Поскольку по второму сельхозпредприятию имеются данные о посевной площади, применяем формулу средней арифметической взвешенной:

$$\bar{X}_2 = \frac{20\cdot 1540+19\cdot 120+18\cdot 460+13\cdot 80}{1540+120+460+80} = \frac{ц/га \cdot га}{га} = 19,27 \text{ ц/га}$$

При этом расчет по формуле средней арифметической и гармонической простой дает верный результат только в том случае, если посевная площадь и валовой сбор каждой группы зерновых культур являются одинаковыми.

Таким образом, средняя урожайность зерновых культур первого сельхозпредприятия больше второго в $\frac{23,31}{19,27} = 1,2$ раза или на 20%.

4.6. Структурные средние величины

Структурные средние применяются для изучения внутреннего строения совокупности, а также для оценки степенной средней, если по имеющимся данным ее расчет не может быть выполнен.

К структурным средним относятся мода и медиана. Мода – это такое значение признака, которое встречается (повторяется) у единиц совокупности наиболее часто. Медиана – это значение признака, которое делит всю совокупность на две равные части, в одной из которых значения признака меньше медианы, а в другой – больше.

Если значения признака представлены упорядоченно в виде группировочных интервалов (от и до), то моду и медиану можно определить по формулам:

$$\mathrm{Mo} = X_{\mathrm{Mo}} + h_{\mathrm{Mo}} \frac{f_{\mathrm{Mo}} - f_{\mathrm{Mo}-1}}{2f_{\mathrm{Mo}} - f_{\mathrm{Mo}-1} - f_{\mathrm{Mo}+1}}, \quad (4.9)$$

$$\mathrm{Me} = X_{\mathrm{Me}} + h_{\mathrm{Me}} \frac{\frac{\sum f}{2} - S_{\mathrm{Me}-1}}{f_{\mathrm{Me}}}, \quad (4.10)$$

где X_{Mo} и X_{Me} – нижняя граница модального и медианного интервала; h_{Mo} и h_{Me} – размах модального и медианного интервала; f_{Mo} и f_{Me} – частота модального и медианного интервала; $f_{\text{Mo}-1}, f_{\text{Mo}+1}$ – частота интервала, предшествующего и следующего за модальным интервалом; $S_{\text{Me}-1}$ – накопленная частота (нарастающим итогом) до медианного интервала.

В случае неравных группировочных интервалов в формуле моды и медианы частоты заменяются плотностями.

Например, определим моду и медиану, используя следующие статистические данные о себестоимости однородной продукции, выпускаемой несколькими промышленными предприятиями:

Таблица 4.2

Себестоимость X, тыс. р./ед.	Число предприятий	
	f	S
1,6-2,0	2	2
2,0-2,4	3	5
2,4-2,8	7	12
2,8-3,2	5	17
3,2-3,6	10	27
3,6-4,0	3	30
Итого:	30	-

Поскольку интервал себестоимости от 3,2 до 3,6 тыс. р./ед. повторяется наиболее часто, т.е. встречается у набольшего числа предприятий, такой интервал является модальным:

$$\text{Mo} = 3,2 + 0,4 \cdot \frac{10-5}{2 \cdot 10 - 5 - 3} = 3,37 \text{ тыс. р./ед.;}$$

Поскольку интервал себестоимости от 2,8 до 3,2 тыс. р./ед. делит все предприятия на две равные части, т.е. частота, определенная нарастающим итогом достигла половины от общего числа предприятий (15), такой интервал является медианным:

$$\text{Me} = 2,8 + 0,4 \cdot \frac{\frac{30}{2} - 12}{5} = 3,04 \text{ тыс. р./ед.}$$

Таким образом, наиболее часто встречаются промышленные предприятия с уровнем себестоимости 3,37 тыс. р./ед., при этом у одной половины уровень себестоимости больше, а у другой – меньше 3,04 тыс. р./ед.

4.7. Понятие и показатели вариации

Вариация – это несовпадение значений одного и того же признака у разных единиц совокупности в силу особенностей их собственного развития, а также конкретных условий, в которых они находятся (например, вариация возраста студентов, роста, веса и т.д.).

Если средняя величина сглаживает индивидуальные различия единиц совокупности, то вариация, наоборот, их подчеркивает и помогает понять сущность изучаемого явления.

Наиболее простым показателем вариации является ее размах, который определяется по формуле

$$H = X_{\max} - X_{\min}, \tag{4.11}$$

где $X_{\max}$ и $X_{\min}$ – наибольшее и наименьшее значение признака.

Поскольку каждое значение признака отличается (отклоняется) от среднего по-разному и в любую сторону (со знаком плюс или минус), необходимо определить показатель среднего отклонения от средней величины. Но, в связи с тем, что сумма положительных и отрицательных отклонений всегда будет равна нулю, необходима нейтрализация знака минус. Для этого существует два способа: линейный и квадратический.

При изучении вариации дополнительно можно установить тип ассиметрии. Для этого применяется коэффициент асимметрии, который определяется по формуле

$$K_{\text{ас}} = \frac{\bar{X} - \text{Mo}}{\sigma}. \qquad (4.12)$$

Значения коэффициента асимметрии колеблются от –3 до 3 и если он равен нулю, то ряд распределения является симметричным, если меньше или больше нуля, то наблюдается соответственно левосторонняя или правосторонняя асимметрия.

4.8. Линейные и квадратические показатели вариации

При использовании линейного способа все отклонения от средней величины принимаются по модулю (без учета знака) и получается среднее линейное отклонение, которое может быть простое и взвешенное:

$$Л = \frac{\sum |X_i - \bar{X}|}{N}; \qquad (4.13)$$

$$Л = \frac{\sum |X_i - \bar{X}| f_i}{\sum f_i}. \qquad (4.14)$$

При использовании квадратического способа все отклонения от средней величины возводятся в квадрат и получается дисперсия, которая также может быть простой и взвешенной:

$$Д = \frac{\sum (X_i - \bar{X})^2}{N}; \qquad (4.15)$$

$$Д = \frac{\sum (X_i - \bar{X})^2 f_i}{\sum f_i}. \qquad (4.16)$$

В статистике, в отличие от математики, дисперсия измеряется в квадратных единицах (в «квадратных» годах, рублях и т.д.), поэтому необходимо также определить среднее квадратическое отклонение как корень квадратный из дисперсии:

$$\sigma = \sqrt{Д}. \qquad (4.17)$$

Однако для сравнения вариации различных признаков или совокупностей нужны относительные величины. Для этой цели применяют коэффициенты вариации, которые также могут быть линейные и квадратические.

Линейный коэффициент вариации определяется по формуле

$$\lambda = \frac{Л}{\bar{X}}; \qquad (4.18)$$

Квадратический коэффициент вариации определяется по формуле

$$\nu = \frac{\sigma}{\bar{X}}. \qquad (4.19)$$

Значения коэффициента вариации обычно колеблются от 0 до 1 и чем ближе к нулю, тем более однородна совокупность относительно изучаемого признака и тем типичнее для нее средняя величина. При этом критерием служит 0,333. Если коэффициент вариации не превышает 0,333, то средняя величина считается типичной, а в противном случае - нетипичной.

4.9. Свойства средней арифметической и дисперсии

Средняя арифметическая и дисперсия как обобщающие показатели совокупности обладают несколькими свойствами:

1. Если каждое значение признака изменить на одно и то же число (прибавить или отнять), то средняя арифметическая изменится на это же число, а дисперсия не изменится.

2. Если каждое значение признака изменить в одинаковое число раз (умножить или разделить), то средняя арифметическая изменится во столько же раз, а дисперсия изменится в квадрат таких раз.

3. Путем простых преобразований можно получить формулу дисперсии методом моментов:

$$Д = \frac{\sum X_i^2 f_i}{\sum f_i} - \left(\frac{\sum X_i f_i}{\sum f_i} \right)^2 = \overline{X^2} - \bar{X}^2. \qquad (4.20)$$

4. Если исходные данные представлены в сгруппированном виде, то дисперсию можно определить по правилу сложения:

$$Д = Д_{\text{м}} + \bar{Д}, \qquad (4.21)$$

где $Д_{\text{м}} = \frac{\sum (\bar{X}_i - \bar{X})^2 f_i}{\sum f_i}$ – межгрупповая дисперсия; $\bar{Д} = \frac{\sum Д_i f_i}{\sum f_i}$ – средняя дисперсия из внутригрупповых; $\bar{X}_i$ – групповая средняя; $\bar{X}$ – общая средняя; f_i – число единиц в каждой группе; $Д_i$ – внутригрупповая дисперсия.

Правило сложения дисперсий используется для количественной оценки взаимосвязи двух признаков, один из которых является факторным, а другой – результативным. Для этого определяется коэффициент детерминации как отношение межгрупповой к общей дисперсии. Его значения варьируют от 0 до 1 и чем ближе к единице, тем более сильная связь между изучаемыми признаками.

4.10. Методические указания

Определим среднюю величину чистой прибыли и установим ее типичность или нетипичность на основании следующих статистических данных:

Таблица 4.3

Чистая прибыль, млн. руб.	18-22	22-26	26-30	30-34	34-38	38-42
Число предприятий	2	4	5	4	1	4

1. Определим середину каждого интервала X_i в первом столбце таблицы 4.4 как полусумму его верхней и нижней границы: $\frac{18+22}{2} = 20;\ \frac{22+26}{2} = 24;\ \frac{26+30}{2} = 28$ и т.д.

2. Заполним частоты каждого интервала f_i во втором столбце таблицы 4.4, используя исходные данные о числе предприятий, а также итог этого столбца: 2 + 4 + 5 + 4 + 1 + 4 = 20.

3. Определим размах вариации по формуле (4.11): H = 42 – 18 = 24 млн. р.

4. Определим расчетные значения $X_i f_i$ в третьем столбце таблицы 4.4: 20·2=40, 24·2=48, 28·5=140 и т.д., а также итог этого столбца: 40 + 96 + 140 + 128 + 36 + 160 = 600.

5. Определим среднюю величину по формуле (4.4): $\overline{X} = \frac{600}{20} = 30$ млн. руб.

6. Определим расчетные значения $X_i - \overline{X}$ в четвертом столбце таблицы 4.4: 20 – 30 = -10; 24 – 30 = -6; 28 – 30 = -2 и т.д.

7. Определим расчетные значения $| X_i - \overline{X} |$ в пятом столбце таблицы 4.4: | -10| = 10; |-6| = 6; |-2| = 2 и т.д.

8. Определим расчетные значения $| X_i - \overline{X} | f_i$ в шестом столбце таблицы 4.4: |10·2 = 20; 6·4 = 24; 2·5 = 10 и т.д., а также итог этого столбца: 108.

9. Определим среднее линейное отклонение по формуле (4.14): $Л = \frac{108}{20} = 5,4$ млн.р.

10. Определим линейный коэффициент вариации по формуле (4.18): $\lambda = \frac{5,4}{30} = 0,180 <$ 0,333 – средняя величина чистой прибыли является типичной.

11. Определим расчетные значения $(X_i - \overline{X})^2$ в пятом столбце таблицы 4.4: $(-10)^2 = 100$; $(-6)^2 = 36$; $(-2)^2 = 4$ и т.д.

12. Определим расчетные значения $(X_i - \overline{X})^2 f_i$ в пятом столбце таблицы 4.4: 100·2 = 200; 36·4 = 144; 4·5 = 20 и т.д., а также итог этого столбца: 816.

13. Определим дисперсию по формуле (4.16): $Д = \frac{816}{20} = 40,8$ млн. р.^{2}

14. Определим среднеквадратическое отклонение по формуле (4.17): $\sigma = \sqrt{40,8} = 6,4$ млн. р.

15. Определим квадратический коэффициент вариации по формуле (4.19): $v = \frac{6,4}{30} = 0,213 <$ 0,333 – средняя величина чистой прибыли является типичной.

Таблица 4.4

X_i	f_i	$X_i f_i$	$X_i - \overline{X}$	$\| X_i - \overline{X} \|$	$\| X_i - \overline{X} \| f_i$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^2 f_i$
20	2	40	-10	10	20	100	200
24	4	96	-6	6	24	36	144
28	5	140	-2	2	10	4	20
32	4	128	2	2	8	4	16
36	1	36	6	6	6	36	36
40	4	160	10	10	40	100	400
Итого:	20	600	–	–	108	–	816

Контрольные задания 4

Вариант 1

Определить средний вес работников и установить его типичность или нетипичность по следующим статистическим данным:

Вес, кг	40-50	50-60	60-70	70-80	80-90	90-100
Число работников, чел.	1	8	3	4	3	1

Вариант 2

Определить среднее время изготовления детали работниками установить его типичность или нетипичность по следующим статистическим данным:

Время, мин.	0-20	20-40	40-60	60-80	80-100	100-120
Число работников, чел.	5	7	1	5	1	1

Вариант 3

Определить средний рост работников и установить его типичность или нетипичность по следующим статистическим данным:

Рост, см	140-150	150-160	160-170	170-180	180-190	190-200
Число работников, чел.	2	4	5	7	1	1

Вариант 4

Определить средний возраст работников и установить его типичность или нетипичность по следующим статистическим данным:

Возраст, лет	20-26	26-32	32-38	38-44	44-50	50-56
Число работников, чел.	3	2	6	5	2	2

Вариант 5

Определить средний стаж работников и установить его типичность или нетипичность по следующим статистическим данным:

Стаж, мес.	0-30	30-60	60-90	90-120	120-150	150-180
Число работников, чел.	5	4	2	7	1	1

Глава 5. Выборочное наблюдение

5.1. Применение и понятия выборочного наблюдения

Выборочное наблюдение применяется, когда сплошное наблюдение физически невозможно из-за огромного объема изучаемой совокупности (например, при обследовании рыночных цен, семейных бюджетов и т.д.) или экономически нецелесообразно (например, при проверке качества продукции, дегустации продуктов питания и т.д.). Выборочное наблюдение используют также для проверки результатов сплошного наблюдения.

Часть единиц, отобранных для наблюдения, называется выборочной совокупностью (выборкой), а совокупность, из которой производится отбор, называется генеральной. Численность выборки обозначается n, а численность генеральной совокупности – N. Отношение n/N называется относительным размером выборки и выражается в процентах.

Качество результатов выборочного наблюдения зависит от репрезентативности, то есть от того, насколько выборка представляет генеральную совокупность. Для обеспечения репрезентативности надо соблюдать принцип случайности отбора единиц из генеральной совокупности. Для этого существует несколько способов.

5.2. Способы формирования выборочной совокупности

1. Случайный отбор – осуществляется с помощью жеребьевки (лото), когда каждой единице генеральной совокупности присваивают порядковый номер, который заносят на определенные предметы (например, шары или карточки), затем они перемешиваются и отбираются наугад или производится выбор случайных чисел (например, с помощью ЭВМ), которые образуют порядковые номера для отбора.

2. Механический отбор, при котором в выборку попадает каждая $N/n-$ая единица генеральной совокупности. Например, если она состоит из 100000 ед., из которых требуется отобрать 1000 ед., то в выборку попадет каждая 100-ая единица. При этом, если они не упорядочены относительно изучаемого признака, то первая единица выбирается наугад из первой сотни, а порядковый номер каждой последующей единицы будет на 100 больше (например, №13, №113, №213, №313 и т.д.).

3. Стратифицированный (расслоенный) отбор – осуществляется из неоднородной совокупности, которую сначала разбивают на однородные группы, а затем производят отбор из каждой группы случайным или механическим способом.

4. Серийный отбор, при котором случайно или механически каждый раз выбирают определенную последовательность единиц (серию), внутри которой проводят сплошное наблюдение.

5.3. Понятие ошибки повторной и бесповторной выборки

Качество результатов выборочного наблюдения зависят от типа выборки – повторная или бесповторная. При повторной выборке каждая отобранная единица (серия) возвращается в генеральную совокупность и имеет шанс еще раз попасть в выборку, а при бесповторной – не возвращается, поэтому для остальных единиц вероятность попасть в выборку увеличивается. Бесповторная выборка дает более точные результаты, и только когда ее провести невозможно, используется повторная выборка (например, при обследовании потребительского спроса, пассажиропотоков, опросы населения т.п.).

По мере отбора единиц из генеральной совокупности или после его завершения производится регистрация предусмотренных программой признаков, а итогом является расчет обобщающих показателей.

Если изучается количественный признак, то рассчитывается выборочная средняя X и генеральная средняя $\bar{X}$, а если изучается альтернативный признак – выборочная доля w и генеральная доля d.

Разность между обобщающими показателями выборочной и генеральной совокупности $\tilde{X} - \bar{X}$ или $w - d$ называется ошибкой выборки, которая подразделяется на ошибку регистрации и ошибку репрезентативности. Первая возникает из-за неправильных или неточных сведений по причине непонимания существа вопросов, невнимательности при заполнении бланков, округления и т.п. Эту ошибку легко обнаружить и устранить путем логического и арифметического контроля собранных статистических данных (например, если при анкетировании при ответе на вопросы лицо в графе «профессия» указало «пенсионер», а в графе «возраст» - «14 лет»).

Вторая означает систематическое или случайное отклонение выборочного обобщающего показателя от генерального из-за несоблюдения принципа случайности отбора. Эту ошибку обнаружить и устранить гораздо труднее, поэтому ей уделяется основное внимание.

5.4. Средняя ошибка повторной и бесповторной выборки

Если из генеральной совокупности сделать несколько выборок, выборочный обобщающий показатель каждый раз будет принимать различные значения в зависимости от состава конкретной выборки, поэтому можно определить среднюю из возможных ошибок в виде среднеквадратического отклонения, которое прямо пропорционально дисперсии и обратно пропорционально численности выборки.

Однако дисперсия при выборочном наблюдении неизвестна, поэтому при достаточно большой численности выборки, когда $n > 30$, среднюю ошибку повторной выборки можно определить по формуле

$$\mu = \sqrt{\frac{Д_в}{n}}, \tag{5.1}$$

где $Д_в$ – дисперсия выборки, определяемая при изучении альтернативного признака по формуле

$$Д_{в} = w(1-w), \tag{5.2}$$

В случае бесповторной выборки вероятность отбора для каждой единицы генеральной совокупности неодинакова, поэтому средняя ошибка бесповторной выборки определяется по формуле

$$\mu = \sqrt{\frac{Д_{в}}{n}\left(1-\frac{n}{N}\right)}. \tag{5.3}$$

При стратифицированном отборе в выборку могут попасть представители всех групп в тех же пропорциях, что и в генеральной совокупности, поэтому ее ошибка будет зависеть только от внутригрупповых дисперсий, а при серийном отборе внутри серий проводят сплошное наблюдение, поэтому ее ошибка будет зависеть только от межгрупповой дисперсии. Таким образом, средняя ошибка для этих способов всегда будет меньше, чем для случайного и механического.

5.5. Предельная ошибка выборки

Для решения практических задач кроме средней нужна предельная ошибка выборки, связанная с гарантирующей ее вероятностью β, а она в свою очередь определяет коэффициент доверия *t*, который характеризует нормированное отклонение выборочного обобщающего показателя от генерального

Эти параметры связаны между собой в соответствии с законом нормального распределения вероятностей и имеются в математических таблицах:

$$\beta(t) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{t} e^{-\frac{t^2}{2}}dt. \tag{5.4}$$

В статистике чаще всего используют следующие сочетания:

Таблица 5.1

β	0,683	0,866	0,954	0,988	0,997	0,999
t	1	1,5	2,0	2,5	3,0	3,5

Например, с вероятностью 0,954 можно утверждать, что предельная ошибка выборки будет в два раза большей средней. Поэтому предельная ошибка выборки определяется по формуле

$$\Delta = t\mu. \tag{5.5}$$

После исчисления предельной ошибки находят границы доверительных интервалов.

Если изучается количественный признак, то генеральная средняя $\bar{X}$ будет находиться в следующем доверительном интервале:

$$\tilde{X} - \Delta \le \bar{X} \le \tilde{X} + \Delta. \tag{5.6}$$

Если изучается альтернативный признак, то генеральная доля d будет находиться в следующем доверительном интервале:

$$w - \Delta \le d \le w + \Delta. \tag{5.7}$$

Таким образом, определяется не одно точное значение генерального обобщающего показателя, а лишь границы доверительного интервала с определенным уровнем вероятности, что является серьезным недостатком выборочного наблюдения.

Доказано, что при малом объеме выборки, когда $n < 30$, выборочную дисперсию необходимо умножить на отношение n к $n - 1$. В этом случае коэффициент доверия t и доверительная вероятность $\beta = 1 - \alpha$, будут подчинены закону распределения Стъюдента, где α – уровень значимости, который принимается равным 0,05 или 5%. Параметром функции плотности распределения вероятностей $S(t)$ является число степеней свободы, которое при определении выборочной дисперсии ограничивает только средняя величина, поэтому $v = n - 1$.

5.6. Определение необходимой численности выборки

При разработке программы выборочного наблюдения обычно устанавливают предельную ошибку и доверительную вероятность. Неизвестной остается численность выборки, необходимая для обеспечения установленной точности.

Формулы для ее определения можно вывести из формулы предельной ошибки повторной или бесповторной выборки.

Численность повторной выборки определяется по формуле:

$$n = \frac{Д_{в} t^2}{\Delta^2}. \qquad (5.8)$$

Численность бесповторной выборки определяется по формуле:

$$n = \frac{Д_{в}}{\Delta^2 / t^2 + Д_{в} / N}. \qquad (5.9)$$

Однако дисперсия до начала выборочного наблюдения неизвестна, поэтому ее находят приближенно одним из следующих способов.

1) берут из предыдущих наблюдений;

2) определяют по правилу «шести сигм», согласно которому в размахе вариации укладывается приблизительно шесть средних квадратических отклонений:

$$H = 6\sigma,\ \sigma = \frac{H}{6},\ Д = \frac{H^2}{36}; \qquad (5.10)$$

3) определяют по правилу «трех сигм», согласно которому в средней величине укладывается приблизительно три средних квадратических отклонения:

$$\bar{X} = 3\sigma,\ \sigma = \frac{\bar{X}}{3},\ Д = \frac{\bar{X}^2}{9}; \qquad (5.11)$$

4) при изучении альтернативного признака, если нет даже приблизительных сведений о выборочной доле, она принимается равной 0,5, поэтому дисперсия будет равна $Д = 0,5 \cdot (1 - 0,5) = 0,25$.

5.7. Методические указания

С вероятностью 0,954 определим средний выпуск продукции и долю промышленных предприятий с выпуском продукции до 10 млн. р. на основании следующих статистических данных, полученных в результате 10%-ой бесповторной выборки:

Выпуск продукции, млн. р.	до 3	3-5	5-10	10-30	30 и более
Число предприятий	5	15	24	40	16

1. Для нахождения неизвестных границ первого и последнего интервала в первом столбце таблицы 5.2 используем принцип «соседа», согласно которого для нахождения неизвестной нижней границы первого интервала (до 3 млн. р.) надо из известной верхней границы отнять размах соседнего интервала, то есть 3 – 2 = 1. Аналогично, для нахождения неизвестной верхней границы последнего интервала (30 и более), надо к известной нижней границе прибавить размах соседнего интервала, то есть 30 + 20 = 50.

2. Заполним частоты каждого интервала f_i во втором столбце таблицы 5.2, используя исходные данные о числе предприятий, а также итог этого столбца: 100.

3. Определим середину каждого интервала X_i в третьем столбце таблицы 5.2 как полусумму его верхней и нижней границы: $\frac{1+3}{2}=2;\ \frac{3+5}{2}=4;\ \frac{5+10}{2}=7,5$ и т.д.

4. Определим расчетные значения $X_i f_i$ в четвертом столбце таблицы 5.2, а также итог этого столбца: 1690.

5. Определим выборочную среднюю количественного признака по формуле (4.4): $\tilde{X}=\frac{1690}{100}=16,9$ млн. р.

6. Определим расчетные значения $X_i-\tilde{X}$, $(X_i-\tilde{X})^2$ и $(X_i-\tilde{X})^2 f_i$ в пятом, шестом и седьмом шестом столбцах таблицы 5.2, а также итог последнего столбца: 14649.

7. Определим выборочную дисперсию по формуле (4.16): $Д_{\text{в}}=\frac{14649}{100}=146,5$ млн. руб.^{2}

Таблица 5.2

X	f_i	X_i	$X_i f_i$	$X_i-\tilde{X}$	$(X_i-\tilde{X})^2$	$(X_i-\tilde{X})^2 f_i$
1-3	5	2	10	–14,9	222,01	1110,05
3-5	15	4	60	–12,9	166,41	2496,15
5-10	24	7,5	180	–9,4	88,36	2120,64
10-30	40	20	800	3,1	9,61	384,40
30-50	16	40	640	23,1	533,61	8537,76
Итого:	100	–	1690	–	–	14649

8. Определим численность генеральной совокупности. Поскольку численность выборки n = 100 или 10%, то N = 100/0,1=1000.

9. Определим среднюю ошибку бесповторной выборки по формуле (5.3):

$$\mu=\sqrt{\frac{146,5}{100}\left(1-\frac{100}{1000}\right)}=1,15 \text{ млн. р.}$$

10. Определим коэффициент доверия по таблице 5.1. Поскольку доверительная вероятность $\beta=0,954$, то $t=2.$

11. Определим предельную ошибку выборки по формуле (5.5): $\Delta=$ 2 · 1,15 = 2,3 млн. р.

12. Определим границы доверительного интервала для генеральной средней по формуле (5.6):

$$16,9-2,3\le\overline{X}\le16,9+2,3 \text{ или } 14,6\le\overline{X}\le19,2.$$

13. Определим выборочную долю промышленных предприятий с выпуском продукции до 10 млн. р. по формуле (1.3):

$$w=\frac{5+15+24}{100}=0,44.$$

14. Определим выборочную дисперсию альтернативного признака по формуле (5.2):

$$Д_{\text{в}}=w(1-w)=0,44\cdot(1-0,44)=0,246.$$

15. Определим среднюю ошибки бесповторной выборки по формуле (5.3):

$$\mu = \sqrt{\frac{0,246}{100}\left(1 - \frac{100}{1000}\right)} = 0,047.$$

16. Определим предельную ошибку выборки по формуле (5.5):

$$\Delta = t\mu = 2 \cdot 0,047 = 0,094.$$

17. Определим границы доверительного интервала для генеральной доли по формуле (5.7):

$$0,44 - 0,094 \le d \le 0,44 + 0,094 \text{ или } 0,346 \le d \le 0,536.$$

Контрольные задания 5

Вариант 1

С вероятностью 0,988 определить средний возраст и долю рабочих в возрасте до 30 лет по результатам 5%-ой бесповторной выборки:

Возраст, лет	до 24	24-26	26-30	30-40	40-50	50 и более
Число рабочих, чел.	10	10	45	15	6	4

Вариант 2

С вероятностью 0,866 определить среднюю заработную плату и долю рабочих с заработной платой до 10 тыс. р. по результатам 4%-ой бесповторной выборки:

Заработная плата, тыс. р.	до 6	6-8	8-10	10-15	15-20	20 и более
Число рабочих, чел.	10	20	12	4	2	2

Вариант 3

С вероятностью 0,954 определить средний стаж работы и долю рабочих со стажем более 5 лет по результатам 5%-ой бесповторной выборки:

Стаж, лет	до 3	3-5	5-10	10-15	15-25	15 и более
Число рабочих, чел.	7	24	35	30	2	2

Вариант 4

С вероятностью 0,954 определить среднюю чистую прибыль и долю предприятий с чистой прибылью до 30 млн. р. по результатам 10%-ой бесповторной выборки:

Чистая прибыль, млн. р.	до 20	20-30	30-60	60-120	120-200	200 и более
Число предприятий	15	35	38	62	7	3

Вариант 5

С вероятностью 0,866 определить средний уставный капитал и долю банков с уставным капиталом более 150 млн. р. по результатам 2%-ой бесповторной выборки:

Уставный капитал, млн. р.	до 10	10-20	20-60	60-150	150-300	300 и более
Число банков	10	12	16	22	25	35

Глава 6. Ряды динамики

6.1. Понятие и классификация рядов динамики

Ряд динамики – это последовательность упорядоченных во времени (расположенных в хронологическом порядке) числовых значений показателя, характеризующих развитие изучаемого явления или процесса.

Конкретные числовые значения показателя называются уровнями и обозначаются Y, а их количество в ряду динамики – n.

Ряды динамики классифицируются по нескольким признакам.

1. По времени – моментные и периодные. Уровень моментного ряда показывает фактическое наличие изучаемого явления на определенный момент времени или дату (например, численность населения страны на начало каждого года), а уровень периодного ряда относится к результату, накопленному за определенный отрезок времени (например, объем произведенной продукции по месяцам года).

При суммировании уровней периодного ряда получается реальный показатель (например, общий выпуск продукции за год), а сумма уровней моментного ряда, никакого смысла, как правило, не имеет (например, если сложить запасы готовой продукции на начало каждого месяца, то общий объем запасов готовой продукции за год не получится).

2. По форме представления – ряды абсолютных, относительных и средних величин.

3. По размаху интервалов – полные (равноотстоящие), когда уровни следуют друг за другом с равными интервалами, и неполные (неравноотстоящие), когда принцип равных интервалов не соблюдается.

4. По числу показателей – изолированные (одномерные), когда во времени анализируется только один показатель (например, цена на конкретную продукцию), и комплексные (многомерные), когда во времени анализируется система показателей (например, цены на основные продукты питания).

6.2. Правила построения рядов динамики

При построении ряда динамики необходимо соблюдать несколько требований.

1. Периодизация развития, то есть разбивка его во времени на качественно однородные этапы, в пределах которых показатель подчиняется одному закону развития.

2. Сопоставимость статистических данных по территории, охвату совокупности, единицам измерения, времени регистрации, ценам, методологии расчета показателей.

3. Соответствие размаха интервалов интенсивности изучаемых процессов. Так, переписи населения достаточно проводить раз в десять лет, учет урожая – раз в год, валютных курсов – ежедневно, температуры воздуха – ежечасно и т.п.

4. Упорядоченность уровней рядов динамики во времени, т.е. без пропусков отдельных уровней, а если они неизбежны, то их заменяют расчетными значениями.

6.3. Периодизация рядов динамики

Существует несколько методов периодизации.

а) Исторический метод, когда периодизация осуществляется на основе «узаконенной» структуры динамики, при этом обращают внимание на значимые и знаменательные даты и события (рождение Христа, войны и революции, принятие управленческих решений, смена руководства и т.п.). Недостаток этого метода в том, что точные временные границы периодов путем качественного анализа удается получить крайне редко.

б) Метод параллельной периодизации, в соответствии с которым предполагается, что существует показатель X, полностью определяющий поведение изучаемого показателя Y, тогда в качестве периодов развития Y можно взять периоды развития X. Недостаток метода заключается в сложности нахождения показателя X.

в) Метод многомерного статистического анализа, который применяется, когда для получения адекватного отображения изучаемого явления только одного показателя недостаточно. Например, для описания такого сложного явления, как здоровье населения нельзя ограничиваться только показателями смертности, продолжительности жизни, заболеваемости и т.д., поскольку необходима система показателей, которая учитывает многообразие изучаемого явления, позволяет устранить негативное воздействие недостоверных и неточных статистических данных, а также повысить обоснованность и надежность статистических выводов и прогнозов.

6.4. Сопоставимость рядов динамики

Сопоставимость по территории означает, что данные по странам и регионам, границы которых изменились, должны быть пересчитаны в старых пределах. Сопоставимость по охвату совокупности (объемная сопоставимость) предполагает сравнение совокупностей с одинаковым числом элементов.

Территориальная и объемная сопоставимость обеспечивается путем смыкания несопоставимых рядов динамики. При этом либо делается пересчет в условные абсолютные уровни путем умножения несопоставимых уровней рядов динамики на соотношение нового и старого уровня, либо абсолютные уровни заменяются относительными, для чего уровни «переломного» момента или периода времени принимаются за 100%.

Например, приведем ряды динамики к сопоставимому виду по следующим статистическим данным:

Наименование	2008	2009	2010	2011	2012	2013	2014
ВВП (в старых границах), млрд. долл.	1690	1780	1616				
ВВП (в новых границах), млрд. долл.			1980	2050	2069	2102	2133

При использовании абсолютного способа определим коэффициент пересчета в новые границы Ксн = 1980 / 1616 = 1,225 или коэффициент пересчета в старые границы $K_{\text{нс}}$ = 1616/1980 = 0,816. Уровни в старых границах определим путем умножения уровней в новых границах на $K_{\text{нс}}$, т.е. $Y=Y'\cdot K_{\text{нс}}$, а уровни в новых границах определим путем умножения уровней в старых границах на $K_{\text{сн}}$, т.е. $Y'=Y\cdot K_{\text{сн}}$.

При использовании относительного способа примем «переломный» период 2011 г. за 100% и определим процентные соотношения каждого уровня к уровню этого периода.

Наименование	2008	2009	2010	2011	2012	2013	2014
ВВП (в старых границах), млрд. долл.	1690	1780	1616	1673	1689	1716	1741
ВВП (в новых границах), млрд. долл.	2071	2181	1980	2050	2069	2102	2133
ВВП, в процентах к уровню 2011 г.	101,0	106,4	96,6	100,0	100,9	102,5	104,0

При обеспечении сопоставимости по единицам измерения особых проблем не возникает, а сопоставимость по ценам (стоимостная сопоставимость) обеспечивается при помощи системы постоянных цен какого-либо базисного периода.

Проблемы могут возникнуть при обеспечении сопоставимости по времени регистрации. Особенно это относится к сезонным явлениям, поскольку в этом случае регистрации на одинаковые даты (например, на начало каждого квартала) бывает недостаточно, поэтому учет таких явлений обычно производится по «нейтральным» датам.

6.5. Абсолютные и относительные показатели ряда динамики

При анализе ряда динамики применяются абсолютные и относительные изменения его уровней во времени, а количество таких изменений можно определить по формуле

$$k = n - 1. \tag{6.1}$$

Любое изменение может определяться базисным и цепным способами. При расчете базисным способом каждый конкретный уровень ряда динамики сравнивается с первым уровнем, а при расчете цепным способом каждый конкретный уровень ряда динамики сравнивается с предыдущим уровнем.

Таким образом, абсолютное изменение уровней можно определить базисным и цепным способом по формулам:

$$\Delta Y^{\text{б}} = Y_i - Y_1; \tag{6.2}$$

$$\Delta Y^{\text{ц}} = Y_i - Y_{i-1}. \tag{6.3}$$

Критерием для абсолютного изменения служит нуль, сравнивая с которым можно сделать вывод о динамике: если оно больше нуля, то имеет место рост; равно нулю – стабильность; если меньше нуля, то наблюдается спад.

Для контроля может использоваться правило, согласно которому сумма последовательных цепных абсолютных изменений должна быть равна последнему базисному абсолютному изменению:

$$\sum_{r=1}^{k} \Delta Y_r^{\text{ц}} = \Delta Y_k^{\text{б}}. \tag{6.4}$$

Относительные изменения (индексы динамики) также можно определить базисным и цепным способом по формулам:

$$i^{б} = \frac{Y_i}{Y_1}; \qquad (6.5)$$

$$i^{ц} = \frac{Y_i}{Y_{i-1}}. \qquad (6.6)$$

Критерием для относительного изменения служит единица: если оно больше единицы, имеет место рост; равно единице – стабильность; если меньше единицы, то наблюдается спад. Путем вычитания из него единицы и умножения на 100 можно получить темп динамики в процентах.

Для контроля может использоваться правило, согласно которому произведение последовательных цепных индексов динамики должно быть равно последнему базисному индексу динамики:

$$\prod_{r=1}^{k} i_r^{ц} = i_k^{б}. \qquad (6.7)$$

6.6. Средние показатели ряда динамики

Обобщающую характеристику ряда динамики дает система средних показателей, которая включает: средний уровень, среднее абсолютное и относительное изменение.

Средний уровень периодного ряда динамики определяется по формуле средней арифметической:

$$\bar{Y} = \frac{\sum Y_i}{n}. \qquad (6.8)$$

Средний уровень моментного равноотстоящего ряда динамики определяется по формуле средней хронологической:

$$\bar{Y} = \frac{\frac{Y_1 + Y_n}{2} + \sum_{i=2}^{n-1} Y_i}{n-1}, \qquad (6.9)$$

где Y_1, Y_n – первый и последний уровни ряда динамики; Y_i – промежуточные уровни.

Средний уровень моментного неравноотстоящего ряда динамики определяется по формуле средней арифметической взвешенной:

$$\bar{Y} = \frac{\sum_{i=1}^{n-1} Y_i t_i}{\sum_{i=1}^{n-1} t_i}, \qquad (6.10)$$

где t_i – размах интервала между соседними уровнями.

Среднее абсолютное изменение можно определить цепным и базисным способом по формуле средней арифметической:

$$\Delta\overline{Y}^{\text{ц}} = \frac{\sum \Delta Y_r^{\text{ц}}}{k}; \; \Delta\overline{Y}^{\text{б}} = \frac{\Delta Y_k^{\text{б}}}{n-1}. \qquad (6.11)$$

Критерием для него служит нуль, сравнивая с которым можно сделать вывод о динамике в среднем: рост, стабильность или спад.

Среднее относительное изменение (средний индекс динамики) можно определить цепным и базисным способом по формуле средней геометрической:

$$\overline{i}^{\text{ц}} = \sqrt[k]{\prod i_r^{\text{ц}}}; \; \overline{i}^{\text{б}} = \sqrt[k]{i_k^{\text{б}}}. \qquad (6.12)$$

Критерием для него служит единица, а путем вычитания из него единицы и умножения на 100, можно получить средний темп динамики в процентах по знаку которого также можно судить о динамике в среднем.

6.7. Проверка ряда динамики на наличие тренда

Всякий ряд динамики теоретически может быть представлен в виде таких составляющих, как

1) тренд – основная тенденция развития ряда динамики к увеличению либо к снижению его уровней;

2) циклические (периодические) колебания, в том числе сезонные;

3) случайные колебания.

Изучение тренда включает два основных этапа: на первом этапе ряд динамики проверяется на наличие тренда, а на втором этапе производится непосредственное выделение тренда и экстраполяция.

При проверке ряда динамики на наличие тренда могут использоваться несколько методов.

1. Метод средних, согласно которому изучаемый ряд динамики разбивают на два равных интервала, для каждого из которых определяется средний уровень, и если они различаются существенно (более 5%), то в ряду динамики признается наличие тренда.

2. Метод Валлиса и Мура, в соответствии с которым наличие тренда в ряду динамики утверждается в том случае, если он не содержит или содержит в незначительном количестве фазы, то есть перемену знака цепного абсолютного изменения (с плюса на минус или с минуса на плюс).

3. Метод Кокса и Стюарта, согласно которому изучаемый ряд динамики разбивают на три равные по числу уровней группы, и если средние уровни первой и последней групп различаются существенно (более 5%), то в ряду динамики признается наличие тренда.

4. Метод серий, в соответствии с которым каждый конкретный уровень ряда динамики считается принадлежащим к одному из двух типов: если он меньше средней величины, считается, что он имеет тип А, а в противном случае – тип Б. Тогда ряд динамики превращается в последовательность типов (например, ААБББАААББАБ). В образовавшейся последовательности определяется число серий R, т.е. последовательностей уровней одинакового типа. В нашем примере $R = 6$.

Если в ряду динамики тренд отсутствует, то число серий является случайной величиной, распределенной приближенно по нормальному закону, и будет находиться в следующем доверительном интервале

$$\overline{R} - t\sigma_R \le R \le \overline{R} + t\sigma_R, \qquad (6.13)$$

где $\overline{R} = (n+1)/2$ – среднее число серий; $\sigma_R = \sqrt{(n-1)}/2$ – среднее квадратическое отклонение числа серий; t – нормированное отклонение, которое устанавливается в зависимости от принятого уровня доверительной вероятности.

Если число серий выходит за пределы доверительного интервала, то в ряду динамики признается наличие тренда.

6.8. Метод укрупнения интервалов

Непосредственное выделение тренда может осуществляться несколькими методами: укрупнения интервалов, скользящей средней и аналитического выравнивания.

В соответствии с методом укрупнения интервалов ряд динамики разбивают на некоторое достаточно большое число равных интервалов, и, если их уровни не позволяют непосредственно выделить основную тенденцию развития, то переходят к расчету уровней за большие промежутки времени путем увеличения размаха интервалов и уменьшения их количества.

Например, произведем непосредственное выделение тренда методом укрупнения интервалов используя следующие статистические данные о номинальном ВВП, млрд. р.

Годы	2006				2007				2008			
Квартал	I	II	III	IV	I	II	III	IV	I	II	III	IV
млрд. р.	5 792,9	6 368,1	7 275,8	7 480,3	6 780,2	7 767,5	8 902,7	9 797,0	8 877,7	10 238,3	11 542,0	10 618,9

Годы	2009				2010				2011			
Квартал	I	II	III	IV	I	II	III	IV	I	II	III	IV
млрд. р.	8 334,6	9 244,8	10 411,3	10 816,4	9 995,8	10 977,0	12 086,5	13 249,3	11 954,2	13 376,4	14 732,9	15 903,7

Годы	2012				2013				2014			
Квартал	I	II	III	IV	I	II	III	IV	I	II	III	IV
млрд. р.	13 677,4	14 971,1	16 280,7	17 247,3	14 577,3	15 827,8	17 450,6	18 334,4	15 454,4	17 299,9	18 722,1	19 930,0

Перейдем от квартальных к годовым уровням ВВП.

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
млрд. р.	26 917,2	33 247,5	41 276,8	38 807,2	46 308,5	55 967,2	62 176,5	66 190,1	71 406,4

Фактический ряд динамики и укрупненные уровни представим в виде статистического графика.

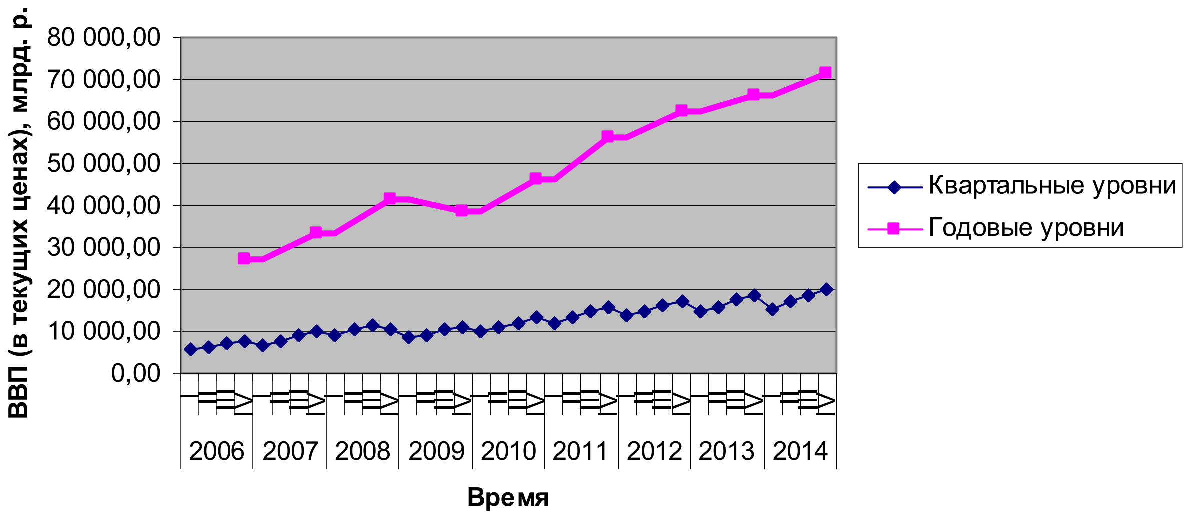


Рис. 6.1. Ряд динамики и линия тренда номинального ВВП, млрд. р.

Таким образом, имеет место основная тенденция к увеличению ВВП в текущих ценах.

6.9. Метод скользящей средней

Метод скользящей средней заключается в том, что исходные уровни ряда динамики заменяются средними величинами, которые получают из каждого уровня и нескольких симметрично его окружающих. При этом число уровней, для которого рассчитывается средняя величина, называют интервалом сглаживания, который может быть как нечетным, так и четным. При нечетном сглаживании полученную среднюю величину закрепляют за серединой интервала, а при четном этого делать нельзя. Поэтому при обработке ряда динамики четными интервалами их искусственно делают нечетными, для чего образуют ближайший нечетный интервал, но из его крайних уровней берут только 50%.

Формулы скользящей средней при нечетном сглаживании выглядят следующим образом:

$$\bar{Y}_i = \frac{Y_{i-1} + Y_i + Y_{i+1}}{3}; \qquad (6.14)$$

$$\bar{Y}_i = \frac{Y_{i-2} + Y_{i-1} + Y_i + Y_{i+1} + Y_{i+2}}{5}. \qquad (6.15)$$

Например, произведем непосредственное выделение тренда методом скользящей средней, используя следующие статистические данные о ежемесячном объеме потребления овощей человеком Y за несколько лет, кг:

Таблица 6.1

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
кг	10,0	10,7	12,0	10,3	12,9	16,3	15,6	17,8	18,0

Определим трехуровневую скользящую среднюю по формуле (6.14):

$$\bar{Y}_1 = \frac{10 + 10,7 + 12}{3} = 10,9;$$

$$\bar{Y}_2 = \frac{10,7 + 12 + 10,3}{3} = 11,0;$$

$$\overline{Y}_3 = \frac{12+10,3+12,9}{3} = 11,8 \text{ и т.д.}$$

Фактический ряд динамики и скользящую среднюю представим в виде статистического графика.

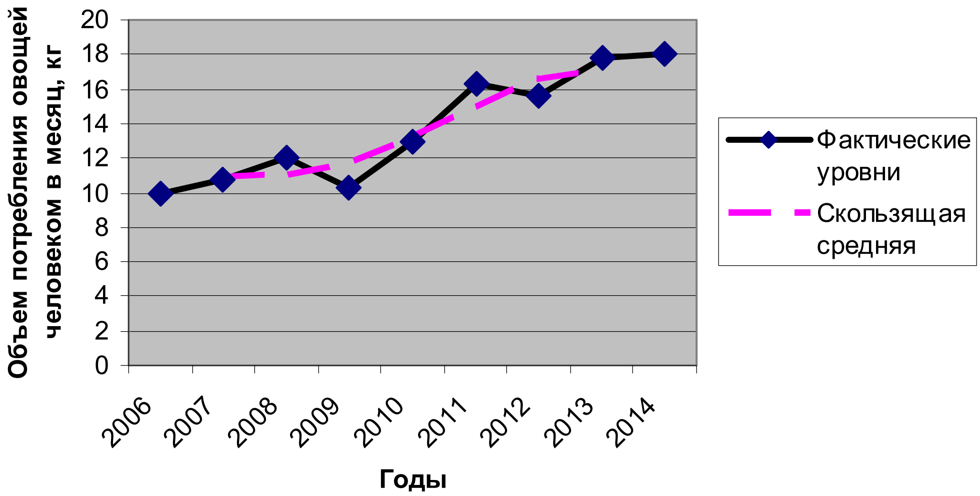


Рис. 6.2. Ряд динамики и линия тренда объема потребления овощей человеком в месяц, кг

Таким образом, имеет место основная тенденция к увеличению ежемесячного объема потребления овощей человеком.

Недостаток метода состоит в невозможности определения нескольких сглаженных уровней в начале и в конце ряда динамики.

6.10. Метод аналитического выравнивания

В соответствии с методом аналитического выравнивания развитие изучаемого явления представляют в зависимости только от течения времени. В итоге получают наиболее общий результат действия всех факторов, а отклонение фактических уровней от общей тенденции развития, объясняют действием факторов, проявляющихся случайно или циклически, т.е. строится трендовая модель следующего вида:

$$Y_t = f(t) + \varepsilon_t, \qquad (6.16)$$

где t – порядковый номер момента или периода времени; Y_t – фактический уровень в момент или период времени t; $f(t)$ – уровень, определяемый основной тенденцией развития; ε_t – случайное или циклическое отклонение от тренда.

Функция $f(t)$ называется уравнением тренда и может быть линейная вида $f(t) = a_0 + a_1 t$; гиперболическая вида $f(t) = a_0 + \frac{a_1}{t}$; параболическая вида $f(t) = a_0 + a_1 t + a_2 t^2$ и т.д. Оценка параметров a_0, a_1, a_2 и т.д. осуществляется методом наименьших квадратов, который обеспечивает наименьшую сумму квадратов отклонений фактических уровней от тренда.

Так, в случае линейной функции необходимо решить следующую систему нормальных уравнений:

$$\begin{cases} a_0 n + a_1 \sum t = \sum Y; \\ a_1 \sum t + a_1 \sum t^2 = \sum Yt. \end{cases} \tag{6.17}$$

Однако эту систему можно существенно упростить, если отсчет времени начать от середины ряда динамики, т.е. при нечетном числе уровней центральный уровень обозначают 0, а уровни до и после него ± 1; ± 2; ± 3 и т.д.; при четном числе уровней два центральных уровня обозначают ± 1, а уровни до и после него ± 3; ± 5; ± 7 и т.д. В этом случае $\sum t = 0,$ поэтому формулы для определения параметров будут иметь следующий вид:

$$a_0 = \frac{\sum Y}{n}; \tag{6.18}$$

$$a_1 = \frac{\sum Yt}{\sum t^2}. \tag{6.19}$$

6.11. Проверка надежности уравнения тренда и экстраполяция

Построив уравнение тренда, проводят оценку его надежности с помощью критерия Фишера:

$$F = \frac{Д_а}{Д_о} \frac{n-m}{m-1} \tag{6.20}$$

где m – число параметров в уравнении тренда; $Д_а = \frac{\sum (f(t) - \overline{Y})^2}{n}$ – аналитическая дисперсия; $Д_о = \frac{\sum (f(t) - Y)^2}{n}$ – остаточная дисперсия.

Если расчетное значение критерия превышает табличное (см. Приложение 1), которое зависит от уровня значимости α = 0,05 или 5% и числа степеней свободы для аналитической и остаточной дисперсии $v_1 = m - 1$ и $v_2 = n - m,$ то можно утверждать, что построенное уравнение тренда адекватно отражает фактическую тенденцию развития изучаемого явления.

В этом случае путем экстраполяции можно дать прогнозную оценку уровня изучаемого явления в виде следующего доверительного интервала:

$$f(t) - t_\alpha \sigma_о \le Y_t' \le f(t) + t_\alpha \sigma_о, \tag{6.21}$$

где $f(t)$ – уровень, определяемый основной тенденцией развития в момент или период времени t; t_α – нормированное отклонение, которое устанавливается для принятого уровня значимости α и числа степеней свободы $v = n - 1$ (см. Приложение 2); $\sigma_{ос} = \sqrt{\frac{\sum (f(t) - Y)^2}{n - m}}$ – остаточное среднеквадратическое отклонение.

6.12. Анализ сезонных колебаний

Если в изучаемом ряду динамики наблюдаются устойчивые отклонения от тренда, то можно предположить наличие в нем одного или нескольких колебательных процессов. Особенно это заметно, когда изучаемые явления имеют сезонный характер, то есть возрастание или убывание уровней повторяется регулярно с интервалом в один год (производство молока и мяса по месяцам года, потребление топлива и электроэнергии для бытовых нужд, сезонная распродажа товаров и т.д.).

Уровень сезонности изучаемого явления оценивается количественно с помощью индексов сезонности или путем гармонического анализа.

Индексы сезонности показывают, во сколько раз фактический уровень ряда динамики в конкретный момент или период времени отличается от среднего уровня или уровня, определяемого по уравнению тренда.

Если тренд отсутствует или он незначителен, то для каждого месяца или квартала изучаемого года индекс сезонности определяется по формуле

$$i_{\text{с}t} = Y_t \big/ \overline{Y}, \tag{6.22}$$

где Y_t – фактический уровень; $\overline{Y}$ – средний уровень.

При наличии тренда для каждого месяца или квартала изучаемого года индекс сезонности определяется по формуле

$$i_{\text{с}t} = Y_t \big/ f(t), \tag{6.23}$$

где $f(t)$ – уровень, определяемый по уравнению тренда.

Для обеспечения устойчивости этого показателя можно взять больший промежуток времени. Тогда для одноименных месяцев или кварталов определяют обобщающий индекс сезонности по формуле

$$I_{\text{с}t} = \frac{i_{\text{с}t}^1 + i_{\text{с}t}^2 + \ldots + i_{\text{с}t}^T}{T}, \tag{6.24}$$

где T – общее число лет.

6.13. Гармонический анализ ряда динамики

Гармонический анализ производят, представляя изучаемый ряд динамики как совокупность гармонических колебательных процессов (гармоник). Для каждого уровня такого ряда динамики будет справедлива следующая формула:

$$Y_t = f(t) + \sum_{k=1}^{n/2} \left(a_k \cos\left(kt\frac{2\pi}{n} \right) + b_k \sin\left(kt\frac{2\pi}{n} \right) \right), \tag{6.25}$$

где $t = 1, 2, \ldots, n$ – порядковый номер момента или периода времени; a_k, b_k – параметры гармонического колебательного процесса с порядковым номером k, которые в совокупности оценивают размах отклонения от тренда и сдвиг гармонических колебаний относительно начального периода или момента времени.

Общее число гармоник, которое можно выделить для ряда динамики будет равно $n/2$, однако обычно ограничиваются меньшим числом наиболее важных гармоник, параметры каждой из которых определяются по формулам:

$$a_k = \frac{2}{n}\sum_{t=1}^{n}\varepsilon_t \cos\left(kt\frac{2\pi}{n}\right); \tag{6.26}$$

$$b_k = \frac{2}{n}\sum_{t=1}^{n}\varepsilon_t \sin\left(kt\frac{2\pi}{n}\right). \tag{6.27}$$

6.14. Анализ взаимосвязанных рядов динамики

Под взаимосвязанными понимаются такие ряды динамики, в которых уровни одного в какой-то степени определяют уровни другого. Например, ряд динамики объема внесенных удобрений связан с рядом динамики урожайности, ряд динамики производительности труда – с рядом динамики заработной платы и т.д.

В простейших случаях для характеристики взаимосвязи рядов динамики их приводят к общему основанию, для чего принимают за базу уровни одного и того же периода или момента времени и определяют индексы опережения (отставания) в развитии одного явления по сравнению с другим.

Для этого путем деления сопоставляют цепных индексы или темпы динамики разных явлений. Например, если производительность труда на предприятии увеличилась на 12%, а средняя зарплата – на 7,5%, то рост производительности труда опережает рост средней зарплаты по индексу динамики на 1,12/1,075 – 1 = 0,042 или 4,2%, а по темпу динамики на 0,12/0,075 – 1 = 0,6 или 60%. Индекс отставания роста средней зарплаты от производительности труда будет обратной величиной.

Однако совпадение общих тенденций развития часто бывает вызвано не взаимной связью, а автокорреляцией, то есть зависимостью каждого последующего уровня ряда динамики от предыдущего, которая объясняется прочими неучтенными факторами. Поэтому прежде чем оценивать взаимосвязь, автокорреляцию из рядов динамики необходимо исключить.

6.15. Методы исключения автокорреляции

Для исключения автокорреляции есть три метода.

1. Исключение тренда из рядов динамики, согласно которому для каждого из взаимосвязанных рядов динамики X и Y составляют уравнение тренда следующего вида:

$$\widehat{X}_t = a_0 + a_1 t; \tag{6.28}$$

$$\widehat{Y}_t = b_0 + b_1 t. \tag{6.29}$$

Затем строят новые ряды динамики из отклонений от трендов:

$$\varepsilon_t = X_t - \widehat{X}_t; \tag{6.30}$$

$$\varepsilon'_t = Y_t - \widehat{Y}_t. \tag{6.31}$$

После этого новые ряды динамики проверяют на автокорреляцию при помощи критерия Дарбина и Уотсона:

$$K_{\text{ДУ}} = \frac{\sum(\varepsilon_{t+1} - \varepsilon_t)^2}{\sum \varepsilon_t^2}. \qquad (6.32)$$

Значения критерия изменяются от 0 до 4: если он равен 0, то имеется полная положительная автокорреляция; если он равен 2, то автокорреляция отсутствует; если он равен 4, то имеется полная отрицательная автокорреляция. Если значение критерия приблизительно равно 2, то такой ряд динамики оставляют без изменения, а в противном случае для такого ряда динамики составляют уравнение взаимосвязи следующего вида:

$$\hat{\varepsilon}_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}; \qquad (6.33)$$

$$\hat{\varepsilon}'_t = \beta_0 + \beta_1 \varepsilon'_{t-1}. \qquad (6.34)$$

Затем подсчитывают новые отклонения и проводят анализ их взаимосвязи.

2. Метод первых разностей, в соответствии с которым от исходных рядов динамики X и Y осуществляется переход к новым рядам динамики:

$$\Delta X_t = X_t - X_{t-1}; \qquad (6.35)$$

$$\Delta Y_t = Y_t - Y_{t-1}. \qquad (6.36)$$

По полученным рядам динамики определяют направление и силу связи.

3. Включение фактора времени в уравнение взаимосвязи, которое в случае линейной зависимости будет иметь следующий вид:

$$Y_t = a_0 + a_1 X_t + a_2 t. \qquad (6.37)$$

Из перечисленных методов наиболее простыми являются второй и третий, однако наиболее эффективен первый метод.

6.15. Методические указания

Например, по статистическим данным о среднемесячном курсе доллара по отношению к рублю, р./долл. определим абсолютные, относительные и средние показатели ряда динамики цепным и базисным способами, а также произведем непосредственное выделение тренда методом аналитического выравнивания.

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
р./долл.	27,08	26,27	23,30	31,68	30,36	29,35	31,06	31,82	38,42

1. По исходным данным заполним периоды и уровни ряда динамики в первом и втором столбце таблицы 6.2.

2. Определим абсолютные изменения уровней базисным способом в третьем столбце таблицы 6.2 по формуле (6.2): 26,27 – 27,08 = -0,81 р./долл.▼; 23,30 – 27,08 = -3,78 р./долл.▼; 31,68 – 27,08 = 4,60 р./долл.▲ и т.д.

3. Определим абсолютные изменения уровней цепным способом в четвертом столбце таблицы 6.2 по формуле (6.3): 26,27 – 27,08 = -0,81 р./долл.▼; 23,30 – 26,27 = -2,97 р./долл.▼; 31,68 – 23,30 = 8,38 р./долл. ▲.

4. Контроль по формуле (6.4): -0,81 + (-2,97) + 8,38 + (-1,32) + (-1,01) + 1,71 + 0,76 + 6,6 = 11,34.

5. Определим относительные изменения уровней и их темпы базисным способом в пятом и седьмом столбце таблицы 6.2 по формуле (6.5): 26,27 / 27,08 = 0,970 ▼ на 3%; 23,30 / 27,08 = 0,860▼ на 14%; 31,68 / 27,08 = 1,170▲ на 17% и т.д.

6. Определим относительные изменения уровней и их темпы цепным способом в шестом и восьмом столбце таблицы 6.2 по формуле (6.6): 26,27 / 27,08 = 0,970 ▼ на 3%; 23,30 / 26,27 = 0,887▼ на 11,3%; 31,68 / 23,30 = 1,360 р./долл. ▲ на 36% и т.д.

7. Контроль по формуле (6.7): 0,970·0,887·1,360·0,958·0,967·1,058·1,024·1,207 = 1,419.

8. Определим средний уровень ряда динамики по формуле (6.8): $\overline{Y} =$(27,08 + 26, 27 + 23,30 + 31,68 + 30,36 + 29,35 + 31,06 + 31,82 + 38,42)/9 = 29,93 р./долл.

9. Определим среднее абсолютное изменение уровней базисным или цепным способом по формуле (6.11): $\overline{\Delta Y} = \dfrac{-0{,}81 + (-2{,}97) + 8{,}38 + (-1{,}32) + (-1{,}01) + 1{,}71 + 0{,}76 + 6{,}6}{9-1} = \dfrac{11{,}34}{8}$ =1,42 р./долл. ▲. Таким образом, имеет место рост среднемесячного курса доллара по отношению к рублю в среднем на 1,42 р./долл. в год.

10. Определим среднее относительное изменение уровней базисным или цепным способом по формуле (6.12): $\bar{i} = \sqrt[8]{0{,}970 \cdot 0{,}887 \cdot 1{,}360 \cdot 0{,}958 \cdot 0{,}967 \cdot 1{,}058 \cdot 1{,}024 \cdot 1{,}207} = \sqrt[8]{1{,}419}$ =1,045 ▲. Таким образом, имеет место рост среднемесячного курса доллара по отношению к рублю в среднем на 4,5% в год.

11. Введем условную нумерацию периодов *t*, начиная от середины ряда динамики в девятом столбце таблицы 6.2., чтобы их сумма была равна нулю.

12. Определим расчетные значения t^2 и Yt в десятом и одиннадцатом столбце таблицы 6.2, а также определим их итоги: 60 и 75,20.

13. Определим параметры линейного уравнения тренда по формулам (6.18) и (6.19): a_0 = (27,08 + 26, 27 + 23,30 + 31,68 + 30,36 + 29,35 + 31,06 + 31,82 + 38,42)/9 = 29,93 и a_1 = 75,20/60 = 1,253.

14. Постоим линейное уравнение тренда $f(t) = 29{,}93 + 1{,}253 \cdot t$.

15. Определяем трендовые уровни *f(t)* в двенадцатом столбце таблицы 6.2, для чего сначала подставляем *t* = -4, тогда получаем $29{,}93 + 1{,}253 \cdot (-4)$ =24,92, затем *t* = -3, тогда получаем $29{,}93 + 1{,}253 \cdot (-3)$ =26,17, затем *t* = -2, тогда получаем $29{,}93 + 1{,}253 \cdot (-2)$ =27,42 и т.д.

Таблица 6.2

Годы	Y	$\Delta Y^б$	$\Delta Y^ц$	$i^б$	$i^ц$	$T^б$	$T^ц$	t	t^2	Yt	$f(t)$	$(f(t)-\overline{Y})^2$	$(f(t)-Y)^2$
2006	27,08	-	-	-	-	-	-	-4	16	-108,32	24,92	8,12	4,67
2007	26,27	-0,81	-0,81	0,970	0,970	-3	-3	-3	9	-78,81	26,17	13,40	0,01
2008	23,30	-3,78	-2,97	0,860	0,887	-14	-11,3	-2	4	-46,6	27,42	43,96	17,01
2009	31,68	4,6	8,38	1,170	1,360	17	36,0	-1	1	-31,68	28,68	3,06	9,02
2010	30,36	3,28	-1,32	1,121	0,958	12,1	-4,2	0	0	0	29,93	0,19	0,18
2011	29,35	2,27	-1,01	1,084	0,967	8,4	-3,3	1	1	29,35	31,18	0,34	3,36
2012	31,06	3,98	1,71	1,147	1,058	14,7	5,8	2	4	62,12	32,44	1,28	1,89
2013	31,82	4,74	0,76	1,175	1,024	17,5	2,4	3	9	95,46	33,69	3,57	3,49
2014	38,42	11,34	6,6	1,419	1,207	41,9	20,7	4	16	153,68	34,94	72,08	12,10
Итого:		-	11,34	-	1,419	-	-	0	60	75,20	–	145,99	51,74

16. Построим ряд динамики и линию тренда в виде статистического графика.

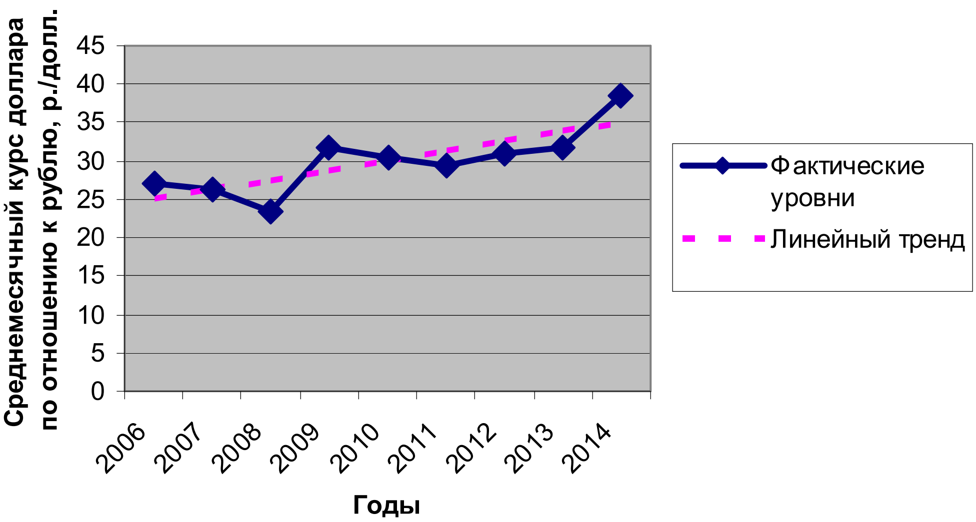


Рис. 6.3. Ряд динамики и линия тренда среднемесячного курса доллара по отношению к рублю, р./долл.

17. Определим расчетное значение $(f(t)-\overline{Y})^2$ в тринадцатом столбце таблицы 6.2: $(27{,}08 - 29{,}93)^2=8{,}12$; $(26{,}27 - 29{,}93)^2=13{,}40$; $(23{,}30 - 29{,}93)^2=43{,}96$ и т.д., а также его итог: 145,99.

18. Определим расчетное значение $(f(t)-Y)^2$ в четырнадцатом столбце таблицы 6.2: $(27{,}08 - 24{,}92)^2=4{,}67$; $(26{,}27 - 26{,}17)^2=0{,}01$; $(23{,}30 - 27{,}42)^2=17{,}01$ и т.д., а также его итог: 51,74.

19. Определим расчетное значение критерия Фишера по формуле (6.20): $F=\dfrac{145{,}99}{51{,}74}\cdot\dfrac{9-2}{2-1}$=19,75.

20. Сравним с табличным значением критерия Фишера 5,59, которое определяем для α = 0,05, а также 1 и 7 степеней свободы (см. Приложение 1). Поскольку расчетное значение больше табличного, построенное уравнение тренда является надежным.

Контрольные задания 6

Вариант 1

Определить абсолютные, относительные и средние показатели ряда динамики цепным и базисным способами, а также произвести непосредственное выделение тренда методом скользящей средней и аналитического выравнивания, используя статистические данные о производстве зерна за несколько лет:

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
млн. т	55,5	66,6	65,2	67,2	78,1	78,2	78,6	81,8	83,6

Вариант 2

Определить абсолютные, относительные и средние показатели ряда динамики цепным и базисным способами, а также произвести непосредственное выделение тренда методом скользящей средней и аналитического выравнивания, используя статистические данные о сумме выданных кредитов коммерческими банками за несколько лет:

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
млрд. руб.	102	118	124	137	175	187	204	222	245

Вариант 3

Определить абсолютные, относительные и средние показатели ряда динамики цепным и базисным способами, а также произвести непосредственное выделение тренда методом скользящей средней и аналитического выравнивания, используя статистические данные о производстве мяса за несколько лет:

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
млн. т	32,3	37,9	42,9	42,4	50,5	51,5	59,7	61,5	64,8

Вариант 4

Определить абсолютные, относительные и средние показатели ряда динамики цепным и базисным способами, а также произвести непосредственное выделение тренда методом скользящей средней и аналитического выравнивания, используя статистические данные о площади введенного жилья за несколько лет:

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
млн. кв. м.	347	350	349	351	345	349	357	359	365

Вариант 5

Определить абсолютные, относительные и средние показатели ряда динамики цепным и базисным способами, а также произвести непосредственное выделение тренда методом скользящей средней и аналитического выравнивания, используя статистические данные о валовом сборе картофеля за несколько лет:

Годы	2006	2007	2008	2009	2010	2011	2012	2013	2014
млн. т	34,0	35,0	32,9	36,7	35,9	37,3	40,3	40,8	41,3

Глава 7. Индексы

7.1. Назначение и типы индексов

Индекс – это относительная величина, которая показывает, во сколько раз уровень изучаемого явления в конкретных условиях отличается от уровня того же явления в других условиях. При этом условия могут различаться во времени, тогда получается динамический индекс, в пространстве, тогда получается территориальный индекс, а также по отношению к планируемому уровню, тогда получаются индексы планового задания и выполнения плана.

В статистическом анализе индексы используются не только для сопоставления уровней, но и для оценки значимости факторов, объясняющих абсолютное различие уровней результативного показателя.

В зависимости от сложности различают три типа индексов: индивидуальные индексы, общие индексы и индексы средних величин.

7.2. Индивидуальные индексы

Индивидуальные индексы применяются, когда не имеет значения структура изучаемого явления. Например, если изучается выручка от продажи товара, которая определяется по формуле

$$Q = qp, \tag{7.1}$$

где q – количество товара; p – цена.

Сравнивая уровни этих показателей в условиях отчетного и базисного периода, получаем соответствующие индивидуальные индексы:

$$i_q = \frac{q_1}{q_0}; \tag{7.2}$$

$$i_p = \frac{p_1}{p_0}; \tag{7.3}$$

$$i_Q = \frac{Q_1}{Q_0} = \frac{q_1 p_1}{q_0 p_0} = i_q i_p \tag{7.4}$$

7.3. Мультипликативные индексные модели

Последняя формула, записанная в виде $Q_1 = i_q i_p Q_0$, представляет собой двухфакторную мультипликативную индексную модель выручки, которая позволяет находить ее абсолютное изменение в отчетном периоде по сравнению с базисным под влиянием каждого фактора в отдельности.

В статистическом анализе часто используются мультипликативные модели с тремя и более факторами. Например, материальные затраты на производство продукции определяются по формуле

$$M = qmp, \tag{7.5}$$

где q – количество продукции; m – расход материала на единицу продукции (удельный расход материала); p – цена материала.

Отсюда можно получить следующую трехфакторную мультипликативную индексную модель материальных затрат

$$M_1 = i_q i_m i_p M_0. \tag{7.6}$$

7.4. Общий индекс результативного показателя

Индекс становится общим, когда изучается неоднородное явление (например, выручка от продажи не одного, а всех или несколько товаров), что делает невозможным суммирование объемного показателя в натуральных единицах.

В этом случае общую выручку можно записать в агрегатном виде как сумму произведений количества товаров на их цены. Сравнивая уровни этого показателя в условиях отчетного и базисного периода, получаем общий индекс выручки:

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0}. \tag{7.7}$$

При анализе динамике общей выручки она также объясняется влиянием двух факторов (количества товаров и цен), которое оценивается с помощью соответствующих общих индексов, формулы для расчета которых были предложены в середине XIX в. немецкими статистиками Ласпейресом и Пааше.

7.5. Общие индексы Ласпейреса, Пааше и Фишера

Если первым фактором считается количество товаров, а вторым – цены, то определяется общий количественный индекс Ласпейреса и общий ценовой индекс Пааше по формулам:

$$I_q^{\text{Л}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \tag{7.8}$$

$$I_p^{\text{П}} = \frac{\sum q_1 p_1}{\sum q_1 p_0}. \tag{7.9}$$

Если первым фактором считаются цены, а вторым – количество товаров, то определяется общий ценовой индекс Ласпейреса и общий количественный индекс Пааше по формулам:

$$I_p^{\text{Л}} = \frac{\sum p_1 q_0}{\sum p_0 q_0}; \tag{7.10}$$

$$I_q^{\text{П}} = \frac{\sum p_1 q_1}{\sum p_1 q_0}. \tag{7.11}$$

Если перемножить общие количественный и ценовой индексы Ласпейреса и Пааше или наоборот, то можно получить общий индекс выручки:

$$I_q^{\text{Л}} I_p^{\text{П}} = I_p^{\text{Л}} I_q^{\text{П}} = \frac{\sum Q_1}{\sum Q_0} = I_Q. \qquad (7.12)$$

Однако как следует из вышеприведенных формул, по одному и тому же фактору общие индексы Ласпейреса и Пааше не равны между собой (эффект Гершенкрона): $I_p^П \neq I_p^Л; I_q^Л \neq I_q^П$. И поскольку в статистике у каждого индекса должно быть только одно числовое значение, американский экономист Фишер предложил применять для этой цели формулу средней геометрической.

Таким образом, общие индексы Фишера определяются по формулам:

$$I_q^{\Phi} = \sqrt{I_q^{\text{Л}} I_q^{\text{П}}}; \qquad (7.13)$$

$$I_p^{\Phi} = \sqrt{I_p^{\text{Л}} I_p^{\text{П}}}. \qquad (7.14)$$

7.6. Общие индексы как средние из индивидуальных индексов

Помимо записи общих индексов в агрегатном виде на практике часто используют формулы их расчета как средних величин из соответствующих индивидуальных индексов.

Например, общий индекс выручки может быть записан как средняя арифметическая взвешенная из индивидуальных индексов выручки. Для этого умножим и разделим его числитель на выручку базисного периода, тогда получаем:

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum (Q_1/Q_0)\cdot Q_0}{\sum Q_0} = \frac{\sum i_Q Q_0}{\sum Q_0}. \qquad (7.15)$$

Общий индекс выручки также может быть записан как средняя гармоническая взвешенная из индивидуальных индексов выручки. Для этого умножим и разделим его знаменатель на выручку отчетного периода, тогда получаем:

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum Q_1}{\sum Q_1 : (Q_1/Q_0)} = \frac{\sum Q_1}{\sum Q_1/i_Q}. \qquad (7.16)$$

Аналогично через индивидуальные индексы количества товара и цены могут быть выражены общие индексы Ласпейреса и Пааше.

7.7. Индексы средних величин

Рассмотренные выше общие индексы могут применяться и при изучении однородных объектов (например, предприятий, реализующих один и тот же товар). В этом случае динамику общего количества товара можно показать непосредственно, т.е. отдельно от динамики цен.

Для этого сначала запишем общий индекс выручки в агрегатном виде, а затем умножим и разделим его числитель и знаменатель на общее количество товара соответственного отчетного и базисного периода, тогда получаем:

$$I_Q = \frac{\sum Q_1}{\sum Q_0} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1}{\sum q_0} \cdot \frac{\dfrac{\sum q_1 p_1}{\sum q_1}}{\dfrac{\sum q_0 p_0}{\sum q_0}} = \frac{\sum q_1}{\sum q_0} \cdot \frac{\sum d_1 p_1}{\sum d_0 p_0} = I_q I_{\bar{p}}, \qquad (7.17)$$

где $d = q / \sum q$ – доля конкретного предприятия в общем количестве товара; $I_q = \dfrac{\sum q_1}{\sum q_0}$ – общий индекс количества товара; $I_{\bar{p}}$ – индекс переменного состава, который характеризует динамику средней цены за счет динамики цен по конкретным предприятиям и их долей в общем количестве товара и определяется по формуле:

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\sum d_1 p_1}{\sum d_0 p_0} \qquad (7.18)$$

Влияние этих факторов количественно оценивается при помощи индекса структурных сдвигов и индекса постоянного состава, которые определяются по формулам:

$$I_d = \frac{\sum d_1 p_0}{\sum d_0 p_0} \qquad (7.19)$$

$$I_p = \frac{\sum d_1 p_1}{\sum d_1 p_0} \qquad (7.20)$$

Если перемножить индексы структурных сдвигов и постоянного состава, то можно получить индекс переменного состава:

$$I_d I_p = I_{\bar{p}}. \qquad (7.21)$$

Подставляя это выражение в формулу общего индекса выручки, можно получить следующую трехфакторную мультипликативную индексную модель:

$$\Delta \sum Q = I_q I_d I_p \sum Q_0. \qquad (7.22)$$

7.8. Факторный индексный анализ

Например, если изучается выручка от продажи конкретного товара (индивидуальная выручка), то ее итоговое абсолютное изменение в отчетном периоде по сравнению с базисным $\Delta Q = Q_1 - Q_0$ можно распределить по факторам с помощью двухфакторной мультипликативной модели.

Если первым фактором считается количество товара, а вторым – цена, то за счет только количественного фактора при постоянной цене $(i_p = 1)$ оно будет равно

$$\Delta Q_q = i_q Q_0 - Q_0 = (i_q - 1) Q_0. \qquad (7.23)$$

Тогда за счет ценового фактора оно будет равно

$$\Delta Q_p = \Delta Q - \Delta Q_q = Q_1 - Q_0 - (i_q - 1) Q_0 = i_q i_p Q_0 - Q_0 - (i_q - 1) Q_0 = i_q (i_p - 1) Q_0. \qquad (7.24)$$

Если поменять факторы местами, то итоговое абсолютное изменение выручки останется прежним, однако за счет каждого из факторов в отдельности оно будет определяться по формулам

$$\Delta Q_p = (i_p - 1)Q_0, \tag{7.25}$$

$$\Delta Q_q = i_p(i_q - 1)Q_0. \tag{7.26}$$

Таким образом, для контроля результатов факторного индексного анализа может использоваться формула:

$$\Delta Q = \Delta Q_q + \Delta Q_p. \tag{7.27}$$

Аналогично, итоговое абсолютное изменение общей выручки в отчетном периоде по сравнению с базисным $\Delta\sum Q = \sum Q_1 - \sum Q_0$ с помощью трехфакторной мультипликативной модели распределяется по факторам следующим образом:

$$\Delta\sum Q_q = (I_q - 1)\sum Q_0; \tag{7.28}$$

$$\Delta\sum Q_d = I_q(I_d - 1)\sum Q_0; \tag{7.29}$$

$$\Delta\sum Q_p = I_q I_d (I_p - 1)\sum Q_0. \tag{7.30}$$

Если поменять факторы местами, то получаться другие формулы, но их сумма всегда будет одинаковой и равна итоговому абсолютному изменению общей выручки:

$$\Delta\sum Q = \Delta\sum Q_q + \Delta\sum Q_d + \Delta\sum Q_p. \tag{7.31}$$

Наряду с этим, итоговое абсолютное изменение общей выручки можно распределять по отдельным предприятиям, для каждого из которых применяется следующая трехфакторная мультипликативная индексная модель:

$$Q_1 = I_q i_d i_p Q_0, \tag{7.32}$$

где i_d – индивидуальный индекс доли конкретного предприятия в общем количестве товара, который можно определить по формуле

$$i_d = \frac{d_1}{d_0}. \tag{7.33}$$

7.9. Территориальные индексы

Территориальные индексы представляют собой разновидность относительных величин сравнения, когда сопоставляются уровни одного и того же показателя, относящегося к одинаковому периоду времени, но к разным территориям (странам, регионам, городам, районам и т.д.).

Построение территориальных индексов рассмотрим на примере показателя товарооборота, который представляет собой общую выручку от продажи товаров для двух районов – А и Б, один из которых (например, Б) принимается за базу сравнения, тогда территориальный индекс товарооборота определяется по формуле:

$$I_{Q.\text{А/Б}} = \frac{\sum Q_{\text{А}}}{\sum Q_{\text{Б}}} = \frac{\sum q_{\text{А}} p_{\text{А}}}{\sum q_{\text{Б}} p_{\text{Б}}}. \tag{7.34}$$

Абсолютное различие товарооборотов этих районов объясняется различием ассортимента и количества товаров, а также различием цен. Влияние этих факторов оценивается при помощи территориальных индексов количества товаров и цен, которые в соответствии с методикой Эджворта-Маршалла можно определить по формулам:

$$I_{q.\text{А/Б}} = \frac{\sum q_{\text{А}} \overline{p}}{\sum q_{\text{Б}} \overline{p}}; \tag{7.35}$$

$$I_{p.\text{А/Б}} = \frac{\sum q p_{\text{А}}}{\sum q p_{\text{Б}}}, \tag{7.36}$$

где $\overline{p} = \dfrac{p_{\text{А}} q_{\text{А}} + p_{\text{Б}} q_{\text{Б}}}{q_{\text{А}} + q_{\text{Б}}}$ – средняя межрайонная цена конкретного товара, $q = q_{\text{А}} + q_{\text{Б}}$ – суммарное по двум районам количество конкретного товара.

Однако условие двухфакторной мультипликативной индексной модели может нарушаться, поскольку $I_{q.\text{А/Б}} \cdot I_{p.\text{А/Б}} \neq I_{Q.\text{А/Б}}$, поэтому использование территориальных индексов для факторного индексного анализа дает приближенные результаты.

7.10. Методические указания

Определим всевозможные индексы и выполним факторный индексный анализ выручки по следующим данным:

Предприятие	Базисный период		Отчетный период	
	Количество товара, тыс. ед.	Цена, руб./ед.	Количество товара, тыс. ед.	Цена, руб./ед.
А	100	20	140	15
Б	150	22	160	25

1. Используя исходные данные о количестве q и цене товара p заполним столбцы второй и третий, шестой и седьмой таблицы 7.1.
2. Определим итоги по количеству товара в базисном и отчетном периоде: 100 + 150 = 250 и 140 + 160 =300.
3. Определим среднюю цену товара в отчетном и базисном периоде по формуле (4.4):

$$\overline{p}_0 = \frac{100 \cdot 20 + 150 \cdot 22}{100 + 150} = 21{,}2 \text{ руб./ед.;}$$

$$\overline{p}_1 == \frac{100 \cdot 20 + 150 \cdot 22}{100 + 150} = 20{,}33 \text{ руб./ед.}$$

4. Определим выручку от продажи товара каждого предприятия в базисном периоде в четвертом столбце таблицы 7.1 по формуле (7.1): Q_0^A = 100·20 = 2000 тыс. р.; $Q_0^Б$ =150·22 =3300 тыс. р., а также ее итог: 2000 + 3300 = 5300. Аналогично можно определить выручку от продажи товара в отчетном периоде.
5. Определим долю каждого предприятия в общем количестве товара в базисном периоде в пятом столбце таблицы 7.1 по формуле (1.3): d_0^A = 100/250=0,4; $d_0^Б$ = 150/250 = 0,6. Аналогично можно определить долю каждого предприятия в общем количестве товара в отчетном периоде.

Таблица 7.1

Предприятие	Базисный период				Отчетный период			
	q_0	p_0	Q_0	d_0	q_1	p_1	Q_1	d_1
А	100	20	2000	0,4	140	15	2100	0,4667
Б	150	22	3300	0,6	160	25	4000	0,5333
Итого	250		5300	1	300		6100	1

6. Определим индивидуальные индексы по предприятию А по формулам (7.2), (7.3), (7.4) и (7.33): $i_q = 140/100 = 1,4;\ i_p = 15/20 = 0,75;\ i_Q = 2100/2000 = 1,05;\ i_d = 0,4667/0,4 = 1,1668.$ Таким образом, имеет место рост количества товара на 40%, снижение цены на 25%, рост выручки на 5%, а также доли предприятия на 16,68%. Аналогично можно определить индивидуальные индексы по предприятию Б (таблица 7.2).

7. Заполним итоги таблицы 7.2. Для этого определим общие индексы количества товара, средней цены и общей выручки по формулам (7.17), (7.18) и (7.4): $I_q = 300/250 = 1,2;$ $I_{\bar{p}} = 20,33/21,2 = 0,9591;$ $I_Q = 6300/5100 = 1,1509.$

Таблица 7.2

Предприятие	Индексы			
	q	p	Q	d
А	1,4	0,75	1,05	1,1668
Б	1,0667	1,1364	1,2121	0,8888
Итого	1,2	0,9591	**1,1509**	-

8. Определим общий количественный индекс Ласпейреса по формуле (7.8): $I_q^Л = \dfrac{140 \cdot 20 + 160 \cdot 22}{5300} = \dfrac{6320}{5300} = 1,1925.$

9. Определим общий ценовой индекс Пааше по формуле (7.9): $I_p^П = \dfrac{6100}{6320} = 0,9652.$

10. Определим общий ценовой индекс Ласпейреса по формуле (7.10): $I_p^Л = \dfrac{15 \cdot 100 + 25 \cdot 150}{5300} = \dfrac{5250}{5300} = 0,9906.$

11. Определим общий количественный индекс Пааше по формуле (7.11): $I_q^П = \dfrac{6100}{5250} = 1,1619.$

12. Определим общий количественный индекс Фишера по формуле (7.13): $I_q^Ф = \sqrt{1,1619 \cdot 1,1925} = 1,1771.$

13. Определим общий ценовой индекс Фишера по формуле (7.14): $I_p^Ф = \sqrt{0,9906 \cdot 0,9652} = 0,9778.$

14. Для контроля используем формулу (7.12): 1,1771 · 0,9778 = 1,1509. Таким образом, общая выручка увеличилась на 15,09% в связи с увеличением общего количества товара на 17,71% и снижением общего уровня цен на 2,22%.

15. Определим общий индекс выручки как по формуле (7.15): $I_Q = \dfrac{2000 \cdot 1,05 + 3300 \cdot 1,2121}{5300} = 1,1509.$

16. Определим общий индекс выручки как по формуле (7.16): $I_Q = \dfrac{6100}{2100/1,05 + 4000/1,2121} = 1,1509.$

17. Определим индекс переменного состава по формуле (7.18): $I_{\bar{p}} = \dfrac{0,4667 \cdot 15 + 0,5333 \cdot 25}{0,4 \cdot 20 + 0,6 \cdot 22} = 0,9591.$

18. Определим индекс структурных сдвигов по формуле (7.19): $I_d = 0,9937.$

19. Определим индекс постоянного состава по формуле (7.20): $I_p = 0,9652.$

Таким образом, имеет место снижение средней цены товара на 4,09%, в том числе на 0,63% за счет структурных сдвигов и на 3,48% за счет динамики цены товара по каждому предприятию.

20. Для контроля используем формулу (7.22): $1,2 \cdot 0,9937 \cdot 0,9652 = 1,1509.$

21. Итоговое абсолютное изменение общей выручки $\Delta\sum Q = 6100 - 5300 = 800$ тыс. р.

22. Факторные абсолютные изменения общей выручки определим по формулам (7.28), (7.29) и (7.30):

$$\Delta\sum Q_q = (1,2-1)\cdot 5300 = 1060 \text{ тыс. р.};$$

$$\Delta\sum Q_d = 1,2\cdot(0,9937-1)\cdot 5300 = -40 \text{ тыс. р.};$$

$$\Delta\sum Q_p = 1,2\cdot 0,9937\cdot(0,9652-1)\cdot 5300 = -220 \text{ тыс. р.}$$

23. Для контроля используем формулу (7.31): $1060 + (-40) + (-220) = 800.$

24. Итоговое абсолютное изменение выручки по предприятию А: $\Delta Q = Q_1 - Q_0 = 2100 - 2000 = 100$ тыс. руб. Аналогично можно определить итоговое абсолютное изменение выручки по предприятию Б.

25. Факторные абсолютные изменения выручки по предприятию А:

$$\Delta Q_q = (1,2-1)\cdot 2000 = 400 \text{ тыс. р.};$$

$$\Delta Q_d = 1,2\cdot(1,1668-1)\cdot 2000 = 400 \text{ тыс. р.};$$

$$\Delta Q_p = 1,2\cdot 1,1668\cdot(0,75-1)\cdot 2000 = -700 \text{ тыс. р.}$$

Аналогично можно определить факторные абсолютные изменения выручки по предприятию Б

26. Для контроля используем формулу (7.31): $400 + 400 + (-700) = 100.$ Результаты факторного анализа в таблице 7.3.

Таблица 7.3

Предприятие	Абсолютное изменение выручки, тыс. руб.	В том числе по факторам:		
		количество товара	структура	цены
А	100	400	400	–700
Б	700	660	–440	480
Итого:	800	1060	–40	–220

Контрольные задания 7

Вариант 1

Определить индексы и выполнить факторный индексный анализ выручки по следующим данным:

Предприятие	Базисный период		Отчетный период	
	Количество товара, тыс. ед.	Цена, руб./ед.	Количество товара, тыс. ед.	Цена, руб./ед.
1	180	32	150	35
2	220	51	270	48
3	200	19	240	15

Вариант 2

Определить индексы и выполнить факторный индексный анализ выручки по следующим данным:

Предприятие	Базисный период		Отчетный период	
	Количество товара, тыс. ед.	Цена, руб./ед.	Количество товара, тыс. ед.	Цена, руб./ед.
1	230	75	260	80
2	120	95	150	85
3	160	80	130	91

Вариант 3

Определить индексы и выполнить факторный индексный анализ выручки по следующим данным:

Предприятие	Базисный период		Отчетный период	
	Количество товара, тыс. ед.	Цена, руб./ед.	Количество товара, тыс. ед.	Цена, руб./ед.
1	460	20	450	19
2	310	22	330	20
3	770	19	750	18

Вариант 4

Определить индексы и выполнить факторный индексный анализ выручки по следующим данным:

Предприятие	Базисный период		Отчетный период	
	Количество товара, тыс. ед.	Цена, руб./ед.	Количество товара, тыс. ед.	Цена, руб./ед.
1	135	26	155	25
2	265	18	225	21
3	345	16	320	17

Вариант 5

Определить индексы и выполнить факторный индексный анализ выручки по следующим данным:

Предприятие	Базисный период		Отчетный период	
	Количество товара, тыс. ед.	Цена, руб./ед.	Количество товара, тыс. ед.	Цена, руб./ед.
1	350	24	500	23
2	240	35	250	44
3	520	65	550	62

Глава 8. Статистическое изучение взаимосвязей

8.1. Классификация взаимосвязей

Формы проявления взаимосвязей очень разнообразны. В наиболее общем виде можно выделить функциональную (полную) и корреляционную (неполную) связи. В первом случае каждому значению независимой переменной строго соответствует одно или несколько значений функции (зависимой переменной). Достаточно часто функциональная связь проявляется в физике, химии, а в экономике ее примером может служить пропорциональная зависимость между производительностью труда и объемом произведенной продукции. Корреляционная связь проявляется в среднем, для массовых наблюдений, когда каждому значению независимой переменной (факторного признака) соответствует некоторый ряд вероятных значений зависимой переменной (результативного признака), случайно распределенных в некотором интервале, что объясняется сложностью взаимосвязей между изучаемыми факторами, на взаимодействие которых оказывают влияние неучтенные случайные величины. Такие зависимости встречаются повсеместно. Например, в сельском хозяйстве – связь между урожайностью и количеством внесенных удобрений, поскольку для каждого конкретного участка, один и тот же объем удобрений вызовет разный прирост урожайности, так как во взаимосвязи находится еще целый ряд факторов (погода, состояние почвы, качество посадочного материала и др.), которые и формируют конечный результат. Однако в среднем такая связь наблюдается, поскольку увеличение количества удобрений ведет к росту урожайности.

По направлению связи бывают прямыми (положительными), когда зависимая переменная растет с увеличением факторного признака, и обратными (отрицательными), при которых рост последнего сопровождается уменьшением результативного признака. По своей аналитической форме связи бывают линейными и нелинейными. По числу изучаемых признаков связи могут быть парные, когда характеризуется связь двух признаков, и множественные, когда отражается связь более чем двух переменных.

Кроме вышеперечисленных различают непосредственные, косвенные и ложные связи. В первом случае факторы взаимодействуют между собой непосредственно. Для косвенной связи характерно участие какой-то третьей переменной, которая опосредует связь между изучаемыми признаками, а ложная связь устанавливается формально и, как правило, только количественными оценками. Однако она не имеет качественной основы или вообще бессмысленна.

8.2. Методы корреляционного и регрессионного анализа

Задачи статистики в области изучения взаимосвязей состоят в количественной оценке их наличия и направления, а также характеристике силы и формы влияния одних признаков на другие.

Для этого применяются две группы методов: 1) корреляционный анализ, задачи которого сводятся к измерению силы связи, определению неизвестных причинных связей и оценке факторов, оказывающих наибольшее влияние на результативный признак; 2) регрессионный анализ, к задачам которого относится установление формы зависимости, определение функции регрессии, а также использование уравнения для оценки неизвестных значений зависимой переменной.

Эти методы широко представлены в пакетах статистических прикладных программах для ЭВМ, однако знание принципов изучения взаимосвязей, а также возможностей и ограничений тех или иных методов является обязательным условием исследования.

Методы корреляционного анализа подразделяются на параметрические и непараметрические. Первые применяются, когда изучаемая совокупность подчиняется закону нормального распределения, а признаки являются количественными. При этом необходимо определить основные параметры распределения (среднюю величину и дисперсию), поэтому они получили название параметрических. Непараметрические методы не накладывают ограничений на характер распределения, и их преимущество заключается в простоте вычислений.

8.3. Парная корреляция

Простейшим приемом выявления связи между двумя признаками является построение корреляционной таблицы:

Таблица 8.1

Y_j / X_i	Y_1	Y_2	…	Y_j	…	Y_z	Итого:	$\bar{Y}_i$
X_1	f_{11}	f_{12}	…	f_{1j}	…	f_{1z}	$\sum f_{1j}$	$\bar{Y}_1$
X_2	f_{21}	f_{22}	…	f_{2j}	…	f_{2z}	$\sum f_{2j}$	$\bar{Y}_2$
…	…	…	…	…	…	…	…	…
X_i	f_{i1}	f_{i2}	…	f_{ij}	…	f_{iz}	$\sum f_{ij}$	$\bar{Y}_i$
…	…	…	…	…	…	…	…	…
X_k	f_{k1}	f_{k2}	…	f_{kj}	…	f_{kz}	$\sum f_{kj}$	$\bar{Y}_k$
Итого:	$\sum_{i=1}^{k} f_{i1}$	$\sum_{i=1}^{k} f_{i2}$	…	$\sum_{i=1}^{k} f_{ij}$	…	$\sum_{i=1}^{k} f_{iz}$	$\sum_{ij} f_{ij} = n$	$\bar{Y}$
$\bar{X}_j$	$\bar{X}_1$	$\bar{X}_2$	…	$\bar{X}_j$	…	$\bar{X}_z$	$\bar{X}$	–

В основу таблицы положена комбинацонная группировка двух изучаемых во взаимосвязи признаков – X и Y, а частоты f показывают количество соответствующих сочетаний значений X и Y. Если они расположены беспорядочно, то можно говорить об отсутствии связи, а если они образуют какое-либо характерное сочетание и концентрируются около одной из двух диагоналей корреляционной таблиц, то можно утверждать наличие прямой или обратной линейной связи.

Наглядным изображением корреляционной таблицы служит корреляционное поле в виде графика, где на оси абсцисс откладываются значения *X*, а по оси ординат – *Y*, и точками показывают сочетание первичных наблюдений *X* и *Y*. По расположению точек и концентрации в определенном направлении также можно судить о наличии связи.

В итогах корреляционной таблицы по строкам и столбцам приводятся два распределения – одно по *X,* другое по *Y*, и для каждого *X* рассчитывается среднее значение *Y* и для каждого *Y* – среднее значение *X* по формуле средней арифметической взвешенной. Последовательность точек с координатами $(X_i;\ \bar{Y}_i)$ или $(Y_j;\ \bar{X}_j)$ на графике образует эмпирическую линию регрессии, которая иллюстрирует зависимость среднего значения результативного признака от факторного.

8.4. Линейный коэффициент корреляции

Корреляционная таблица, корреляционное поле и эмпирическая линия регрессии характеризуют взаимосвязь, когда выбраны факторный и результативный признаки и необходимо сформулировать предположение о форме и направлении связи. Однако количественная оценка силы связи требует дополнительных расчетов. На практике для этого широко используется линейный коэффициент корреляции, который определяется как отношение средней величины из произведений парных отклонений (ковариации) *X* и *Y* к произведению их среднеквадратических отклонений:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n\sigma_X \sigma_Y} = \frac{\sum XY - \dfrac{\sum X \sum Y}{n}}{\sqrt{\sum X^2 - \dfrac{\left(\sum X\right)^2}{n}}\sqrt{\sum Y^2 - \dfrac{\left(\sum Y\right)^2}{n}}}. \qquad (8.1)$$

Коэффициент корреляции принимает значения в интервале от –1 до 1: если $|r| < 0{,}3$, то связь слабая; если $0{,}3 \le |r| < 0{,}7$ – умеренная; если $|r| \ge 0{,}7$ – тесная; если $|r| = 1$ – связь функциональная, а если $r = 0$, то линейная связь между *X* и *Y* отсутствует. Однако в этом случае возможно нелинейное взаимодействие, что требует дополнительной проверки и расчета других показателей.

8.5. Парная линейная регрессия

Модель парной линейной регрессии имеет следующий вид:

$$Y_i = a_0 + a_1 X_i + \varepsilon_i, \qquad (8.2)$$

где i = 1, 2, ... , n; n – общее число наблюдений (единиц совокупности); a_0, a_1 – неизвестные параметры уравнения регрессии; ε_i – случайная ошибка зависимой переменной *Y*.

Таким образом, уравнение парной линейной регрессии будет иметь следующий вид:

$$Y_X = a_0 + a_1 X, \tag{8.3}$$

где Y_X – теоретическое значение регрессии в виде оценки среднего (ожидаемого) значения результативного признака Y для конкретного значения факторного признака X.

Параметры a_0 и a_1 оцениваются методом наименьших квадратов, из которого вытекают формулы:

$$a_1 = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sum X^2 - \frac{\left(\sum X\right)^2}{n}} = r\frac{\sigma_Y}{\sigma_X}; \tag{8.4}$$

$$a_0 = \frac{\sum Y}{n} - a_1 \frac{\sum X}{n}. \tag{8.5}$$

Параметр a_0 – это постоянная величина в уравнении регрессии, который интерпретируют как исходное значение Y. Параметр a_1 – это коэффициент регрессии, который показывает, на сколько единиц в среднем изменится Y при изменении X на единицу и обладает размерностью отношения Y к X. Например, если по данным о стоимости оборудования X и производительности труда Y получено уравнение регрессии вида $Y = -12{,}14 + 0{,}208X$, то коэффициент регрессии означает, что увеличение стоимости оборудования на 1 тыс. р. ведет в среднем к росту производительности труда на 208 р.

8.6. Множественная линейная регрессия

Парную корреляцию и парную регрессию можно рассматривать как частный случай отражения связи зависимой переменной, с одной стороны, и одной из множества независимых переменных – с другой, поэтому, прежде всего, необходимо определить перечень независимых переменных X, включаемых в уравнение регрессии путем теоретического анализа взаимосвязи, который на практике должен формально подкрепляться парными коэффициентами корреляции между зависимой и независимой переменными. Наиболее значимые факторы, тесно связанных с результативным признаком, обычно отбираются с помощью ЭВМ. При этом решается вопрос о форме уравнения регрессии. Такой метод получил называние пошаговой регрессии.

Первоначально стоится модель множественной линейной регрессии следующего вида:

$$Y_i = a_0 + a_1 X_{i1} + a_2 X_{i2} + \ldots + a_k X_{ik} + \varepsilon_i \tag{8.6}$$

Таким образом, уравнение множественной линейной регрессии будет иметь следующий вид:

$$Y_{X_1,\, X_2,\, \ldots,\, X_k} = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_k X_k. \tag{8.7}$$

где $Y_{X_1, X_2, \ldots, X_k}$ – теоретическое значение регрессии, которое представляет собой оценку ожидаемого значения Y при фиксированных значениях $X_1, X_2, \ldots, X_k$; $a_1, a_2, \ldots, a_k$ – коэффициенты регрессии, каждый из которых показывает, на сколько единиц изменится Y при изменении соответствующего факторного признака X на единицу при условии, что остальные факторные признаки останутся на прежнем уровне.

Параметры уравнения множественной регрессии находятся методом наименьших квадратов и, как правило, с помощью ЭВМ. При этом необходимо установить насколько линейная форма связи соответствует реально существующей зависимости между Y, с одной стороны, и множеством X – с другой.

8.7. Нелинейная регрессия

Представление связи через линейную функцию там, где на самом деле существуют нелинейные зависимости, вызовет ошибку аппроксимации и в конечном итоге – упрощенные или ложные положения и выводы. Поэтому вопрос о нелинейности необходимо решать на стадии теоретического анализа. Однако на практике допускается и другое решение – нелинейность формулируется как предположение и очерчивается круг возможных уравнений регрессии, а затем их форма и вид уточняются с помощью ЭВМ.

В наиболее общем виде можно выделить два класса нелинейных уравнений регрессии. К первому относятся уравнения нелинейные относительно переменных, но линейные по параметрам. Например, полиномы в случае парной регрессии имеют следующий вид:

$$Y = a_0 + a_1 X + a_2 X^2 + a_3 X^3 + \ldots; \qquad (8.8)$$

Возможно также применение гиперболы и других функций, а также любого их сочетания.

Второй класс нелинейных уравнений регрессии отличается нелинейностью по параметрам. Например, степенная функция, которая в случае парной регрессии имеет следующий вид:

$$Y = a_0 X^{a_1}; \qquad (8.9)$$

По рассмотренным выше примерам можно судить о широком спектре аналитических представлений нелинейной формы связи, использование которых ограничивает лишь сложность процедур оценки параметров. Относительно просто такая задача решается для функций, преобразуемых к линейному виду. Например, степенную функцию можно прологарифмировать, получив линейную зависимость Y от X в логарифмах.

Однако интерпретация коэффициента регрессии как в линейном уравнении для нелинейной зависимости не подходит. Поэтому для характеристики влияния изменения одной переменной на изменение другой используют частную производную, взятую по соответствующему фактору X, что позволяет определить коэффициент эластичности, который показывает, на сколько процентов изменится Y при изменении X на 1% и определяется по формуле:

$$Э = \frac{\partial Y}{\partial X} \cdot \frac{X}{Y}. \tag{8.10}$$

где $Y'(X)$ – частная производная, взятая по соответствующему факторному признаку X.

Например, в случае парной линейной регрессии коэффициент эластичности будет равен

$$Э = \frac{a_1 X}{Y} = \frac{a_1 X}{a_0 + a_1 X}. \tag{8.11}$$

Использование коэффициентов эластичности в качестве дополнения к коэффициентам регрессии расширяет возможности сопоставления и экономической интерпретации результатов.

8.8. Множественная корреляция

Оценки силы связи (корреляции) дополняют результаты регрессионного анализа, дают представление о взаимосвязи изучаемых факторов и аппроксимации фактических данных аналитической функцией, поэтому расчет показателей множественной корреляции предполагает оценку уравнений регрессии.

Коэффициент множественной корреляции, который в случае нелинейной связи называется индексом корреляции, отражает силу связи между вариацией зависимой переменной и вариациями всех включенных в уравнение регрессии независимых переменных и определяется по формуле

$$R = \sqrt{\frac{Д_{ф}}{Д_{Y}}} = \sqrt{1 - \frac{Д_{ос}}{Д_{Y}}} = \sqrt{\frac{\sum (Y_X - \bar{Y})^2}{\sum (Y - \bar{Y})^2}}; \tag{8.12}$$

где $Д_Y = Д_ф + Д_{ос}$ – общая дисперсия результативного признака как сумма факторной и остаточной дисперсии, которые характеризуют вариацию результативного признака за счет факторов соответственно включенных и не учтенных в уравнении регрессии.

Значения коэффициента множественной корреляции колеблются от 0 до 1 и чем ближе к 1, тем более тесная связь между Y и множеством X.

Этот же показатель используется для измерения точности аппроксимации фактических данных аналитической функцией. При этом, если $R \le 0{,}3$, то можно утверждать, что либо не все важнейшие факторы учтены, либо выбрана неподходящая форма уравнения регрессии.

Квадрат коэффициента множественной корреляции называется коэффициентом детерминации $\eta = R^2$, который показывает, какая часть вариации зависимой переменной объясняется включенными в регрессионную модель факторами.

8.9. Оценка значимости параметров взаимосвязи

Получив оценки корреляции и регрессии, необходимо проверить их значимость особенно при малом объеме выборочной совокупности, когда $n \le 30$, то есть установить соответствие истинным параметрам взаимосвязи в генеральной совокупности.

Значимость коэффициента корреляции проверяется путем сопоставления с его среднеквадратической ошибкой. Для этого определяется критерий Стьюдента по формуле

$$S_r = |r| \sqrt{\frac{n-2}{1-r^2}}. \qquad (8.13)$$

Если расчетное значение критерия больше табличного в зависимости от уровня значимости α и числа степеней свободы $v = n-2$, то можно утверждать, что коэффициент корреляции является значимым параметром взаимосвязи.

Аналогично рассчитывают стандартные ошибки, а затем критерии Стьюдента для каждого параметра уравнения регрессии. Например, в для парной линейной регрессии используются формулы:

$$S_0 = |a_0| \frac{\sqrt{n-2}}{\sigma_{\text{ос}}}; \qquad (8.14)$$

$$S_1 = |a_1| \frac{\sigma_X \sqrt{n-2}}{\sigma_{\text{ос}}}. \qquad (8.15)$$

Вывод о правильности выбора вида взаимосвязи и значимости всего уравнения регрессии можно получить с помощью критерия Фишера:

$$F = \frac{Д_{\text{ф}}}{Д_{\text{ос}}} \cdot \frac{n-m}{m-1} = \frac{R^2}{1-R^2} \cdot \frac{n-m}{m-1}, \qquad (8.16)$$

где n – общее число наблюдений (единиц совокупности); m – число параметров в уравнении регрессии.

Расчетное значение критерия должно быть больше табличного для уровня значимости α = 0,05 и числа степеней свободы $v_1 = m-1$ и $v_2 = n-m$. В противном случае следует пересмотреть форму уравнения, перечень переменных и т.д.

8.10. Коэффициент взаимной сопряженности

Обобщающим показателем, характеризующим силу связи между двумя неколичественными признаками, позволяющим сравнить ее проявление в разных совокупностях, является коэффициент взаимной сопряженности, который может определяться по формулам Пирсона и Чупрова:

$$K_{\text{П}} = \sqrt{\frac{\varphi^2}{1+\varphi^2}}; \qquad (8.17)$$

$$K_{ч} = \sqrt{\frac{\varphi^2}{\sqrt{(K_1-1)(K_2-1)}}}, \qquad (8.18)$$

где K_1 и K_2 – число групп по каждому из неколичественных признаков; φ^2 – показатель средней квадратической сопряженности, определяемый путем вычитания единицы из суммы отношений квадратов частот каждой клетки корреляционной таблицы к произведению частот соответствующего столбца и строки:

$$\varphi^2 = \sum_{ij} \frac{f_{ij}^2}{f_i f_j} - 1; \; f_i = \sum_i f_{ij}\, ; \; f_j = \sum_j f_{ij}\, . \qquad (8.19)$$

Значения коэффициента взаимной сопряженности колеблются от 0 до 1 и интерпретируются в соответствии с общепринятыми критериями силы связи.

Например, определим коэффициенты взаимной сопряженности, используя статистические данные о распределении выпускников средних школ и их родителей по сферам занятости:

Таблица 8.2

Занятия детей / Занятия родителей	промышленность и строительство	сельское хозяйство	обслуживание	интеллектуальный труд	Итого:
1. Промышленность и строительство	40	5	7	39	91
2. Сельское хозяйство	34	29	13	12	88
3. Обслуживание	16	6	15	19	56
4. Интеллектуальный труд	24	5	9	72	110
Итого:	114	45	44	142	345

$$\varphi^2 = \left(\frac{40^2}{114} + \frac{5^2}{45} + \frac{7^2}{44} + \frac{39^2}{142}\right) : 91 + \left(\frac{34^2}{114} + \frac{29^2}{45} + \frac{13^2}{44} + \frac{12^2}{142}\right) : 88 + \ldots - 1 = 1,264 - 1 = 0,264;$$

$$K_{П} = \sqrt{\frac{0,264}{1+0,264}} = 0,457;$$

$$K_{ч} = \sqrt{\frac{0,264}{\sqrt{(4-1)(4-1)}}} = 0,296.$$

Таким образом, коэффициенты свидетельствует о наличии умеренной связи между изучаемыми признаками.

8.11. Коэффициенты ассоциации и контингенции

Некоторые особенности имеет анализ взаимосвязи между двумя альтернативными признаками, который производится с помощью комбинационной группировки единиц совокупности в виде «четырехклеточной» таблицы взаимной сопряженности, которая имеет следующий вид:

Таблица 8.3

X_i \ Y_j	Y_1	Y_2	Итого:
X_1	a	b	$a+b$
X_2	c	d	$c+d$
Итого:	$a+c$	$b+d$	$a+b+c+d$

Оценить силу связи между такими признаками можно с помощью коэффициентов взаимной сопряженности, но это можно сделать проще с помощью коэффициентов ассоциации и контингенции:

$$K_{\text{а}} = \frac{ad - bc}{ad + bc}, \tag{8.20}$$

$$K_{\text{к}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}; \tag{8.21}$$

Например, определим коэффициенты контингенции и ассоциации, используя статистические данные о распределении работников предприятий по полу и характеру труда:

Таблица 8.4

Пол \ Характер труда	сезонный	несезонный	Итого
Мужчины	187	265	452
Женщины	307	272	579
Итого	494	537	1031

$$K_{\text{к}} = \frac{187 \cdot 272 - 265 \cdot 307}{\sqrt{452 \cdot 579 \cdot 494 \cdot 537}} = -0{,}116;$$

$$K_{\text{а}} = \frac{187 \cdot 272 - 265 \cdot 307}{187 \cdot 272 + 265 \cdot 307} = -0{,}231.$$

Таким образом, коэффициенты свидетельствуют о наличии слабой связи между изучаемыми признаками.

8.12. Ранговые коэффициенты корреляции

В статистических исследованиях часто встречаются ситуации, когда признак не имеет непосредственного числового выражения, однако единицы совокупности можно ранжировать, т.е. упорядочить относительно изучаемого признака. Например, студентов можно ранжировать по способностям к точным и гуманитарным наукам, людей по уровню образования, творческим способностям и т.д. При этом каждой единице совокупности присваивается ранг, то есть порядковый номер, а если они совпадают у нескольких единиц совокупности, то каждой из них присваивается средний порядковый номер. Например, если у 5-й и 6-й единиц совокупности значения признаков одинаковы, то каждая из них получает ранг, равный 5,5.

Измерение силы связи в таких случаях производится с помощью ранговых коэффициентов корреляции, которые могут применяться и при изучении количественных признаков.

Для этого единицы совокупности ранжируют по возрастанию факторного признака X и для каждой единицы совокупности определяют ранг с точки зрения результативного признака Y. При этом, если связь прямая, то с увеличением ранга признака X, ранг признака Y также, как правило, будет возрастать и наоборот. Затем определяются ранговые коэффициенты корреляции по формулам Спирмена и Кендэла:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}, \tag{8.22}$$

$$\tau = \frac{\sum Q - \sum P}{\sum Q + \sum P}, \tag{8.23}$$

где $d = R_X - R_Y$ – разность рангов признаков X и Y, n – число единиц совокупности, Q и P – число баллов, т.е. случаев, когда ранг признака Y у всех последующих единиц совокупности соответственно больше и меньше, чем у конкретной единицы совокупности.

Например, определим ранговые коэффициенты корреляции, используя статистические данные о распределении семей А, Б, В, Г, Д, Е, Ж, З, И, К по уровню доходов соответственно 16,3; 10,8; 5,2; 6,9; 9,7; 9,0; 5,7; 7,7; 10,4; 14,9 и доле расходов на продукты питания соответственно 30; 28; 33; 49; 41; 46; 52; 38; 50; 44.

Построим единицы совокупности по возрастанию уровня доходов X с соответствующими наименованиями семей в первом столбце, а также с присвоением порядковых номеров R_X и R_Y во втором и третьем столбце таблицы 8.5.

Таблица 8.5

Семья	Ранги семьи				Баллы	
	R_X	R_Y	d	d^2	Q	P
В	1	3	–2	4	7	2
Ж	2	10	–8	64	0	8
Г	3	8	–5	25	1	6
З	4	4	0	0	4	2
Е	5	7	–2	4	1	4
Д	6	5	1	1	2	2
И	7	9	–2	4	0	3
Б	8	1	7	49	2	0
К	9	6	3	9	0	1
А	10	2	8	64	0	0
Итого	55	55	–	224	17	28

Определим расчетные значения d и d^2 в четвертом и пятом столбцах таблицы 8.5, а также итог последнего: 224.

Определим ранговый коэффициент корреляции Спирмена по формуле (8.22):

$$\rho = 1 - \frac{6 \cdot 224}{10 \cdot (10^2 - 1)} = -0{,}358.$$

Определим расчетные значения баллов Q и P в шестом и седьмом столбцах таблицы 8.5, а также их итоги: 17 и 28.

Определим ранговый коэффициент корреляции Кендэла по формуле (8.23):

$$\tau = \frac{-11}{45} = -0{,}244.$$

Таким образом, коэффициенты свидетельствуют об обратной слабой связи между изучаемыми признаками.

8.12. Коэффициент конкордациии

Ранговые коэффициента корреляции Спирмена и Кендэла применяются для решения одних и тех же задач, однако преимуществом последнего является возможность его использования в многофакторном анализе. Например, при двухфакторном анализе частный коэффициент корреляции между признаками Y и X_1 при исключении факторного признака X_2 будет равен

$$r_{YX_1/X_2} = \frac{\tau_{YX_1} - \tau_{YX_2}\tau_{X_1X_2}}{\sqrt{\left(1-\tau^2_{YX_2}\right)\left(1-\tau^2_{X_1X_2}\right)}}, \quad (7.24)$$

где $\tau_{YX_1}, \tau_{YX_2}, \tau_{X_1X_2}$ – парные коэффициенты ранговой корреляции между Y и X_1, Y и X_2, X_1 и X_2.

Если число изучаемых признаков больше двух, то для измерения силы связи между ними может использоваться коэффициент конкордации:

$$W = \frac{12S}{m^2(n^3 - n)}, \quad (7.25)$$

где S – сумма квадратов отклонений суммы рангов от средней суммы рангов; m – число ранжируемых признаков; n – число ранжируемых единиц.

После определения коэффициента конкордации необходимо проверить его значимость с помощью критерия Пирсона:

$$\chi^2 = \frac{12S}{mn(n-1)}. \quad (7.26)$$

Если расчетное значение больше табличного, которое зависит от принятого уровня значимости α и числа степеней свободы $\nu = n-1$, то коэффициент конкордации считается значимым.

Например, определим коэффициент конкордации по следующим статистическим данным:

Банк	Активы, млн. руб.	Депозиты, млн. руб.	Собственный капитал, млн. руб.
1	507,2	448,1	209
2	506,6	451,9	201
3	487,7	447,9	177
4	496,0	444,3	136
5	493,6	443,2	175
6	458,9	411,7	88
7	429,3	328,6	60

Построим единицы совокупности по возрастанию активов X с порядковыми номерами банков в первом столбце, а также рангов активов, депозитов и собственного капитала R_X и R_Y и R_Z во втором, третьем и четвертом столбцах таблицы 8.6.

Определим сумму рангов для каждого банка, а также их итоги в пятом столбце таблицы 8.6: 84.

Определим среднюю величину суммы рангов: $\overline{R}$ = 84/7 = 12.

Определим расчетные значения отклонений суммы рангов для каждого банка от средней величины $\sum R_i - \overline{R}$ в шестом столбце таблицы 8.6.

Определим расчетные значения квадратов отклонений суммы рангов для каждого банка от средней величины в седьмом столбце таблицы 8.6., а также определим итог этого столбца: 134.

Таблица 8.6

Банк	Ранги				$\sum R_i - \overline{R}$	$(\sum R_i - \overline{R})^2$
	R_X	R_Y	R_Z	Итого		
7	1	1	1	3	-9	81
6	2	2	2	6	-6	36
3	3	5	5	13	1	1
5	4	3	4	11	-1	1
4	5	4	3	12	-	-
2	6	7	6	19	7	49
1	7	6	7	20	8	64
Итого	-	-	-	84	-	134

Определим коэффициент конкордации по формуле (7.25): $W = \dfrac{12 \cdot 34}{3^2 \cdot (7^3 - 7)} =$ 0,532.

Таким образом, коэффициент свидетельствуют о умеренной связи между изучаемыми признаками.

8.13. Методические указания

Определим линейный коэффициент корреляции и построим уравнение регрессии, используя следующие статистические данные по промышленным предприятиям:

Основные фонды, млн. р.	330	400	480	550	600	700	750	850	870	940
Выпуск продукции, млн. р.	9,9	10,8	11,5	12,0	12,4	12,9	13,1	13,5	13,6	13,9

1. По исходным данным заполним порядковые номера в первом столбце, а также числовые значения факторного и результативного признака X и Y в во втором и третьем столбце таблицы 8.7., а также итоги этих столбцов.

2. Определим расчетные значения XY, X^2 и Y^2 в четвертом, пятом и шестом столбце таблицы 8.7, а также их итоги.

3. Определим линейный коэффициент корреляции по формуле (8.1):

$$r = \frac{82375 - \dfrac{6470 \cdot 123{,}6}{10}}{\sqrt{\left(4577300 - \dfrac{6470^2}{10}\right)\left(1543{,}1 - \dfrac{123{,}6^2}{10}\right)}} = \frac{2405{,}8}{\sqrt{391210 \cdot 15{,}404}} = 0{,}980.$$

Таким образом, коэффициент свидетельствуют о прямой тесной связи между изучаемыми признаками.

4. Определим параметры уравнения парной линейной регрессии по формулам (8.4) и (8.5):

$$a_1 = \frac{2405{,}8}{391210} = 0{,}006; \; a_0 = \frac{123{,}6}{10} - \frac{6470}{10} \cdot 0{,}006 = 8{,}381.$$

5. Постоим уравнение парной линейной регрессии по формуле (8.3) $Y_X = 8{,}381 + 0{,}006 \cdot X.$

6. Определяем теоретические значения регрессии Y_X в седьмом столбце таблицы 8.7.

7. Построим фактические значения результативного признака и функцию регрессии в виде статистического графика (эмпирическая и теоретическая линия регрессии).

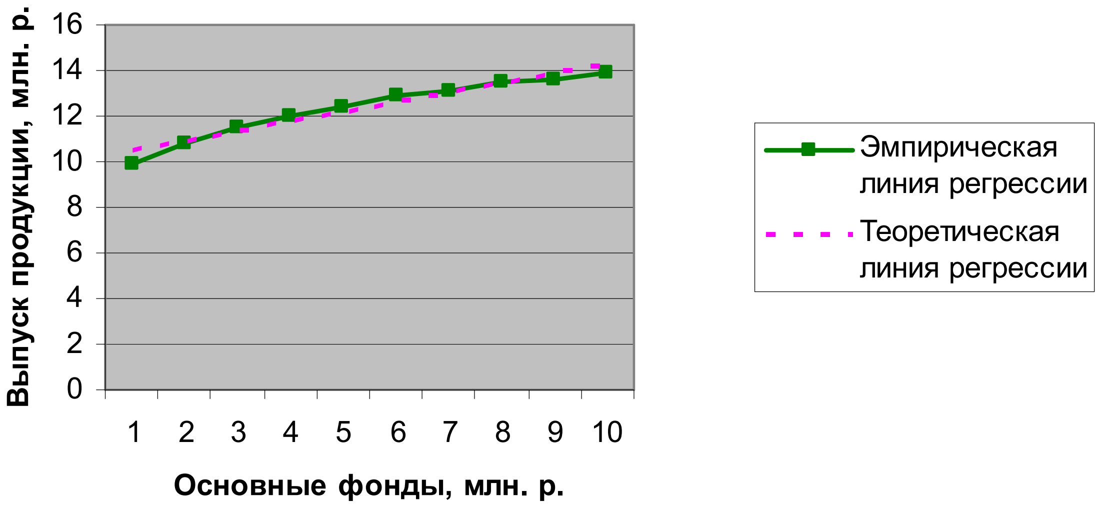


Рис. 8.1. Эмпирическая и теоретическая линия регрессии

8. Определим расчетные значения $(Y_X - \overline{Y})^2$ и $(Y_X - Y)^2$ в восьмом и девятом столбце таблицы 8.7, а также их итоги.

Таблица 8.7

№	X	Y	XY	X^2	Y^2	Y_X	$(Y_X - \overline{Y})^2$	$(Y_X - Y)^2$
1	330	9,9	3267	108900	98,01	10,4	3,8416	0,25
2	400	10,8	4320	160000	116,64	10,8	116,64	-
3	480	11,5	5520	230400	132,25	11,3	127,69	0,04
4	550	12,0	6600	302500	144	11,8	139,24	0,04
5	600	12,4	7440	360000	153,76	12,1	146,41	0,09
6	700	12,9	9030	490000	166,41	12,7	161,29	0,04
7	750	13,1	9825	562500	171,61	13,0	169	0,01
8	850	13,5	11475	722500	182,25	13,6	184,96	0,01
9	870	13,6	11832	756900	184,96	13,7	187,69	0,01
10	940	13,9	13066	883600	193,21	14,2	201,64	0,09
Итого:	6470	123,6	82375	4577300	1543,1	-	1438,4016	0,58

9. Определим расчетное значение критерия Стьюдента для линейного коэффициента корреляции по формуле (8.13): $S_r = 0,98\sqrt{\dfrac{10-2}{1-0,98^2}} = 13,929.$

10. Сравним расчетное значение критерия Стьюдента с табличным 2,306, которое определяем для α = 0,05, а также 8 степеней свободы (см. Приложение 2). Поскольку расчетное значение больше табличного, линейный коэффициент корреляции является значимым параметром взаимосвязи.

11. Определим расчетные значения критерия Стьюдента для каждого из параметров уравнения парной линейной регрессии по формулам (8.14) и (8.15): $S_0 = 8,381\dfrac{\sqrt{10-2}}{\sqrt{0,58/10}} = 98,430$ и $S_1 = 0,006\dfrac{\sqrt{10-2}}{\sqrt{0,58/10}} = 13,936.$

12. Сравним расчетные значения критерия Стьюдента с табличным 2,306, которое определяем также для α = 0,05, а также 8 степеней свободы (см. Приложение 2). Поскольку расчетные значения больше табличного, каждый из параметров уравнения парной линейной регрессии является значимым параметром взаимосвязи.

13. Определим расчетное значение критерия Фишера по формуле (8.16) $F = \dfrac{1438,4016}{0,58} \cdot \dfrac{10-2}{2-1}$ = 19840,02.

14. Сравним расчетное значение критерия Фишера с табличным 5,32, которое определяем для α = 0,05, а также 1 и 8 степеней свободы (см. Приложение 1). Поскольку расчетное значение больше табличного, можно сделать вывод о правильности выбора вида взаимосвязи, а также значимости построенного уравнения регрессии.

Контрольные задания 8

Вариант 1

1. Определить линейный коэффициент корреляции и построить уравнение регрессии, используя следующие статистические данные по торговым предприятиям:

Товарооборот, млн. р.	70	100	150	200	300	450	600	800	800	910
Издержки обращения, млн. р.	10,7	10,9	11,1	11,2	11,9	12,6	13,2	14,0	13,6	14,5

2. Определить ранговые коэффициенты корреляции по следующим статистическим данным:

Предприятие	Цена спроса на акцию, р./шт.	Цена предложения акции, р./шт.
А	83,6	60,6
Б	83,0	40,7
В	30,3	33,8
Г	13,6	22,1
Д	13,9	33,8
Е	26,5	33,8
Ж	18,1	20,9
З	28,7	35,9
И	19,8	21,7
К	19,0	22,5

Вариант 2

1. Определить линейный коэффициент корреляции и построить уравнение регрессии, используя следующие статистические данные по промышленным предприятиям:

Оборотные средства, млн. р.	120	171	150	178	222	271	254	278	300	352
Выпуск продукции, млн. р.	10,9	7,8	13,1	12,4	13,4	17,1	23,1	23,5	29,6	33,1

2. Определить ранговые коэффициенты корреляции по следующим статистическим данным:

Банк	Собственный капитал, тыс. руб.	Активы, млн. р.
А	10,8	28
Б	16,3	30
В	10,4	38
Г	14,9	33
Д	9,7	41
Е	9,0	46
Ж	7,7	44
З	6,9	49
И	5,7	52
К	5,2	50

Вариант 3

1. Определить линейный коэффициент корреляции и построить уравнение регрессии, используя следующие статистические данные по сельскохозяйственным предприятиям:

Выпуск продукции, млн. руб.	310	320	380	420	450	450	460	510	650	700
Балансовая прибыль, млн. руб.	15,9	20,1	22,2	26,5	35,4	42,3	40,5	48,8	50,4	55,5

2. Определить ранговые коэффициенты корреляции по следующим статистическим данным:

Семья	Доходы, тыс. руб.	Доля расходов на продукты питания, %

А	10,8	28
Б	16,3	30
В	10,4	38
Г	14,9	33
Д	9,7	41
Е	9,0	46
Ж	7,7	44
З	6,9	49
И	5,7	52
К	5,2	50

Вариант 4

1. Определить линейный коэффициент корреляции и построить уравнение регрессии, используя следующие статистические данные по коммерческим банкам:

Балансовая прибыль, млн. руб.	840	820	800	770	760	750	730	710	650	600
Выдано кредитов, млн. руб.	99,5	97,9	87,5	71,1	69,5	67,6	64,3	61,5	60,2	60,1

2. Определить ранговые коэффициенты корреляции по следующим статистическим данным:

Предприятие	Кредиторская задолженность, тыс. р.	Собственный капитал, млн. р.
А	3176	2496
Б	3066	1962
В	1888	783
Г	2941	900
Д	3221	943
Е	1997	1319
Ж	1865	1522
З	3220	1142
И	1194	658
К	518	311

Вариант 5

1. Определить линейный коэффициент корреляции и построить уравнение регрессии, используя следующие статистические данные по страховым компаниям:

Собственные средства, млн. р.	620	600	510	520	480	480	440	350	350	300
Страховые взносы, млн. р.	43,5	47,1	41,5	44,4	36,6	36,4	34,3	24,9	25,3	18,0

2. Определить ранговые коэффициенты корреляции по следующим статистическим данным:

Банк	Выплачено дивидендов, млн. р.	Чистая прибыль, млн. р.
А	23,4	228
Б	14,8	310
В	18,7	418
Г	17,5	331
Д	34,1	421
Е	35,1	496
Ж	17,6	239
З	18,7	419
И	27,2	299
К	13,2	317

Список литературы

1. Федеральный закон № 282-ФЗ от 29.11.2007 «Об официальном статистическом учете и системе государственной статистики в Российской Федерации»
2. Постановление Правительства РФ от 02.06.2008 №420 «Об утверждении положения о Федеральной службе государственной статистики»
3. Российский статистический ежегодник, 2009 : Стат. сб. / Федеральная служба государственной статистики (Росстат). – М., 2009.
4. Воронин В.Ф., Жильцова Ю.В. Статистика. – М.: Экономистъ, 2004. – 301 с.
5. Гусаров В.М., Проява С.М. Общая теория статистики. – М.: ЮНИТИ, 2008. – 207 с.
6. Едронова В.Н., Малафеева М.В. Общая теория статистики. – 2-е изд., перераб. – М.: Магистр, 2007. – 606 с.
7. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник. – М.: Финансы и статистика, 2008. – 423 с.
8. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. – 2-ое изд., испр. и доп. – М.: ИНФРА-М, 2005. – 416 с.
9. Ефимова М.Р., Рябцев В.Н. Общая теория статистики. – М.: Финансы и статистика, 2009. – 304 с.
10. Статистика: Учебник / Л.П.Харченко, В.Г. Ионин, В.В. Глинский и др.; Под ред. канд. экон. наук, проф. В.Г. Ионина. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2009. – 445 с.
11. Салин В.Н., Шпаковская Е.П. Социально-экономическая статистика. Учебник. – М.: Юристъ, 2008. – 378 с.
12. Теория статистики: Учеб. пособие для экон. вузов/Под. ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 2007. – 656 с.
13. Экономическая статистика: Учебник. – 3-е изд., перераб. и доп./ Под ред. проф. Ю.Н. Иванова. – М.: ИНФРА-М, 2010. – 736 с.

Приложение 1. Таблица значений критерия Фишера

(уровень значимости 0,05)

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	161,5	200	215,7	224,6	230,2	234	238,9	243,9	249	254,3
2	18,5	19	19,16	19,25	19,3	19,33	19,37	19,41	19,45	19,5
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,9	2,71
10	4,96	4,1	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,2	3,09	2,95	2,79	2,61	2,4
12	4,75	3,88	3,49	3,26	3,11	3	2,85	2,69	2,5	2,3
13	4,67	3,8	3,41	3,18	3,02	2,92	2,77	2,6	2,42	2,21
14	4,6	3,74	3,34	3,11	2,96	2,85	2,7	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,9	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,2	2,96	2,81	2,7	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,9	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,1	2,87	2,71	2,6	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,3	3,44	3,05	2,82	2,66	2,55	2,4	2,23	2,03	1,78
23	4,28	3,42	3,03	2,8	2,64	2,53	2,38	2,2	2	1,76
24	4,26	3,4	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,6	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,3	2,13	1,93	1,67
28	4,2	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,7	2,54	2,43	2,28	2,1	1,9	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2	1,79	1,52
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,4	2,29	2,13	1,95	1,72	1,44
60	4	3,15	2,76	2,52	2,37	2,25	2,1	1,92	1,7	1,39
70	3,98	3,13	2,74	2,5	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,1	2,71	2,47	2,32	2,2	2,04	1,86	1,64	1,28
100	3,94	3,09	2,7	2,46	2,3	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,6	1,21
150	3,9	3,06	2,66	2,43	2,27	2,16	2	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,8	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79.	1,55	1,1
400	3,86	3,02	2,63	2,4	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3	2,61	2,38	2,22	2,1	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,6	2,37	2,21	2,09	1,94	1,75	1,52	

Приложение 2. Таблица значений критерия Стьюдента

ν	α			ν	α		
	0,1	0,05	0,01		0,1	0,05	0,01
1	6,314	12,706	63,66	18	1,734	2,101	2,878
2	2,92	4,3027	9,925	19	1,729	2,093	2,861
3	2,353	3,1825	5,841	20	1,725	2,086	2,845
4	2,132	2,7764	4,604	21	1,721	2,08	2,831
5	2,015	2,5706	4,032	22	1,717	2,074	2,819
6	1,943	2,4469	3,707	23	1,714	2,069	2,807
7	1,895	2,3646	3,5	24	1,711	2,064	2,797
8	1,86	2,306	3,355	25	1,708	2,06	2,787
9	1,833	2,2622	3,25	26	1,706	2,056	2,779
10	1,813	2,2281	3,169	27	1,703	2,052	2,771
11	1,796	2,201	3,106	28	1,701	2,048	2,763
12	1,782	2,1788	3,055	29	1,699	2,045	2,756
13	1,771	2,1604	3,012	30	1,697	2,042	2,75
14	1,761	2,1448	2,977	40	1,684	2,021	2,705
15	1,753	2,1315	2,947	60	1,671	2	2,66
16	1,746	2,1199,	2,921	120	1,658	1,98	2,617
17	1,74	2,1098	2,898	∞	1,645	1,96	2,576

Приложение 3. Таблица значений критерия Пирсона

ν \ α	0,990	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	0,0002	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0201	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,1148	0,3519	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	12,8382
4	0,2971	0,7107	1,0636	1,9226	3,3567	5,3853	7,7794	9,4877	11,1433	13,2767	14,8603
5	0,5543	1,1455	1,6103	2,6746	4,3515	6,6257	9,2364	11,0705	12,8325	15,0863	16,7496
6	0,8721	1,6354	2,2041	3,4546	5,3481	7,8408	10,6446	12,5916	14,4494	16,8119	18,5476
7	1,2390	2,1674	2,8331	4,2549	6,3458	9,0372	12,0170	14,0671	16,0128	18,4753	20,2777
8	1,6465	2,7326	3,4895	5,0706	7,3441	10,2189	13,3616	15,5073	17,5346	20,0902	21,9550
9	2,0879	3,3251	4,1682	5,8988	8,3428	11,3888	14,6837	16,9190	19,0228	21,6660	23,5894
10	2,5582	3,9403	4,8652	6,7372	9,3418	12,5489	15,9872	18,3070	20,4832	23,2093	25,1882
11	3,0535	4,5748	5,5778	7,5841	10,3410	13,7007	17,2750	19,6751	21,9201	24,7250	26,7569
12	3,5706	5,2260	6,3038	8,4384	11,3403	14,8454	18,5494	21,0261	23,3367	26,2170	28,2995
13	4,1069	5,8919	7,0415	9,2991	12,3398	15,9839	19,8119	22,3620	24,7356	27,6883	29,8195
14	4,6604	6,5706	7,7895	10,1653	13,3393	17,1169	21,0641	23,6848	26,1190	29,1412	31,3194
15	5,2294	7,2609	8,5468	11,0365	14,3389	18,2451	22,3071	24,9958	27,4884	30,5779	32,8013
16	5,8122	7,9617	9,3122	11,9122	15,3385	19,3689	23,5418	26,2962	28,8454	31,9999	34,2672
17	6,4078	8,6718	10,0852	12,7919	16,3382	20,4887	24,7690	27,5871	30,1910	33,4087	35,7185
18	7,0149	9,3905	10,8649	13,6753	17,3379	21,6049	25,9894	28,8693	31,5264	34,8053	37,1565
19	7,6327	10,1170	11,6509	14,5620	18,3377	22,7178	27,2036	30,1435	32,8523	36,1909	38,5823
20	8,2604	10,8508	12,4426	15,4518	19,3374	23,8277	28,4120	31,4104	34,1696	37,5662	39,9969
21	8,8972	11,5913	13,2396	16,3444	20,3372	24,9348	29,6151	32,6706	35,4789	38,9322	41,4011
22	9,5425	12,3380	14,0415	17,2396	21,3370	26,0393	30,8133	33,9244	36,7807	40,2894	42,7957
23	10,1957	13,0905	14,8480	18,1373	22,3369	27,1413	32,0069	35,1725	38,0756	41,6384	44,1813
24	10,8564	13,8484	15,6587	19,0373	23,3367	28,2412	33,1962	36,4150	39,3641	42,9798	45,5585
25	11,5240	14,6114	16,4734	19,9393	24,3366	29,3389	34,3816	37,6525	40,6465	44,3141	46,9279
26	12,1982	15,3792	17,2919	20,8434	25,3365	30,4346	35,5632	38,8851	41,9232	45,6417	48,2899
27	12,8785	16,1514	18,1139	21,7494	26,3363	31,5284	36,7412	40,1133	43,1945	46,9629	49,6449
28	13,5647	16,9279	18,9392	22,6572	27,3362	32,6205	37,9159	41,3371	44,4608	48,2782	50,9934
29	14,2565	17,7084	19,7677	23,5666	28,3361	33,7109	39,0875	42,5570	45,7223	49,5879	52,3356
30	14,9535	18,4927	20,5992	24,4776	29,3360	34,7997	40,2560	43,7730	46,9792	50,8922	53,6720

Виктор Александрович **Матвеев**

СТАТИСТИКА

Учебно-методическое пособие

Государственное образовательное учреждение высшего
профессионального образования
«Нижегородский государственный университет им. Н.И.Лобачевского».
603950, Нижний Новгород, пр. Гагарина, 23.

Формат 60x84 1/16.
Бумага офсетная. Печать офсетная. Гарнитура Таймс.
Усл. печ. л. 4,0. Уч.-изд. л..
Заказ № . Тираж 300 экз.

Отпечатано в типографии Нижегородского госуниверситета
им. Н.И. Лобачевского
603600, г. Нижний Новгород, ул. Большая Покровская, 37
Лицензия ПД № 18-0099 от 14.05.01