Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Комсомольский-на-Амуре государственный технический университет»

**А. В. Инзарцев**

# СТАТИСТИКА В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Утверждено в качестве учебного пособия
Ученым советом Федерального государственного бюджетного
образовательного учреждения высшего профессионального образования
«Комсомольский-на-Амуре государственный технический университет»

Комсомольск-на-Амуре
2013

УДК 311:004(07)
ББК 60.6я7
И631

***Рецензенты:***

В. А. Ханов, канд. техн. наук, зам. начальника отдела информационных технологий ОАО «Амурский судостроительный завод»;
кафедра «Информационные системы, компьютерные технологии и физика» ФГБОУ ВПО «Амурский гуманитарно-педагогический государственный университет», зав. кафедрой канд. физ.-мат. наук, доцент М. А. Савунов

**Инзарцев, А. В.**
И631 Статистика в информационных системах : учеб. пособие / А. В. Инзарцев. – Комсомольск-на-Амуре : ФГБОУ ВПО «КнАГТУ», 2013. – 115 с.

ISBN 978-5-7765-1060-1

Рассмотрены общие вопросы теории статистики и экономической статистики, знание которых необходимо для проведения анализа массивов информации, содержащейся в информационных системах. Подробно излагаются разделы, связанные с группировкой статистической информации, анализом больших массивов статистической информации, однородности совокупностей и выборочными исследованиями. Представлены типовые примеры с решениями, методические указания к выполнению расчётно-графического задания.

Пособие предназначено для студентов, обучающихся по основной образовательной программе бакалавров по направлениям 230700 – «Прикладная информатика», 080500 – «Бизнес-информатика».

УДК 311:004(07)
ББК 60.6я7

ISBN 978-5-7765-1060-1

## ОГЛАВЛЕНИЕ

# ВВЕДЕНИЕ

Решение экономических задач на всех уровнях хозяйствования невозможно без использования статистики. Статистика обеспечивает сбор, научную обработку, обобщение и анализ разнообразной информации, характеризующей состояние экономики, культуры, народонаселения и других аспектов жизни страны.

Термин «***статистика***» употребляется в четырёх значениях:

1) *наука*, изучающая с количественной стороны массовые явления и их закономерности;

2) *отрасль практической деятельности людей* по сбору, обработке и анализу массовых данных о различных явлениях и процессах;

3) *совокупность числовых данных*, характеризующих состояние массовых явлений и процессов в том числе экономической жизни; статистические данные, накапливаемые в информационных системах организаций, данные, публикуемые в сборниках, справочниках, размещённые на сайтах;

4) *статистические методы*, применяемые при сборе, анализе и интерпретации данных.

В процессе развития статистики как науки из неё выделились такие дисциплины, как: теория статистики, экономическая статистика, отраслевые статистики, социально-демографическая статистика и её отрасли.

*Теория статистики* изучает и разрабатывает наиболее общие приёмы и методы исследования, вводит основные понятия и категории статистической науки, разрабатывает общую методологию статистических исследований.

*Экономическая статистика и отраслевые статистики* (статистика промышленности, сельского хозяйства, транспорта, статистика цен, финансов и т.д.) разрабатывают и используют для анализа системы статистических показателей, характеризующих как национальную экономику в целом, так и состояние соответствующих отраслей.

Знание статистики необходимо специалисту для сбора информации и принятия решений в условиях, когда анализируемые явления подвержены влиянию множества случайных факторов, для анализа происходящих в экономике процессов, в финансовом менеджменте, прогнозировании.

Поскольку теория статистики является методологической основой экономической и отраслевых статистик, то в данном учебном пособии основное внимание уделено именно вопросам теории статистики. Углублённые знания в этой области позволят легко ориентироваться в информационных системах, предназначенных для сбора и обработки любой статистической информации, независимо от конкретной области применения.

# 1. ОСНОВНЫЕ ПОНЯТИЯ

## 1.1. Предмет и метод статистики

Статистика как наука имеет свой предмет исследования.

*Предметом* изучения статистики являются, как правило, *массовые явления*, состоящие из множества отдельных элементов или фактов. Каждый элемент такого множества обладает индивидуальными признаками, отличающими его от других элементов массива. В то же время элементы массива обладают и общими (сходными) признаками, позволяющими объединить их в один массив. Так, например, изучая выпуск хлебобулочной продукции в крае, рассматривают множество хлебопекарных предприятий, каждое из которых имеет свои индивидуальные признаки, такие как: ассортимент выпускаемой продукции, численность работников, стоимость основных фондов и т.д. В то же время все эти предприятия, как единицы массива, имеют общий признак – это хлебопекарные предприятия (а не сельскохозяйственные, строительные и т.д.). При этом показатели каждого из элементов изучаемого массива полностью или частично независимы и выражаются некоторыми количественными величинами. Поскольку количественная сторона явления с течением времени может меняться (например, изменяется число работников предприятия), то количественную сторону явлений изучают в конкретных условиях места и времени.

В статистике разработаны и применяются специфические приемы, совокупность которых образует *методологию статистики* (методы массовых наблюдений, группировок, обобщающих показателей, динамических рядов, индексный метод и др.). Применение того или иного конкретного метода зависит от решаемой задачи и от характера исходной информации.

Статистические методы используются в комплексе, поскольку типичное экономико-статистическое исследование состоит из трех основных стадий:

1) сбор первичной статистической информации;
2) статистическая сводка и обработка первичной информации;
3) анализ и интерпретация статистической информации.

На первой стадии статистического исследования применяется метод массового статистического наблюдения, обеспечивающий полноту и представительность (репрезентативность) полученной первичной информации.

На второй стадии собранная информация подвергается обработке методом статистических группировок, позволяющим выделить в изучаемой совокупности основные социально-экономические типы; совершается переход от единичных фактов к характеристике данных, объединенных в группы величин.

На третьей стадии проводится анализ статистической информации, позволяющий выявить причинно-следственные связи изучаемых явлений и процессов, определить влияние и взаимодействия различных факторов, прогнозировать дальнейшее развитие изучаемых явлений.

## 1.2. Основные категории статистики

При изучении своего предмета статистика использует определенные категории, т.е. понятия, отражающие существенные, общие свойства и основные отношения явлений действительности.

Объектом конкретного статистического исследования является статистическая совокупность.

***Статистическая совокупность*** – это множество единиц (объектов, явлений), объединенных в соответствии с задачей исследования единой качественной основой. Например, совокупность предприятий (строительных, промышленных, сельскохозяйственных), производящих однотипную продукцию, но различающихся между собой объемами производства, численностью работающих и финансовым ресурсами. Единой качественной основой в данном случае будет тип производимой продукции (строительная, промышленная или сельскохозяйственная), в то время как по другим признакам, таким как объем производства, численность работающих, финансовые ресурсы, совокупности будут разнородны.

Единая качественная основа формируемой статистической совокупности, обеспечивающая её однородность, а также состав статистической совокупности устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями.

Статистическая совокупность образуется из множества неделимых первичных элементов, называемых ***единицами совокупности***. Так, например, при переписи населения единицей совокупности будет отдельный человек, если же изучаются совокупность семей, то единицами совокупности будут отдельные семьи.

Отдельные единицы статистической совокупности характеризуются общими для них свойствами, называемыми в статистике признаками.

***Признак*** – характерное свойство изучаемого явления, отличающее его от других явлений. Например, единица статистической совокупности "сельскохозяйственное предприятие" может быть охарактеризована следующими признаками: размер посевных площадей, объем произведённой продукции, численность работников, стоимость основных фондов и т.д. Поскольку входящие в рассматриваемую статистическую совокупность сельскохозяйственные предприятия имеют различный размер посевных площадей, различную численность работников, то можно говорить, что эти признаки варьируют или имеют некоторую вариацию.

***Вариация*** – различия в значениях рассматриваемого признака у отдельных единиц, составляющих данную совокупность. Значения рассматриваемого признака у отдельных единиц совокупности вследствие вариации могут быть различными. Конкретные значения признака, которые он принимает у отдельных единиц рассматриваемой совокупности, называются ***вариантами***. Обозначим варианты как $x_1$, $x_2$, … $x_i$ … $x_n$. Например, каждый из работников некоторого предприятия имеет такой признак, как стаж работы. Признак этот является варьирующим и имеет у трёх различных работников значения 1 год, 5 лет, 7,5 лет, т.е. $x_1 = 1$, $x_2 = 5$, $x_3 = 7{,}5$.

Варьирующий признак называется ***количественным***, если его варианты выражаются числовыми значениями (возраст, стаж работы, заработная плата и т.п.). Варьирующий признак называется ***атрибутивным (качественным),*** если он не имеет числового выражения и выражается смысловым понятием (национальность, профессия, цвет и т.д.).

Если варианты признака могут принимать только одно из двух противоположных значений ("да" или "нет"), то признак называют ***альтернативным.*** Так, например, продукция может быть годной ("да") или бракованной, негодной ("нет").

Количественные признаки могут быть ***дискретными***, когда признак принимает отдельные фиксированные значения (чаще всего – целые), и ***непрерывными***, когда признак принимает любые, в том числе дробные, значения.

Статистическая совокупность состоит из большого числа отдельных единиц. Для установления общих свойств единиц совокупности, изучения имеющиеся взаимосвязей и закономерностей развития используются статистические показатели.

***Статистический показатель*** – это количественная оценка некоторого свойства изучаемой совокупности в целом. Статистический показатель характеризует всю совокупность в целом, в отличие от индивидуальных значений, которые называются признаками. Например, средний размер начисляемых пенсий по стране – статистический показатель, а размер пенсии конкретного пенсионера – признак.

Статистический показатель может рассчитываться как сумма абсолютных значений признака (объём нефтедобычи всеми предприятиями страны, численность безработных в крае), как среднее значение признака по совокупности (средняя зарплата, средняя урожайность с гектара) или как относительная величина (индексы цен, темпы роста промышленного производства).

Статистический показатель рассчитывается в условиях конкретного места и времени и даёт не только качественную, но и количественную характеристику социально-экономических процессов и явлений.

Так, например, за 2010 г. в г. Комсомольске-на-Амуре выпущено 1319 т мороженого. Качественная сторона этого показателя – выпуск мороженого, а количественная сторона выражается числом 1319 и единицей измерения (тонны). Рассмотренный показатель является абсолютным и выражается именованными числами. Статистический показатель указывает на конкретное место («в г. Комсомольске-на-Амуре») и время («за 2010 г.»).

Важнейшей категорией статистики является статистическая закономерность. Закономерностью вообще принято называть повторяемость, последовательность и порядок изменений в явлениях.

Статистические показатели можно условно подразделить на *учётно-оценочные* (первичные, объемные, количественные) и *аналитические* (вторичные, производные, качественные).

Учётно-оценочные показатели характеризуют либо общее число единиц совокупности, либо сумму значений какого-либо признака (объём розничного товарооборота, численность пенсионеров). Аналитические показатели обычно выражаются средними или относительными величинами и используются для анализа статистической информации (темп прироста производства продукции, индекс инфляции).

Для комплексного изучения статистических совокупностей и выявления статистических закономерностей применяются системы статистических показателей, представляющие собой совокупность взаимосвязанных показателей, охватывающих все стороны изучаемого явления. Системы показателей весьма разнообразны и зависят от решаемых задач и сложности изучаемых объектов и явлений.

При массовом наблюдении действует закон больших чисел, согласно которому совместное действие множества случайных факторов приводит к результату, практически не зависящему от случая. Таким образом, количественные закономерности массовых явлений отчётливо проявляются лишь в достаточно большом их числе. Так, например, реально располагаемые денежные доходы населения в Российской Федерации в 2012 г. выросли на 4,4 %, но это не означает, что доходы выросли у каждого без исключения гражданина России.

Закономерность – важная категория статистики. Закономерностью принято называть повторяемость, последовательность и порядок изменений в явлениях. Закономерность, выявленная на основе массового наблюдения, называют ***статистической закономерностью***. Она проявляется не в каждом отдельном, индивидуальном явлении, а в массе однородных явлений, при обобщении данных статистической совокупности, т.е. в среднем.

# 2. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

## 2.1. Понятие о статистической информации

Слово "информация" в переводе с латинского языка означает "осведомление, доведение сведений о чем-либо".

***Статистическая информация*** (статистические данные) – первичный статистический материал, формирующийся в процессе статистического наблюдения, который затем подвергается систематизации, сводке, обработке, анализу и обобщению.

Достоверность информации является необходимым условием для эффективного использования статистики в различных сферах деятельности. Трудоемкая работа по обеспечению необходимых для этих целей данных является важной государственной задачей, которую выполняет государственная (официальная) статистика. При этом микроданные должны держаться в тайне, в то время как макроданные должны быть доступны для широкого круга пользователей.

## 2.2. Статистическое наблюдение

***Статистическое наблюдение*** – первая стадия статистического исследования. Она представляет собой научно-организованную работу по сбору массовых первичных данных об изучаемых явлениях и процессах общественной жизни. Статистическое наблюдение должно быть массовым, систематическим и проводиться по заранее разработанным плану и программе.

План статистического исследования включает вопросы методологии, организации, техники и сбора информации, контроля её достоверности и оформления итоговых результатов. В плане статистического наблюдения указываются время и вместо наблюдения. При этом устанавливаются *критический момент* (дата) или интервал времени и определяется срок (период) наблюдения.

Статистические показатели могут характеризовать исследуемое явление либо на определенный момент времени, либо за определенный период времени. Так, например, численность населения города может быть представлена на определенный момент (на начало года), а данные о количестве родившихся в городе детей могут быть представлены только за определенный интервал времени (год, квартал).

Результаты всего статистического исследования во многом зависят от достоверности первичных данных статистического наблюдения, т.е. от их соответствия фактическому положению. На достоверность данных влияет множество причин, включая тщательность подготовки статистического наблюдения, содержание анкет, качество инструкций по их заполнению, личную заинтересованность участников наблюдения в его результатах и т.д.

Полученные от отдельных единиц наблюдения (предприятий, граждан и т.д.) данные должны быть сопоставимы друг с другом, что делает возможным их последующее обобщение. Сопоставимость данных обеспечивается единством сроков наблюдения (например, численность населения определяется на начало календарного года), методов регистрации данных, программ наблюдения.

### *2.2.1. Программно-методологические вопросы статистического наблюдения*

При подготовке к проведению статистического наблюдения решается ряд вопросов. К группе программно-методологических вопросов статистического наблюдения относятся вопросы, связанные с установлением цели наблюдения, определением объекта и единицы наблюдения, разработкой программы наблюдения, а также выбором вида и способа наблюдения.

Основной целью статистического наблюдения является получение достоверной информации для выявления статистических закономерностей развития явлений и процессов. В зависимости от цели статистического исследования выбирается объект статистического наблюдения.

***Объект статистического наблюдения*** – некоторая статистическая совокупность, в которой протекают исследуемые социально-экономические процессы и явления.

Так, например, при проведении Всероссийской переписи населения 2010 г. объектом наблюдения являлось всё население, постоянно (обычно) проживающие в Российской Федерации.

Кроме объекта статистического наблюдения, необходимо также установить единицу наблюдения.

***Единица наблюдения*** – составной элемент объекта статистического наблюдения, являющийся носителем признаков, подлежащих регистрации. При переписи населения 2010 г. объектом наблюдения являлось население страны, а единицей наблюдения – домохозяйства, т.е. совокупность лиц, проживающих в одном жилом помещении или его части, совместно обеспечивающих себя необходимыми средствами к существованию, объединяя полностью или частично свои доходы.

Исходя цели и задач статистического наблюдения, отбираются основные признаки, характеризующие объект наблюдения, т.е. разрабатывается программа наблюдения.

***Программа статистического наблюдения*** - перечень показателей, подлежащих регистрации. Иначе говоря, это перечень вопросов, на которые должны быть получены достоверные ответы по каждой из единиц наблюдения. Так, например, программа переписи населения содержит вопросы о семейном положении, возрасте, образовании, жилищных условиях и др.

### *2.2.2. Организационные формы, виды и способы наблюдения*

Организационными формами статистического наблюдения являются отчетность, специально организованные статистические наблюдения и регистры.

***Статистическая отчетность*** – основная форма статистического наблюдения за деятельностью предприятий и организаций. Органы государственной статистики получают информацию из организаций в виде установленных в законодательном порядке отчетных документов, подписанных ответственными лицами. Отчётность предоставляется в установленные адреса и сроки в обязательном порядке.

***Специально организованное статистическое наблюдение*** проводится с целью получения сведений, по которым отсутствует отчётность. Сбор необходимых сведений проводится с помощью переписей, единовременных учетов и обследований (сельскохозяйственная перепись, перепись населения, социологические исследования).

Регистровая форма наблюдения получила распространение с развитием вычислительной техники и информационных систем. ***Регистровое наблюдение*** предполагает непрерывное статистическое наблюдение за долговременными процессами с использованием регистров. *Регистр* – поименованный и постоянно уточняемый перечень единиц наблюдения, в котором каждая единица наблюдения характеризуется совокупностью показателей. При этом значения некоторых показателей остаются неизменными в течение всего времени наблюдения, значения других показателей обновляются по мере их изменения либо с заранее известным периодом обновления. Примерами таких регистров являются Единый государственный реестр юридических лиц (ЕГРЮЛ), содержащий сведения о создании, реорганизации и ликвидации юридических лиц или Единый государственный реестр прав на недвижимое имущество и сделок с ним (ЕГРП) – государственный информационный ресурс, содержащий данные об объектах недвижимого имущества на территории РФ, данные о существующих и прекращенных правах на эти объекты, а также сведения о правообладателях, наличии обременений и т.д.

Статистическое наблюдение подразделяется на группы в зависимости от времени регистрации данных и от степени охвата единиц совокупности.

По ***времени регистрации фактов*** бывает *непрерывное* (текущее наблюдение), например, отчетность, *периодическое* (регистрация по мере надобности) и *единовременное*. Примером текущего наблюдения является регистрация дорожно-транспортных происшествий, примерами периодического наблюдения служат переписи населения, которые проводятся раз в 10 лет.

По ***степени охвата единиц совокупности*** различают сплошное и несплошное наблюдение.

***Сплошное наблюдение*** охватывает все без исключения единицы изучаемой совокупности. К такому виду наблюдения относятся переписи населения или сбор данных в форме отчетности.

***Несплошное наблюдение*** охватывает не все единицы изучаемой совокупности, а только заранее обусловленную их часть. Переход от сплошного наблюдения к несплошному позволяет сократить затрачиваемое время, снизить затраты более детально изучить единицы совокупности и т.д. Можно выделить несколько видов несплошного наблюдения.

При *выборочном* наблюдении характеристика всей совокупности даётся по некоторым её единицам, отобранным в случайном порядке. При правильной организации такое наблюдение даёт достаточно точные результаты.

Метод *основного массива* состоит в том, что сбор данных осуществляется только по тем единицам совокупности, о которых заранее известно, что они делают основной вклад в характеристику изучаемого явления. Те единицы совокупности, которые не вносят существенного вклада в общую характеристику совокупности, исключаются из наблюдения. С использованием этого метода, например, изучается работа городских рынков.

При проведении *монографического наблюдения* подробному описанию и изучению подвергаются отдельные единицы, типичные для данной совокупности. С помощью такого наблюдения получают данные, дополняющие и углубляющие результаты массового наблюдения. Обычно монографическое наблюдение проводится в целях выявления и изучения имеющихся или намечающихся тенденций развития. С помощью монографического наблюдения, например, получают данные о бюджетах отдельных, специальным образом отобранных, семей.

При *анкетном обследовании* сбор данных основан на принципе добровольного заполнения адресатами анкет. Недостатком этого простого и удобного метода является невозможность проверки достоверности полученных данных.

Наибольшее признание и распространение в статистической практике получило выборочное наблюдение.

# 3. СТАТИСТИЧЕСКАЯ СВОДКА И ГРУППИРОВКА

## 3.1. Понятие сводки и группировки статистических данных

Полученная в процессе статистического наблюдения информация представляет собой большое количество первичных, разрозненных сведений об отдельных единицах объекта исследования. Поскольку этот материал не позволяет непосредственно выявить определенные статистические закономерности и охарактеризовать сущность социально-экономических явлений, то необходимо перейти ко второй стадии статистического исследования – сводке и группировке.

***Статистическая сводка*** – систематизация единичных фактов, позволяющая перейти к обобщающим показателям, характеризующим типичные черты и закономерности, присущие изучаемому явлению в целом.

Различают простую и сложную сводку.

***Простой сводкой*** называется операция по подсчёту общих итогов по совокупности единиц наблюдения.

***Сложная сводка*** представляет собой комплекс операций, включающих группировку единиц наблюдения, подсчёт итогов по каждой группе и по всему объекту в целом.

При проведении сводки отдельные единицы статистической совокупности объединяются в группы при помощи метода группировок.

***Группировкой*** называют процесс расчленения статистической совокупности на группы по определённым, существенным для них признакам. Например, группировка работников по размеру начисленной заработной платы, группировка предприятий по формам собственности и т.д.

На основе метода статистических группировок в дальнейшем рассчитываются сводные показатели по группам, что позволяет сравнивать группы между собой, анализировать причины различий между группами, изучать взаимосвязи между признаками. Таким образом, группировка является основой для последующей сводки и анализа данных.

С помощью метода группировок решаются задачи:

- выделения социально-экономических типов явлений;
- изучения структуры явления и структурных сдвигов, происходящих в нем;
- выявления и изучения связей и зависимостей между отдельными признаками явления.

В зависимости от решаемых задач группировки делятся на: типологические, структурные и аналитические.

Признаки, по которым производится распределение единиц изучаемой совокупности на группы, называются ***группировочными признаками***, ***или основанием группировки***. Правильный выбор группировочных признаков возможен только после четкой формулировки задачи статистического исследования.

С помощью *типологических группировок* решаются задачи выявления и характеристики основных социально-экономических типов. При этом совокупность разделяется на классы, социально-экономические типы, однородные группы единиц в соответствии с правилами научной группировки. Примером типологической группировки может служить группировка хозяйствующих субъектов по организационно-правовой форме (группы обществ с ограниченной ответственностью, акционерных обществ, государственных предприятий, учреждений и т.д.).

Выделенные с помощью типологической группировки типы явлений могут быть изучены с точки зрения их структуры и состава. С помощью *структурной группировки* разделяют полученные группы на однородные совокупности, характеризующие их структуру по какому-либо варьирующему признаку. Примерами структурных группировок являются группировки акционерных обществ по размеру уставного капитала или группировка учреждений по числу работников.

Анализируя структурные группировки, взятые за ряд периодов или моментов времени, можно выявить изменение структуры изучаемых явлений во времени – структурные сдвиги. Структурные сдвиги отражают закономерности развития общественных явлений и позволяют лучше понять происходящие социально-экономические процессы.

Для выявления взаимосвязи между изучаемыми явлениями и их признаками используются *аналитические группировки*. При их построении можно установить связь между двумя или несколькими признаками. Основанием аналитической группировки является один или несколько факторных признаков, а каждая выделенная группа характеризуется средними значениями результативного признака. По результатам группировки можно установить, как факторные признаки влияют на результативные признаки.

Например, если имеется достаточно большое количество данных о количестве комнат в каждой из продаваемых на рынке квартир и рыночной стоимости этих квартир, можно установить влияние числа комнат на стоимость квартиры. Группируя квартиры по числу комнат и вычисляя среднюю стоимость квартиры по каждой из групп, можно определить, как влияет число комнат на среднюю рыночную стоимость квартиры.

В зависимости от цели статистического исследования группировки могут производиться по одному или нескольким признакам. Если группы образуются по одному признаку, группировка называется *простой*, группировка по двум или нескольким признакам, при которой каждая из групп делится ещё и на подгруппы, называется *сложной*.

## 3.2. Образование групп и определение интервалов группировки

Группировочный признак определяется в зависимости от цели исследования. В качестве основания группировки следует использовать существенные признаки, которые могут быть определены на основании теоретического анализа. При этом важно правильно определить число групп, на которые будет разбита статистическая совокупность.

При построении группировки по атрибутивному (качественному) признаку количество групп, как правило, будет совпадать с количеством градаций, состоянием группировочного признака. Так, например, при исследовании производительности труда на однотипных предприятиях в зависимости от формы собственности, вся совокупность предприятий будет разбита на три группы: частные предприятия, государственные предприятия, муниципальные предприятия.

В случаях, когда атрибутивный признак имеет большое число градаций (состояний) и образовать соответствующее число групп не представляется целесообразным, разрабатывают и используют классификацию. ***Классификацией*** называют систематизированное распределение явлений и объектов на определенные группы, классы на основе их сходства и различия. Классификации служат статистически нормативом, они устанавливаются органами статистики и широко используются в информационных системах. Классификации устойчивы, они остаются неизменными в течение длительного времени, однако при необходимости с течением времени в них вносятся дополнения и изменения. Примерами классификаторов могут служить Общероссийский классификатор основных фондов (ОКОФ, утвержден Постановлением Госстандарта РФ от 26 декабря 1994 г. № 359) или Общероссийский классификатор управленческой документации (ОКУД, утверждён Постановлением Госстандарта РФ от 30 декабря 1993 г. № 299).

При построении группировки по альтернативному признаку, принимающему только одно из двух противоположных значений ("да" или "нет"), получим только две группы. Например, группировка населения по половому признаку: мужчины и женщины.

При построении группировки по количественному признаку принимают во внимание численность изучаемой совокупности и степень колеблемости группировочного признака. Также имеет значение, является ли группировочный признак дискретной или непрерывной величиной. Признак является *дискретной* величиной, если может принимать только фиксированные (обычно – целые) значения, и *непрерывной* величиной, если в пределах вариации может принимать любые значения, отличающиеся друг от друга на сколь угодно малую величину.

Если варьирующий признак является дискретной величиной и может принимать ограниченное число значений, то количество групп, как правило, принимают равным количеству возможных значений признака. Например, при группировке квартир по числу комнат.

Если варьирующий признак является непрерывной величиной, или дискретной величиной, принимающей большое число значений (например, население города), то весь диапазон изменения признака разбивается на интервалы.

Под ***интервалом*** будем понимать количественное значение, отделяющее одну группу от другой, т.е. интервал определяет количественные границы групп. Как правило, величина интервала равна разности между максимальным и минимальным возможными значениями признака в каждой группе.

При решении вопроса о числе групп и величине интервала необходимо исходить из целей исследования, количества единиц в исследуемой совокупности, значений изучаемого признака и т.д. Большое число групп позволяет более точно воспроизвести характер исследуемой совокупности. Однако чрезмерное количество групп затрудняет выявление закономерностей. Кроме того, при увеличении числа групп будет уменьшаться численность каждой из них, в результате малочисленные группы перестанут быть представительными.

При автоматизированной обработке данных группировка производится с помощью стандартных процедур. При этом для расчёта оптимального количества групп может быть использована формула американского ученого Стерджесса:

$$n = 1 + 3{,}322\, lg\, N,$$

где $n$ – число групп; $N$ – число единиц исследуемой совокупности.

Формула Стерджесса даёт хорошие результаты при условии, что распределение единиц совокупности по группировочному признаку близко к нормальному, а интервалы в группах равны друг другу.

Интервалы группировки в зависимости от их величины могут быть равными и неравными.

Неравные интервалы делятся на прогрессивно возрастающие, прогрессивно убывающие, произвольные и специализированные. Неравные интервалы применяются в случаях, когда значения признака варьируют в широком диапазоне, а количественные изменения размера признака имеют неодинаковые значения в низших и высших по размеру признака группах.

Так, например, города по численности населения принято делить на группы с неравными интервалами (тыс. чел.): менее 50, 50 – 100, 100 – 250, 250 – 500, 500 – 1000, 1000 и более.

Равные интервалы применяются в случаях, когда значения признака варьируют в сравнительно узких границах, а распределение носит достаточно равномерный характер. В этом случае величина равного интервала $i$ будет:

$$i = \frac{x_{\max} - x_{\min}}{n}, \tag{3.1}$$

где $x_{max}$, $x_{min}$ – соответственно наибольшее и наименьшее значения признака в совокупности; $n$ – число образуемых групп.

Так, например, если месячная заработная плата 30 работников организации колеблется в пределах от 18 000 до 32 500 р., то группировку по размеру зарплаты проведём в следующем порядке.

1) Рассчитываем необходимое число групп:

$$N = 1 + 3{,}322\, lg\, 30 = 5{,}9 \approx 6.$$

2) Определяем величину интервала, р.:

$$i = \frac{32500 - 18000}{6} = 2417.$$

Если в результате деления получится не круглое число, то для удобства рационально его округлить, при этом, как правило, в бо́льшую сторону. Округлим длину интервала с 2 417 до 2 500 р.

3) Прибавляем к минимальному значению признака, равному 18 000 р., величину интервала 2 500 р. и получаем верхнюю границу первой группы: 18 000 + 2 500 = 20 500.

4) Прибавляем величину интервала к верхней границе первой группы и получаем верхнюю границу второй группы: 20 500 + 2 500 = 23 000.

Продолжая процесс, найдём все группы работников по размеру заработной платы, р.: 18 000 – 20 500; 20 500 – 23 000; 23 000 – 25 500; 25 500 – 28 000; 28 000 – 30 500, 30 500 – 33 000.

Если работник с заработком в 20 500 р. попадает на границу двух групп, то возникает вопрос, к какой группе его относить – к первой или второй? В этом случае открывают один из крайних интервалов или используют принцип единообразия – нижняя граница включает в себя обозначенное значение, а верхняя – не включает. В этом случае работник, получающий 20 500 р., должен быть отнесен ко второй группе. Такой же порядок применяется в отношении всех остальных групп.

Вернёмся к приведённому выше примеру о группировке городов по численности населения (тыс. чел.): менее 50, 50 – 100, 100 – 250, 250 – 500, 500 – 1 000, 1 000 и более. В данном примере имеются закрытые и открытые интервалы.

*Закрытыми* называются интервалы, у которых указаны и нижняя, и верхняя границы (например, 50 – 100). У *открытых* интервалов указана лишь одна из границ – только верхняя или только нижняя. В нашем примере это первая группа "менее 50" и последняя "1 000 и более".

Открытые интервалы позволяют более определённо решить вопрос об отнесении к той или иной группе тех единиц совокупности, которые попадают на границы интервалов. Так, слова "менее 50" указывают на то, что города с численностью населения 50 тыс. чел. к первой группе отнесены быть не могут, они должны быть включены во вторую группу 50 – 100. Далее, исходя из принципа единообразия, все значения, равные верхней границе интервала, не входят в данную группу, а значения, равные нижней границе, – входят. По этой же причине значение 1 000 не входит в предпоследнюю группу, но входит в последнюю "1 000 и более".

## 3.3. Статистические ряды распределения

После определения группировочного признака, границ и численности каждой из групп, строится ряд распределения.

***Статистический ряд распределения*** – это упорядоченное распределение единиц изучаемой совокупности на группы по группировочному варьирующему признаку.

Ряды распределения, построенные по атрибутивным (качественным) признакам, называются ***атрибутивными***. Например, распределение населения по уровню образования (от высшего образования до "не имеет образования").

Ряды распределения, построенные по количественному признаку, называются ***вариационными***. Например, распределение населения по возрасту или по размеру дохода.

Главными элементами вариационных рядов распределения являются варианты и частоты.

***Варианты*** – отдельные числовые значения признака, которые он принимает в вариационном ряду распределения, т.е. конкретные значения варьирующего признака. Обозначим варианты $x_i$.

***Частоты*** – численности отдельных вариантов или каждой группы вариационного ряда, т.е. это числа, показывающие, сколько раз встречаются те или иные варианты в ряду распределения. Обозначим частоты $f_i$. Сумма всех частот называется *объемом* совокупности и равна числу элементов всей совокупности.

Иногда наряду с частотами используют частости. ***Частости*** – это частоты, выраженные в долях единицы или процентах к итогу:

$$d_i = \frac{f_i}{\sum_{i=1}^{n} f_i}, \qquad (3.2)$$

где $d_i$ – частость; $\sum_{i=1}^{n} f_i -$ объём совокупности, шт.

Сумма частостей равна единице или 100 %. Использование частостей позволяет сопоставлять вариационные ряды с разным числом наблюдений.

Вариационные ряды в зависимости от характера вариации подразделяются на дискретные и интервальные. *Дискретные вариационные ряды* строятся для дискретных признаков, принимающих целые значения. Ряды распределения принято оформлять в виде таблицы, пример дискретного вариационного ряда приведён в табл. 3.1.

Таблица 3.1

Распределение квартир, предлагаемых к продаже в декабре 2011 г. через риэлтерские конторы г. Комсомольска-на-Амуре, по количеству комнат

| Количество комнат в квартире ($x_i$) | Число квартир | |
|---|---|---|
| | всего, единиц ($f_i$) | в % к итогу ($d_i$) |
| 1 | 655 | 19,5 |
| 2 | 1 226 | 36,5 |
| 3 | 1 146 | 34,2 |
| 4 | 294 | 8,8 |
| 5 | 34 | 1,0 |
| Итого | 3 355 | 100 |

В данном примере в графе "количество комнат в квартире" расположены варианты ($x_i$), ранжированные по возрастанию значения изучаемого варьирующего признака. Во второй графе "число квартир, всего" проставлены частоты ($f_i$), соответствующие каждой из вариант. В третьей графе "в % к итогу" расположены частости ($d_i$).

*Интервальные вариационные ряды* строятся для непрерывных признаков, принимающих любые (в том числе и дробные) значения, либо для дискретных, представленных в виде интервалов "от – до".

Пример интервального вариационного ряда приведён в табл. 3.2.

Ряды распределения удобно изучать с помощью графического метода путём построения полигона или гистограммы.

***Полигон*** – графическое изображение дискретного вариационного ряда. По оси абсцисс откладываются значения варьирующего признака (варианты), по оси ординат наносится шкала для выражения величины частот.

Построим полигон распределения квартир (рис. 3.1) по количеству комнат на основе данных табл. 3.1.

На полигоне хорошо видно, что на рынке недвижимости г. Комсомольска-на-Амуре преобладают двух- и трёхкомнатные квартиры.

Таблица 3.2

Распределение квартир, предлагаемых к продаже в декабре 2011 г.
через риэлтерские конторы г. Комсомольска-на-Амуре,
по размеру общей площади квартиры

| Общая площадь квартиры, м$^2$ | Число квартир | |
|---|---|---|
| | всего, ед. | в % к итогу |
| 22 – 29 | 38 | 1,1 |
| 29 – 36 | 483 | 14,4 |
| 36 – 43 | 127 | 3,8 |
| 43 – 50 | 534 | 15,9 |
| 50 – 57 | 458 | 13,7 |
| 57 – 64 | 966 | 28,8 |
| 64 – 71 | 305 | 9,1 |
| 71 – 78 | 114 | 3,4 |
| 78 – 85 | 229 | 6,8 |
| 85 – 92 | 38 | 1,1 |
| 92 – 99 | 38 | 1,1 |
| 99 – 106 | 15 | 0,4 |
| 106 – 113 | 10 | 0,3 |
| Всего | 3 355 | 100 |

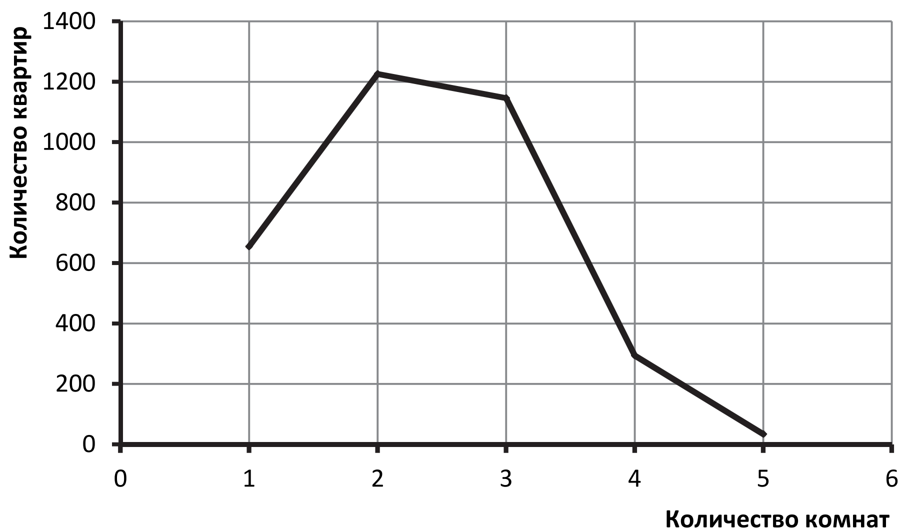


Рис. 3.1. Полигон распределения квартир по количеству комнат

***Гистограмма*** – графическое изображение интервального вариационного ряда. На оси абсцисс откладываются величины интервалов, а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. Высота прямоугольника в случае равных интервалов равна частоте соответствующего интервала.

Построим гистограмму распределения квартир (рис. 3.2) по размеру общей площади квартиры на основе данных табл. 3.2.

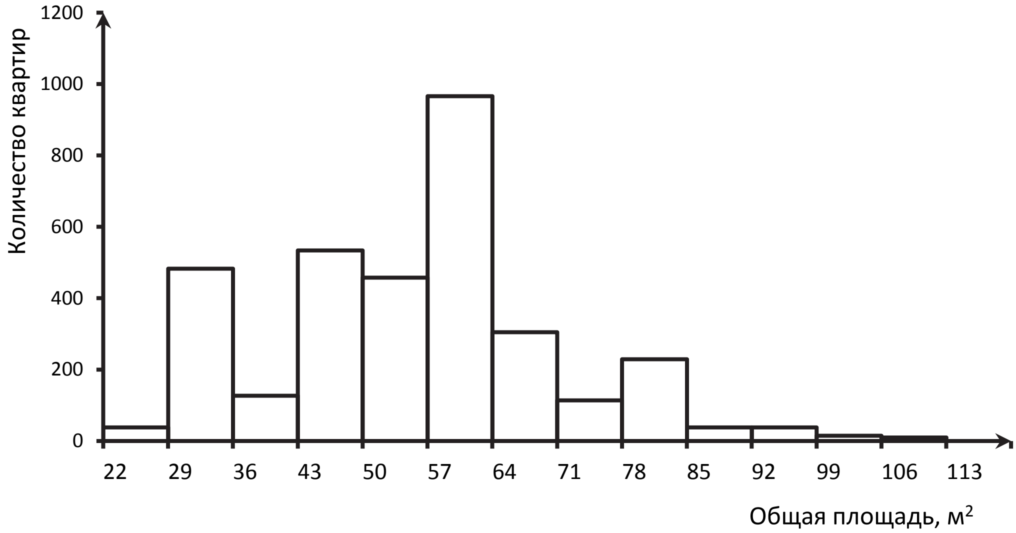


Рис. 3.2. Гистограмма распределения квартир по размеру общей площади квартиры

На гистограмме выделяются интервалы 29-36, 43-50, 57-64 и 78-85, соответствующие типовым площадям одно-, двух-, трёх-, четырёхкомнатных квартир соответственно.

## 4. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ

Результатом сводки статистических данных являются обобщающие статистические показатели. В зависимости от методов расчёта обобщающие показатели могут быть абсолютными, относительными или средними величинами.

Исходной формой выражения статистических показателей, отражающих уровень развития или объём явления, служат абсолютные величины.

### 4.1. Абсолютные статистические величины

***Абсолютные*** суммарные обобщающие показатели характеризуют размеры (уровни, объемы) изучаемых явлений в конкретных условиях места и времени. Например, численность постоянного населения Российской Федерации на 1 января 2013 г. составила 143 347 тыс. чел.

Абсолютные показатели делятся на индивидуальные и суммарные.

*Индивидуальные* абсолютные показатели получаются непосредственно в процессе статистического наблюдения и характеризуют размеры признака у отдельных единиц совокупности (возраст отдельного человека, продажная цена квартиры).

*Суммарные* абсолютные показатели подсчитываются по определенной совокупности объектов, охваченных статистическим наблюдением. При этом абсолютные величины как обобщающие показатели либо представляют собой сумму количества единиц изучаемой совокупности (численность совокупности), либо являются результатом суммирования значений варьирующего признака всех единиц совокупности (объем варьирующего признака).

Абсолютные статистические величины представляют собой именованные числа, т.е. имеют какую-либо единицу измерения, присущую тем или иным явлениям. Выражаются абсолютные величины в натуральных, стоимостных и трудовых единицах измерения.

*Натуральные* единицы измерения выражаются в физических мерах длины (метры, футы), веса (граммы, тонны), объёма (литры, кубометры) и т.д. Используются также и сложные натуральные единицы, представляющие собой комбинации двух разноименных величин. Так, например, потребление электроэнергии измеряется в киловатт-часах, грузооборот транспорта – в тонно-километрах.

В статистике применяют и абсолютные показатели, выраженные в условно-натуральных единицах измерения. Такие единицы получают, приводя различные натуральные единицы к одной, принятой за эталон. Так, например, разные виды топлива пересчитываются в условное топливо, тракторный парк – в эталонные тракторы, банки консервов – в условные банки консервов.

Так, например, в консервной промышленности емкость банки, равной 353,4 см$^3$, принята за условную банку. Если завод выпустил 100 тыс. банок ёмкостью 858,0 см$^3$, то объём производства составил 242,8 тыс. условных банок консервов (858,0 см$^3$ : 353,4 см$^3$ × 100 тыс.).

Большое значение имеют *стоимостные* единицы измерения, позволяющие соизмерить в денежной форме величины, несоизмеримые в натуральной форме. Так, например, объем выпуска разнородной продукции и объём товарооборота исчисляют в рублях (тыс. р., млн р.). Поскольку с течением времени уровень цен изменяется, то для сопоставимости значений показателя используются "неизменные" или "сопоставимые" цены.

*Трудовые* единицы измерения (человеко-часы или человеко-дни) используют для расчёта трудоемкость отдельных операций технологического цикла, трудоёмкости отдельных видов продукции или общих затрат труда на предприятии.

## 4.2. Относительные статистические величины

При анализе статистической информации в ряде случаев удобно использовать относительные величины. Относительные величины используют при изучении структуры явления, соотношения между отдельными его частями, развития явления во времени.

***Относительная величина*** представляет собой частное от деления одного абсолютного показателя на другой и характеризует количественное соотношение между ними. При этом в числителе дроби помещают *сравниваемый* (или *текущий*) показатель, а в знаменателе – показатель, принимаемый за *базу сравнения* (или *основание*).

Если сравниваемая величина существенно больше базы сравнения, то относительная величина является коэффициентом, и показывает, во сколько раз изучаемая величина больше основания. Если сравниваемая величина сопоставима с базой сравнения или меньше её, то относительный показатель может быть представлен в долях единицы: десятых; сотых (т.е. процентах, %); тысячных (промилле, ‰); десятитысячных (продецимилле, $^{0}/_{000}$).

В случаях, когда сопоставляемые величины являются разноименными, то их наименования образуются от наименований сравниваемых величин (чел./км$^2$; ц/га, р./м).

Рассмотрим основные виды относительных величин.

*Относительные величины структуры* характеризуют долю отдельных частей изучаемой совокупности во всем ее объеме:

$$\begin{matrix}\textit{Относительная величина}\\ \textit{структуры, \%}\end{matrix} = \frac{\textit{Величина изучаемой части совокупности}}{\textit{Величина всей совокупности}} * 100\ \%.$$

В качестве примера относительной величины структуры могут служить данные об удельном весе занятого в экономике населения в общей численности населения России по итогам переписи 2010 г. – 46,1 %.

*Относительная величина динамики* характеризует изменение какого-либо явления во времени. При расчёте показателя сравнивается уровень явления в определенный период или момент времени с уровнем этого же явления в прошлом:

$$\begin{matrix}\textit{Относительная величина}\\ \textit{динамики, \%}\end{matrix} = \frac{\textit{Текущий показатель}}{\textit{Предшествующий или базисный показатель}} * 100\ \%.$$

Относительные величины динамики, выраженные в процентах, называют *темпами роста*, а если величина выражена в единицах (долях единицы) – коэффициентом роста. Выбор или предшествующего, или базисного уровня явления в качестве базы сравнения определяется целью исследования.

Так, например, если в 2011 г. в России родилось 1 796,6 тыс. детей, а в 2012 г. – 1 902,1 тыс. детей, то темп роста составит

$$\frac{1\,902{,}1}{1\,796{,}6} * 100\,\% = 105{,}8\,\%.$$

Относительная величина выполнения плана рассчитывается как отношение фактически достигнутого в данном периоде уровня к запланированному уровню.

*Относительная величина сравнения* характеризует количественное соотношение одноимённых показателей, относящихся к двум разным статистическим совокупностям, являющимся объектами наблюдения.

Так, например, на 1 января 2012 г. численность населения Дальневосточного федерального округа (ДФО) составляла 6 266 тыс. чел., а население г. Москва – 11 613 тыс. чел. Принимая за базу сравнения ДФО, получаем:

$$\frac{11\,613}{6\,266} = 1{,}85.$$

Таким образом, население Москвы в 1,85 раза больше, чем население всего Дальневосточного округа.

*Относительные величины координации* рассчитывают как соотношение между частями одного целого. Например, в Российской Федерации на 1 января 2012 г. соотношение женщин и мужчин составляло 76,9 млн женщин на 66,1 млн мужчин, т.е. 76,9 : 66,1 или приблизительно 7 : 6.

*Относительные величины интенсивности* характеризуют степень распространения или уровень развития того или иного явления в определенной среде. При этом сравниваются разноименные величины, находящиеся в определенной связи между собой. Эти показатели являются именованными числами и показывают, сколько единиц одной совокупности приходится на единицу (100, 1000 и т.д. единиц другой совокупности). Так, например, плотность населения, выражающаяся средним числом жителей на одном квадратном километре территории, составила в Хабаровском крае в 2010 г. 1,8 чел./км$^2$. Таким же образом можно рассчитать уровень преступности. В 2012 г. в г. Комсомольске-на-Амуре было совершено 5 423 преступления при среднегодовой численности населения 259 тыс. чел. В расчёте на 1000 человек населения уровень преступности в городе составил:

$$\frac{5\,423}{259} = 20{,}9\ (\text{преступлений на 1000 чел.}).$$

Относительные показатели дополняют и углубляют показатели абсолютные. Например, что можно сказать о преступности в г. Комсомольске-на-Амуре, если в 2012 г. в городе было совершено 5 423 преступления, а в Хабаровском крае – 26 518 преступлений? Только то, что в крае число преступлений гораздо больше. Но ведь в крае и население больше, чем Комсомольске. Рассчитав число преступлений на 1000 человек населения для Хабаровского края, получим уровень преступности в крае 19,8 преступлений на 1000 человек. Сравнивая уровни преступности в городе 20,9 и в крае 19,8, приходим к выводу, что преступность в г. Комсомольске-на-Амуре несколько выше, чем в Хабаровском крае в целом.

Как видно из примера, сочетание абсолютных и относительных величин позволяет более глубоко проанализировать различные явления социально-экономической жизни.

# 5. СРЕДНИЕ ВЕЛИЧИНЫ

## 5.1. Сущность и значение средних величин

Среди обобщающих показателей, характеризующих статистические совокупности, большое распространение имеют средние величины.

***Средняя величина*** – это обобщающая характеристика множества индивидуальных значений некоторого количественного признака статистической совокупности в конкретных условиях места и времени.

Средняя – один из распространенных приемов обобщения. Средний показатель отражает типичный размер явления в изучаемой совокупности, игнорируя различия отдельных единиц. При исчислении средних в силу действия закона больших чисел случайности взаимопогашаются, что позволяет абстрагироваться от случайности отдельных значений и выявить типичный размер признака, сформированный под действием основных факторов и присущий изучаемой совокупности.

При вычислении средних множество различных индивидуальных значений признака заменяется одним средним показателем, характеризующим всю совокупность в целом. Это даёт возможность выявить закономерности, присущие массовым явлениям, которые сложно распознать, изучая единичные явления.

Средняя отражает характерный, типичный, реальный уровень изучаемых явлений, характеризует эти уровни и позволяет изучать их изменения во времени и в пространстве. Так, например, знание средних многолетних температур зимних месяцев позволяет правильно спланировать проведение отопительного сезона.

При вычислении средних показателей необходимо помнить, что средняя величина будет объективно отражать типичный уровень признака, если она рассчитывается по массовым данным для качественно однородной совокупности.

Так, например, средний уровень дохода работников малого частного предприятия будет фиктивной величиной, поскольку работники весьма сильно различаются по уровню дохода, а численность такой совокупности невелика. Средний размер начисляемых в регионе пенсий будет типичной величиной, т.к. размер пенсии варьируется в узких пределах, а количество пенсионеров в регионе исчисляется сотнями тысяч.

## 5.2. Виды средних

### *5.2.1. Средняя арифметическая*

Наиболее распространенным видом средних величин является средняя арифметическая. Она применяется в тех случаях, когда объем осредняемого признака для всей совокупности образуется как сумма его значений у отдельных ее единиц. Например, валовой сбор зерна в хозяйстве рассчитывается как сумма произведенной продукций со всех посевных площадей хозяйства.

Средняя арифметическая рассчитывается в форме простой средней и взвешенной средней.

Известная из школы ***средняя арифметическая простая (невзвешенная)*** рассчитывается для несгруппированных индивидуальных значений признака. Она равна простой сумме отдельных значений осредняемого признака ($x_i$), деленной на общее число этих значений:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{5.1}$$

где $\bar{x}$ – среднее значение величины; $x_1$, $x_2$, … $x_n$ – индивидуальные значения варьирующего признака (варианты); $n$ – число единиц совокупности.

**Пример.**

Известно количество деталей, изготовленных за смену каждым из 14 фрезеровщиков бригады, шт.:

9; 8; 9; 11; 8; 9; 10; 9; 12; 9; 10; 8; 9; 8.

Найти среднюю выработку одного фрезеровщика.

Средняя арифметическая простая рассчитывается по формуле (5.1)

$$\bar{x} = \frac{9+8+9+11+8+9+10+9+12+9+10+8+9+8}{14} = \frac{129}{14} = 9{,}21 \approx 9\,шт.$$

Если одно и то же значение признака в совокупности встречается несколько раз, то объединив данные в группы по величине признака и подсчитав численность каждой из образованных групп, построим вариационный ряд. Средняя из вариантов, которые повторяются различное число раз, называется *взвешенной*. В качестве весов выступают численности групп вариационного ряда.

***Средняя арифметическая взвешенная*** вычисляется для сгруппированных величин по формуле

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i}, \tag{5.2}$$

где $f_1, f_2, \dots f_n$ – частоты (число повторений одинаковых значений варьирующего признака); $x_1$, $x_2$, ... $x_n$ – индивидуальные значения варьирующего признака (варианты); $f_i$ – количество единиц в каждой из групп; $m$ – количество групп.

Рассчитаем среднюю арифметическую взвешенную, используя данные предыдущего примера. Для этого сгруппируем фрезеровщиков по числу изготовленных каждым из них деталей (табл. 5.1).

Таблица 5.1

Распределение фрезеровщиков по выработке деталей

| Выработка деталей за смену одним фрезеровщиком, шт. ($x_i$) | Число фрезеровщиков (веса) ($f_i$) | $x_i \times f_i$ |
|---|---|---|
| 8 | 4 | 32 |
| 9 | 6 | 54 |
| 10 | 2 | 20 |
| 11 | 1 | 11 |
| 12 | 1 | 12 |
| Итого | 14 | 129 |

Подставляем итоговые значения в формулу средней арифметической взвешенной (5.2):

$$\bar{x} = \frac{32 + 54 + 20 + 11 + 12}{14}; \quad \bar{x} = \frac{129}{14} = 9{,}21 \approx 9\,шт.$$

Если значения осредняемого признака заданы в виде интервальных рядов распределения ("от" – "до"), то при расчете средней арифметической в качестве значений признаков в группах принимают середины этих интервалов $x_i'$.

**Пример.**

Известно распределение сотрудников организации по размеру заработной платы (табл. 5.2). Рассчитать средний размер заработной платы сотрудников в организации.

Таблица 5.2

Распределение сотрудников по размеру заработной платы

| Исходные данные | | Расчетные значения | |
|---|---|---|---|
| Размер заработной платы, тыс. р. | Число сотрудников, чел. ($f_i$) | Серединное значение, тыс. р. ($x_i'$) | $x_i' f_i$ |
| До 20,0 | 6 | 19,0 | 114 |
| 20,0 – 22,0 | 10 | 21,0 | 210 |
| 22,0 – 24,0 | 21 | 23,0 | 483 |
| 24,0 – 26,0 | 25 | 25,0 | 625 |
| 26,0 и более | 18 | 27,0 | 486 |
| Итого | 80 | | 1918 |

Рассчитаем серединные значения интервалов.

В *закрытых* интервалах серединное значение найдём как полусумму значений верхней и нижней границ интервала. Например, для интервала 20,0 – 22,0 серединное значение будет (20 + 22) / 2 = 21.

Для *открытых* интервалов задача осложняется тем, что длина открытого интервала не определена. В этом случае предполагается, что длина открытого интервала условно равна длине смежного с ним закрытого интервала. Так, для интервала "26 и более" смежным будет интервал 24,0 – 26,0 с длиной интервала 26 – 24 = 2. Тогда условная верхняя граница открытого интервала "26,0 и более" будет 26 + 2 = 28, а его условная середина (26 + 28) / 2 = 27.

После того как найдены середины интервалов $x'_i$, вычисления проводят так же, как и в дискретном ряду, варианты $x'_i$ умножают на частоты $f_i$ и сумму произведений делят на сумму частот:

$$\bar{x} = \frac{\sum x_i' f_i}{\sum f_i}; \quad \bar{x} = \frac{1918}{80} = 23{,}98 \text{ тыс. р.}$$

Для простоты изложения в дальнейшем у знака суммы пределы суммирования не указываются.

### *5.2.2. Средняя гармоническая*

Средняя арифметическая применяется в тех случаях, когда известны варианты варьирующего признака $x_i$ и их частоты $f_i$.

В случаях, когда имеющаяся информация не содержит частот $f_i$ по отдельным вариантам $x_i$ совокупности, а представлена как их произведение $x_i f_i$, используется формула средней гармонической взвешенной. Обозначим $x_i f_i = w_i$, тогда $f_i = w_i/x_i$. Подставляя найденные величины в формулу средней арифметической взвешенной (5.2), получим формулу средней гармонической взвешенной:

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{w_1 + w_2 + ... + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + ........ \frac{w_n}{x_n}}. \tag{5.3}$$

Таким образом, средняя гармоническая является преобразованной формой арифметической средней и тождественна ей. Вместо гармонической всегда можно рассчитать среднюю арифметическую, предварительно определив веса отдельных значений признака как $f_i = w_i/x_i$.

Средняя гармоническая применяется в случаях, когда неизвестны действительные веса $f_i$, а известно $w_i = x_i f_i$.

В качестве примера по данным (табл. 5.3) определим среднюю продажную цену 1 кг сахара.

Таблица 5.3

Цена и выручка от реализации по трем магазинам

| Магазин | Исходные данные | | Расчетные значения |
|---|---|---|---|
| | Цена сахара, р./кг, $x_i$ | Выручка от реализации, р., $w_i$ | Частота (количество реализованных единиц), кг |
| А | 40 | 8 000 | 8 000:40 = 200 |
| Б | 44 | 6 160 | 6 160:44 = 140 |
| В | 43 | 4 300 | 4 300:43 = 100 |
| Итого | | 18 460 | 4 400 |

Расчет средней цены проведём по формуле

$$\textit{Средняя цена, р.} = \frac{\textit{Выручка от реализации, р.}}{\textit{Количество реализованных единиц, кг}}$$

Выручка от реализации $w_i$ известна (числитель). Количество реализованных единиц по каждому магазину найдём как частное от деления выручки на продажную цену. Тогда средняя продажная цена 1 кг сахара по трем магазинам может быть исчислена по формуле (5.3) средней гармонической

$$\bar{x} = \frac{\sum w}{\sum \frac{w}{x}}; \quad \bar{x} = \frac{8000 + 6160 + 4300}{\frac{8000}{40} + \frac{6160}{44} + \frac{4300}{43}} = \frac{18460}{4400} = 41{,}95 \text{ р.}$$

### 5.2.3. Средняя геометрическая

Среднюю геометрическую удобно использовать в случаях, когда индивидуальные значения признака представляют собой, относительные величины.

***Средняя геометрическая*** представляет собой корень $n$-й степени из произведений отдельных значений – вариантов признака $x$:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = \sqrt[n]{\Pi x_i} \, ,$$

где $n$ – число вариантов; $\Pi$ – знак произведения.

Средняя геометрическая, в частности, используется для вычисления средних темпов роста в рядах динамики.

### 5.2.4. Средняя квадратическая и средняя кубическая

Если размер осредняемого признака выражен в квадратных или кубических единицах измерения, то в ряде случаев применяются средняя квадратическая (например, для вычисления средних диаметров труб, стволов деревьев и т.п.) и средняя кубическая (например, при определении средней длины стороны $n$ кубов).

***Средняя квадратическая простая*** является квадратным корнем из частного от деления суммы квадратов отдельных значений признака на их число:

$$\bar{x} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\sum x_i^2}{n}} \, .$$

Средняя квадратическая взвешенная:

$$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}} \, ,$$

где $f_i$ – веса.

Формулы для расчета средней кубической аналогичны:

Средняя кубическая простая

$$\bar{x} = \sqrt[3]{\frac{\sum x_i^3}{n}} \, .$$

Средняя кубическая взвешенная

$$\bar{x} = \sqrt[3]{\frac{\sum x_i^3 f_i}{\sum f_i}} \, .$$

Средние квадратическая и кубическая в статистике используются достаточно редко. В частности, средняя квадратическая применяется при расчете показателей вариации.

### *5.2.5. Структурные средние*

При изучении структуры рядов распределения используются особые виды средних величин – структурные средние. К таким показателям относятся мода и медиана.

***Мода*** (*Мо*) – чаще всего встречающийся вариант в дискретном вариационном ряду, т.е. вариант, имеющий наибольшую частоту.

Например, в табл. 5.1 наибольшей частотой является число 6. Этой частоте соответствует модальное значение признака, т.е. выработка деталей за смену. Мода свидетельствует, что в данном примере чаще всего встречаются фрезеровщики, изготавливающие за смену девять деталей.

В интервальных рядах распределения с равными интервалами мода вычисляется по формуле

$$M_0 = X_{M_0} + i_{M_0} \frac{f_{M_0} - f_{M_{0-1}}}{\left(f_{M_0} - f_{M_{0-1}}\right) + \left(f_{M_0} - f_{M_{0+1}}\right)},$$

где $X_{Mo}$ – нижняя граница модального интервала; $i_{Mo}$ – длина модального интервала; $f_{Mo}$ – частота модального интервала; $f_{Mo-1}$ – частота интервала, предшествующего модальному; $f_{Mo+1}$ – частота интервала, следующего за модальным.

Модальный интервал – интервал с наибольшей частотой.

По данным табл. 5.2 рассчитаем моду. Модальным будет интервал 24,0 – 26,0 с частотой 24, длина интервала в этом примере равна 2:

$$M_0 = 24{,}0 + 2 \times \frac{25-21}{(25-21)+(25-18)} = 24{,}73 \text{ } \textit{тыс.р.}$$

Модальным значением заработной платы сотрудников является зарплата, равная 24,73 тыс. р.

Мода широко используется при изучении покупательского спроса, типовых цен на товары и т.п.

***Медиана*** (*Ме*) – это вариант, который приходится в середину ранжированного ряда. Медиана делит ряд на две равные по числу единиц части, одна из них имеет значения признака меньше медианы, а другая – больше медианы. Процедуру нахождения медианы для дискретного признака рассмотрим на примере.

Пусть известны тарифные разряды каждого из девяти фрезеровщиков бригады:

5; 2; 4; 6; 3; 6; 3; 5; 3.

Проведём ранжирование, для чего упорядочим совокупность от меньшего к большему:

2; 3; 3; 3; <u>4</u>; 5; 5; 6; 6.

Вычислим номер медианы, который для нечетного числа единиц совокупности вычисляется по формуле

$$N_{Me} = \frac{n+1}{2},$$

где $n$ – число членов ряда.

В рассматриваемом примере номер медианы равен 5, а медиана – 4, т.е. одна половина фрезеровщиков имеет разряд менее 4-го (2; 3; 3; 3), а другая – более 4-го (5; 5; 6; 6).

Если изучаемая совокупность имеет чётное число единиц, то медиана равна средней из двух вариантов, находящихся в середине ряда.

Так, например, если совокупность состоит из восьми единиц

2; 3; 3; 3; 4; 5; 5; 6,

то медиана будет (3 + 4)/2 = 3,5.

В интервальных рядах распределения медианное значение расположено в одном из интервалов признака *x*. Этот медианный интервал характерен тем, что его накопленная сумма частот равна или превышает полусумму всех частот ряда. Значение медианы вычисляется линейной интерполяцией по формуле

$$M_e = X_{Me} + i_{Me} \times \frac{\frac{\sum f}{2} - S_{Me-1}}{f_{Me}}, \qquad (5.4)$$

где $X_{Me}$ – нижняя граница медианного интервала; $i_{Me}$ – величина медианного интервала; $\frac{\sum f}{2}$ – полусумма частот ряда; $S_{Me-1}$ – сумма накопленных частот, предшествующих медианному интервалу; $f_{Me}$ – частота медианного интервала.

**Пример.** Рассчитать медиану по данным табл. 5.2.

Найдем медианный интервал. Полусумма частот ряда будет 80:2 = 40. Далее рассчитываем суммы накопленных частот:

- первый интервал – 6;
- второй интервал – 6 + 10 = 16;
- третий интервал – 6 + 10 + 21 = 37;
- четвёртый интервал – 6 + 10 + 21 + 25 = 62;
- пятый интервал – 6 + 10 + 21 + 25 + 18 = 80.

Сумма накопленных частот для четвёртого интервала равна 62, что больше полусуммы частот ряда, т.е. 62 > 40. Таким образом, четвёртый интервал 24 – 26 тыс. р. является медианным интервалом и содержит сороковую по счёту единицу ранжированной совокупности. Нижняя граница медианного интервала равна 24, его частота – 25, сумма накопленных частот предшествующего (третьего) интервала – 37.

По формуле (5.4) находим значение медианы:

$$M_e = 24 + 2 \times \left[ \left( \frac{80}{2} - 37 \right) / 25 \right] = 24{,}24 \ тыс.р.$$

Полученный результат говорит о том, что из 80 сотрудников 40 сотрудников имеют уровень оплаты менее 24,24 тыс. р. в месяц, а 40 сотрудников – более.

Медиана находит практическое применение вследствие особого свойства – сумма абсолютных отклонений чисел ряда от медианы есть величина наименьшая:

$$\sum |x - M_e| \ \to \min.$$

Мода и медиана совпадают со средней только в случае симметричного распределения частот вариационного ряда. Анализ соотношения моды, медианы и средней арифметической позволяет оценить асимметрию ряда распределения.

# 6. ПОКАЗАТЕЛИ ВАРИАЦИИ

## 6.1. Понятие вариации

***Вариация*** – различия в значениях рассматриваемого признака у отдельных единиц, составляющих данную совокупность в один и тот же период или момент времени. Так, например, студенты одной учебной группы различаются по росту, весу, цвету глаз и т.д.

Вариация является результатом воздействия на каждую из единиц совокупности множества разнообразных факторов, которые по-разному сочетаются в каждом отдельном случае, в результате чего каждая из единиц совокупности имеет свое собственное, индивидуальное значение изучаемого признака.

Рассмотренные выше средние величины дают обобщающую характеристику признака изучаемой совокупности. Однако средняя величина не показывает, как располагаются около нее варианты изучаемого признака, сосредоточены ли они вблизи средней или значительно отклоняются от нее. Две совокупности могут иметь одинаковую среднюю величину, но в одной из них значения признака у отдельных единиц могут незначительно отличаются от средней, а в другой совокупности эти отличия могут быть большими. Таким образом, в одном случае вариация признака мала, а в другом – велика, что имеет весьма важное значение для характеристики надежности средней величины.

Чем больше варианты отдельных единиц совокупности различаются между собой, тем больше они отличаются от своей средней величины. Чем меньше варианты отличаются друг от друга, тем меньше они отличаются от своей средней величины, и средняя величина в этом случае будет более реально представлять всю совокупность. По этой причине в дополнение к средним необходимы и другие показатели, характеризующие отклонения отдельных значений от общей средней.

## 6.2. Мера вариации

Поясним сказанное примером. Две бригады овощеводов одного и того же хозяйства выращивают огурцы в стандартных теплицах. Урожайность огурцов в каждой из теплиц составила, кг/м$^2$:

- в первой бригаде – 28, 30, 32 ($\bar{x}$ = 30 кг/м$^2$);

- во второй бригаде – 20, 30, 40 ($\bar{x}$ = 30 кг/м$^2$).

Средняя урожайность в обеих бригадах одинакова и составляет $\bar{x}$ = 30 кг/м$^2$, однако различия урожайности в первой бригаде значительно меньше, чем во второй, что говорит о её более стабильной работе.

Для оценки вариации признака в совокупностях используют ряд показателей вариации.

### *6.2.1. Абсолютные и средние показатели вариации*

***Размах вариации R*** является наиболее простым показателем вариации признака и рассчитывается как разность между максимальным $x_{max}$ и минимальным $x_{min}$ значениями признака в совокупности:

$$R = x_{max} - x_{min}. \tag{6.1}$$

В рассмотренном примере размах вариации урожайности (6.1) составляет в первой бригаде $R_1$ = 32 – 28 = 4 (кг/м$^2$), во второй бригаде $R_2$ = 40 – 20 = 20 (кг/м$^2$), что в 5 раз больше.

Несмотря на простоту вычисления, размах вариации показывает лишь крайние, возможно, даже и аномальные, отклонения признака и не отражает отклонений всех вариант в ряду. Для получения обобщающей характеристики распределения отклонений используются иные показатели вариации. Наиболее простым из них является среднее линейное отклонение.

***Среднее линейное отклонение*** $\bar{d}$ рассчитывается как средняя арифметическая абсолютных значений отклонений отдельных вариант от их средней арифметической.

Среднее линейное отклонение для несгруппированных данных:

$$\bar{d} = \frac{\sum|x_i - \bar{x}|}{n}, \tag{6.2}$$

где $n$ – число единиц совокупности; $\bar{x}$ – средняя арифметическая.

Среднее линейное отклонение для сгруппированных данных:

$$\bar{d} = \frac{\sum|x_i - \bar{x}|}{\sum f_i}, \tag{6.3}$$

где $\sum f_i$ – сумма частот вариационного ряда.

Поскольку сумма отклонений значений признака от средней величины всегда равна нулю, то разность в числителе формул (6.2) и (6.3) взята по модулю.

Среднее линейное отклонение для оценки величины вариации признака применяется, в частности, при анализе ритмичности производства и оборота внешней торговли. Однако чаще всего в качестве меры вариации используется дисперсия.

***Дисперсия*** $\boldsymbol{\sigma^2}$ признака рассчитывается как средний квадрат отклонений вариант от их средней величины:

Простая дисперсия для несгруппированных данных:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}. \tag{6.4}$$

Взвешенная дисперсия для сгруппированных данных:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}. \tag{6.5}$$

В ряде случаев для вычисления дисперсии удобно пользоваться модифицированной формулой

$$\sigma^2 = \overline{x^2} - \bar{x}^2, \tag{6.6}$$

где $\overline{x^2}$ – средняя из квадратов; $\bar{x}^2$ – квадрат средней.

***Среднее квадратическое отклонение*** $\boldsymbol{\sigma}$ вычисляется как корень квадратный из дисперсии:

$$\sigma = \sqrt{\sigma^2}. \tag{6.7}$$

Среднее квадратическое отклонение, как и дисперсия, является мерой вариации признака в совокупности, оно показывает, насколько в среднем отклоняются отдельные варианты от их среднего значения. Поскольку среднее квадратическое отклонение выражается в тех же единицах, что и варианты исследуемого признака, то оно хорошо интерпретируется.

**Пример.** Рассчитать дисперсию и среднее квадратическое отклонение по несгруппированным данным (табл. 6.1).

Таблица 6.1

Расчет дисперсии

| Исходные данные | | Расчётные данные | |
|---|---|---|---|
| Киоск | Дневная выручка, тыс. р. ($x_i$) | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
| *1* | *2* | *3* | *4* |
| 1 | 35 | 4 | 16 |
| 2 | 30 | -1 | 1 |
| 3 | 28 | -3 | 9 |
| 4 | 30 | -1 | 1 |
| 5 | 32 | 1 | 1 |
| Итого | 155 | 0 | 28 |

1) Рассчитываем среднюю дневную выручку по формуле средней арифметической простой (5.1)

$$\bar{x} = \frac{155}{5} = 31 \text{ тыс. р.}$$

2) В соответствии с формулой (6.4) находим отклонения $x_i$ от $\bar{x}$ и записываем их в графу 3.

3) Возводим отклонения в квадрат, вписываем в графу 4. Определяем их сумму, она равна 28.

4) Рассчитаем дисперсию, для чего найденное значение подставим в формулу (6.4) либо (6.6)

$$\sigma^2 = \frac{28}{5} = 5{,}6 \text{ тыс. р.}^2$$

Обратите внимание на размерность дисперсии.

5) Извлекая из дисперсии квадратный корень по формуле (6.7), находим среднее квадратическое отклонение дневной выручки киоска $\sqrt{5{,}6} = 2{,}37$ тыс. р.

Теперь рассмотрим **пример** расчёта дисперсии и среднего квадратического отклонения по сгруппированным данным (табл. 6.2).

Таблица 6.2

Распределение машиностроительных предприятий по стоимости основных производственных фондов

| Исходные данные | | Расчётные данные | | | | |
|---|---|---|---|---|---|---|
| Группы предприятий по стоимости фондов, млн р. | Число предприятий ($f_i$) | Середина интервала, тыс. р. ($x_i'$) | $x_i' f_i$ | $x_i' - \bar{x}$ | $(x_i' - \bar{x})^2$ | $(x_i' - \bar{x})^2 \times f_i$ |
| *1* | *2* | *3* | *4* | *5* | *6* | *7* |
| До 200 | 6 | 100 | 600 | -450 | 202 500 | 1 215 000 |
| 200 – 400 | 10 | 300 | 3 000 | -250 | 62 500 | 625 000 |
| 400 – 600 | 21 | 500 | 10 500 | -50 | 2 500 | 52 500 |
| 600 – 800 | 12 | 700 | 8 400 | 150 | 22 500 | 270 000 |
| 800 – 1000 | 8 | 900 | 7 200 | 350 | 122 500 | 980 000 |
| 1000 – 1200 | 3 | 1100 | 3 300 | 550 | 302 500 | 907 500 |
| Итого | 60 | | 33 000 | | | 4 050 000 |

1) Определяем середины интервалов, вписываем в графу 3.

2) Рассчитываем среднюю дневную выручку по формуле средней арифметической взвешенной (5.2)

$$\bar{x} = \frac{33000}{60} = 550 \text{ млн р.}$$

3) В соответствии с формулой (6.5) находим отклонения $x_i'$ от $\overline{x}$ и записываем их в графу 5.

4) Возводим отклонения в квадрат, вписываем в графу 6, умножаем квадраты на соответствующие частоты, вписываем в графу 7. Определяем их сумму, она равна 4 050 000.

5) Рассчитаем дисперсию, для чего найденное значение подставим в формулу (6.5)

$$\sigma^2 = \frac{4050000}{60} = 67500.$$

6) Извлекая из дисперсии квадратный корень, находим среднее квадратическое отклонение стоимости основных производственных фондов $\sqrt{67\,500} = 259{,}8$ млн р.

***Дисперсия альтернативного признака $\sigma^2_p$.***

Альтернативный признак у каждой из единиц совокупности принимает одно из двух противоположных значений. Каждая из единиц совокупности либо обладает интересующим нас признаком – "да", что можно обозначить как 1, либо не обладает – "нет", т.е. 0.

Обозначим долю единиц, обладающих данным альтернативным признаком в совокупности как $p$, а долю единиц, не обладающих данным признаком, как $q$. Естественно, что $p + q = 1$.

Доля единиц, обладающих данным альтернативным признаком, рассчитывается как

$$p = \frac{b}{n}, \tag{6.8}$$

где $b$ – количество единиц совокупности, обладающих данным альтернативным признаком; $n$ – общее количество единиц совокупности.

Используя формулу (6.5), найдём среднее значение альтернативного признака и его дисперсию. Среднее значение альтернативного признака:

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{1 \times p + 0 \times q}{p + q} = p.$$

При расчёте дисперсии альтернативного признака учтём, что $q = 1 - p$

$$\sigma_p^2 = \frac{\sum \left(x_i - \overline{x}\right)^2 f_i}{\sum f_i} = \frac{(1-p)^2 p + (0-q)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq = p(1-p). \tag{6.9}$$

Таким образом, дисперсия альтернативного признака $\sigma^2_p$ равна произведению доли единиц совокупности $p$, обладающих данным признаком, на долю единиц $q$, не обладающих данным признаком, т.е. $\sigma^2_p = pq$.

Пример. Из проживающих в городе 5000 детей дошкольного возраста посещают детские дошкольные учреждения 3900 детей.

Рассчитаем долю детей, посещающих дошкольные учреждения, по формуле (6.8)

$$p = \frac{3900}{5000} = 0{,}78.$$

Рассчитаем долю детей, не посещающих дошкольные учреждения

$$q = 1 - 0{,}78 = 0{,}22.$$

Тогда дисперсия альтернативного признака будет

$$\sigma^2_p = pq; \quad \sigma^2_p = 0{,}78 \times 0{,}22 = 0{,}172.$$

Чем меньше значение имеют дисперсия и среднее квадратическое отклонение, тем однороднее будет совокупность и тем типичней будет средняя величина. Однако при попытке использовать абсолютные показатели для сравнения двух совокупностей по величине вариации признака, возникают трудности: совокупности могут иметь разные значения средней величины, а могут и вовсе иметь разную размерность.

В таких случаях используют относительные показатели вариации.

### *6.2.2. Относительные показатели вариации*

Относительные показатели вариации используют для оценки однородности совокупности, а также для сравнения колеблемости одного и того же признака в совокупностях с различным значением средней арифметической.

Из относительных показателей вариации чаще всего используют ***коэффициент вариации V***. Коэффициент вариации представляет собой отношение среднего квадратического отклонения к средней арифметической:

$$V = \frac{\sigma}{\bar{x}}, \text{ или в процентах, } V = \frac{\sigma}{\bar{x}} \times 100\,\%. \qquad (6.10)$$

Коэффициент вариации может быть использован как для сравнительной оценки вариации нескольких разных совокупностей, так и для характеристики однородности совокупности. Совокупность принято считать количественно-однородной, если коэффициент вариации не превышает 0,33 (33 %).

## 6.3. Правило сложения дисперсий

Вариация признака обусловлена множеством факторов, влияние отдельных факторов на совокупную вариацию можно оценить, если статистическую совокупность разбить на однородные группы по исследуемому признаку. В простейшем случае, когда совокупность расчленена на группы по одному фактору, влияние данного фактора на величину вариации изучается путём исчисления и анализа трех видов дисперсий: общей, межгрупповой и средней из внутригрупповых.

***Общая дисперсия*** $\sigma_o^2$ измеряет вариацию признака по всей совокупности, являющуюся результатом влияния всех действующих на совокупность факторов. Она может быть вычислена как простая дисперсия по формуле (6.4) или взвешенная дисперсия по формуле (6.5). Для сгруппированных данных формула запишется как

$$\sigma_o^2 = \frac{\sum (x_i - \overline{x_o})^2 f_i}{\sum f_i}, \qquad (6.11)$$

где $\overline{x_o}$ – общая средняя для всей изучаемой совокупности.

***Межгрупповая дисперсия*** $\delta^2$ характеризует систематическую вариацию результативного признака, обусловленную влиянием признака-фактора, положенного в основание группировки. Она равна среднему квадрату отклонений групповых средних от общей средней

$$\delta^2 = \frac{\sum (\overline{x_j} - \overline{x_o})^2 f_j}{\sum f_j}, \qquad (6.12)$$

где $\overline{x_j}$ – средняя по каждой из отдельных групп, на которые была расчленена исследуемая совокупность; $f_j$ – численность каждой из отдельных групп, на которые была расчленена исследуемая совокупность; $\overline{x_o}$ – общая средняя для всей изучаемой совокупности.

***Средняя внутригрупповых дисперсий*** $\overline{\sigma_j^2}$ отражает случайную вариацию, т.е. часть вариации, обусловленную влиянием неучтенных факторов и не зависящую от признака-фактора, положенного в основание группировки. Рассчитывается на основании внутригрупповых дисперсий по каждой из групп, на которые была расчленена исследуемая совокупность:

$$\overline{\sigma_j^2} = \frac{\sum \sigma_j^2 f_j}{\sum f_j}, \qquad (6.13)$$

где $\sigma_j^2$ – дисперсия по каждой из отдельных групп, на которые была расчленена исследуемая совокупность; $f_j$ – численность каждой из отдельных групп, на которые была расчленена исследуемая совокупность.

Между дисперсиями общей, межгрупповой и средней из внутригрупповых существует простая связь, выраженная ***правилом сложения дисперсий:*** общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий:

$$\sigma_o^2 = \overline{\sigma_j^2} + \delta^2.$$

Рассмотрим методику применения дисперсий для анализа влияния группировочного фактора на общую вариацию признака.

Пусть транспортная фирма Альфа имеет производственные подразделения в городах А и Б. Требуется оценить влияние места работы водителей (город А или город Б) на различия в их заработной плате.

1) Разделим всех водителей фирмы Альфа на группы, в основание группировки положим признак, влияние которого на различия в зарплате оценивается, т.е. сгруппируем водителей по месту работы (город А или город Б). Далее приступим к расчёту общей дисперсии зарплаты водителей по всей фирме Альфа.

2) Используя табл. 6.3, рассчитываем среднюю зарплату водителей по всей фирме (формула (5.2)):

$$\bar{x}_o = \frac{\sum x_i' f_o}{\sum f_o}; \quad \bar{x}_o = \frac{2530}{100} = 25{,}3 \text{ тыс. р.}$$

Таблица 6.3

Расчёт общей дисперсии

| Заработная плата, тыс. р. ($x_i$) | Число водителей | | | Расчёт общей дисперсии | | | | |
|---|---|---|---|---|---|---|---|---|
| | Город А ($f_a$) | Город Б ($f_в$) | Всего ($f_o$) | Середина интервала ($x_i'$) | $x_i' f_o$ | $x'_i - \bar{x}_o$ | $(x'_i - \bar{x}_o)^2$ | $(x'_i - \bar{x}_o)^2 \times f_o$ |
| 20-22 | 9 | 1 | 10 | 21 | 210 | -4,3 | 18,49 | 184,90 |
| 22-24 | 25 | 4 | 29 | 23 | 667 | -2,3 | 5,29 | 153,41 |
| 24-26 | 14 | 8 | 22 | 25 | 550 | -0,3 | 0,09 | 1,98 |
| 26-28 | 8 | 10 | 18 | 27 | 486 | 1,7 | 2,89 | 52,02 |
| 28-30 | 4 | 13 | 17 | 29 | 493 | 3,7 | 13,69 | 232,73 |
| 30-32 | 0 | 4 | 4 | 31 | 124 | 5,7 | 32,49 | 129,96 |
| Итого | 60 | 40 | 100 | | 2 530 | | | 755,00 |

3) Используя табл. 6.3, рассчитываем общую дисперсию зарплаты водителей (формула (6.11)).

$$\sigma_o^2 = \frac{755}{100} = 7{,}55.$$

При необходимости проверка совокупности на однородность может быть проведена по формуле (6.10).

4) Проводим аналогичные расчёты для подразделения в городе А, исходные данные для табл. 6.4 возьмём из табл. 6.3.

Используя табл. 6.4, рассчитываем среднюю зарплату водителей по подразделению в городе А:

$$\bar{x}_a = \frac{\sum x_i' f_a}{\sum f_a}; \quad \bar{x}_a = \frac{1446}{60} = 24{,}1 \text{ тыс.р.}$$

Таблица 6.4

Расчёт дисперсии по подразделению А

| Заработная плата, тыс. р. ($x_i$) | Число водителей город А ($f_a$) | Расчёт общей дисперсии | | | | |
|---|---|---|---|---|---|---|
| | | Середина интервала ($x_i'$) | $x_i' f_a$ | $x'_i - \bar{x}_a$ | $(x'_i - \bar{x}_a)^2$ | $(x'_i - \bar{x}_a)^2 \times f_a$ |
| 20-22 | 9 | 21 | 189 | -3,1 | 9,61 | 86,49 |
| 22-24 | 25 | 23 | 575 | -1,1 | 1,21 | 30,25 |
| 24-26 | 14 | 25 | 350 | 0,9 | 0,81 | 11,34 |
| 26-28 | 8 | 27 | 216 | 2,9 | 8,41 | 67,28 |
| 28-30 | 4 | 29 | 116 | 4,9 | 24,01 | 96,04 |
| 30-32 | 0 | 31 | 0 | 6,9 | 47,61 | 0,00 |
| Итого | 60 | | 1 446 | | | 291,40 |

Используя табл. 6.4, рассчитываем дисперсию зарплаты водителей по подразделению в городе А:

$$\sigma_a^{\ 2} = \frac{291{,}4}{60} = 4{,}86 \text{ тыс. р.}$$

5) Проводим расчёты для подразделения в городе Б, исходные данные для табл. 6.5 возьмём из табл. 6.3.

Используя табл. 6.5, рассчитываем среднюю зарплату и дисперсию зарплаты водителей по подразделению в городе Б:

$$\bar{x}_б = \frac{\sum x_i' f_б}{\sum f_б}; \quad \bar{x}_б = \frac{1084}{40} = 27{,}1 \text{ тыс. р.}$$

$$\sigma_б^{\ 2} = \frac{247{,}6}{40} = 6{,}19 \text{ тыс. р.}$$

Таблица 6.5

Расчёт дисперсии по подразделению Б

| Заработная плата, тыс. р. ($x_i$) | Число водителей город Б ($f_б$) | Расчёт общей дисперсии | | | | |
|---|---|---|---|---|---|---|
| | | Середина интервала ($x_i'$) | $x_i' f_б$ | $x'_i - \bar{x}_б$ | $(x'_i - \bar{x}_б)^2$ | $(x'_i - \bar{x}_б)^2 \times f_б$ |
| 20-22 | 1 | 21 | 21 | -6,1 | 37,21 | 37,21 |
| 22-24 | 4 | 23 | 92 | -4,1 | 16,81 | 67,24 |
| 24-26 | 8 | 25 | 200 | -2,1 | 4,41 | 35,28 |
| 26-28 | 10 | 27 | 270 | -0,1 | 0,01 | 0,10 |
| 28-30 | 13 | 29 | 377 | 1,9 | 3,61 | 46,93 |
| 30-32 | 4 | 31 | 124 | 3,9 | 15,21 | 60,84 |
| Итого | 40 | | 1 084 | | | 247,60 |

Внутригрупповые дисперсии $\sigma_а^2$ и $\sigma_б^2$ показывают вариации зарплаты в каждой группе, вызванные действием всех возможных факторов (техническое состояние автопарка, обеспеченность инструментами, горюче-смазочными материалами, стаж водителей, интенсивность труда и т.д.), кроме различий в месте работы (внутри группы все водители работают в одном и том же подразделении).

6) Рассчитаем среднюю из внутригрупповых дисперсий по формуле (6.13)

$$\overline{\sigma_j^2} = \frac{4{,}86 \times 60 + 6{,}19 \times 40}{60 + 40} = 5{,}39 \text{ тыс. р.}$$

Средняя из внутригрупповых дисперсий отражает вариацию заработной платы, обусловленную всеми факторами, кроме места работы водителей, но в среднем по всей совокупности.

7) Исчислим межгрупповую дисперсию по формуле (6.12)

$$\delta^2 = \frac{(24{,}1 - 25{,}3)^2 \times 60 + (27{,}1 - 25{,}3)^2 \times 40}{60 + 40} = 2{,}16 \text{ тыс. р.}$$

Межгрупповая дисперсия характеризует вариацию групповых средних, обусловленную различиями групп водителей по месту работы.

8) Проверим результат, используя правило сложения дисперсий:

$$\sigma_o^2 = \overline{\sigma_j^2} + \delta^2, \text{ т.е. } 7{,}55 = 5{,}39 + 2{,}16.$$

Чем больше доля межгрупповой дисперсии в общей дисперсии, тем сильнее влияние группировочного признака (место работы) на изучаемый признак (заработная плата).

***Эмпирический коэффициент детерминации $\eta^2$*** характеризует силу влияния группировочного признака на общую вариацию всей совокупности. Этот показатель представляет долю межгрупповой дисперсии в общей дисперсии результативного признака:

$$\eta^2 = \frac{\delta^2}{\sigma_0^2}. \qquad (6.14)$$

Эмпирический коэффициент детерминации показывает долю вариации результативного признака (зарплаты) под влиянием факторного признака (места работы). Используя формулу (6.14), в нашем примере

$$\eta^2 = \frac{2{,}16}{7{,}55} = 0{,}286 \quad \text{или } 28{,}6\,\%.$$

Это означает, что на 28,6 % различия в заработной плате родителей обусловлены различиями в месте их работы (подразделение А или подразделение Б), и на 71,4 % – влиянием всех прочих факторов вместе взятых.

## 6.4. Теоретические распределения в анализе вариационных рядов

При анализе вариационных рядов выявляются закономерности распределения, которые могут быть описаны с помощью некоторых теоретических кривых распределения. Характер распределения выявляется при большом числе наблюдений и малой длине интервала вариационного ряда. Если длина интервала вариационного ряда стремится к нулю, то графическое изображение эмпирического вариационного ряда (полигон, гистограмма) принимают вид плавной кривой, именуемой ***кривой распределения.*** Кривая распределения, выражающая общую закономерность данного типа распределения, называется ***теоретической кривой распределения***.

В практике статистических исследований встречаются различные распределения: нормальное, биномиальное, распределение Пуассона и др. Каждое распределение имеет свою специфику и область применения. Чаще всего в качестве теоретического распределения используется ***нормальное распределение.*** Распределение непрерывной случайной величины называют нормальным, если соответствующая ей плотность распределения вероятности выражается формулой

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}},$$

где $x$ – значение изучаемого признака; $\bar{x}$ – средняя арифметическая ряда; $\sigma^2$ – дисперсия значений изучаемого признака; $\sigma$ – среднее квадратическое отклонение изучаемого признака; $\pi$ = 3,14159 – отношение длины окружности к её диаметру; $e$ = 2,7182 – основание натурального логарифма.

Плотность распределения рассчитывается как отношение частоты $f_i$ к длине интервала вариационного ряда $i$ (формула (3.1)). Плотность распределения вероятности – отношение частости $d_i$ (формула (3.2)) к длине интервала вариационного ряда $i$.

Для удобства вычислений вероятностей случайные величины нормируются, для чего полагают $\sigma = 1$, начало координат переносят в точку $x = \bar{x}$, а $(x - \bar{x})/\sigma$ обозначают как $t$. В результате получается нормированная функция $\varphi(t)$, которая используется в табулированном виде:

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

Особенности кривой нормального распределения:

1) Кривая симметрична.

2) Кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности.

3) Площадь между кривой и осью $t$ равна единице, т.е. 100 % исследуемых единиц (частот) лежит в интервале от $-\infty$ до $+\infty$.

4) Площадь между ординатами, проведёнными на расстоянии 0 ± 1 ($\overline{x} \pm \sigma$), составляет 0,683 (область на рис. 6.1 выделена серым цветом). Это означает, что 68,3 % всех исследуемых единиц (частот) отклоняется от средней арифметической не более чем на $\sigma$, т.е. находится в пределах ($\overline{x} \pm \sigma$). В промежутке 0 ± 2 ($\overline{x} \pm 2\sigma$) находится 95,4 %, а в промежутке 0 ± 3 ($\overline{x} \pm 3\sigma$) соответственно – 99,7 % всех единиц исследуемой совокупности.

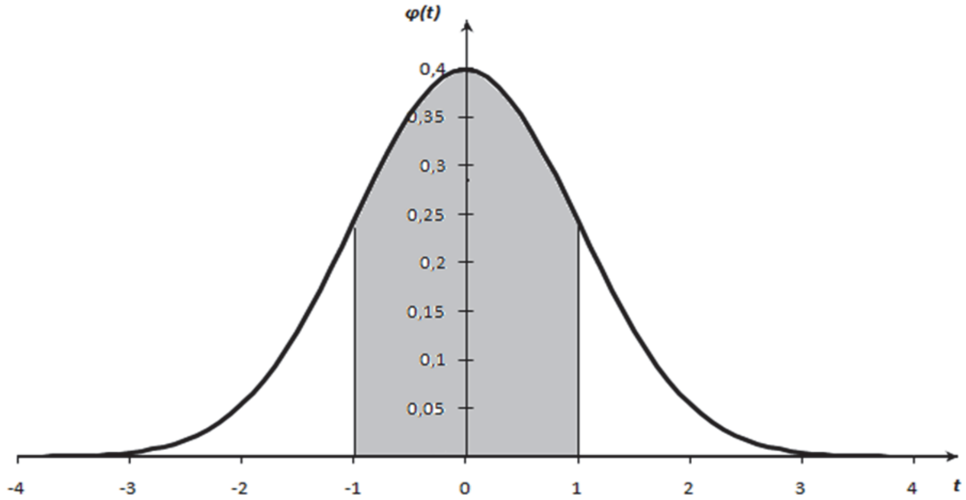


Рис. 6.1. Кривая нормального распределения

В кривой нормального распределения выражается закономерность, возникающая при взаимодействии множества случайных причин, действующих независимо друг от друга, когда ни одна из причин не имеет преобладающего влияния над другими.

Кривая нормального распределения находит широкое применение в математической статистике и в теории выборочного метода.

# 7. ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

## 7.1. Понятие о выборочном наблюдении

Статистическое наблюдение можно осуществлять как сплошное и несплошное. Как было отмечено в п. 2.2.2, сплошное наблюдение предусматривает обследование всех единиц изучаемой совокупности без исключения, что в ряде случаев связано с большими трудовыми и материальными затратами. Несплошное наблюдение позволяет судить о свойствах всей совокупности в целом на основе изучения не всех единиц совокупности, а лишь некоторой её части. В статистической практике самым распространенным из несплошных наблюдений является выборочное наблюдение.

***Выборочное наблюдение*** – несплошное наблюдение, при котором отбор подлежащих обследованию единиц проводится в случайном порядке, отобранная часть изучается, а результаты распространяются на всю исходную совокупность.

Подлежащая изучению статистическая совокупность, из которой производится отбор единиц, называется ***генеральной совокупностью***, и все ее обобщающие показатели – *генеральными*.

Совокупность отобранных из генеральной совокупности единиц называют ***выборочной совокупностью или выборкой***, и все ее обобщающие показатели – *выборочными*.

Основные достоинства выборочного наблюдения: экономия времени и средств в результате сокращения объема работы, уменьшение ошибок регистрации. При проведении ряда исследований выборочный метод является единственно возможным, например, при контроле качества продукции, если проверка сопровождается уничтожением единиц продукции или их частей (установление носкости обуви, испытание электрических лампочек на продолжительность горения, определение сахаристости фруктов, проверка консервов на доброкачественность и т.д.).

Выборочное наблюдение позволяет на основе характеристик выборочной совокупности (выборочной средней и выборочной доли) получить достоверные суждения о показателях генеральной средней и генеральной доли в генеральной совокупности.

Поскольку изучаемая статистическая совокупность состоит из единиц с варьирующими признаками, то состав выборочной совокупности не полностью воспроизводит состав генеральной совокупности. В результате возникает расхождение между значениями показателей, полученных по выборке, и значениями показателей этих же величин, которые были бы получены при проведенном с одинаковой степенью точности сплошном наблюдении – ошибка репрезентативности (представительства).

При проведении выборочного наблюдения ошибки репрезентативности исключить невозможно, но их можно оценить. Ошибка репрезентативности будет минимальна, если при отборе единиц в выборочную совокупность обеспечена равная возможность попадания в выборку каждой из единиц генеральной совокупности, т.е. соблюдён принцип случайности отбора.

По методу отбора различают повторную и бесповторную выборки.

При повторной выборке каждую из единиц совокупности, попавшую в выборку, после регистрации значения исследуемого признака возвращают в генеральную совокупность, и она имеет возможность вновь попасть в выборку. Повторная выборка встречается редко, чаще применяется схема бесповторной выборки.

При бесповторной выборке отобранные единицы совокупности, после регистрации значения исследуемого признака в генеральную совокупность не возвращают и численность единиц генеральной совокупности в процессе исследования уменьшается.

Введём обозначения основных характеристик параметров генеральной и выборочной совокупностей:

$N$ – объем генеральной совокупности (число единиц генеральной совокупности);

$n$ – объем выборки (число обследованных единиц);

$\overline{x}$ – среднее значение признака в генеральной совокупности (генеральная средняя);

$\tilde{x}$ – среднее значение признака в выборочной совокупности (выборочная средняя);

$p$ – генеральная доля (доля единиц, обладающих данным значением признака в генеральной совокупности);

$w$ – выборочная доля (доля единиц, обладающих данным значением признака в выборочной совокупности);

$\sigma_o^2$ – генеральная дисперсия (дисперсия признака в генеральной совокупности);

$\sigma^2$ – выборочная дисперсия (дисперсия того же признака в выборочной совокупности);

$\sigma_o$ – среднее квадратическое отклонение в генеральной совокупности;

$\sigma$ – среднее квадратическое отклонение в выборке.

При выборочном наблюдении должна быть обеспечена случайность отбора единиц, при этом каждая единица должна иметь равную с другими возможность быть отобранной. Именно на этом основывается наиболее часто применяемая собственно-случайная выборка.

Собственно-случайная выборка формируется путём отбора единиц из всей генеральной совокупности с помощью жеребьевки, с использованием таблицы случайных чисел или иного подобного способа. Принцип случайности предполагает, что на включение или исключение объекта из выборки не может повлиять какой-либо фактор, кроме случая. Примером собственно-случайного отбора могут служить тиражи лотерей, таких как: "Спортлото", "Золотой ключ" и т.д.

В выборку отбирают не произвольное число единиц, а заранее обусловленную их часть – долю выборки. Доля выборки определяется как отношение числа единиц выборочной совокупности к числу единиц генеральной совокупности:

Так, при 5%-й выборке из партии деталей в 200 ед. объем выборки и составляет 10 ед., а при 10%-й выборке – 20 ед.

При выборочном исследовании обычно используют два основных вида обобщающих показателей: среднюю величину количественного признака и относительную величину альтернативного признака.

Выборочная доля $w$ определяется отношением числа единиц, обладающих изучаемым признаком $b$ к общему числу единиц выборочной совокупности $n$:

$$w = b/n. \tag{7.1}$$

Например, если из 100 деталей выборки ($n$ = 100) 97 деталей оказались стандартными ($b$ = 97), то выборочная доля будет:

$$w = 97/100 = 0{,}97\ (97\ \%).$$

## 7.2. Ошибка выборки

Поскольку выборочная совокупность не полностью воспроизводит генеральную совокупность, то между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения, существуют расхождения – ***ошибка выборки – Δ***.

Для средней количественного признака

$$\Delta_{\tilde{x}} = |\overline{x} - \tilde{x}|. \tag{7.2}$$

Для доли (альтернативного признака)

$$\Delta_{w} = |p - w|. \tag{7.3}$$

Выборочная средняя и выборочная доля являются случайными величинами, которые могут принимать различные значения в зависимости от того, какие из единиц генеральной совокупности случайно попали в выборку. Следовательно, ошибки выборки также являются случайными величинами и могут принимать различные значения. Поэтому определяют среднюю из возможных ошибок – ***среднюю ошибку выборки – $\mu$***.

Средняя ошибка выборки $\mu$ для повторного отбора может быть рассчитана по формуле

$$\mu \approx \sqrt{\frac{\sigma^2}{n}}, \tag{7.4}$$

где $\sigma^2$ – выборочная дисперсия (дисперсия, рассчитанная по выборочной совокупности); $n$ – объём выборочной совокупности.

При случайном бесповторном отборе в формуле (7.4) необходимо подкоренное выражение умножить на (1 – $n/N$), поскольку в процессе бесповторной выборки уменьшается численность единиц генеральной совокупности. Для бесповторной выборки формула расчёта средней ошибки выборки $\mu$ примет вид:

$$\mu \approx \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}, \tag{7.5}$$

где $N$ – объём генеральной совокупности.

Формулы (7.4) и (7.5) применимы как для количественных, так и для альтернативных признаков. Для количественного признака выборочная дисперсия рассчитывается по формулам (6.4) и (6.5).

Для несгруппированных признаков

$$\sigma^2 = \sigma_x^2 = \frac{\sum(x_i - \tilde{x})^2}{n}, \qquad (7.6)$$

где $\tilde{x}$ – среднее значение признака в выборочной совокупности (выборочная средняя); $x_1$, $x_2$ … $x_n$ – индивидуальные значения варьирующего признака (варианты).

Для сгруппированных признаков

$$\sigma^2 = \sigma_x^2 = \frac{\sum(x_i' - \tilde{x})^2 f_i}{\sum f_i}, \qquad (7.7)$$

где $x'_i$ – середины интервалов либо варианты; $f_i$ – частоты (численность каждой из групп).

Для альтернативного признака дисперсия рассчитывается на основе формулы (6.9) путём замены $p$ на выборочную долю $w$:

$$\sigma^2 = \sigma_w^2 = w(1 - w). \qquad (7.8)$$

## 7.3. Распространение результатов выборочного наблюдения на генеральную совокупность

На основе выборочных результатов устанавливаются характеристики генеральной совокупности. Выборочные средние и относительные величины распространяют на генеральную совокупность с учетом предела их возможной ошибки.

В каждой конкретной выборке расхождение между выборочной средней и генеральной средней $|\overline{x} - \tilde{x}|$ (7.2) может быть меньше средней ошибки выборки $\mu$, равно ей или больше ее.

Поскольку каждое из расхождений имеет различную вероятность (объективную возможность появления события), то фактические расхождения между выборочной средней и генеральной средней $|\overline{x} - \tilde{x}|$ можно рассматривать как некую предельную ошибку, связанную со средней ошибкой и гарантируемую с определенной вероятностью $F$.

***Предельная ошибка выборки*** для средней величины количественного признака ($\Delta_{\tilde{x}}$) при бесповторном отборе рассчитывается по формуле

$$\Delta_{\tilde{x}} = t\mu_x = t\sqrt{\frac{\sigma_x^2}{n}\left(1 - \frac{n}{N}\right)}, \qquad (7.9)$$

где $t$ – коэффициент доверия, зависящий от вероятности, с которой гарантируется предельная ошибка выборки; $\mu_x$ – средняя ошибка выборки для количественного признака.

Аналогично, предельная ошибка выборки для доли $\mu_w$ при бесповторном отборе (7.3) будет:

$$\Delta_w = t\mu_w = t\sqrt{\frac{\sigma_w^2}{n}\left(1-\frac{n}{N}\right)}\ . \qquad (7.10)$$

В формулах (7.9), (7.10) значения выборочных дисперсий $\sigma_x^{\ 2}$ и $\sigma_w^{\ 2}$ рассчитываются по формулам (7.6) – (7.8).

При случайном повторном отборе формулы расчета предельных ошибок выборки имеют вид:

$$\Delta_{\tilde{x}} = t\mu_x = t\sqrt{\frac{\sigma_x^2}{n}}, \quad \Delta_w = t\mu_w = t\sqrt{\frac{\sigma_w^2}{n}}\ .$$

Формула предельной ошибки выборки вытекает из основных положений теории выборочного метода, сформулированных в ряде теорем теории вероятностей, отражающих закон больших чисел. Русский математик А.М. Ляпунов (1857-1918) дал выражение конкретных значений коэффициента доверия $t$ для различных степеней вероятности в виде функции

$$F(t) = \frac{1}{\sqrt{2\pi}}\int_{-t}^{+t} e^{-\frac{t^2}{2}}\,dt\ . \qquad (7.11)$$

На практике пользуются готовыми таблицами функции (7.11), которые вычислены для различных значений $t$ применительно к случаю нормально распределённой совокупности. В табл. 7.1 приведены некоторые их значения.

Таблица 7.1

Некоторые значения интеграла вероятностей нормального закона распределения

| Кратность ошибки $t$ | 1,0 | 1,5 | 1,96 | 2,0 | 2,5 | 2,58 | 3,0 |
|---|---|---|---|---|---|---|---|
| Вероятность $F(t)$ | 0,683 | 0,866 | 0,950 | 0,954 | 0,988 | 0,990 | 0,997 |

Из формул (7.9) и (7.10) следует, что величина предельной ошибки выборки зависит не только от средней ошибкой выборки μ, но и от коэффициента доверия $t$, значение которого определяет доверительную вероятность $F$, с которой гарантируется точность выборочного исследования. Как видно из табл. 7.1, при $t = 1$ предельная ошибка составит $\Delta = \mu$. Следовательно, с вероятностью 0,683 можно утверждать, что разность между выборочными и генеральными показателями не превысит одной средней ошибки выборки. Иначе говоря, в 68,3 % случаев отклонение выборочного показателя от генерального показателя не выйдет за пределы $\pm 1\mu$. При $t = 2$ с вероятностью 0,954 она не выйдет за пределы $\pm 2\mu$, при $t = 3$ с вероятностью 0,997 – за пределы $\pm 3\mu$ и т.д. В практических расчетах, как правило, задаются вероятностью 0,95 или 0,99. Из приведённых в табл. 7.1

значений функции $F(t)$ видно, что вероятность появления ошибки, равной или большей утроенной средней ошибки выборки, т.е. $\Delta \geq 3\mu$, крайне мала и равна 1 – 0,997 = 0,003 или 0,3 %. Такие маловероятные события считаются практически невозможными, а потому величину $\Delta = 3\mu$ можно принять за предел возможной ошибки выборки.

Рассчитав предельную ошибку выборки, можно провести оценку значений генеральной доли и генеральной средней:

- генеральная средняя

$$\overline{x} = \tilde{x} \pm \Delta_{\overline{x}} = \tilde{x} \pm t\mu_x; \tag{7.12}$$

- генеральная доля

$$p = w \pm \Delta_w = w \pm t\mu_w. \tag{7.13}$$

На основе (7.12) и (7.13) ***доверительные интервалы*** генеральной средней и генеральной доли выразятся следующим образом:

- для средней

$$\tilde{x} - t\mu_x \leq \overline{x} \leq \tilde{x} + t\mu_x; \tag{7.14}$$

- для доли

$$w - t\mu_w \leq p \leq w + t\mu_w. \tag{7.15}$$

Таким образом, можно с любой наперёд заданной вероятностью утверждать, что значение генеральной средней лежит в пределах от $\tilde{x} - t\mu_x$ до $\tilde{x} + t\mu_x$, значение генеральной доли лежит в пределах от $w - t\mu_w$ до $w + t\mu_w$.

**Пример.** Для обследования жилищных условий из 100 000 жителей города методом случайного отбора была сформирована выборочная совокупность в количестве 500 человек.

В результате выборочного обследования жилищных условий 500 жителей города, осуществлённого на основе собственно-случайной выборки, установлено, что из числа обследованных 42 человека проживают в жилых помещениях, не имеющих канализации. Данные об общей площади жилых помещений, приходящейся на одного человека из числа обследованных, приведены в табл. 7.2.

С вероятностью 99 % определить доверительные интервалы, в которых находится среднее значение общей площади жилого помещения, приходящейся на одного жителя города. Определить долю жителей города, не имеющих канализации, гарантируя результат с вероятностью 95 %.

Таблица 7.2

Исходные данные

| Общая площадь на одного человека, м$^2$ | До 10 | 10,0-15,0 | 15,0-20,0 | 20,0-25,0 | 25,0-30,0 | 30,0 и более |
|---|---|---|---|---|---|---|
| Число человек | 23 | 68 | 147 | 172 | 58 | 32 |

Решение.

1) Рассчитаем выборочную среднюю $\tilde{x}$, т.е. определим среднюю площадь, приходящуюся на одного человека из числа обследованных. Для удобства расчёты проведём с использованием табл. 7.3.

Таблица 7.3

Расчёт дисперсии общей площади, приходящейся на одного человека

| Исходные данные | | Расчётные данные | | | |
|---|---|---|---|---|---|
| Общая площадь на одного человека, м$^2$ | Число человек ($f_i$) | Середина интервала ($x_i'$) | $x_i' f_i$ | $x - \tilde{x}$ | $(x_i - \tilde{x})^2 \times f_i$ |
| До 10,0 | 23 | 7,5 | 172,5 | -12,7 | 3 709,7 |
| 10,0-15,0 | 68 | 12,5 | 850,0 | -7,7 | 4 031,7 |
| 15,0-20,0 | 147 | 17,5 | 2 572,5 | -2,7 | 1 071,6 |
| 20,0-25,0 | 172 | 22,5 | 3 870,0 | 2,3 | 909,9 |
| 25,0-30,0 | 58 | 27,5 | 1 595,0 | 7,3 | 3 090,8 |
| 30,0 и более | 32 | 32,5 | 1 040,0 | 12,3 | 4 841,3 |
| Итого | 500 | | 10 100,0 | | 17 655,0 |

Используя формулу (5.2), получим:

$$\tilde{x} = \frac{10\,100}{500} = 20{,}2 \text{ м}^2.$$

2) Определим среднюю ошибку выборки $\mu_x$ площади, для чего предварительно рассчитаем выборочную дисперсию общей площади квартиры $\sigma^2_x$.

По формуле (7.7)

$$\sigma_x^2 = \frac{17\,655}{500} = 35{,}31 \text{ м}^2.$$

Тогда из (7.5) получаем

$$\mu_x = \sqrt{\frac{35{,}1}{500}\left(1 - \frac{500}{100\,000}\right)} = 0{,}265 \text{ м}^2.$$

3) Для определения доверительных интервалов используем формулу (7.14). Коэффициент доверия $t$ определим по табл. 7.1.

По условию задачи доверительные интервалы, в которых находится среднее значение общей площади жилого помещения, необходимо определить с вероятностью 99 %. По указанной таблице доверительной вероятности 0,990 соответствует значение $t$ = 2,58. Подставляя найденные значения в (7.14), получаем доверительные интервалы

$$\tilde{x} - t\mu_x \leq \overline{x} \leq \tilde{x} + t\mu_x;$$

$$20{,}2 - 2{,}58 \times 0{,}265 \leq \tilde{x} \leq 20{,}2 + 2{,}58 \times 0{,}265;$$

$$19{,}516 \leq \tilde{x} \leq 20{,}884.$$

Таким образом, с вероятностью 99 % можно утверждать, что средний размер общей жилой площади, приходящейся на одного жителя города, лежит в пределах от 19,516 до 20,884 м$^2$.

4) Переходим к определению генеральной доли $p$.

По формуле (7.1) рассчитываем долю жителей из числа обследованных, проживающих в жилых помещениях без канализации (выборочную долю $w$):

$$w = b/n = 42/500 = 0{,}084\ (8{,}4\ \%).$$

5) Определим среднюю ошибку выборки $\mu_w$ доли, для чего предварительно рассчитаем выборочную дисперсию $\sigma^2_w$ доли жителей, не имеющих канализации.

По формуле (7.8)

$$\sigma_w^2 = 0{,}084 \times (1 - 0{,}084) = 0{,}0769.$$

Далее из (7.5) получаем

$$\mu_w = \sqrt{\frac{0{,}0769}{500}\left(1 - \frac{500}{100\,000}\right)} = 0{,}0124.$$

6) Для определения доверительных интервалов используем формулу (7.15). Коэффициент доверия $t$ определим по табл. 7.1.

Доверительной вероятности 0,950 соответствует значение $t$ = 1,96. Подставляя найденные значения в (7.15), получаем доверительные интервалы

$$w - t\mu_w \leq p \leq w + t\mu_w;$$

$$0{,}084 - 1{,}96 \times 0{,}0124 \leq p \leq 0{,}084 + 1{,}96 \times 0{,}0124;$$

$$0{,}0597 \leq p \leq 0{,}1083.$$

Таким образом, с вероятностью 95 % можно утверждать, что в целом по городу доля жителей, проживающих в жилых помещениях, не имеющих канализации, находится в пределах от 5,97 до 10,83 %.

## 7.4. Малая выборка

Приведённые в подразделах 7.2 и 7.3 формулы выведены для случая большого числа единиц выборочной совокупности ($n$ > 100).

На практике приходится сталкиваться с небольшими по объёму так называемыми малыми выборками. Под ***малой выборкой*** понимается несплошное статистическое обследование, при котором выборочная совокупность образуется из сравнительно небольшого числа генеральной совокупности. Объём малой выборки не превышает 30 единиц и может доходить до 4 – 5 единиц. Величина ошибки малой выборки вычисляется по формулам, отличным от формул выборочного наблюдения со сравнительно большим объёмом выборки. Средняя ошибка малой выборки $\mu_{мв}$ вычисляется по формуле

$$\mu_{мв} = \sqrt{\frac{\sigma^2}{n-1}}. \tag{7.16}$$

Величина $\sigma^2$ вычисляется на основе данных выборочного наблюдения по формуле

$$\sigma^2 = \frac{\sum(x_i - \tilde{x})^2}{n}.$$

Предельная ошибка малой выборки $\Delta_{мв}$ определяется по формуле

$$\Delta_{мв} = t\mu_{мв}. \tag{7.17}$$

При этом значение коэффициента доверия $t$ зависит не только от заданной доверительной вероятности, но и от числа единиц выборки $n$, т.к. при малой выборке действует особый закон распределения. Для отдельных значений $t$ и $n$ доверительная вероятность малой выборки определяется по специальным таблицам Стьюдента ("Стьюдент" – псевдоним английского статистика В.С. Госсета, начавшего разработку теории малых выборок). Приведём выдержку из таблицы вероятностей Стьюдента в форме, которая обычно используется при решении практических задач (табл. 7.4).

Таблица 7.4

Некоторые значения $t$-критерия Стьюдента

| Объём выборки $n$ | | 4 | 5 | 6 | 8 | 10 | 15 | 20 | 30 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|
| Доверительная вероятность | 0,95 | 3,18 | 2,78 | 2,57 | 2,36 | 2,26 | 2,14 | 2,09 | 2,05 | 1,96 |
| | 0,99 | 5,84 | 4,60 | 4,03 | 3,50 | 3,25 | 2,97 | 2,86 | 2,76 | 2,58 |

При проведении малых выборочных обследований важно иметь в виду, что чем меньше объём выборки, тем больше различие между распределением Стьюдента и нормальным распределением. При минимальном объёме выборки это различие весьма существенно, что указывает на уменьшение точности результатов малой выборки. Так, например, при большом объёме выборки для доверительной вероятности 0,99 коэффициент доверия $t$ = 2,58 (последний столбец табл. 7.1). Для малой выборки объёмом $n$ = 4 единицы для той же доверительной вероятности коэффициент доверия $t$ = 5,84, т.е. в два с лишним раза больше.

**Пример.** Для определения влажности поступившего на предприятие материала было взято десять проб, при этом получены следующие данные о влажности материала в пробах, %: 14, 18, 17, 13, 18, 15, 19, 17, 14, 15.

По данным выборочного обследования установить с вероятностью 95 % пределы, в которых находится средний % влажности материала в данной партии.

1) Рассчитываем среднюю влажность материала в малой выборке (выборочную среднюю $\tilde{x}$):

$$\tilde{x} = \frac{\sum x_i}{n}; \quad \tilde{x} = \frac{160}{10} = 16\ \%.$$

Для удобства расчётов составим табл. 7.5.

Таблица 7.5

Расчет дисперсии

| Номер пробы | Влажность, %, ($x_i$) | $x_i - \tilde{x}$ | $(x_i - \tilde{x})^2$ |
|---|---|---|---|
| 1 | 14 | -2 | 4 |
| 2 | 18 | 2 | 4 |
| 3 | 17 | 1 | 1 |
| 4 | 13 | -3 | 9 |
| 5 | 18 | 2 | 4 |
| 6 | 15 | -1 | 1 |
| 7 | 19 | 3 | 9 |
| 8 | 17 | 1 | 1 |
| 9 | 14 | -2 | 4 |
| 10 | 15 | -1 | 1 |
| Итого | 160 | | 38 |

2) С помощью табл. 7.5 рассчитываем выборочную дисперсию

$$\sigma^2 = \frac{\sum(x_i - \tilde{x})^2}{n}; \quad \sigma^2 = \frac{38}{10} = 3{,}8\ \%.$$

3) По формуле (7.16) определяем среднюю ошибку малой выборки

$$\mu_{\text{мв}} = \sqrt{\frac{3{,}8}{10-1}} = 0{,}65\ \%.$$

4) По табл. 7.4 находим коэффициент доверия $t$. Для доверительной вероятности 0,95 и численности выборки $n$ = 10 значение коэффициента доверия $t$ = 2,26.

5) По формуле (7.17) рассчитываем предельную ошибку малой выборки $\Delta_{мв} = t\mu_{мв}$; $\Delta_{мв}$ = 2,26×0,65 = 1,47 %.

6) Определяем доверительный интервал $\bar{x} = \tilde{x} \pm \Delta_{мв}$; $\bar{x}$ = 16 ± 1,469, т.е. 16 – 1,47 ≤ $\bar{x}$ ≤ 16 + 1,47 или 14,53 ≤ $\bar{x}$ ≤ 17,47.

Таким образом, с вероятностью 95 % можно утверждать, что во всей партии материала средняя влажность находится в пределах от 14,53 до 17,47 %.

# 8. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ

## 8.1. Понятие о статистических рядах динамики

В статистике под ***динамикой*** понимается процесс развития, движения социально-экономических явлений во времени.

***Ряд динамики (динамический ряд, хронологический ряд)*** представляет собой ряд изменяющихся во времени значений статистического показателя, расположенных в хронологическом порядке. Основными элементами ряда динамики являются время $t$ и конкретное значение показателя (уровень ряда) $y$.

В качестве показателей *времени* в рядах динамики выступают либо определённые даты (моменты) времени, либо отдельные периоды времени (годы, месяцы, сутки).

Уровни ряда динамики – показатели, числовые значения которых составляют динамический ряд.

Ряды динамики используются для выявления закономерностей развития общественных явлений во времени. Эти закономерности не проявляются на каждом конкретном уровне, а лишь в тенденции на достаточно длительном отрезке времени. На основную тенденцию динамики накладываются другие, прежде всего случайные, иногда и сезонные влияния.

Постоянно действующие факторы оказывают на изучаемое явление определяющее влияние и формируют в рядах динамики основную тенденцию развития, именуемую ***трендом***. Другие факторы действуют периодически или случайно, вызывая отклонения от тренда.

Ряды динамики разделяются на моментные и интервальные.

***Моментные ряды*** динамики отображают состояние изучаемых явлений на определенные даты (моменты времени).

Пример моментного ряда приведен в табл. 8.1.

Таблица 8.1

Численность населения Российской Федерации
на 1 января соответствующего года

| Годы | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|
| Население, млн чел. | 143,2 | 142,8 | 142,7 | 141,7 | 142,9 | 142,9 | 143,1 |

Примечание: данные за 2010 и 2011 гг. уточнены с учётом предварительных результатов переписи населения 2010 г.

Поскольку в каждом последующем уровне содержится полностью или частично значение предыдущего уровня, суммировать уровни моментного ряда не следует, так как это приводит к повторному счету.

***Интервальные ряды*** динамики отображают итоги развития изучаемых явлений за отдельные периоды времени (год, квартал, месяц).

Пример интервального ряда приведен в табл. 8.2.

Таблица 8.2

Число родившихся в Российской Федерации

| Годы | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|---|
| Родилось, тыс. чел. | 1502 | 1457 | 1480 | 1610 | 1714 | 1762 | 1789 |

Значения уровней интервального ряда в отличие от уровней моментного ряда не содержатся в предыдущих или последующих показателях, их можно просуммировать, что позволяет получать ряды динамики более укрупненных периодов. Например, суммирование числа рождений за каждый год по данным, приведенным выше, позволяет определить число родившихся за все семь лет в целом и в среднем за год.

По расстоянию между уровнями ряды динамики подразделяются на ряды с *равноотстоящими и неравноотстоящими* уровнями по времени. Так, приведенные данные о численности населения Российской Федерации представляют собой ряд динамики с равностоящими уровнями, поскольку данные представлены через равные, следующие друг за другом интервалы времени длиной в один год.

Если в рядах динамики прерывающиеся или неравномерные интервалы времени, то такие ряды являются неравноотстоящими.

В статистике широко используются графические изображения рядов динамики, позволяющих наглядно представить развитие явления во времени и способствующих проведению анализа уровней. Наиболее распространенным видом графического изображения для аналитических целей является линейная диаграмма, которая строится в прямоугольной системе координат: на оси абсцисс отмечается время, а на оси ординат – уровни ряда (рис. 8.1).

Для графического изображения рядов динамики широко используются также и столбиковые диаграммы (рис. 8.2) и другие виды диаграмм (фигурные, квадратные, полосовые и т.п.).

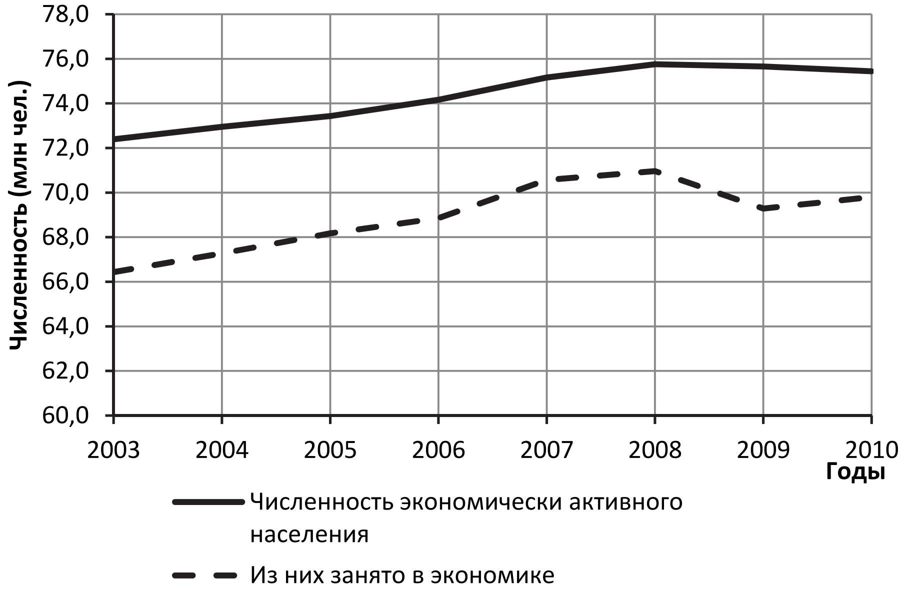


Рис. 8.1. Численность экономически активного населения в Российской Федерации, в среднем за год

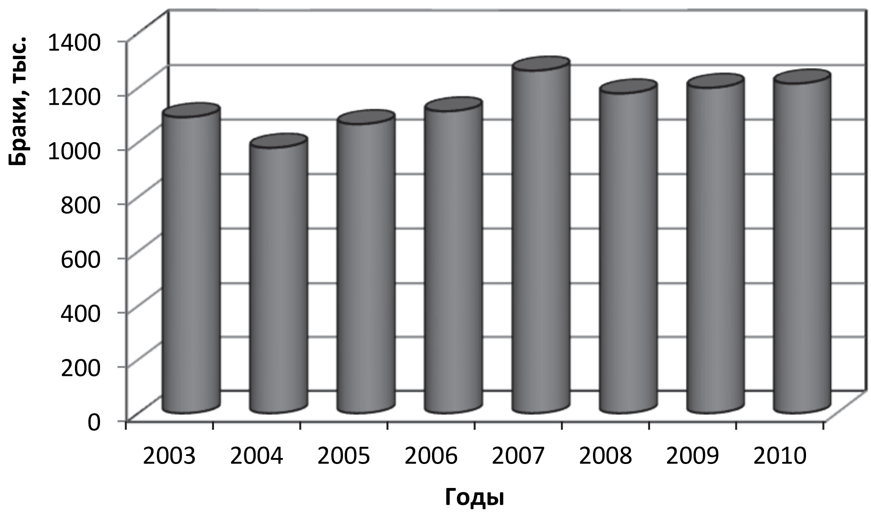


Рис. 8.2. Количество браков, заключённых в Российской Федерации

## 8.2. Сопоставимость в рядах динамики

Основным условием для получения правильных выводов при анализе рядов динамики является сопоставимость его элементов.

Статистические данные должны быть сопоставимы по методологии расчета, территории, кругу охватываемых объектов, единицам измерения, продолжительности показаний времени, ценам и пр.

При определении уровней динамического ряда необходимо использовать единую методологию их расчета. Например, до 1958 г. уровень производительности труда в промышленности определялся в расчете на одного рабочего, а с 1958 г. – на одного работающего (т.е. с включением подсобных рабочих, ИТР и служащих). Поэтому для динамического анализа уровни производительности труда, рассчитанные до 1958 г., необходимо пересчитывать по новой методологии.

Сопоставимость по территории предполагает одни и те же границы территории. Изменение границ влияет на численность населения, объем продукции.

Сопоставимость по кругу охватываемых объектов требует, чтобы сопоставляемые показатели динамического ряда были однородны по экономическому содержанию и границам объекта, который они характеризуют. Например, несопоставимость может возникнуть вследствие перехода ряда объектов (например, предприятий отрасли) из одного подчинения в другое.

При изучении динамики несопоставимость возникает при неодинаковой продолжительности показаний времени (месяцев, кварталов, полугодий), продолжительность високосного года (366 дней) отличается от продолжительности обычного года (365 дней).

При проведении к сопоставимому виду продукции, измеренной в стоимостных показателях, необходимо учитывать как непрерывное изменение цен, так и факт существования нескольких видов цен. Для устранения влияния изменения цен количество продукции, произведенной в разные периоды, оценивают в ценах одного и того же базисного периода, которые называют неизменными, или сопоставимыми ценами.

В ряде случаев несопоставимость может быть устранена с помощью так называемого ***смыкания рядов динамики***. Этот прием позволяет преодолеть несопоставимость данных, возникающую вследствие изменения во времени круга охватываемых объектов, методологии расчета показателей, территориальных границ. Покажем смыкание рядов динамики на примере. Пусть имеются данные о себестоимости единицы продукции, производимой некоторым предприятием. В течение анализируемого периода в 2009 г. произошло изменение методики расчёта себестоимости (табл. 8.3).

Сомкнём два ряда в один, для чего пересчитаем данные за 2007 – 2008 гг. по новой методике. На основе данных о себестоимости единицы продукции в 2009 г., рассчитанных по старой и новой методике, находим соотношение между ними: 288 : 240 = 1,2. Умножая данные за 2007 – 2008 гг. на полученный коэффициент, приводим ряд к сопоставимому виду: $265 \times 1{,}2 = 318$, $245 \times 1{,}2 = 294$ (см. табл. 8.3).

Таблица 8.3

Смыкание рядов динамики

| Годы | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|
| Себестоимость продукции, р.<br>- по старой методике<br>- по новой методике | <br><br>265<br>- | <br><br>245<br>- | <br><br>240<br>288 | <br><br>-<br>280 | <br><br>-<br>298 |
| Сомкнутый ряд абсолютных величин, р. | 318 | 294 | 288 | 280 | 298 |
| Сопоставимый ряд относительных величин, в % к 2009 г. | 265/240= =110,4 | 245/240= =102,1 | 240/240= =100,0 | 280/288= =97,2 | 298/288= =103,5 |

Другой способ смыкания рядов состоит в том, что уровни года, в котором произошли изменения (в нашем примере 2009 г.), принимаются за 100 %, а остальные пересчитываются в процентах по отношению к этим уровням (см. табл. 8.3).

Смыкание рядов дает возможность устранить несопоставимость уровней и составить представление о динамике за весь период, однако при этом следует иметь в виду, что полученные результаты являются приближенными.

При проведении анализа рядов динамики следует предварительно убедиться в сопоставимости их уровней и, если сопоставимость отсутствует, привести ряд к сопоставимому виду, когда это возможно.

## 8.3. Аналитические показатели динамики

Анализ скорости и интенсивности изменения во времени осуществляется с помощью статистических показателей, получаемых в результате сравнения уровней ряда динамики между собой. К таким показателям относятся: абсолютный прирост, темп роста, темп прироста.

Показатели динамики могут вычисляться на постоянной и переменных базах сравнения. При этом сравниваемый уровень называют *отчетным*, а уровень, с которым производится сравнение, – *базисным*.

Для расчета показателей анализа динамики на постоянной базе сравнения каждый уровень ряда сравнивается с одним и тем же базисным уровнем. В качестве базисного обычно выбирается начальный уровень в ряду динамики либо уровень, с которого начинается какой-либо новый этап развития явления. Исчисляемые при этом показатели называются ***базисными***.

Для расчета показателей анализа динамики на переменной базе сравнения каждый последующий уровень ряда сравнивается с предыдущим уровнем, вычисленные таким образом показатели динамики называются ***цепными***.

Простейшим показателем динамики является *абсолютный прирост* (сокращение), который показывает увеличение или уменьшение уровня ряда за определенный промежуток времени.

***Абсолютный прирост базисный***

$$\Delta y_{бi} = y_i - y_o, \tag{8.1}$$

где $y_i$ – уровень сравниваемого (отчётного) периода; $y_o$ – уровень базисного периода.

***Абсолютный прирост цепной***

$$\Delta y_{цi} = y_i - y_{i-1}, \tag{8.2}$$

где $y_{i-1}$ – уровень предшествующего периода.

В табл. 8.4 базисные и цепные абсолютные приросты, рассчитанные по (8.1) – (8.2), показывают абсолютное изменение по сравнению с 2007 г. валового сбора зерновых и зернобобовых культур в РФ и прирост (сокращение) по годам.

Таблица 8.4

Показатели динамики валового сбора зерновых
и зернобобовых культур в РФ

| Показатель | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|
| Валовой сбор, млн т, $y_i$ | 81,5 | 108,2 | 97,1 | 61,0 |
| Абсолютный базисный прирост $\Delta y_{бi}$ | - | 108,2-81,5= 26,7 | 97,1-81,5= 15,6 | 61,0-81,5= -20,5 |
| Абсолютный цепной прирост $\Delta y_{цi}$ | - | 108,2-81,5= 26,7 | 97,1-108,2= -11,1 | 61,0-97,1= -36,1 |
| Темп роста базисный, %, $Tp_{бi}$ | - | 108,2/81,5*100 =132,8 | 97,1/81,5*100= 119,1 | 61,0/81,5*100= 74,8 |
| Темп роста цепной, %, $Tp_{цi}$ | - | 108,2/81,5*100 =132,8 | 97,1-108,2*100= 89,7 | 61,0-97,1*100= 62,8 |
| Темп прироста базисный, %, $Tп_{бi}$ | - | 26,7/81,5*100= 32,8 | 15,6/81,5*100= 19,1 | -20,5/81,5*100= -25,2 |
| Темп прироста цепной, %, $Tп_{цi}$ | - | 26,7/81,5*100= 32,8 | -11,1/108,2*100= -10,3 | -36,1/97,1*100= -37,2 |

Цепные и базисные абсолютные приросты связаны между собой: алгебраическая сумма последовательных цепных абсолютных приростов равна базисному приросту за весь промежуток времени, т.е.

$$\sum \Delta y_{цi} = \Delta y_{бm}.$$

По данным табл. 8.4: 26,7 – 11,1 – 36,1 = –20,5 (млн т).

Для оценки интенсивности изменения уровня ряда динамики за какой-либо период времени рассчитывают темпы роста (снижения). Интенсивность изменения уровня оценивается отношением отчетного уровня к базисному. Показатель интенсивности изменения уровня ряда, выраженный в процентах, называется темпом роста, а долях единицы – коэффициентом роста.

Если коэффициент роста (снижения) больше единицы, то он показывает, во сколько раз сравниваемый уровень больше уровня, с которым производится сравнение. Если коэффициент роста (снижения) меньше единицы, то следует говорить не о росте, а о снижении. В этом случае коэффициент показывает, какую часть уровня, с которым производится сравнение, составляет сравниваемый уровень.

***Коэффициент роста базисный***

$$Тр_{бi} = y_i / y_o. \tag{8.3}$$

***Коэффициент роста цепной***

$$Кр_{цi} = y_i / y_{i-1}. \tag{8.4}$$

Если отношение выражено в процентах, то говорят о темпе роста.

***Темп роста базисный***

$$Тр_{бi} = (y_i / y_o) \cdot 100~\%. \tag{8.5}$$

***Темп роста цепной***

$$Тр_{цi} = (y_i / y_{i-1}) \cdot 100~\%. \tag{8.6}$$

Несложно заметить, что $Тр = Кр \times 100~\%$.

Базисные и цепные темпы роста, характеризующие интенсивность изменения валового сбора зерновых и зернобобовых культур в РФ по годам и за весь период, исчислены в табл. 8.4.

Связь цепных и базисных коэффициентов роста: произведение последовательных цепных коэффициентов роста равно базисному коэффициенту роста за весь период: $\prod Кр_{цi} = Кр_n$, а частное от деления последующего базисного темпа роста на предыдущий – соответствующему цепному темпу роста.

На примере табл. 8.4: 1,328 × 0,897 × 0,628 = 0,748.

*Темп прироста (сокращения)* показывает, на сколько процентов сравниваемый уровень больше или меньше уровня, принятого за базу сравнения. Вычисляется как отношение абсолютного прироста к абсолютному уровню, принятому за базу сравнения.

Темп прироста может быть положительным (положительная динамика), отрицательным (отрицательная динамика) или равным нулю. Темп прироста выражается в процентах. Темп прироста, выраженный в долях единицы, называют коэффициентом прироста.

***Темп прироста базисный***

$$Tn_{бi} = \frac{\Delta y_{бi}}{y_o} = \frac{y_i - y_o}{y_o} \cdot 100\ \%. \tag{8.7}$$

***Темп прироста цепной***

$$Tn_{цi} = \frac{\Delta y_{цi}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\ \%. \tag{8.8}$$

Темп прироста и коэффициент прироста связаны с темпом роста и коэффициента роста простыми и удобными в практических расчётах соотношениями:

$$Tn = Tp - 100\ \%; \quad Kn = Kn - 1.$$

Пример расчёта базисных и цепных темпов прироста (формулы (8.3) – (8.8)), характеризующих интенсивность изменения валового сбора зерновых и зернобобовых культур в РФ, приведён в табл. 8.4.

## 8.4. Средние по рядам динамики

Средние показатели рядов динамики дают обобщающую характеристику динамики исследуемого явления.

*Средний уровень* ряда рассчитывается по средней хронологической, т.е. по средней исчисленной из значений, изменяющихся во времени.

Средний уровень для интервального и моментного рядов динамики рассчитывается по-разному.

Для интервальных рядов динамики средний уровень за период времени определяется по формуле средней арифметической. При равных интервалах применяется средняя арифметическая простая:

$$\overline{y} = \frac{\sum y_i}{n} = \frac{y_1 + y_2 + y_3 \ldots + y_n}{n}, \tag{8.9}$$

где $y_i$ – абсолютные уровни ряда; $n$ – число уровней ряда.

При неравных интервалах используется средняя арифметическая взвешенная:

$$\overline{y} = \frac{\sum y_i t_i}{\sum t_i} = \frac{y_1 t_1 + y_2 t_2 + y_3 t_3 \ldots + y_n t_n}{t_1 + t_2 + t_3 \ldots + t_n}, \tag{8.10}$$

где $y_1 \ldots y_n$ – уровни ряда динамики, сохраняющиеся без изменения в течение промежутка времени $t_1 \ldots t_n$; $t_1 \ldots t_n$ – период действия уровней $y_1 \ldots y_n$.

Средний уровень валового сбора зерновых и зернобобовых культур в РФ в 2007 – 2010 гг. (см. табл. 8.4) находим по формуле (8.9), так как исследуемый ряд динамики представляет собой интервальный ряд с одинаковыми интервалами:

$$\overline{y} = \frac{81{,}5 + 108{,}2 + 97{,}1 + 61{,}0}{4} = 86{,}95\ (\text{млн т}).$$

Расчет среднего уровня для интервального ряда динамики с неравноотстоящими уровнями рассмотрим на примере.

**Пример.** Если известно, что с 1-го по 15-е число месяца на предприятии работало 40 человек, с 16-го по 25-е – 54 человека, а с 26-го по 30-е – 60 человек, то среднесписочное число работников за месяц в соответствии с формулой (8.10) составит, чел.:

$$\overline{y} = \frac{40 \cdot 15 + 54 \cdot 10 + 60 \cdot 5}{15 + 10 + 5} = 48.$$

Средний уровень моментного ряда динамики с равностоящими уровнями определяется по формуле средней хронологической моментного ряда:

$$\overline{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}, \qquad (8.11)$$

где $y_i$ … $y_n$ – уровни периода, за который делается расчёт; $n$ – число уровней ряда.

**Пример.** Известны товарные остатки магазина на 1-е число каждого месяца, тыс. р., рассчитать среднемесячный товарный остаток на I квартал.

| 1 января | 1 февраля | 1 марта | 1 апреля |
|---|---|---|---|
| 170 | 140 | 130 | 190 |

По формуле (8.11)

$$\overline{y} = \frac{\frac{170}{2} + 140 + 130 + \frac{190}{2}}{4-1} = 150 \text{ (тыс. р.).}$$

*Средний абсолютный прирост* (убыль) по цепным данным об абсолютных приростах за ряд периодов рассчитывается как средняя арифметическая простая:

$$\overline{\Delta y_{ц}} = \frac{\sum \Delta y_{цi}}{n}. \qquad (8.12)$$

Определим средний абсолютный прирост (снижение), используя формулу (8.12) и данные табл. 8.4 о цепных абсолютных приростах валового сбора зерновых и зернобобовых культур в РФ:

$$\overline{\Delta y_{ц}} = \frac{26{,}7 - 11{,}1 - 36{,}1}{3} = -6{,}83 \text{ (млн т.).}$$

Средний абсолютный прирост может определяться по абсолютным уровням ряда динамики. Для этого определяется разность между конечным $y_n$ и базисным $y_1$ уровнями изучаемого периода, которая делится на $n$-1 субпериодов:

$$\overline{\Delta y_{ц}} = \frac{y_n - y_o}{n-1}. \qquad (8.13)$$

Определим средний абсолютный прирост (снижение) по формуле (8.13), используя данные табл. 8.4:

$$\overline{\Delta y_ц} = \frac{61,0 - 81,5}{4 - 1} = -6,83 \text{ (млн т)}.$$

*Средний темп роста* (снижения) для цепных показателей рассчитывается по формуле

$$\overline{Tp} = \sqrt[n]{Tp_1 \cdot Tp_2 \ldots \cdot Tp_n} = \sqrt[n]{\prod Tp_i}. \qquad (8.14)$$

Определим средний цепной темп роста (снижения) по формуле (8.14), используя данные табл. 8.4:

$$\overline{Tp} = \sqrt[3]{132,8 \cdot 89,7 \cdot 62,8} = 90,79\ \%.$$

Следовательно, с 2007 по 2010 гг. валовой сбор зерновых и зернобобовых культур в России снижался в среднем на 9,21 % в год (100 % – – 90,79 % = 9,21 %).

Если известны уровни динамического ряда, то расчет среднего коэффициента роста может быть определён по формуле

$$\overline{Kp} = \sqrt[n-1]{{y_n}/{y_1}}, \qquad (8.15)$$

где $n$ – число уровней ряда динамики в изучаемом периоде, включая базисный.

Подставляем значения в формулу (8.15)

$$\overline{Kp} = \sqrt[3]{61,0/\cdot 81,5} = 0,9079.$$

Результат совпадает с предыдущим.

*Средние темпы прироста* (сокращения) рассчитываются на основе средних темпов роста вычитанием из последних 100 %. Соответственно при исчислении средних коэффициентов прироста из значений коэффициентов роста вычитается единица:

$$\overline{Tn} = \overline{Tp} - 100\ \%; \quad \overline{Kn} = \overline{Kp} - 1.$$

Если уровни ряда динамики снижаются (отрицательная динамика), то средний темп роста будет меньше 100 %, а средний темп прироста – отрицательной величиной. Отрицательный темп прироста $\overline{Tn}$ представляет собой средний темп сокращения и характеризует среднюю относительную скорость снижения уровня.

Так, в нашем примере среднегодовой темп прироста валового сбора зерновых и зернобобовых культур характеризуется отрицательным значением (90,79 % – 100 % = – 9,21 %), что свидетельствует о ежегодном сокращении сбора зерновых и зернобобовых культур.

## 8.5. Изучение основной тенденции развития

Важным направлением исследований закономерностей динамики является определение в рядах динамики общей тенденции развития изучаемого явления. В некоторых случаях закономерность изменения явления, общая тенденция его развития явно и отчетливо отражается уровнями динамического ряда (уровни на изучаемом периоде непрерывно растут или непрерывно снижаются).

Однако часто приходится встречаться с такими рядами динамики, в которых уровни ряда претерпевают самые различные изменения (то возрастают, то убывают), и общая тенденция развития неясна. На развитие явления во времени оказывают влияние факторы, различные по характеру и силе воздействия. Некоторые из них оказывают практически постоянное воздействие и формируют в рядах динамики определенную тенденцию развития. Воздействие же других факторов может быть кратковременным или носить случайный характер. Поэтому при анализе динамики речь идет не просто о тенденции развития, а об основной тенденции, достаточно стабильной (устойчивой) на протяжении изученного этапа развития.

***Основной тенденцией развития (трендом)*** называется плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний.

Задача состоит в том, чтобы выявить общую тенденцию в изменении уровней ряда, освобожденную от действия различных случайных факторов. На практике наиболее распространённым методом статистического изучения тренда являются: укрупнение интервалов, сглаживание скользящей средней, аналитическое выравнивание.

***Метод укрупнения интервалов*** основан на укрупнении периодов времени, к которым относятся уровни ряда динамики, одновременно уменьшается количество интервалов. Например, ряд ежесуточного выпуска продукции заменяется рядом месячного выпуска продукции и т.д. Рассмотрим применение метода укрупнения интервалов на ежемесячных данных об объёме реализации металлопродукции со склада (табл. 8.5).

Таблица 8.5

Объём реализации металлопродукции со склада

| Месяц | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Объём реализации, тыс. т | 2,5 | 3,1 | 2,6 | 2,6 | 3,4 | 3,2 | 2,9 | 3,7 | 3,5 | 3,2 | 3,8 | 3,6 |

Различные направления изменений уровней ряда по отдельным месяцам затрудняют выводы об основной тенденции реализации. Если соответствующие месячные уровни объединить в квартальные и вычислить реализацию металлопродукции по кварталам (табл. 8.6), т.е. укрупнить интервалы, то решение задачи упрощается.

Таблица 8.6

Укрупнение интервалов

| Квартал | I | II | III | IV |
|---|---|---|---|---|
| Объём реализации металлопродукции, тыс. т | 8,2 | 9,2 | 10,1 | 10,6 |

После укрупнения интервалов основная тенденция роста реализации металлопродукции стала очевидной:

$$8{,}4 < 9{,}4 < 10{,}1 < 10{,}6 \text{ тыс. р.}$$

***Метод скользящей (подвижной) средней.*** Сущность его заключается в том, что исчисляется средний уровень из определенного числа, обычно нечетного (3, 5, 7 и т.д.), первых по счету уровней ряда, затем – из такого же числа уровней, но начиная со второго по счету, далее – начиная с третьего и т.д. Таким образом, средняя как бы "скользит" по ряду динамики, передвигаясь на один срок.

Расчет скользящей средней по данным об объёме реализации металлопродукции со склада приведён в табл. 8.7, графическая иллюстрация – на рис. 8.3. Сглаживание скользящими средними выявляет устойчивую тенденцию к увеличению объёмов реализации.

Таблица 8.7

Расчет скользящей средней

| Месяц | Объём реализации, тыс. т | Скользящая средняя трёхчленная |
|---|---|---|
| 1 | 2,5 | - |
| 2 | 3,1 | $\overline{y_2} = (y_1 + y_2 + y_3) = (2{,}5 + 3{,}1 + 2{,}5) / 3 = 2{,}7$ |
| 3 | 2,5 | $\overline{y_3} = (y_2 + y_3 + y_4) = (3{,}1 + 2{,}5 + 3{,}0) / 3 = 2{,}87$ |
| 4 | 3,0 | $\overline{y_4} = (y_3 + y_4 + y_5) = (2{,}5 + 3{,}0 + 3{,}1) / 3 = 2{,}87$ |
| 5 | 3,1 | $\overline{y_5} = (y_4 + y_5 + y_6) = (3{,}0 + 3{,}1 + 2{,}9) / 3 = 3{,}0$ |
| 6 | 2,9 | $\overline{y_6} = (y_5 + y_6 + y_7) = (3{,}1 + 2{,}9 + 3{,}1) / 3 = 3{,}03$ |
| 7 | 3,1 | $\overline{y_7} = (y_6 + y_7 + y_8) = (2{,}9 + 3{,}1 + 3{,}6) / 3 = 3{,}20$ |
| 8 | 3,6 | $\overline{y_8} = (y_7 + y_8 + y_9) = (3{,}1 + 3{,}6 + 3{,}7) / 3 = 3{,}47$ |
| 9 | 3,7 | $\overline{y_9} = (y_8 + y_9 + y_{10}) = (3{,}6 + 3{,}7 + 3{,}4) / 3 = 3{,}57$ |
| 10 | 3,4 | $\overline{y_{10}} = (y_9 + y_{10} + y_{11}) = (3{,}7 + 3{,}4 + 3{,}9) / 3 = 3{,}67$ |
| 11 | 3,9 | $\overline{y_{11}} = (y_{10} + y_{11} + y_{12}) = (3{,}4 + 3{,}9 + 3{,}8) / 3 = 3{,}70$ |
| 12 | 3,8 | - |

Сглаженный ряд реализации металлопродукции при сглаживании скользящими трёхчленными средними короче фактического на один член ряда в начале и в конце.

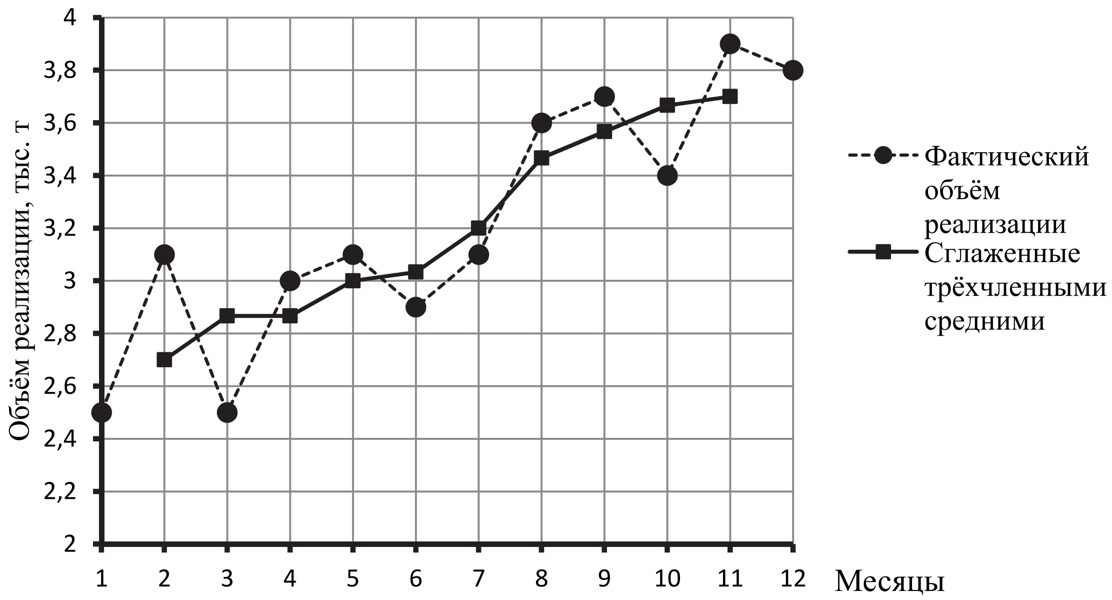


Рис. 8.3. Сглаживание трёхчленными средними

Как видно из табл. 8.7, при сглаживании трёхчленными скользящими средними средние уровни исчисляются по формуле

$$\overline{y_i} = \frac{y_{i-1}+y_i+y_{i+1}}{3},$$

где $\overline{y_i}$ – сглаженное значение $i$-го уровня; $y_i$ – фактическое значение $i$-го уровня.

При сглаживании пятичленными скользящими средние уровни исчисляются по формуле, при этом сглаженный ряд короче фактического ряда на два члена в начале и конце ряда:

$$\overline{y_i} = \frac{y_{i-2}+y_{i-1}+y_i+y_{i+1}+y_{i+2}}{5}.$$

Сглаживание по 7, 9 и т.д. членам ряда проводится аналогично.

При сглаживании по чётному числу членов ряда (4, 6, 8 и т.д.) используется несколько иная формула. Так, при сглаживании по четырём членам ряда она имеет вид:

$$\overline{y_3} = \frac{\frac{y_1}{2}+y_2+y_3+y_4+\frac{y_5}{2}}{4}, \ldots \overline{y_i} = \frac{\frac{y_{i-2}}{2}+y_{i-1}+y_i+y_{i+1}+\frac{y_{i+2}}{2}}{4}.$$

Сглаживание по 6, 8 и т.д. членам ряда проводится аналогично.

Недостатком сглаживания ряда является потеря информации вследствие "укорачивания" сглаженного ряда по сравнению с фактическим.

Укрупнение интервалов и метод скользящей средней позволяют определить лишь общую тенденцию развития явления, в некоторой степени освобождённую от случайных и волнообразных колебаний. Однако получить обобщенную статистическую модель тренда посредством этих методов нельзя.

Для получения количественной модели, выражающей основную тенденцию изменения уровней ряда динамики во времени, используется аналитическое выравнивание ряда динамики.

Основным содержанием ***метода аналитического выравнивания*** в рядах динамики является то, что общая тенденция развития рассчитывается как функция времени:

$$\hat{y}_t = f(t).$$

Определение теоретических (расчетных) уровней $\hat{y}_t$ производится на основе так называемой адекватной математической модели, которая наилучшим образом отображает (аппроксимирует) основную тенденцию ряда динамики.

Выбор типа модели зависит от цели исследования и должен быть основан на теоретическом анализе, выявляющем характер развития явления, а также на графическом изображении ряда динамики (линейной диаграмме).

На практике чаще всего используются следующие эталонные типы трендов:

1) Линейная функция

$$\hat{y}_t = a_0 + a_1 t. \quad (8.16)$$

Описывает равномерное развитие, применяется для рядов динамики со стабильными абсолютными цепными приростами, т.е. в случае, когда $\Delta y_{цi} \approx const.$

Параметр $a_0$ является коэффициентом регрессии, определяющим направление развития. Если $a_0 > 0$, то уровни рядов динамики равномерно возрастают, при $a_0 < 0$ происходит равномерное снижение.

2) Парабола второго порядка

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2. \quad (8.17)$$

Описывает равноускоренное (равнозамедленное) развитие, применяется для сглаживания рядов динамики со стабильными цепными темпами прироста, т.е. в случае, когда $Tп_{цi} \approx const.$

3) Парабола третьего порядка

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

Применяется для описания развития с переменным ускорением (замедлением).

4) Показательная функция

$$\hat{y}_t = a_0 a_1^{\,t}. \quad (8.18)$$

Применяется в случаях развития процесса по экспоненте, эффективна для сглаживания рядов динамики со стабильными цепными темпами роста, т.е. в случае, когда $Тр_{цi} \approx const$.

5) Логарифмическая функция

$$\hat{y}_t = a_0 + a_1 lg\, t. \tag{8.19}$$

Хорошо описывает процессы развития с замедлением роста в конце периода, эффективна для сглаживания рядов динамики с убывающими абсолютными цепными приростами, т.е. в случае, когда $\Delta y_{цп} \to 0$.

Для описания рядов динамики, кроме (8.16) – (8.19), используются также степенная $\hat{y}_t = a_0 t^{a_1}$, гиперболическая $\widehat{y_t} = a_o + a_1 \frac{1}{t}$ и другие математические функции.

Расчет параметров $a_o$, $a_1$, $a_2$ адекватной математической функции обычно производится методом наименьших квадратов, в котором в качестве решения принимается точка минимума суммы квадратов разностей между теоретическими $\hat{y}_{ti}$ и фактическими (эмпирическими) $y_i$ уровнями:

$$\sum(\hat{y}_{ti} - y_i)^2 \to min, \tag{8.20}$$

где $\hat{y}_{ti}$ – теоретические (выровненные) уровни; $y_i$ – фактические уровни.

В случае использовании для описания ряда динамики прямой вида $\hat{y}_t = a_0 + a_1 t$, после подстановки $\hat{y}_{ti}$ из (8.16) в уравнение (8.20) получаем

$$\sum(a_0 + a_1 t_i - y_i)^2 \to min.$$

После нахождения частных производных по искомым параметрам уравнения $a_0$ и $a_1$ приходим к системе нормальных линейных уравнений

$$\begin{cases} a_0 n + a_1 \sum t_i = \sum y_i \\ a_0 \sum t_i + a_1 \sum t_i^2 = \sum y_i t_i \,, \end{cases}$$

где $n$ – количество уровней ряда; $t_i$ – время (порядковый номер периода или момента времени).

Решая полученную систему методом определителей, получаем искомые значения параметров линейного тренда $\hat{y}_t = a_0 + a_1 t$:

$$a_0 = \frac{\sum y_i \sum t_i^2 - \sum(y_i t_i) \sum t_i}{n \sum t_i^2 - (\sum t_i)^2}\,, \tag{8.21}$$

$$a_1 = \frac{n \sum(y_i t_i) - \sum y_i \sum t_i}{n \sum t_i^2 - (\sum t_i)^2}\,. \tag{8.22}$$

**Пример.** Известны данные о производстве молока и молокопродуктов в Российской Федерации (табл. 8.8). Провести выравнивание ряда динамики с помощью линейной функции.

Расчёт параметров линейного тренда проведём по формулам (8.21) и (8.22), необходимые для подстановки в формулы данные вычислим в табл. 8.8.

Таблица 8.8

Данные о производстве молока и молокопродуктов
в Российской Федерации

| Исходные данные | | Расчётные данные | | | |
|---|---|---|---|---|---|
| Год | Производства молока и молокопродуктов (млн т) $y_i$ | $t_i$ | $t_i^2$ | $t_i y_i$ | $\hat{y}_{ti}$ |
| 2005 | 30,8 | 1 | 1 | 30,8 | 31,06 |
| 2006 | 31,1 | 2 | 4 | 62,2 | 31,35 |
| 2007 | 32,0 | 3 | 9 | 96,0 | 31,65 |
| 2008 | 32,4 | 4 | 16 | 129,6 | 31,95 |
| 2009 | 32,6 | 5 | 25 | 163,0 | 32,25 |
| 2010 | 31,9 | 6 | 36 | 191,4 | 32,54 |
| Итого | 190,8 | 21 | 91 | 673,0 | 190,80 |

Для упрощения вычислений обозначим 2005 г. как $t_1 = 1$, 2006 г. как $t_2 = 2$ и т.д. Подставляя в формулы (8.21) и (8.22) рассчитанные в табл. 8.8 данные $\sum y_i = 190{,}8$, $\sum t_i = 21$, $\sum t_i^2 = 91$, $\sum y_i t_i = 673{,}0$, $n = 6$, получаем:

$$a_0 = \frac{190{,}8 \times 91 - 673{,}0 \times 21}{6 \times 91 - 21^2} = 30{,}76, \quad a_1 = \frac{6 \times 673{,}0 - 190{,}8 \times 21}{6 \times 91 - 21^2} = 0{,}297.$$

Уравнение прямой, представляющее собой трендовую модель искомой функции, будет иметь вид: $\hat{y}_t = a_0 + a_1 t = 30{,}76 + 0{,}297\,t$.

Подставляя в найденное уравнение значения $t_1 = 1$, $t_2 = 2$ и т.д., находим теоретические уровни тренда $\hat{y}_{t1} = 31{,}06$, $\hat{y}_{t2} = 31{,}35$ и т.д. Теоретические уровни тренда вписаны в последний столбец табл. 8.8. Как видно из таблицы, $\sum y_i = \sum \hat{y}_{ti} = 190{,}8$, что свидетельствует о правильности решения. Полученные результаты проиллюстрируем рис. 8.4.

Для упрощения расчётов используется способ *отсчёта времени от условного начала*. Он основан на обозначении в ряду динамики показаний времени таким образом, чтобы $\Sigma t_i = 0$. При этом в ряду динамики с нечётным числом уровней порядковый номер уровня, находящегося в середине ряда, обозначают через нулевое значение и принимают его за условное начало отсчёта времени с интервалом +1 для всех последующих уровней и –1 для предыдущих уровней. Например, при $n = 5$ обозначения времени будут: –2, –1, 0 ,+1, +2. При чётном числе уровней порядковые номера берутся через 2. Например, $n = 6$ обозначения времени будут: –5, –3, –1, +1, +3, +5.

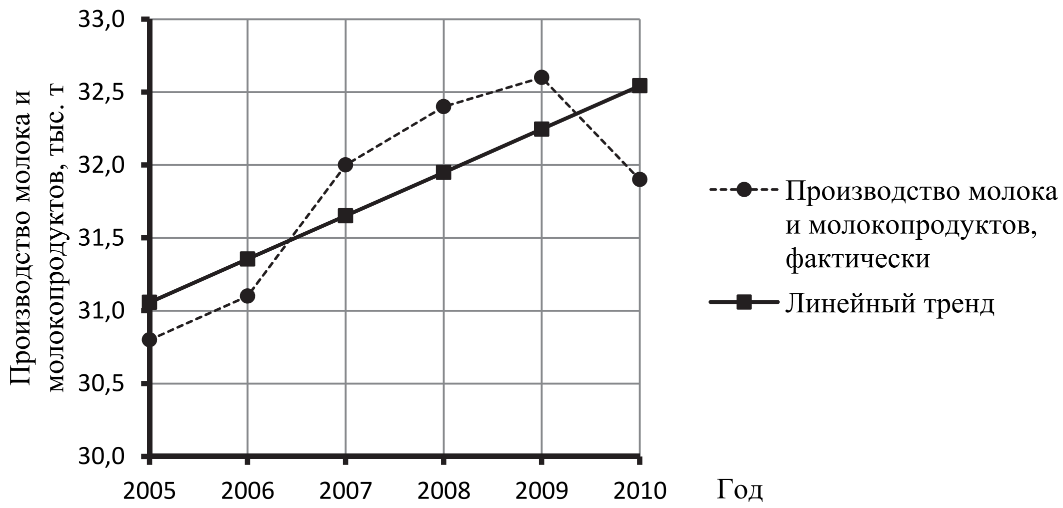


Рис. 8.4. Выравнивание с использованием линейной функции

При использовании способа отсчёта времени от условного начала, когда $\Sigma t_i = 0$, формулы для расчёта параметров $a_o$, $a_1$, $a_2$ примут вид:

а) Для линейной функции $\hat{y}_t = a_0 + a_1 t$ (при $\Sigma t_i = 0$):

$$a_0 = \frac{\sum y_i}{n}, \qquad a_1 = \frac{\sum(y_i t_i)}{\sum t_i^2}.$$

б) Для параболы второго порядка $\hat{y}_t = a_0 + a_1 t + a_2 t^2$ (при $\Sigma t_i = 0$):

$$a_0 = \frac{\sum t_i^4 \sum y_i - \sum t_i^2 \sum(t_i^2 y_i)}{n\sum t_i^4 - (\sum t_i^2)^2}, \; a_1 = \frac{\sum(y_i t_i)}{\sum t_i^2}, \; a_2 = \frac{n\sum(t_i^2 y_i) - \sum t_i^2 \sum y_i}{n\sum t_i^4 - (\sum t_i^2)^2}.$$

в) Для показательной функции $\hat{y}_t = a_0 a_1^t$ (при $\Sigma t_i = 0$):

$$\lg a_0 = \frac{\sum \lg y_i}{n}, \qquad \lg a_1 = \frac{\sum(t_i \lg y_i)}{\sum t_i^2}.$$

При изучении социально-экономических явлений приходится иметь дело со сложным механизмом взаимодействия факторов, формирующих тренд, а потому выбор адекватной математической функции весьма затруднён. В качестве показателя адекватности математической функции может использоваться стандартизованная ошибка аппроксимации $\sigma_{yi}$:

$$\sigma_{yi} = \sqrt{\frac{\sum(\hat{y}_{ti} - y_i)^2}{n}}. \qquad (8.23)$$

За наиболее адекватную из нескольких рассчитанных функций принимается та, у которой стандартизованная ошибка аппроксимации минимальна.

**Пример.** Известны данные о производстве молока и молокопродуктов в Российской Федерации (табл. 8.9). Подобрать адекватную математическую функцию, провести выравнивание ряда динамики с её помощью.

Таблица 8.9

Расчет параметров

| Исходные данные | | Расчётные данные | | | | | Линейный тренд | | Параболический тренд | |
|---|---|---|---|---|---|---|---|---|---|---|
| Год | Производство (тыс. т) $y_i$ | $t_i$ | $t_i^2$ | $t_i^4$ | $t_iy_i$ | $t_i^2y_i$ | $\hat{y}_{ti}$ | $(y_{ti} - y_i)^2$ | $\hat{y}_{ti}$ | $(y_{ti} - y_i)^2$ |
| 2005 | 30,8 | -5 | 25 | 625 | -154,0 | 770,0 | 31,06 | 0,0661 | 30,59 | 0,043 |
| 2006 | 31,1 | -3 | 9 | 81 | -93,3 | 279,9 | 31,35 | 0,0647 | 31,45 | 0,121 |
| 2007 | 32,0 | -1 | 1 | 1 | -32,0 | 32,0 | 31,65 | 0,1215 | 32,02 | 0,001 |
| 2008 | 32,4 | +1 | 1 | 1 | 32,4 | 32,4 | 31,95 | 0,2038 | 32,32 | 0,006 |
| 2009 | 32,6 | +3 | 9 | 81 | 97,8 | 293,4 | 32,25 | 0,1255 | 32,34 | 0,068 |
| 2010 | 31,9 | +5 | 25 | 625 | 159,5 | 797,5 | 32,54 | 0,4133 | 32,08 | 0,032 |
| Итого | 190,8 | 0 | 70 | 1414 | 10,4 | 2205,2 | 190,8 | 0,9949 | 190,8 | 0,271 |

Используем способ отсчёта времени от условного начала, для чего обозначим в ряду динамики показания времени таким образом, чтобы $\Sigma t_i = 0$. Для расчёта параметров уравнений используем соответствующие формулы.

Линейная функция $\hat{y}_t = a_0 + a_1 t$ .

$$a_0 = \frac{\sum y_i}{n} = \frac{190,8}{6} = 31,8, \qquad a_1 = \frac{\sum(y_i t_i)}{\sum t_i^2} = \frac{10,4}{70} = 0,1486.$$

Уравнение прямой будет иметь вид: $\hat{y}_t = 31,8 + 0,1486\,t$. Подставляя в найденное уравнение $t_1 = -5$, $t_2 = -3$, $t_3 = -1$ и т.д., определим теоретические (расчетные) уровни для линейного тренда и рассчитаем стандартизованную ошибку аппроксимации по формуле (8.23):

$$\sigma_{yi} = \sqrt{\frac{\sum(\hat{y}_{ti} - y_i)^2}{n}}; \quad \sigma_{yi} = \sqrt{\frac{0,9949}{6}} = 0,4072 \text{ (тыс. т).}$$

Параболическая функция второго порядка $\hat{y}_t = a_0 + a_1 t + a_2 t^2$:

$$a_0 = \frac{\sum t_i^4 \sum y_i - \sum t_i^2 \sum(t_i^2 y_i)}{n \sum t_i^4 - (\sum t_i^2)^2}; \qquad a_0 = \frac{1414 \times 190,8 - 70 \times 2205,2}{6 \times 1414 - 70 \times 70} = 32,206,$$

$$a_1 = \frac{\sum(y_i t_i)}{\sum t_i^2}; \qquad a_1 = \frac{10,4}{70} = 0,1486,$$

$$a_2 = \frac{n \sum(t_i^2 y_i) - \sum t_i^2 \sum y_i}{n \sum t_i^4 - (\sum t_i^2)^2}; \qquad a_2 = \frac{6 \times 2205,2 - 70 \times 190,8}{6 \times 1414 - 70 \times 70} = -0,0348.$$

Уравнение параболы будет иметь вид: $\hat{y}_t = 32,206 + 0,1486\,t - 0,0348\,t^2$. Определив теоретические (расчетные) уровни для параболического тренда, рассчитаем стандартизованную ошибку аппроксимации:

$$\sigma_{yi} = \sqrt{\frac{\sum(\hat{y}_{ti} - y_i)^2}{n}}; \quad \sigma_{yi} = \sqrt{\frac{0{,}271}{6}} = 0{,}2124 \text{ (тыс. т)}.$$

Сравнив стандартизованные ошибки аппроксимации, делаем вывод, что параболическая функция более адекватно описывает производство молока и молокопродуктов в Российской Федерации, что хорошо видно на рис. 8.5.

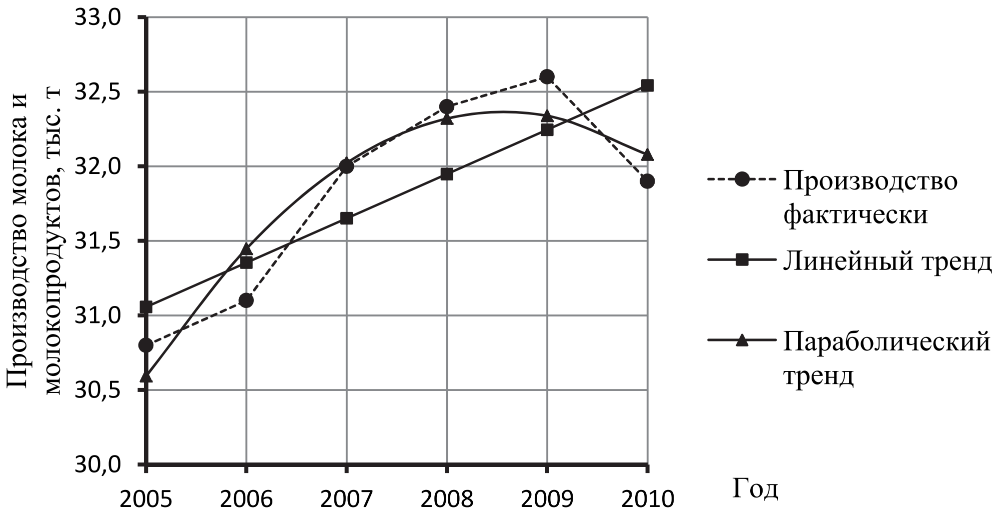


Рис. 8.5. Сглаживание с использованием линейной и параболической функций

## 8.6. Экстраполяция в рядах динамики и прогнозирование

Под ***экстраполяцией*** понимается распространение выявленных при анализе рядов динамики закономерностей развития изучаемого явления на будущее. При этом предполагается, что закономерность, действующая внутри анализируемого ряда динамики, сохранится и в дальнейшем.

Установление сроков прогнозирования зависит от задачи исследования. При этом следует иметь в виду, что чем короче сроки упреждения прогноза, тем надёжнее результаты экстраполяции.

Если ряд динамики имеет постоянные цепные абсолютные приросты ($\Delta y_{цi} \approx const$), то для экстраполяции примеряется формула

$$y_{n+l} = y_n + \overline{\Delta y_{ц}} \times l,$$

где $y_n$ – конечный уровень базового ряда динамики; $l$ – количество периодов упреждения прогноза.

**Пример.** По данным табл. 8.6 дадим прогноз объёма реализации металлопродукции на I и II кварталы следующего года. По формуле (8.13) найдём средний абсолютный прирост

$$\overline{\Delta y_{ц}} = \frac{y_n - y_o}{m-1}; \quad \overline{\Delta y_{ц}} = \frac{10,6-8,2}{4-1} = 0,8 \text{ (тыс. т)}.$$

Тогда прогноз на I квартал года, следующего за базовым, будет $y_{n+1} = y_n + \overline{\Delta y_{ц}} \times 1 = 10,6 + 0,8 = 11,4$ (тыс. т). Прогноз на II квартал следующего года $y_{n+2} = y_n + \overline{\Delta y_{ц}} \times 2 = 10,6 + 0,8 \times 2 = 12,2$ (тыс. т).

При экстраполяции уровня развития изучаемого явления на базе ряда динамики со стабильными темпами роста ($Tp_{цi} \approx const$) применяется формула

$$y_{n+l} = y_n (\overline{T_p})^l.$$

Если для ряда динамики уже проведено аналитическое выравнивание, то, используя трендовую модель, несложно получить точечный прогноз за пределами исследованного базового ряда. Для этого достаточно подставить в трендовую модель соответствующие значения $t_i$.

**Пример.** Используя данные табл. 8.9 и построенную трендовую модель $\hat{y}_t = 32,206 + 0,1486\, t - 0,0348\, t^2$, дать точечный прогноз производства молока и молокопродуктов на 2011 и 2012 гг.

Рассчитаем прогнозные уровни молока и молокопродуктов, подставив в уравнение $t_7 = 7$ и $t_8 = 9$, что соответствует 2011 и 2012 гг.:

$$\hat{y}_{t7} = 32,206 + 0,1486 \times 7 - 0,0348 \times 7^2 = 31,54 \text{ (млн т)};$$

$$\hat{y}_{t8} = 32,206 + 0,1486 \times 9 - 0,0348 \times 9^2 = 30,72 \text{ (млн т)}.$$

Прогноз называется точечным, т.к. не учитывает влияние факторов, вызывающих отклонение значения исследуемого признака от выявленной линии тренда. Поскольку в действительности фактическое значение признака в той или иной степени отклонится от полученной прогнозной оценки, то обычно дают не точечную, а интервальную оценку, для чего используют методы математической статистики и эконометрики.

# 9. СТАТИСТИЧЕСКИЕ ИНДЕКСЫ

## 9.1. Индексы и их виды

Среди методов статистического анализа важное место занимает индексный метод. Слово "индекс" (*index*) в переводе с латинского означает указатель, показатель. В статистике под ***индексом*** понимается относительная величина, характеризующая изменение величины какого-либо явления во времени, пространстве или по сравнению с любым эталоном (нормативом, планом, прогнозом и т.д.).

При сопоставлении уровней изучаемого явления во времени, говорят об индексах динамики, при сопоставлении в пространстве – о территориальных индексах, при сопоставлении с уровнем, например, договорных обязательств – об индексах выполнения обязательств и т.д.

Основным элементом индексного отношения является индексируемая величина. *Индексируемая величина* – значение признака статистической совокупности, изменение которой является объектом изучения. Поскольку объекты изучения индексов весьма разнообразны, то они широко применяются в экономической практике.

По содержанию изучаемых величин индексы разделяют на индексы количественных (объемных) и индексы качественных показателей.

***Индексы количественных показателей*** – индексы физического объема промышленной и сельскохозяйственной продукции, физического объема розничного товарооборота, национального дохода и др. Все индексируемые показатели этих индексов являются объемными, поскольку они характеризуют общий, суммарный размер (объем) того или иного явления и выражаются абсолютными величинами. При расчете таких индексов количества оцениваются в одинаковых, сопоставимых ценах.

***Индексы качественных показателей*** – индексы курса валют, цен, себестоимости, производительности труда, заработной платы, урожайности и др. Индексируемые показатели этих индексов характеризуют уровень явления в расчете на ту или иную единицу совокупности: цена за единицу продукции, себестоимость единицы продукции, выработка в единицу времени (или на одного работника), заработная плата одного работника, урожайность с одного гектара и т.д. Такие показатели называются качественными. Расчет данных индексов производится на базе одинаковых, неизменных количеств продукции.

По степени охвата единиц совокупности индексы делятся на два класса: индивидуальные и общие.

***Индивидуальные индексы*** служат для характеристики изменения отдельных элементов сложного явления (например, изменение объёма выпуска телевизоров определенной марки, рост или падение цен на акции какого-либо акционерного общества и т.д.).

***Общий индекс*** отражает изменение всех элементов сложного явления. При этом под сложным явлением понимают такую статистическую совокупность, отдельные элементы которой непосредственно не подлежат суммированию (физический объем продукции, включающий разноименные товары, цены на разные группы продуктов и т.д.).

## 9.2. Индивидуальные и общие индексы

Индивидуальные индексы обозначаются буквой $i$ и снабжаются подстрочным знаком индексируемого показателя: так $i_q$ – индивидуальный индекс объема произведенной продукции отдельного вида или количества (объема) проданного товара данного вида; $i_p$ – индивидуальный индекс цен и т.д.

Индивидуальные индексы относятся к одному элементу (явлению) и не требуют суммирования данных. Они представляют собой относительные величины динамики, выполнения обязательств, сравнения. Выбор базы сравнения определяется целью исследования.

Расчет индивидуальных индексов прост, их определяют вычислением отношения двух индексируемых величин:

1) Индивидуальный индекс физического объема продукции $i_q$ рассчитывается по формуле

$$i_q = \frac{q_1}{q_0}, \quad (9.1)$$

где $q_1$ – количество (объем) произведенного одноименного товара в текущем (отчетном) периоде; $q_0$ – количество (объем) произведенного одноименного товара в базисном периоде.

В знаменателе может быть плавное значение ($q_{пл}$), договорное ($q_{дог}$) или эталонное ($q_{эт}$) значение, принятые за базу сравнения.

2) Индивидуальный индекс цен:

$$i_p = \frac{p_1}{p_0}, \quad (9.2)$$

где $p_1$, $p_0$ – цена единицы одноименной продукции в отчетном и базисном периодах соответственно.

3) Индивидуальный индекс себестоимости единицы продукции:

$$i_z = \frac{z_1}{z_0},$$

где $z_1$, $z_0$ – себестоимость единицы продукции в отчетном и базисном периодах соответственно.

Индивидуальные индексы других показателей строятся аналогично.

С аналитической точки зрения индивидуальные индексы (9.1) – (9.2) характеризуют изменения индексируемой величины в текущем периоде по сравнению с базисным, т.е. во сколько раз она возросла (уменьшилась) или сколько процентов составляет ее рост (снижение). Значения индексов выражают в коэффициентах или процентах. Если из значения индекса, выраженного в процентах, вычесть 100 %, т.е. $i$ – 100, то полученная разность покажет, на сколько процентов возросла (уменьшилась) индексируемая величина.

Так, если 31 декабря 2009 г. килограмм сахара стоил 36 р., а 31 декабря 2010 г. – 49 р., то в соответствии с (9.2) $i_p$ = 49/36 = 1,36 или 136 %, т.е. цены на сахар повысились за год на 36 % (136 – 100 = 36).

Методика расчета общих индексов сложнее, чем индивидуальных, и различна в зависимости от характера индексируемых показателей, наличия исходных данных и целей исследования.

Любые общие индексы могут быть построены двумя способами: как агрегатные и как средние из индивидуальных. Последние в свою очередь делятся на средние арифметические и средние гармонические. Агрегатные индексы качественных показателей могут быть рассчитаны как индексы переменного состава и индексы постоянного (фиксированного) состава. В индексах переменного состава сопоставляются показатели, рассчитанные на базе изменяющихся структур явлений, в индексах постоянного состава – на базе неизменной структуры явлений.

Общий индекс обозначается буквой $I$ и также сопровождается подстрочным знаком индексируемого показателя: Например, $I_p$ – общий индекс цен, $I_z$ – общий индекс себестоимости.

## 9.3. Агрегатные индексы

***Агрегатный индекс*** – сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из несоизмеримых элементов.

Числитель и знаменатель агрегатного индекса представляют собой набор – "агрегат" непосредственно несоизмеримых и не поддающихся суммированию элементов – сумму произведений двух величин, одна из которых меняется (индексируется), а другая остается неизменной в числителе и знаменателе (вес индекса). ***Вес индекса*** служит для целей соизмерения индексируемых величин.

Построим три индекса: стоимости продукции, физического объёма продукции и цен.

Стоимость продукции – это произведение количества продукции в натуральном выражении ($q$) на её цену ($p$).

***Индекс стоимости продукции*** или ***товарооборота*** $I_{pq}$ представляет собой стоимость продукции текущего периода $\Sigma(p_1q_1)$ к стоимости продукции в базисном периоде $\Sigma(p_0q_0)$:

$$I_{pq} = \frac{\Sigma(p_1q_1)}{\Sigma(p_0q_0)}\,. \qquad (9.3)$$

Этот индекс показывает, во сколько раз возросла (уменьшилась) стоимость продукции (товарооборота) отчётного периода по сравнению с базисным. Если из значения индекса в процентах вычесть 100 % ($I_{pq}$ – 100), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным. Разность числителя и знаменателя $\Sigma(p_1q_1) - \Sigma(p_0q_0)$ показывает, на сколько рублей увеличилась (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным.

Рассмотрим применение индексного метода на примере изучения динамики показателей некоторого предприятия, производящего товары А, Б, В.

Рассчитаем индекс стоимости продукции (товарооборота) по данным табл. 9.1 и формуле (9.3):

$$I_{pq} = \frac{25\times6300+40\times3000+35\times700}{20\times6000+40\times2500+50\times500} = \frac{302000}{245000} = 1{,}233 \text{ или } 123{,}3\ \%.$$

Таблица 9.1

Отгрузка готовой продукции предприятием

| Продукция | Единица измерения | 2010 г. (базовый период) | | 2011 г. (отчётный период) | | Индивидуальные индексы | |
|---|---|---|---|---|---|---|---|
| | | Цена за единицу измерения, р. $p_0$ | Количество $q_0$ | Цена за единицу измерения, р. $p_1$ | Количество $q_1$ | цен $i_p$ | физического объёма $i_q$ |
| А | кг | 20 | 6 000 | 25 | 6 300 | 1,25 | 1,05 |
| Б | м | 40 | 2 500 | 40 | 3 000 | 1 | 1,2 |
| В | шт. | 50 | 500 | 35 | 700 | 0,7 | 1,4 |

Таким образом, в 2011 г. стоимость отгруженной продукции возросла по сравнению с 2010 г. в 1,233 раза (рост составил 123,3 %). Стоимость отгруженной продукции увеличилась на 23,3 % (123,3 % – 100 %) или на 57 000 р. (302 000 – 245 000). Увеличение стоимости отгруженной продукции может быть обусловлено как увеличением физического объёма отгрузки, так и ростом отпускных цен. Индивидуальные индексы физического объёма по каждому виду продукции (см. табл. 9.1) указывают на однозначный рост физического объёма отгрузки. Однако индивидуальные индексы цен такого однозначного ответа не дают: продукция А подорожала на 25 %, стоимость продукции Б не изменилась, а продукция В стала дешевле на 30 %. Произошло ли увеличение отпускных цен в целом по предприятию, или нет, индивидуальные индексы цен не говорят. Для выявления полной картины построим ещё два индекса: физического объёма продукции и цен.

При выборе веса индекса можно руководствоваться правилом: если строится индекс количественного показателя, то веса берутся за базисный период, при построении индекса качественного показателя используются веса отчётного периода.

***Индекс физического объёма продукции.***

Сложность при построении этого индекса заключается в том, что объемы разных видов продукции и товаров в натуральном выражении несоизмеримы и непосредственно суммироваться не могут. Нельзя, например, складывать килограммы хлеба с литрами молока, метрами ткани и парами обуви. Причиной несоизмеримости здесь является неоднородность – различие натуральной формы и свойств.

В связи с этим для разнородных продуктов или товаров сводный индекс физического объема (количества) нельзя построить и вычислить как отношение простых сумм, т.е. как $\sum q_1 / \sum q_0$. Здесь требуется использование специальных приемов индексного метода.

Единство различных видов продукции или разных товаров состоит в том, что они являются продуктами общественного труда, имеют определенную стоимость и ее денежный соизмеритель – цену ($p$). Умножая объем продукции каждого вида $q$ на соответствующую цену единицы продукции, получают сравнимые показатели, которые можно суммировать.

Коэффициенты соизмерения обеспечивают количественную сравнимость, позволяют учитывать "вес" продукта в реальном экономическом процессе. Поэтому их показатели-сомножители, связанные с индексируемыми величинами, принято называть весами индексов, а умножение на них – взвешиванием.

Умножая количество произведенной продукции (проданных товаров) на цены (которые, как правило, выступают в качестве соизмерителя неоднородной продукции), получаем стоимостное ("ценностное") выражение продукции каждого вида, которое допускает суммирование.

Стоимость продукции представляет собой произведение количества продукции в натуральном выражении $q$ на цену единицы продукции $p$.

Для того чтобы индекс охарактеризовал изменение только одного фактора (физического объёма продукции), нужно устранить (элиминировать) в формуле влияние другого фактора (цены), зафиксировав его как в числителе, так и в знаменателе на уровне одного и того же периода.

Поскольку индекс физического объёма продукции является индексом количественного показателя, то в соответствии с правилом веса берутся за базисный период. Формула индекса принимает вид:

$$I_q = \frac{\sum(q_1 p_0)}{\sum(q_0 p_0)}. \tag{9.4}$$

Заметим, что примененная в формуле (9.3) последовательность записей символов $q$ и $p$ определяется тем, что первым сомножителем в индексных отношениях является индексируемая величина, а вторым сомножителем – ее вес – измеритель.

Индекс физического объема продукции показывает, во сколько раз увеличился (уменьшился) физический объем продукции или сколько процентов составляет его рост (снижение) в отчетном периоде по сравнению с базисным периодом.

Если из значения индекса физического объема продукции вычесть 100 %, то разность ($I_q$ – 100) покажет, на сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным из-за роста (снижения) объема ее производства.

Абсолютное изменение физического объема продукции вычисляется как разность между числителем и знаменателем формулы (9.3) $\Sigma(q_1\ p_0)$ – $\Sigma(q_0\ p_0)$.

Экономически эта разность показывает, на сколько денежных единиц (рублей) изменилась стоимость продукции в результате роста (уменьшения) ее физического (т.е. натурального) объёма $q$. Изменение цен на продукцию в текущем периоде по сравнению с базисным не влияет на значение индекса.

Использование неизменных цен в зависимости от объекта исследования дает возможность изучить динамику выпуска совокупности произведенных товаров на отдельном предприятии, в отраслях промышленности и промышленности в целом. Если объектом исследования является какой-то регион, то индекс рассчитывается по товарам, произведенным предприятиями региона.

Рассчитаем индекс физического объёма по данным табл. 9.1 и формуле (9.4)

$$I_q = \frac{6300\times20+3000\times40+700\times50}{6000\times20+2500\times40+500\times50} = \frac{281000}{245000} = 1{,}147 \text{ или } 114{,}7\ \%.$$

Таким образом, в 2011 г. стоимость отгруженной продукции возросла за счёт физического увеличения объёма производства в 1,147 раза (рост составил 114,7 %). Объём отгрузки продукции вырос на 14,7 % (114,7 % – 100 %), за счёт чего предприятие получило дополнительно 36 000 р. (281 000 – 245 000).

***Индекс цен*** – это индекс качественного показателя. Индексируемой величиной будет цена товара. Влияние количества произведённых (проданных) товаров должно быть устранено, а это возможно только в том случае, если количество товаров неизменно в оба периода, т.е. количество товаров одного из периодов принято в качестве весов индекса.

В соответствии с правилом в качестве весов индекса берут количество товаров, произведённых (проданных) в текущем периоде.

Агрегатный индекс цен с отчетными весами впервые предложен в 1874 г. немецким экономистом Г. Пааше и носит его имя. Формула агрегатного индекса цен Пааше:

$$I_p = \frac{\Sigma(p_1 q_1)}{\Sigma(p_0 q_1)}. \qquad (9.5)$$

Такое исчисление индекса цен позволяет определить не только относительное изменение цен путем деления числителя индекса $\Sigma(p_1\ q_1)$ на знаменатель $\Sigma(p_0\ q_1)$, но и абсолютную экономию (–) или абсолютный перерасход (+) денежных средств покупателей в результате изменения цен на эти товары (как разность между числителем и знаменателем индекса): $\Sigma(p_1\ q_1) - \Sigma(p_0\ q_1)$.

Индекс цен Пааше показывает, во сколько раз возрос (уменьшился) в среднем уровень цен на массу товара, реализованную в отчетном периоде, или сколько процентов составляет его рост (снижение) в отчетном периоде по сравнению с базисным периодом.

Если из значения индекса цен *Ip* вычесть 100 %, т.е. (*Ip* – 100), то разность покажет, на сколько процентов в среднем возрос (уменьшился) за это время уровень цен на массу товаров, реализованную в отчетном периоде. При таком методе, рассчитав индекс цен по формуле (9.4), можно подсчитать экономический эффект от изменения цен.

Рассчитаем индекс цен по данным табл. 9.1 и формуле (9.5)

$$I_p = \frac{25\times6300+40\times3000+35\times700}{20\times6300+40\times3000+50\times700} = \frac{302000}{281000} = 1{,}075 \text{ или } 107{,}5\ \%.$$

Таким образом, в 2011 г. стоимость отгруженной продукции возросла за счёт увеличения отпускных цен в 1,075 раза (рост составил 107,5 %). Предприятие увеличило отпускные цены на свою продукцию в среднем на 7,5 % (107,5 % – 100 %), за счёт чего выручило дополнительно 21 000 р. (302 000 – 281 000).

Стоимость продукции можно определить как произведение количества товара на его цену. Точно такая же связь существует между индексами стоимости, физического объёма и цен, т.е. $I_{pq} = I_p \times I_q$. Справедливость данного утверждения легко проверить на рассмотренном выше примере: 1,233 = 1,075 × 1,147.

Таким образом, на основании анализа приведённых в табл. 9.1 данных можно сделать два вывода:

1) Произошедшее в 2001 г. увеличение выручки предприятия на 23,3 % обусловлено двумя факторами: увеличением физического объёма отгрузки продукции на 14,7 % и ростом отпускных цен на 7,5 %.

2) Увеличение на 57 000 р. выручки предприятия в 2011 г. объясняется увеличением физического объёма отгрузки продукции на 36 000 р. и ростом отпускных цен на 21 000 р.

В современных условиях особое место отводится *индексу потребительских цен*. С помощью индекса потребительских цен осуществляются оценка динамики цен на товары конечного потребления, пересчет важнейших стоимостных показателей из фактических цен в сопоставимые цены. Индекс может исчисляться по городу, району, субъекту федерации, стране в целом. Индекс потребительских цен является общим измерителем инфляции, используется при корректировке законодательно устанавливаемого минимального размера оплаты труда, установлении ставок налогов и т.д.

В соответствии с требованиями международной методологии индекс потребительских цен в Российской Федерации рассчитывается Федеральной службой государственной статистики по формуле, предложенной в 1864 г. немецким экономистом Э. Ласпейресом. Формула агрегатного индекса цен Ласпейреса отличается от формулы Пааше тем, что в качестве весов берётся количество произведённых (реализованных) товаров неотчётного $q_1$ базисного $q_0$ периода:

$$I_p = \frac{\Sigma(p_1 q_0)}{\Sigma(p_0 q_0)}.$$

Индекс цен Ласпейреса удобен для оперативной (недельной, месячной, квартальной) информации об изменении цен на определённый фиксированный набор товаров, когда пересчёт каждый раз на текущий набор (количество) товаров сопряжён с большими затратами труда и времени.

Индекс потребительских цен измеряет общее изменение стоимости фиксированного набора потребительских товаров и услуг, называемых "потребительской корзиной". В набор товаров и услуг, разработанный для наблюдения за ценами, включены товары и услуги массового спроса, а также отдельные товары и услуги необязательного пользования (раствор для инъекций "Диклофенак", легковой автомобиль импортный новый, экскурсионная поездка во Францию и т.д.). Отбор позиций произведён с учётом их важности для потребления населением, представительности с точки зрения отражения динамики цен на однородные товары, устойчивого наличия в продаже. Набор потребительской корзины регулярно пересматривается. В 2011 г. потребительский набор включал 464 товаров и услуг: 112 позиций – продовольственные товары, 242 – непродовольственные товары и 110 – платные услуги. Наблюдение за ценами осуществляется во всех субъектах Российской Федерации, на предприятиях торговли и услуг всех форм собственности.

Расчёт индекса потребительских цен производится с недельной, месячной, квартальной периодичностью, а также нарастающим итогом за период с начала года. Результаты публикуются в официальных изданиях и размещаются на сайте Росстата (www.gks.ru).

Сводный индекс потребительских цен рассчитывается на федеральном уровне для всего населения, а также для отдельных групп населения.

## 9.4. Средние индексы

Помимо агрегатных индексов в статистике применяется и другая форма общих индексов – средневзвешенные индексы. К их исчислению прибегают тогда, когда имеющаяся в распоряжении информация не позволяет рассчитать общий агрегатный индекс. Так, если неизвестны количества произведенных отдельных видов продукции в натуральных измерителях, но известны индивидуальные индексы $i_q = q_1/q_0$, то $q_1$ легко находится: $q_1 = i_q\, q_0$.

Подставляя найденное значение в агрегатную формулу индекса физического объёма продукции (9.4), получаем общий индекс физического объема в форме *среднего арифметического индекса физического объема продукции*, где весами служит стоимость отдельных видов продукции в базисном периоде ($q_o p_o$)

$$I_q = \frac{\sum(q_1 p_0)}{\sum(q_0 p_0)} = \frac{\sum(i_q q_0 p_0)}{\sum(q_0 p_0)} . \qquad (9.6)$$

При выборе весов следует иметь в виду, что средний индекс должен быть тождественен агрегатному, который является основной формой индекса.

Если известны данные, позволяющие исчислить только числитель агрегатного индекса физического объема по формуле (9.4), то, аналогично, выражая продукцию базисного периода как $q_0 = q_1 / i_q$, производим замену в знаменателе агрегатной формы индекса:

$$I_q = \frac{\sum(q_1 p_0)}{\sum(q_0 p_0)} = \frac{\sum(q_1 p_0)}{\sum\frac{(q_1 p_0)}{i_q}} .$$

В результате получаем общий индекс физического объема в форме среднего гармонического взвешенного индекса физического объема продукции, где весами служит стоимость продукции отчетного периода в базисных (или сопоставимых) ценах ($q_1 p_o$).

В форме средней гармонической взвешенной индекс физического объема используется только в аналитических целях.

Следовательно, применение той или иной формулы индекса физического объема (агрегатного, среднего арифметического или среднего гармонического) зависит от имеющихся в нашем распоряжении конкретных данных и цели исследования.

Построение *среднего гармонического индекса цен* на основе агрегатного индекса цен в форме Пааше также не представляет особого труда. Выражая из индивидуального индекса цены $p_0 = p_1 / i_p$, получим из (9.5):

$$I_p = \frac{\sum(p_1 q_1)}{\sum(p_0 q_1)} = \frac{\sum(p_1 q_1)}{\sum\frac{(p_1 q_1)}{i_p}}. \qquad (9.7)$$

Проверим найденные зависимости на примере данных табл. 9.1.

Рассчитаем индекс физического объема отгруженной предприятием продукции в форме среднего арифметического индекса по формуле (9.6)

$$I_q = \frac{1{,}05\times6000\times20+1{,}2\times2500\times40+1{,}4\times500\times50}{6000\times20+2500\times40+500\times50} = \frac{281000}{245000} = 1{,}147.$$

Теперь рассчитаем средний гармонический индекс цен по формуле (9.7)

$$I_p = \frac{25\times6300+40\times3000+35\times700}{\frac{25\times6300}{1{,}25}+\frac{40\times3000}{1{,}0}+\frac{35\times700}{0{,}7}} = 1{,}075.$$

Полученные значения полностью совпадают с результатами расчётов по формулам (9.4) и (9.5).

# 10. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

## 10.1. Функциональные и стохастические связи

Изучение объективно существующих связей между явлениями – важная задача статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявить факторы (признаки), оказывающие основное влияние на изучаемые процессы и явления.

Причинно-следственные отношения – это связь явлений и процессов, когда изменение одного из них – причины – ведёт к изменению другого – следствия. Причина – это совокупность условий, обстоятельств, действия которых приводят к появлению следствия. Причина всегда должна предшествовать следствию, однако не каждое предшествующее событие следует считать причиной, а последующее – следствием.

Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих явлений необходимо выделить главные, основные причины, абстрагируясь от второстепенных.

Признаки по их значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменение других, связанных с ними признаков, называются ***факторными***, или просто ***факторами***. Признаки, изменяющиеся под действием факторных признаков, называются ***результативными***.

В статистике различают функциональную связь и стохастическую зависимость. ***Функциональной*** называют связь, при которой определённому значению факторного признака соответствует одно и только одно значение результативного признака, т.е. изменение результативного признака $y$ всецело обусловлено действием факторного признака $x$:

$$y = f(x).$$

Функциональная связь проявляется во всех случаях наблюдения, для каждой конкретной единицы изучаемой совокупности. Примером функциональной связи может служить известная из курса физики зависимость веса тела $P$ от его массы $m$ и ускорения свободного падения $g$:

$$P = m \cdot g.$$

Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений, то такая зависимость называется ***стохастической***. Частным случаем стохастической зависимости является корреляционная связь.

При ***корреляционной*** связи изменение результативного признака $y$ обусловлено влиянием факторного признака $x$ не всецело, а лишь частично, т.к. возможно влияние прочих факторов $\varepsilon$:

$$y = \psi(x) + \varepsilon.$$

В отличие от функциональных связей, при наличии между признаками корреляционной связи одному и тому же значению учтённого факторного признака могут соответствовать различные значения результативного признака. Это обусловлено наличием других, неучтённых факторов, которые могут быть различными по составу, направлению и силе воздействия на отдельные (индивидуальные) единицы статистической совокупности. Поэтому для изучаемой совокупности в целом здесь устанавливается такое соотношение, при котором определённому значению факторного признака соответствует некоторое *среднее* значение результативного признака.

Таким образом, характерной особенностью корреляционных связей является то, что они проявляются не в единичных случаях, а в массе. Поэтому изучаются корреляционные связи по так называемым *эмпирическим* данным, т.е. данным, полученным в результате статистического наблюдения. В таких данных отображается совокупное действие всех причин и условий на изучаемый показатель.

Наличие корреляционных связей присуще многим социально-экономическим явлениям. Так, например, известно, что увеличение количества внесённых удобрений ведёт к увеличению урожайности. Однако это не означает, что на двух одинаково удобренных участках будет одинаковая урожайность одной и той же сельскохозяйственной культуры. Вероятнее всего, урожайности будут различаться. Кроме того, существует вероятность, что более высокая урожайность может наблюдаться и на менее удобренных участках, постольку на урожайность влияет не только количество внесённых удобрений, но и другие, неучтённые факторы: агротехника земледелия, количество внесённых семян, предшествующие культуры, рельеф местности, сроки и качество посева и уборки и т.д. Но если в анализ включить достаточно большое число площадей, то обнаружится прямая корреляционная зависимость между количеством внесённых удобрений и уровнем урожайности.

По направлению выделяют связь *прямую* и *обратную*. При прямой связи с увеличением или уменьшением значений факторного признака происходит соответственно увеличение или уменьшение значений результативного признака. Так, например, увеличение количества внесённых удобрений ведёт к увеличению урожайности. В случае обратной связи увеличение значений факторного признака приводит к уменьшению значений результативного признака, и наоборот, уменьшение значений факторного признака вызывает увеличение значений результативного признака. Так, например, с увеличением производительности труда снижается себестоимость продукции.

По аналитическому выражению выделяют связи *прямолинейные* (или просто *линейные*) и *нелинейные*. Если статистическая связь между явлениями может быть приближенно выражена уравнением прямой линии, то её называют *линейной* связью; если же она выражается уравнением какой-либо кривой линии (параболы, гиперболы, степенной и т.д.), то такую связь называют *нелинейной* (*криволинейной*).

По количеству факторов, действующих на результативный признак, связи делятся на *однофакторные* (один фактор) и *многофакторные* (два или более факторов). Однофакторные (простые) связи обычно называют *парными* (так как рассматривается пара признаков). Например, корреляционная связь между производительностью труда и прибылью. В случае многофакторной (множественной) связи рассматривается зависимость результативного признака от нескольких комплексно действующих факторов. Например, зависимость производительности труда от уровня автоматизации производства, квалификации рабочих и уровня организации труда.

## 10.2. Простейшие методы изучения корреляционных связей

***Метод приведения параллельных*** данных основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о её характере. Для этого единицы наблюдения располагают по возрастанию значений факторного признака $x$ и затем сравнивают с ним поведение результативного признака $y$. Сравним изменение двух величин:

| | | | | | | |
|---|---|---|---|---|---|---|
| $x$ | 1 | 2 | 4 | 7 | 8 | 11 |
| $y$ | 4 | 6 | 11 | 16 | 20 | 27 |

С увеличением величины $x$ значение $y$ также возрастает. Связь между ними прямая, и описать её можно уравнением прямой, или уравнением параболы второго порядка.

***Метод аналитических группировок*** предполагает произведение группировки единиц совокупности по факторному признаку, после чего для каждой группы вычисляется среднее или относительное значение результативного признака. Сопоставляя затем изменение результативного признака по мере изменения факторного, можно выявить направление, характер и тесноту связи.

Примером такой группировки могут служить данные о зависимости между себестоимостью зерна и урожайностью зерновых по условным данным 90 хозяйств.

В последней графе табл. 10.1 приведены средние величины, рассчитанные на основе индивидуальных данных об урожайности зерновых в отдельных хозяйствах. Из таблицы видно, что с увеличением урожайности средняя себестоимость 1 ц зерна уменьшается.

Таблица 10.1

Распределение хозяйств по урожайности зерновых и себестоимости 1 ц зерна

| Урожайность зерновых, ц/га | Число хозяйств | Средняя себестоимость 1 ц зерна, р. |
|---|---|---|
| До 15 | 6 | 310 |
| 15-17 | 7 | 296 |
| 17-19 | 9 | 292 |
| 19-21 | 17 | 280 |
| 21-23 | 35 | 276 |
| 23-25 | 11 | 269 |
| Свыше 25 | 5 | 253 |

***Графически*** взаимосвязь двух признаков изображается с помощью поля корреляции. В системе координат по оси абсцисс откладываются значения факторного признака, а по оси ординат – результативного. Каждое пересечение линий обозначается точкой. При отсутствии связи имеет место беспорядочное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определённой линии, выражающей форму связи (рис. 10.1).

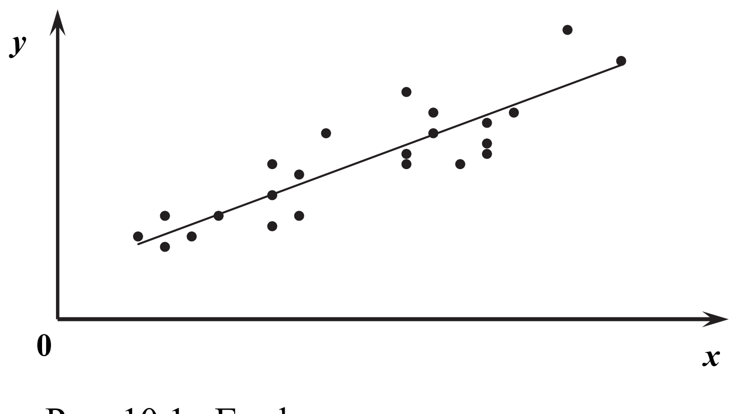


Рис. 10.1. График корреляционного поля

Приведённый на рис. 10.1 график показывает, что с увеличением величины $x$ значение $y$ в среднем также возрастает.

## 10.3. Метод корреляционно-регрессионного анализа связи показателей

***Корреляционный анализ*** имеет своей задачей количественное определение тесноты связи между двумя признаками (при парной связи) или между результативным и множеством факторных признаков (при многофакторной связи).

Теснота связи количественно выражается величиной коэффициентов (индексов) корреляции.

***Регрессионный анализ*** решает задачу определения аналитического выражения выявленной корреляционной связи. При этом влияние множества прочих факторов, оказывающих влияние на результативный признак, но не включённых в число факторных признаков, принимается за постоянные и средние значения. Регрессия может быть однофакторной (парной) и многофакторной (множественной).

Основной предпосылкой регрессионного анализа является то, что результативный признак $y$ подчиняется нормальному закону распределения, а факторные признаки $x_1, x_2 \ldots x_n$ могут иметь произвольный закон распределения.

***Корреляционно-регрессионный анализ,*** как общее понятие, включает в себя измерение тесноты, направления связи и установления формы этой связи.

По форме зависимости различают:

а) линейную регрессию, которая описывается уравнением прямой линии, т.е. линейной функцией вида $\hat{y}_x = a_0 + a_1x$;

б) нелинейную регрессию, для описания которой применяются такие математические функции, как:

- параболическая функция второго порядка $\hat{y}_x = a_0 + a_1x + a_2x^2$;
- параболическая функция третьего порядка $\hat{y}_x = a_0 + a_1x + a_2x^2 + a_3x^3$;
- показательная функция $\hat{y}_x = a_0a_1^x$;
- логарифмическая функция $\hat{y}_x = a_0 + a_1lg\,x$;
- степенная функция $\hat{y}_x = a_0x^{a_1}$ и т.д.

В парной регрессии выбор вида математической функции, используемой для описания зависимости $y$ от $x$, может быть осуществлен тремя методами:

- графическим;
- аналитическим;
- экспериментальным.

При использовании графического метода вид уравнения регрессии выбирается в зависимости от формы поля корреляции. Так, на рис. 10.1 видно, что наиболее подходящая для описания данного корреляционного поля функция – линейная.

Аналитический метод основан на изучении материальной природы связи исследуемых признаков. Так, например, количество израсходованного грузовым автомобилем горючего прямо пропорционально расстоянию перевозки груза. Зависимость расхода горючего от расстояния перевозки также можно описать линейной функцией.

При компьютерной обработке информации выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т.е. путём сравнения величины остаточной дисперсии, рассчитанной при разных уравнениях регрессии.

Оценка параметров $a_0$, $a_1$, $a_2$ уравнений регрессии производится с использованием метода наименьших квадратов. Сущность метода наименьших квадратов заключается в нахождении параметров уравнений $a_0$, $a_1$, $a_2$, при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака $y_i$ от теоретических (расчётных), полученных по выбранному уравнению регрессии $\hat{y}_{xi}$:

$$\sum(\hat{y}_{xi} - y_i)^2 \to min, \qquad (10.1)$$

где $\hat{y}_{xi}$ – теоретические (расчетные) значения результативного признака; $y_i$ – фактические значения результативного признака.

### *10.3.1. Парная линейная корреляционная связь*

Линейная зависимость – наиболее простая и часто используемая форма связи между двумя коррелируемыми признаками. Как указывалось выше, при парной корреляции она выражается уравнением прямой:

$$\hat{y}_x = a_0 + a_1 x. \qquad (10.2)$$

Из уравнения (10.2) следует, что для $x = x_i$ соответствующее значение результативного признака $\hat{y}_{xi}$ будет $\hat{y}_{xi} = a_0 + a_1 x_i$.

Подставляя найденное значение $\hat{y}_{xi}$ в (10.1), получаем уравнение для нахождения параметров $a_0$ и $a_1$:

$$\sum(a_0 + a_1 x - y_i)^2 \to min.$$

Решая данное уравнение с помощью изложенной в разделе 8 методики, получаем искомые значения параметров уравнения линейной регрессии $\hat{y}_x = a_0 + a_1 x$:

$$a_0 = \frac{\sum y_i \sum x_i^2 - \sum(y_i x_i) \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}, \qquad (10.3)$$

$$a_1 = \frac{n \sum(x_i y_i) - \sum y_i \sum x_i}{n \sum x_i^2 - (\sum x_i)^2}, \qquad (10.4)$$

где $n$ – объём исследуемой совокупности (число единиц наблюдений).

В уравнениях линейной регрессии параметр $a_0$ показывает усреднённое влияние на результативный признак неучтённых (не выделенных для исследования) факторов. Параметр $a_1$ – коэффициент регрессии – показывает, на сколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

**Пример.** По условным растениеводческим сельскохозяйственным предприятиям некоторой области имеется информация о посевных площадях и валовом выпуске продукции (табл. 10.2).

Таблица 10.2

Расчет параметров

| Исходные данные | | | Расчётные данные | | |
|---|---|---|---|---|---|
| Номер предприятия | Посевные площади, тыс. га ($x_i$) | Валовой выпуск продукции, млн р. ($y_i$) | $x_i^2$ | $x_i y_i$ | $\hat{y}_{xi}$ |
| 1 | 0,8 | 75 | 0,64 | 60,0 | 77,08 |
| 2 | 1,4 | 118 | 1,96 | 165,2 | 124,32 |
| 3 | 1,2 | 125 | 1,44 | 150,0 | 108,57 |
| 4 | 1,6 | 160 | 2,56 | 256,0 | 140,06 |
| 5 | 2,0 | 190 | 4,00 | 380,0 | 171,55 |
| 6 | 1,1 | 90 | 1,21 | 99,0 | 100,70 |
| 7 | 2,1 | 146 | 4,41 | 306,6 | 179,42 |
| 8 | 0,7 | 71 | 0,49 | 49,7 | 69,21 |
| 9 | 1,2 | 102 | 1,44 | 122,4 | 108,57 |
| 10 | 1,8 | 150 | 3,24 | 270,0 | 155,80 |
| 11 | 2,0 | 165 | 4,00 | 330,0 | 171,55 |
| 12 | 2,3 | 210 | 5,29 | 483,0 | 195,17 |
| Итого | 18,2 | 1 602 | 30,68 | 2 671,9 | 1 602,00 |

Построить уравнение регрессии. Дать точечный прогноз валового выпуска продукции предприятия, располагающего посевной площадью 1,5 тыс. га.

Решение. Очевидно, что выпуск продукции растениеводческим сельскохозяйственным предприятиям зависит от располагаемых посевных площадей, но никак не наоборот. Принимая размер посевных площадей за факторный признак $x$, а валовой выпуск продукции за результативный признак $y$, построим график корреляционного поля (рис. 10.2). По форме графика явно видно, что связь между посевной площадью $x$ и валовым выпуском продукции $y$ может быть описана линейной функцией $\hat{y}_x = a_0 + a_1 x$.

Для нахождения параметров уравнения $a_0$ и $a_1$ используем формулы (10.3) и (10.4). Необходимые для расчёта суммы $\Sigma x_i$, $\Sigma y_i$, $\Sigma x_i^2$ и $\Sigma(x_i y_i)$ получим используя табл. 10.2:

$$a_0 = \frac{1602 \times 30{,}68 - 2671{,}9 \times 18{,}2}{12 \times 30{,}68 - 18{,}2^2} = 14{,}11\ , \quad a_1 = \frac{12 \times 1671{,}9 - 1602 \times 18{,}2}{12 \times 30{,}68 - 18{,}2^2} = 78{,}72.$$

Подставив найденные значения $a_0$ и $a_1$ в уравнение линейной регрессии (10.2), получим математическую модель зависимости выпуска продукции растениеводческими сельскохозяйственными предприятиями от располагаемых посевных площадей

$$\hat{y}_{xi} = 14{,}11 + 78{,}72\,x. \qquad (10.5)$$

Параметр $a_1$ (коэффициент регрессии) равен 78,72, из чего можно сделать вывод, что увеличение посевных площадей на 1 тыс. га приводит к увеличению валового выпуска продукции предприятием в среднем на 78,72 млн р.

Определим теоретические (расчётные) значения выпуска продукции растениеводческими сельскохозяйственными предприятиями для указанных в табл. 10.2 посевных площадей, для чего подставим в найденное уравнение регрессии соответствующие значения $x_i$ вместо $x$:

$$\hat{y}_{x1} = 14{,}11 + 78{,}72 \times 0{,}8 = 77{,}08,$$

$$\hat{y}_{x2} = 14{,}11 + 78{,}72 \times 1{,}4 = 124{,}32 \text{ и т.д.}$$

Рассчитанные значения $\hat{y}_{xi}$ занесём в последний столбец табл. 10.2. Найденные теоретические значения $\hat{y}_{xi}$ нанесём на график (рис. 10.2) в виде звёздочек и соединим звёздочки прямой линией. Как видно из рис. 10.2, построенный график линейной регрессии достаточно хорошо описывает зависимость выпуска продукции от посевных площадей.

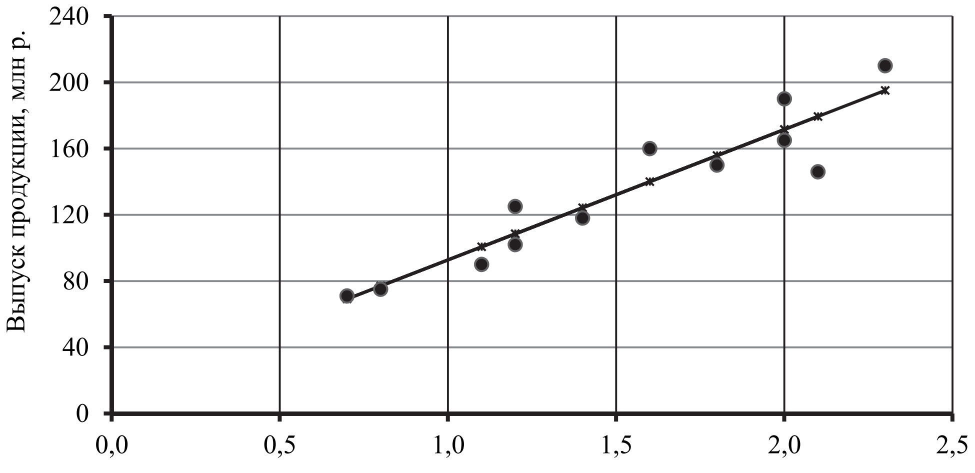


Рис. 10.2. Зависимость валового выпуска продукции от размера посевных площадей

Для нахождения точечного прогноза валового выпуска продукции предприятия, располагающего посевной площадью 1,5 тыс. га, достаточно подставить в построенное уравнение регрессии значение $x = 1{,}5$:

$$\hat{y}_x = 14{,}11 + 78{,}72 \times 1{,}5 = 132{,}19 \text{ (млн р.).}$$

Полученный точечный прогноз является математическим ожиданием размера валового выпуска продукции предприятием с посевной площадью 1,5 тыс. га.

В действительности фактическое значение выпуска продукции будет в той или иной мере отличаться от прогноза, поскольку построенная нами модель не учитывает отклонения, вызываемые различиями в составе высеваемых культур, количестве внесённых семян, удобрений, рельефе местности, сроках и качестве посева и уборки, обработке полей гербицидами и т.д. В результате действия указанных факторов фактическое значение результативного признака $y_i$ в каждом конкретном случае будет являться случайной величиной, подчиняющейся нормальному закону распределения. Как следствие, вычисленные на их основе параметры уравнения регрессии $a_0$ и $a_1$ также будут являться случайными величинами, равно как и полученный на основе этого уравнения точечный прогноз валового выпуска продукции. Естественным образом возникает вопрос об *адекватности* построенной математической модели, т.е. её соответствия фактическим статистическим данным.

Корреляционный и регрессионный анализ обычно проводятся для ограниченной по объёму совокупности. Поэтому параметры уравнения регрессии могут быть результатом стечения случайных обстоятельств. При численности объектов анализа до 30 единиц возникает необходимость проверки значимости (существенности) каждого параметра уравнения регрессии. При этом выясняют, насколько вычисленные параметры характерны для отображения всего комплекса условий: не являются ли полученные значения параметров результатом действия случайных величин.

Значимость коэффициентов линейной регрессии осуществляется с помощью ***t-критерия Стьюдента***. При этом вычисляют фактические значения $t$-критерия:

- для параметра $a_0$

$$t_{a0} = |a_0| \frac{\sqrt{n-2}}{\sigma_\varepsilon}; \tag{10.6}$$

- для параметра $a_1$

$$t_{a1} = |a_1| \frac{\sqrt{n-2}}{\sigma_\varepsilon} \sigma_x. \tag{10.7}$$

В формулах (10.6) и (10.7):

$n$ - объём исследуемой совокупности;

$\sigma_\varepsilon = \sqrt{\frac{\sum(y_i - \hat{y}_{xi})^2}{n}}$ - среднее квадратическое отклонение результативного признака $y$ от выровненных значений $\hat{y}_{xi}$ (рассчитанных по уравнению регрессии для соответствующих значений $x_i$);

$\sigma_x = \sqrt{\dfrac{\sum(x_i - \bar{x})^2}{n}}$ - среднее квадратическое отклонение факторного признака $x$ от общей средней $\bar{x}$.

Вычисленные по формулам (10.6) и (10.7) фактические значения $t_{a0}$ и $t_{a1}$ сравнивают с критическим $t_k$, которое определяют по таблице Стьюдента с учётом принятого уровня значимости α и числа степеней свободы $v$.

Уровень значимости применительно к проверке статистических гипотез – это вероятность, с которой может быть опровергнута гипотеза о том или ином законе распределения. Так, двум доверительным вероятностям 0,95 и 0,99 соответствуют уровни значимости α = 0,05 и α = 0,01.

Число степеней свободы $v$ характеризует число свободно варьирующих элементов совокупности. Применительно к рассматриваемому случаю

$$v = n - m, \tag{10.8}$$

где $n$ – число единиц совокупности; $m$ – количество параметров в уравнении регрессии.

В социально-экономических исследованиях уровень значимости обычно принимают равным 0,05. Параметр признаётся значимым (существенным), если $t$ фактическое больше $t$ критического, т.е $t_{a0} > t_k$ или $t_{a1} > t_k$.

Оценим значимость параметров уравнения регрессии (10.5), описывающей зависимости выпуска продукции растениеводческими сельскохозяйственными предприятиями от располагаемых посевных площадей для уровня значимости α = 0,05, что соответствует доверительной вероятности 0,95. Для проверки значимости параметров исчислим соответствующие значения $t$-критерия Стьюдента. Для проведения расчётов составим вспомогательную таблицу (табл. 10.3).

Таблица 10.3

Расчет параметров

| Исходные данные | | | Расчётные данные | | | | | |
|---|---|---|---|---|---|---|---|---|
| № | $x_i$ | $y_i$ | $\hat{y}_{xi}$ | $y_i - \hat{y}_{xi}$ | $(y_i - \hat{y}_{xi})^2$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $y_i^2$ |
| 1 | 0,8 | 75 | 77,08 | -2,08 | 4,34 | -0,717 | 0,514 | 5 625 |
| 2 | 1,4 | 118 | 124,32 | -6,32 | 39,89 | -0,117 | 0,014 | 13 924 |
| 3 | 1,2 | 125 | 108,57 | 16,43 | 269,90 | -0,317 | 0,100 | 15 625 |
| 4 | 1,6 | 160 | 140,06 | 19,94 | 397,60 | 0,083 | 0,007 | 25 600 |
| 5 | 2,0 | 190 | 171,55 | 18,45 | 340,45 | 0,483 | 0,234 | 36 100 |
| 6 | 1,1 | 90 | 100,70 | -10,70 | 114,48 | -0,417 | 0,174 | 8 100 |
| 7 | 2,1 | 146 | 179,42 | -33,42 | 1 116,96 | 0,583 | 0,340 | 21 316 |
| 8 | 0,7 | 71 | 69,21 | 1,79 | 3,20 | -0,817 | 0,667 | 5 041 |
| 9 | 1,2 | 102 | 108,57 | -6,57 | 43,18 | -0,317 | 0,100 | 10 404 |
| 10 | 1,8 | 150 | 155,80 | -5,80 | 33,69 | 0,283 | 0,080 | 22 500 |
| 11 | 2,0 | 165 | 171,55 | -6,55 | 42,89 | 0,483 | 0,234 | 27 225 |
| 12 | 2,3 | 210 | 195,17 | 14,83 | 220,07 | 0,783 | 0,614 | 44 100 |
| Итого | 18,2 | 1 602 | 1 602,00 | | 2 626,64 | | 3,077 | 235 560 |

Среднее квадратическое отклонение результативного признака $y$ от выровненных значений (см. табл. 10.3)

$$\sigma_\varepsilon = \sqrt{\frac{\sum(y_i - \hat{y}_{xi})^2}{n}}; \quad \sigma_\varepsilon = \sqrt{\frac{2626{,}64}{12}} = 14{,}79.$$

Для расчёта среднего квадратического отклонения факторного признака от общей средней предварительно рассчитаем общую среднюю $\bar{x}$, для чего используем данные табл. 10.2:

$$\bar{x} = \frac{\sum x_i}{n}; \quad \bar{x} = \frac{18{,}2}{12} = 1{,}517 \text{ (га)}.$$

Далее с использованием табл. 10.3 рассчитываем среднее квадратическое отклонение факторного признака $x$ от общей средней $\bar{x}$:

$$\sigma_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}; \quad \sigma_x = \sqrt{\frac{3{,}077}{12}} = 0{,}506.$$

Теперь по формулам (10.6) и (10.7) вычисляем фактические значения $t$-критерия Стьюдента:

$$t_{a0} = 14{,}11\frac{\sqrt{12-2}}{14{,}79} = 3{,}01;$$

$$t_{a1} = 78{,}72\frac{\sqrt{12-2}}{14{,}79}0{,}506 = 8{,}52.$$

Критическое значение $t$-критерия найдём по таблице распределения Стьюдента, которая в сокращённой форме представлена в табл. 10.4.

Таблица 10.4

Значения $t$-критерия Стьюдента при уровне значимости α, равного 0,05 и 0,01

| $\alpha$ | Число степеней свободы $v$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 10 | 15 | 20 | 30 | ∞ |
| 0,05 | 2,571 | 2,447 | 2,365 | 2,306 | 2,228 | 2,132 | 2,086 | 2,042 | 1,96 |
| 0,01 | 4,032 | 3,707 | 3,499 | 3,355 | 3,169 | 2,947 | 2,845 | 2,750 | 2,576 |

Число степеней свободы $v$ определим по формуле (10.8). В рассматриваемом примере уравнение регрессии содержит два параметра $a_0$ и $a_1$, а потому $m = 2$. Тогда число степеней свободы $v$ будет

$$v = n - m; \quad v = 12 - 2 = 10.$$

По табл. 10.4 для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $v = 10$ находим критическое значение $t$-критерия: $t_k = 2{,}228$.

Поскольку $t_{a0} > t_k$ (3,01 > 2,228) и $t_{a1} > t_k$ (8,52 > 2,228), то оба параметра $a_0$ и $a_1$ признаются значимыми (отклоняется гипотеза о том, что каждый из этих параметров в действительности равен нулю, и лишь в силу случайных обстоятельств оказался равным проверяемой величине).

Проверка адекватности регрессионной модели может быть дополнена корреляционным анализом. Для этого необходимо определить *тесноту* корреляционной связи между переменными $x$ и $y$. При линейной форме уравнения для этого обычно применяется линейный коэффициент корреляции. ***Линейный коэффициент корреляции r*** характеризует тесноту и направление связи между двумя коррелируемыми признаками:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}.$$

Для практических расчётов удобно применять другую формулу

$$r = \frac{\sum (x_i y_i) - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right] \cdot \left[\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right]}} \ . \qquad (10.9)$$

Линейный коэффициент корреляции принимает значения в интервале: $-1 \leq r \leq +1$. При $-1 \leq r < 0$ связь между $x$ и $y$ обратная, при $0 < r \leq +1$ связь прямая. При $r = 0$ связь между $x$ и $y$ отсутствует. Чем ближе коэффициент корреляции по абсолютной величине к единице, тем теснее связь между признаками. При $r = \pm 1$ связь функциональная.

Рассчитаем линейный коэффициент корреляции по формуле (10.9), для чего используем данные табл. 10.2 и 10.3:

$\sum (x_i y_i) = 2671{,}9; \sum x_i = 18{,}2; \sum y_i = 1602; \sum x_i^2 = 30{,}68; \sum y_i^2 = 235560;$

$$r = \frac{2671{,}9 - \frac{18{,}2 \cdot 1602}{12}}{\sqrt{\left[30{,}68 - \frac{(18{,}2)^2}{12}\right] \cdot \left[235560 - \frac{(1602)^2}{12}\right]}} = 0{,}937.$$

Для получения выводов о практической значимости полученной математической модели показателю тесноты связи даётся качественная оценка. Это осуществляется на основе шкалы Чеддока (табл. 10.5).

Таблица 10.5

Шкала Чеддока

| Показания тесноты связи | 0,1-0,3 | 0,3-0,5 | 0,5-0,7 | 0,7-0,9 | 0,9-0,99 |
|---|---|---|---|---|---|
| Характеристика силы связи | Слабая | Умеренная | Заметная | Высокая | Весьма высокая |

Полученное значение коэффициента корреляции, равное 0,937, означает, что в соответствии со шкалой Чеддока установленная по уравнению регрессии (10.5) связь между выпуском продукции растениеводческими сельскохозяйственными предприятиями и располагаемыми посевными площадями весьма высокая.

Показатели тесноты связи, исчисленные по данным сравнительно небольшой статистической совокупности, могут искажаться действием случайных причин. Это вызывает необходимость проверки их существенности, дающей возможность распространить выводы по результатам выборки на генеральную совокупность.

Оценка значимости коэффициента корреляции $r$ осуществляется по $t$-критерию Стьюдента. При линейной однофакторной связи $t$-критерий можно рассчитать по формуле

$$t_r = |r| \times \sqrt{\frac{n-2}{1-r^2}}. \qquad (10.10)$$

Полученное значение $t_r$ сравнивается с критическим $t_k$, которое берётся из таблицы Стьюдента с учётом заданного уровня значимости $\alpha$ и числа степеней свободы $v$. Если $t_r > t_k$, то величина коэффициента корреляции признаётся существенной.

В рассматриваемом примере $t$-критерий определим по формуле (10.10)

$$t_r = 0{,}937 \times \sqrt{\frac{12-2}{1-0{,}937^2}} = 8{,}52.$$

Полученное значение $t_r$ = 8,52 значительно больше найденного по табл. 10.4 значения $t_k$ = 2,228, что свидетельствует о значимости коэффициента корреляции и существенности связи между выпуском продукции и посевными площадями.

Квадрат линейного коэффициента корреляции $r^2$ называется линейным ***коэффициентом детерминации.*** В нашем примере $r^2 = 0{,}937^2 = 0{,}878$. Из этого следует, что 87,8 % общей вариации выпуска продукции обусловлено вариацией фактора – размером посевных площадей, и только 12,2 % общей вариации обусловлено вариацией всех остальных признаков. Таким образом, полученная регрессионная зависимость может быть использована для практических целей.

### *10.3.2. Парная нелинейная корреляционная связь*

Как отмечалось выше, для описания нелинейной корреляционной связи могут быть использованы такие функции как параболическая ($\hat{y}_x = a_0 + a_1x + a_2x^2$), показательная ($\hat{y}_x = a_0a_1^x$), логарифмическая ($\hat{y}_x = a_0 + a_1 lgx$) и др. Для нахождения параметров $a_0$, $a_1$, $a_2$ в этом случае также используется метод наименьших квадратов.

Для описания нелинейной связи в статистике часто применяется логарифмическая функция

$$\hat{y}_x = a_0 + a_1 lg\, x. \qquad (10.11)$$

На примере этой функции покажем, как описываются парные нелинейные корреляционные связи.

Подставляя в (10.1) $\hat{y}_{xi} = a_0 + a_1 lgx$, получаем уравнение для нахождения параметров $a_0$ и $a_1$:

$$\sum(a_0 + a_1 lgx_i - y_i)^2 \to min.$$

Используя описанную в разделе 8 методику нахождения параметров $a_0$, $a_1$, приходим к системе нормальных уравнений

$$\begin{cases} a_0 n + a_1 \sum \lg x_i = \sum y_i \\ a_0 \sum \lg x_i + a_1 \sum (\lg x_i)^2 = \sum (y_i \lg x_i). \end{cases}$$

Данная система имеет решение:

$$a_0 = \frac{\sum y_i \sum[\lg(x_i)]^2 - \sum[y_i \cdot \lg(x_i)] \sum \lg(x_i)}{n \sum[\lg(x_i)]^2 - [\sum \lg(x_i)]^2}, \qquad (10.12)$$

$$a_1 = \frac{n \sum[y_i \cdot \lg(x_i)] - \sum y_i \sum \lg(x_i)}{n \sum[\lg(x_i)]^2 - [\sum \lg(x_i)]^2}. \qquad (10.13)$$

Для примера с условными растениеводческими сельскохозяйственными предприятиями расчётная таблица примет вид табл. 10.6.

Таблица 10.6

Расчет параметров

| Исходные данные | | | Расчётные данные | | | |
|---|---|---|---|---|---|---|
| Номер пред-приятия | Посевные площади, тыс. га ($x_i$) | Валовой выпуск продукции, млн р. ($y_i$) | $lg(x_i)$ | $[lg(x_i)]^2$ | $y_i \cdot lg(x_i)$ | $\hat{y}_{xi}$ |
| 1 | 0,8 | 75 | -0,0969 | 0,0094 | -7,27 | 72,2 |
| 2 | 1,4 | 118 | 0,1461 | 0,0214 | 17,24 | 131,7 |
| 3 | 1,2 | 125 | 0,0792 | 0,0063 | 9,90 | 115,3 |
| 4 | 1,6 | 160 | 0,2041 | 0,0417 | 32,66 | 146,0 |
| 5 | 2,0 | 190 | 0,3010 | 0,0906 | 57,20 | 169,7 |
| 6 | 1,1 | 90 | 0,0414 | 0,0017 | 3,73 | 106,1 |
| 7 | 2,1 | 146 | 0,3222 | 0,1038 | 47,04 | 174,9 |
| 8 | 0,7 | 71 | -0,1549 | 0,0240 | -11,00 | 58,0 |
| 9 | 1,2 | 102 | 0,0792 | 0,0063 | 8,08 | 115,3 |
| 10 | 1,8 | 150 | 0,2553 | 0,0652 | 38,29 | 158,5 |
| 11 | 2,0 | 165 | 0,3010 | 0,0906 | 49,67 | 169,7 |
| 12 | 2,3 | 210 | 0,3617 | 0,1308 | 75,96 | 184,6 |
| Итого | 18,2 | 1 602 | 1,8395 | 0,5917 | 321,50 | 1 602,0 |

Для нахождения параметров уравнения $a_0$ и $a_1$ используем формулы (10.12) и (10.13). Необходимые для расчёта суммы $\Sigma lg(x_i)$, $\Sigma y_i$, $\Sigma[lg(x_i)]^2$ и $\Sigma[y_i \cdot lg(x_i)]$ получим используя табл. 10.6:

$$a_0 = \frac{1602 \times 0{,}5917 - 321{,}5 \times 1{,}8395}{12 \times 0{,}5917 - 1{,}8395^2} = 95{,}925,$$

$$a_1 = \frac{12 \times 321{,}5 - 1602 \times 1{,}8395}{12 \times 0{,}5917 - 1{,}8395^2} = 245{,}12.$$

Подставив найденные значения $a_0$ и $a_1$ в уравнение логарифмической регрессии (10.11), получим вторую (нелинейную) математическую модель зависимости выпуска продукции растениеводческими сельскохозяйственными предприятиями от располагаемых посевных площадей

$$\hat{y}_{xi} = 95{,}925 + 245{,}12 \cdot lg(x). \qquad (10.14)$$

Определим теоретические (расчётные) значения выпуска продукции растениеводческими сельскохозяйственными предприятиями для указанных в табл. 10.6 посевных площадей, для чего подставим в найденное уравнение регрессии соответствующие значения $x_i$ вместо $x$:

$$\hat{y}_{x1} = 95{,}925 + 245{,}12 \cdot lg(0{,}8) = 72{,}2,$$

$$\hat{y}_{x2} = 95{,}925 + 245{,}12 \cdot lg(1{,}4) = 131{,}7 \text{ и т.д.}$$

Рассчитанные значения $\hat{y}_{xi}$ занесём в последний столбец табл. 10.6. Найденные теоретические значения $\hat{y}_{xi}$ нанесём на график (рис. 10.3) в виде звёздочек и соединим звёздочки кривой линией. Как видно из рис. 10.3, построенный график логарифмической регрессии также хорошо описывает зависимость выпуска продукции от посевных площадей.

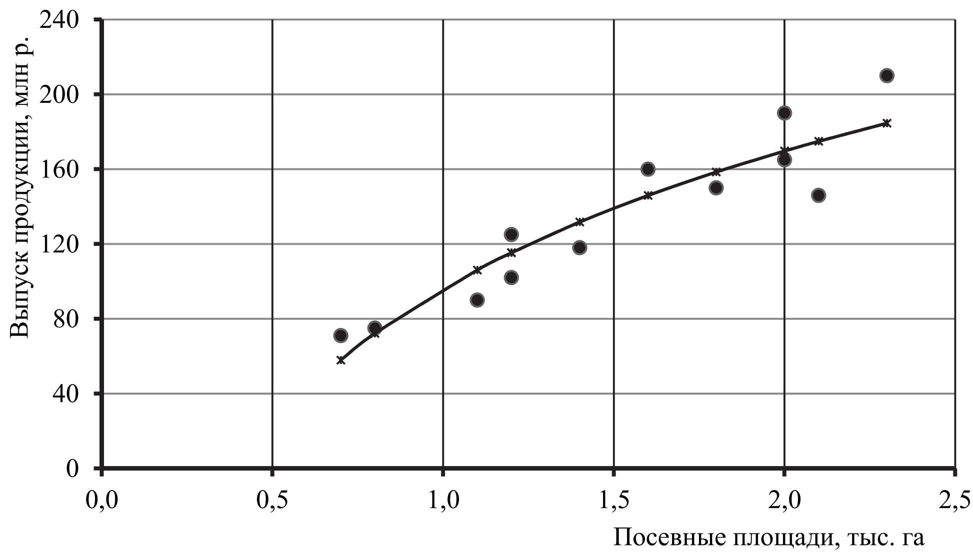


Рис. 10.3. Математическая модель по уравнению логарифмической функции

Проверка типичности параметров модели осуществляется по формулам (10.6) и (10.7). При $\sigma_x$ = 0,506 величину $\sigma_\varepsilon$ (среднее квадратическое отклонение результативного признака $y$ от выровненных значений $\hat{y}_{xi}$) определим по использованной ранее формуле

$$\sigma_\varepsilon = \sqrt{\frac{\Sigma(y_i - \hat{y}_{xi})^2}{n}}. \tag{10.15}$$

Для проведения расчётов составим вспомогательную таблицу (табл. 10.7).

Таблица 10.7

Расчет параметров

| Исходные данные | | | Расчётные данные | | | |
|---|---|---|---|---|---|---|
| № | $x_i$ | $y_i$ | $\hat{y}_{xi}$ | $y_i$ - $\hat{y}_{xi}$ | $(y_i$ - $\hat{y}_{xi})^2$ | $y_i^2$ |
| 1 | 0,8 | 75 | 72,17 | 2,83 | 8,01 | 5 625 |
| 2 | 1,4 | 118 | 131,74 | -13,74 | 188,91 | 13 924 |
| 3 | 1,2 | 125 | 115,33 | 9,67 | 93,42 | 15 625 |
| 4 | 1,6 | 160 | 145,96 | 14,04 | 197,13 | 25 600 |
| 5 | 2,0 | 190 | 169,71 | 20,29 | 411,49 | 36 100 |
| 6 | 1,1 | 90 | 106,07 | -16,07 | 258,29 | 8 100 |
| 7 | 2,1 | 146 | 174,91 | -28,91 | 835,71 | 21 316 |
| 8 | 0,7 | 71 | 57,96 | 13,04 | 170,17 | 5 041 |
| 9 | 1,2 | 102 | 115,33 | -13,33 | 177,81 | 10 404 |
| 10 | 1,8 | 150 | 158,50 | -8,50 | 72,22 | 22 500 |
| 11 | 2,0 | 165 | 169,71 | -4,71 | 22,23 | 27 225 |
| 12 | 2,3 | 210 | 184,59 | 25,41 | 645,51 | 44 100 |
| Итого | 18,2 | 1 602 | 1 602,00 | | 3 080,91 | 235 560 |

Среднее квадратическое отклонение результативного признака $y$ от выровненных значений по формуле (10.15) составит:

$$\sigma_\varepsilon = \sqrt{\frac{3080,91}{12}} = 16,02.$$

По формулам (10.6) и (10.7) вычисляем фактические значения $t$-критерия Стьюдента:

$$t_{a0} = 95,925\frac{\sqrt{12-2}}{16,02} = 18,9\ ;$$

$$t_{a1} = 245,12\frac{\sqrt{12-2}}{16,02}0,506 = 24,5.$$

По табл. 10.4 для уровня значимости α = 0,05 и числа степеней свободы $v$ = 10 критическое значение $t$-критерия: $t_k$ = 2,228.

Поскольку $t_{a0} > t_k$ (18,9 > 2,228) и $t_{a1} > t_k$ (24,5 > 2,228), то оба параметра $a_0$ и $a_1$ признаются значимыми.

Оценим *тесноту* корреляционной связи между переменными $x$ и $y$. Поскольку применяется нелинейная модель, то использованный для линейного уравнения регрессии линейный коэффициент корреляции $r$ (формула (10.9)) в данном случае неприменим. Для оценки математических моделей, построенных на основе нелинейных уравнений корреляционной связи, вместо линейного коэффициента корреляции используется ***индекс корреляции R***:

$$R = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}}\,, \tag{10.16}$$

где $n$ – объём исследуемой совокупности; $\sigma_\varepsilon^2$ – остаточная дисперсия, отображающая вариацию результативного признака $y$ от всех прочих, кроме $x$, факторов (10.5); $\sigma_y^2$ – общая дисперсия результативного признака $y$, отображающая совокупное влияние всех факторов.

Общая дисперсия результативного признака $\sigma_y^2$ может быть вычислена по формуле (6.2), которая для нашего случая примет вид $\sigma_y^2 = \frac{\sum(y_i - \bar{y})^2}{n}$, либо по ранее не использовавшейся формуле для вычисления общей дисперсии:

$$\sigma_y^2 = \overline{y^2} - (\bar{y})^2\,. \tag{10.17}$$

Значение $\sigma_\varepsilon^2$ = 16,02 рассчитано ранее. Для расчёта $\sigma_y^2$ используем данные табл. 10.7:

$$\bar{y} = \sum \frac{y_i}{n}; \quad \bar{y} = \frac{1602}{12} = 133{,}5,$$

тогда по формуле (10.17)

$$\sigma_y^2 = \frac{235560}{12} - (133{,}5)^2 = 1807{,}75.$$

Индекс корреляции, рассчитанный по формуле (10.16), будет:

$$R = \sqrt{1 - \frac{16{,}02}{1807{,}75}} = 0{,}926.$$

Полученное значение индекса корреляции, равное 0,926, означает, что в соответствии со шкалой Чеддока установленная по уравнению логарифмической регрессии (10.14) связь между выпуском продукции растениеводческими сельскохозяйственными предприятиями и располагаемыми посевными площадями весьма высокая.

Для оценки значимости индекса корреляции $R$ для нелинейных моделей применяется *F-критерий Фишера*. Фактическое значение критерия $F_R$ определяется по формуле

$$F_R = \frac{R^2}{1-R^2} \times \frac{n-m}{m-1}, \qquad (10.18)$$

где $m$ – число параметров уравнения регрессии.

Величина $F_R$ сравнивается с критическим значением $F_k$, которое определяется по таблице $F$-критерия Фишера с учётом принятого уровня значимости α и числа степеней свободы $k_1 = m - 1$ и $k_2 = n - m$.

Если $F_R > F_k$, то корреляция признаётся существенной.

Используя формулу (10.18), рассчитаем фактическое значение критерия $F_R$ для нашего примера. Объём исследуемой совокупности – 12 единиц, т.е. $n = 12$. Уравнение логарифмической регрессии содержит два параметра $a_0$ и $a_1$, а потому $m = 2$:

$$F_R = \frac{0{,}926^2}{1-0{,}926^2} \times \frac{12-2}{2-1} = 60{,}41.$$

Критическое значение $F_k$ найдём по таблице критических значений $F$-критерия Фишера, которая для наиболее употребляемого уровня значимости $\alpha = 0{,}05$ в сокращённой форме представлена в табл. 10.8.

Таблица 10.8

Критические значения F-критерия Фишера
при уровне значимости α = 0,05

| $k_1$ | $k_2$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 10 | 15 | 20 | 30 | ∞ |
| 1 | 6,61 | 5,99 | 5,59 | 5,32 | 4,96 | 4,45 | 4,35 | 4,17 | 3,84 |
| 2 | 5,79 | 5,14 | 5,59 | 4,46 | 4,10 | 3,68 | 3,49 | 3,32 | 2,99 |

В нашем примере число степеней свободы $k_1 = m - 1 = 2 - 1 = 1$ и $k_2 = n - m = 12 - 2 = 10$. По таблице для $k_1 = 1$ и $k_2 = 10$ находим, что критическое значение $F_k = 4{,}96$. Поскольку $60{,}41 > 4{,}96$, то $F_R > F_k$ и показатель тесноты связи $R = 0{,}926$ признаётся существенным. Таким образом, полученная с использованием уравнения логарифмической функции регрессионная зависимость может быть использована для практических целей.

Квадрат индекса корреляции $R^2$ называется ***индексом детерминации.*** В нашем примере $R^2 = 0{,}926^2 = 0{,}857$. Из этого следует, что 85,7 % общей вариации выпуска продукции обусловлено вариацией фактора – размером посевных площадей, а только 14,3 % общей вариации – вариацией всех остальных признаков.

Нами было построено два уравнения регрессии – линейное и логарифмическое. Какое из них более точно описывает корреляционную зависимость?

Для линейного уравнения регрессии линейный коэффициент корреляции $r = 0{,}937$, для логарифмического уравнения регрессии индекс корреляции $R$ = 0,926. Поскольку 0,937 > 0,926, то для нашего примера линейное уравнение более точно описывает связь между выпуском продукции растениеводческими сельскохозяйственными предприятиями и располагаемыми посевными площадями.

В тех случаях, когда для описания корреляционных зависимостей используются иные нелинейные функции (гиперболическая, параболическая, показательная и т.д.), расчёт их параметров и оценка значимости производятся по аналогичной методике.

## 11. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задача 1**

Имеются данные Росстата по 28 субъектам РФ о среднегодовой численности занятого в экономике населения и основных фондах в экономике по полной учётной стоимости на конец 2010 г.

| | Субъект РФ | Занято в экономике, тыс. чел. | Основные фонды, млн р. |
|---|---|---|---|
| 1 | Воронежская область | 1054,3 | 788 059 |
| 2 | Костромская область | 321,5 | 295 091 |
| 3 | Курская область | 573,9 | 435 966 |
| 4 | Орловская область | 391,9 | 258 382 |
| 5 | Рязанская область | 502,8 | 577 233 |
| 6 | Смоленская область | 495,8 | 477 280 |
| 7 | Тульская область | 771,1 | 562 328 |
| 8 | Ярославская область | 643,9 | 821 370 |
| 9 | Республика Карелия | 336,7 | 393 841 |
| 10 | Республика Коми | 467,5 | 1 245 540 |
| 11 | Калининградская область | 471,4 | 398 091 |
| 12 | Ленинградская область | 741,1 | 1 210 457 |
| 13 | Мурманская область | 434,8 | 799 454 |
| 14 | Краснодарский край | 2274,2 | 2 139 060 |
| 15 | Астраханская область | 447,7 | 623 538 |
| 16 | Волгоградская область | 1229,7 | 1 200 283 |
| 17 | Республика Мордовия | 385 | 353 809 |
| 18 | Кировская область | 664,2 | 541 725 |
| 19 | Нижегородская область | 1710,9 | 1 578 659 |
| 20 | Оренбургская область | 1070,9 | 1 047 515 |
| 21 | Пензенская область | 667,3 | 579 673 |
| 22 | Курганская область | 409 | 474 482 |
| 23 | Челябинская область | 1665,7 | 1 687 909 |
| 24 | Красноярский край | 1439,3 | 1 628 414 |
| 25 | Иркутская область | 1140,2 | 1 641 443 |
| 26 | Камчатский край | 189,1 | 194 506 |
| 27 | Приморский край | 980,1 | 912 847 |
| 28 | Хабаровский край | 729,4 | 809 105 |

1) Рассчитать число групп.

2) Провести группировку субъектов РФ по среднегодовой численности занятого в экономике населения.

3) По каждой группе подсчитать численность занятого в экономике населения, стоимость основных фондов, среднюю численность занятого в экономике населения и среднюю стоимость основных фондов.

4) Построить график связи средней стоимости основных фондов и численности занятого в экономике населения.

**Задача 2**

Имеются данные Росстата по 28 субъектам РФ о численности населения и валовом региональном продукте в 2009 г.

| Субъект РФ | | Население, тыс. чел. | Валовой продукт, млн р. |
|---|---|---|---|
| 1 | Белгородская область | 1532,5 | 304 343 |
| 2 | Ивановская область | 1061,1 | 86 572,8 |
| 3 | Московская область | 7104 | 1 530 623 |
| 4 | Тверская область | 1350,3 | 197 892 |
| 5 | Республика Коми | 899,7 | 301 410 |
| 6 | Архангельская область | 1225,3 | 323 026,5 |
| 7 | Краснодарский край | 5229,2 | 857 527,3 |
| 8 | Ростовская область | 4276,4 | 556 194,2 |
| 9 | Республика Дагестан | 2981,4 | 265 092,1 |
| 10 | Ставропольский край | 2785,3 | 277 466,8 |
| 11 | Республика Татарстан | 3787,4 | 884 232,9 |
| 12 | Нижегородская область | 3307,6 | 545 940,1 |
| 13 | Оренбургская область | 2031,3 | 414 537,2 |
| 14 | Свердловская область | 4297,5 | 823 833 |
| 15 | Республика Дагестан | 2981,4 | 265 092,1 |
| 16 | Челябинская область | 3478,1 | 564 671 |
| 17 | Республика Бурятия | 972,2 | 124 610,3 |
| 18 | Республика Тыва | 308,1 | 26 918,9 |
| 19 | Красноярский край | 2829,1 | 748 512,1 |
| 20 | Иркутская область | 2427,9 | 455 529,2 |
| 21 | Республика Саха (Якутия) | 958 | 329 679,6 |
| 22 | Камчатский край | 321,3 | 95 591 |
| 23 | Приморский край | 1953,5 | 367 698,3 |
| 24 | Хабаровский край | 1343,3 | 274 983,5 |
| 25 | Амурская область | 827,8 | 151 750,4 |
| 26 | Магаданская область | 156,5 | 48 128,4 |
| 27 | Сахалинская область | 496,6 | 392 311,7 |
| 28 | Еврейская автономная область | 176,3 | 25 345,1 |

1) Рассчитать число групп.

2) Провести группировку субъектов РФ по численности населения.

3) По каждой группе подсчитать численность населения, объём валового регионального продукта, среднюю численность населения и средний объём валового регионального продукта.

4) Построить график зависимости объёма валового регионального продукта от численности населения.

**Задача 3**

Имеются данные о потреблении электроэнергии в микрорайоне в январе месяце. Вычислить среднее потребление электроэнергии одним человеком, моду и медиану.

| Потребление электроэнергии в кВт/ч на человека | 10-40 | 40-60 | 60-80 | 80-100 | 100-150 |
|---|---|---|---|---|---|
| Число человек | 450 | 2800 | 1500 | 420 | 180 |

**Задача 4**

Имеются данные о потреблении холодной воды в микрорайоне в январе месяце. Вычислить среднее потребление холодной воды одним человеком.

| Потребление холодной воды в литрах на человека | До 140 | 140-220 | 220-300 | 300-460 | Свыше 460 |
|---|---|---|---|---|---|
| Число человек | 250 | 1200 | 3600 | 1300 | 580 |

**Задача 5**

Имеются данные по оптовой продаже сахара.

| III квартал | | | IV квартал | | |
|---|---|---|---|---|---|
| Номер партии | Цена за тонну, тыс. р. | Объём партии, т | Номер партии | Цена за тонну, тыс. р. | Объём партии, т |
| 1 | 23,0 | 160 | 1 | 24,2 | 200 |
| 2 | 26,8 | 20 | 2 | 26,7 | 50 |
| 3 | 25,5 | 88 | 3 | 25,9 | 100 |
| 4 | 25,2 | 180 | | | |

Рассчитать средние цены сахара в III и IV кварталах. На сколько процентов изменилась средняя цена сахара в IV квартале по отношению к III кварталу?

**Задача 6**

Имеются данные по оптовой продаже сахара.

| III квартал | | | IV квартал | | |
|---|---|---|---|---|---|
| Номер партии | Цена за тонну, тыс. р. | Выручка, тыс. р. | Номер партии | Цена за тонну, тыс. р. | Выручка, тыс. р. |
| 1 | 13,0 | 2 180 | 1 | 14,2 | 2 840 |
| 2 | 14,5 | 1 950 | 2 | 16,7 | 2 505 |
| 3 | 15,0 | 1 500 | 3 | 15,9 | 795 |
| | | | 4 | 16,0 | 1 950 |

Рассчитать средние цены сахара в III и IV кварталах. В каком квартале средняя цена была выше?

**Задача 7**

Имеются данные о дневном пробеге такси в таксопарке (в километрах).

Определить необходимое число групп, длину интервалов, построить гистограмму. По гистограмме определить типичный пробег такси в таксопарке.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 101 | 263 | 142 | 242 | 110 | 174 | 225 | 120 |
| 194 | 154 | 191 | 151 | 219 | 171 | 206 | 170 |
| 231 | 223 | 143 | 166 | 200 | 130 | 184 | 153 |
| 183 | 138 | 199 | 151 | 208 | 162 | 185 | 212 |
| 165 | 236 | 122 | 219 | 169 | 182 | 147 | 191 |
| 148 | 169 | 128 | 230 | 178 | 232 | 185 | 211 |

**Задача 8**

Имеются данные о выручке хлебных киосков на ул. Комсомольской и ул. Шиханова (тыс. р.).

По каждому киоску рассчитать дисперсию, среднее квадратическое отклонение, коэффициент вариации. Определить, какой из киосков имеет более стабильную выручку.

| Киоск на ул. Комсомольской | | | Киоск на ул. Шиханова | | |
|---|---|---|---|---|---|
| 9,4 | 10,4 | 9,8 | 31,1 | 29,0 | 33,6 |
| 9,1 | 11,2 | 10,2 | 32,0 | 30,2 | 31,4 |
| 8,3 | 8,8 | 9,9 | 31,8 | 29,7 | 34,2 |

**Задача 9**

При обследовании 800 частных легковых автомобилей из 80 000, имеющихся в городе, были получены данные о годовых затратах на ремонт одного автомобиля.

| Затраты, тыс. р. | 0 – 2 | 2 – 5 | 5 – 10 | 10 – 20 | Свыше 20 |
|---|---|---|---|---|---|
| Число автомобилей | 90 | 130 | 380 | 120 | 80 |

С вероятностью 0,955 определить доверительные пределы, в которых находятся средние затраты на ремонт одного автомобиля.

**Задача 10**

При обследовании 1000 частных легковых автомобилей из 60 000, имеющихся в городе, были получены данные о среднемесячном расходе бензина одним автомобилем.

| Расход, л | До 50 | 50 – 75 | 75 – 100 | 100 – 200 | Свыше 200 |
|---|---|---|---|---|---|
| Число автомобилей | 100 | 400 | 350 | 100 | 50 |

Определить:

- коэффициент вариации;

- доверительные пределы, в которых находятся средний расход бензина одним автомобилем с вероятностью 0,997.

**Задача 11**

При выборочном обследовании 400 автомобилей из 40 000, имеющихся в городе, было установлено, что 27 из них не имеют аптечек. С вероятностью 99 % оценить, какая доля из имеющихся в городе автомобилей ездит без аптечек.

**Задача 12**

При выборочном обследовании 500 автомобилей из 100 000, имеющихся в городе, было установлено, что 47 из них технически неисправны. С вероятностью 95,5 % оценить, какая доля из имеющихся в городе автомобилей технически неисправна (и сколько это в штуках).

**Задача 13**

В автопарке 200 такси. Известен дневной пробег восьми из них (км).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 469 | 451 | 477 | 474 | 480 | 449 | 494 | 472 |

С вероятностью 99 % установить пределы, в которых находятся средний дневной пробег одной машины по автопарку в целом.

**Задача 14**

В автопарке 400 такси. Известна суточная выручка шести из них (тыс. р.).

| | | | | | |
|---|---|---|---|---|---|
| 6,5 | 7,3 | 6,4 | 6,8 | 6,3 | 6,9 |

С вероятностью 95 % установить пределы, в которых находится суточная выручка одной машины по автопарку в целом.

**Задача 15**

Имеются данные об объёме перевалки грузов транспортным терминалом (млн т).

| 2007 г. | 2008 г. | 2009 г. | 2010 г. | 2011 г. |
|---|---|---|---|---|
| 192 | 216 | 188 | 196 | 220 |

Вычислить цепные и базисные показатели ряда динамики: а) абсолютные приросты; б) темпы роста; в) темпы прироста; г) средние темп роста и темп прироста за 2007 – 2011 гг.

**Задача 16**

Имеются данные о среднемесячной начисленной зарплате работников предприятия по годам (тыс. р.).

| 2007 г. | 2008 г. | 2009 г. | 2010 г. | 2011 г. |
|---|---|---|---|---|
| 19,2 | 21,9 | 23,5 | 26,4 | 29,0 |

Вычислить цепные показатели ряда динамики: а) абсолютные приросты; б) коэффициенты роста; в) коэффициенты прироста; г) средний абсолютный прирост, средние коэффициенты роста и прироста за 2007 – 2011 гг. Дать прогноз среднемесячной начисленной зарплаты на 2012 г.

**Задача 17**

Имеются статистические данные о населении Хабаровского края на 1 января соответствующего года, тыс. чел.

| 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. |
|---|---|---|---|---|
| 1412,2 | 1405,5 | 1403,7 | 1401,9 | 1400,5 |

Вычислить цепные и базисные показатели ряда динамики: а) абсолютные приросты; б) темпы роста; в) темпы прироста; г) средний абсолютный прирост за 2005 – 2009 гг.

**Задача 18**

Имеются статистические данные о численности населения Хабаровского края на 1 января соответствующего года, тыс. человек. Определить основную тенденцию, используя сглаживание методом скользящих средних (трёхчленных).

| 2001 г. | 2002 г. | 2003 г. | 2004 г. | 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. |
|---|---|---|---|---|---|---|---|---|
| 1446,1 | 1434,2 | 1427,0 | 1420,2 | 1412,2 | 1405,5 | 1403,7 | 1401,9 | 1400,5 |

**Задача 19**

Имеются данные по объёмам реализации продукции. Определить основную тенденцию, используя сглаживание методом скользящих средних (трёхчленных).

| Месяц | Объём реализ., тыс. р. | Месяц | Объём реализ., тыс. р. | Месяц | Объём реализ., тыс. р. | Месяц | Объём реализ., тыс. р. |
|---|---|---|---|---|---|---|---|
| 1 | 130 | 4 | 230 | 7 | 140 | 10 | 110 |
| 2 | 210 | 5 | 175 | 8 | 100 | 11 | 150 |
| 3 | 180 | 6 | 190 | 9 | 125 | 12 | 140 |

**Задача 20**

Имеются данные о выпуске продукции городским пищекомбинатом. Вычислить агрегатный индекс цен по формулам Пааше и Ласпейреса.

| Наименование продукции | Объём производства | | Отпускная цена р. за ед. | |
|---|---|---|---|---|
| | 2010 г. | 2011 г. | 2010 г. | 2011 г. |
| Напиток "Сливка", дал* | 12 000 | 19 000 | 600 | 660 |
| Квас "Летний", дал* | 13 000 | 5 000 | 250 | 250 |
| Паста яблочная, кг | 2 500 | 4 000 | 190 | 210 |
| Набор подарочный, шт. | 1 000 | 4 600 | 320 | 330 |
| *Декалитр | | | | |

**Задача 21**

Имеются данные о выпуске продукции городским пищекомбинатом. Вычислить индивидуальные индексы цен и индивидуальные индексы физического объёма производства по каждому виду продукции. Вычислить агрегатный индекс стоимости продукции.

| Наименование продукции | Объём производства | | Отпускная цена р. за ед. | |
|---|---|---|---|---|
| | 2010 г. | 2011 г. | 2010 г. | 2011 г. |
| Настойка "Сливка", дал* | 12 000 | 19 000 | 600 | 660 |
| Квас "Летний", дал* | 13 000 | 5 000 | 250 | 250 |
| Паста яблочная, кг | 2 500 | 4 000 | 190 | 210 |
| Набор подарочный, шт. | 1 000 | 4 600 | 320 | 330 |
| *Декалитр | | | | |

**Задача 22**

Имеются данные о реализации строительных материалов торговой фирмой. Вычислить индивидуальные индексы цен и индивидуальные индексы физического объёма продаж по каждому виду товара. Вычислить агрегатный индекс физического объёма продаж.

| Наименование товара | Объём продаж | | Цена, р. за ед. | |
|---|---|---|---|---|
| | I кв. | II кв. | I кв. | II кв. |
| Краска масляная, кг | 4 800 | 5 300 | 70 | 74 |
| Растворитель, л | 3 200 | 3 000 | 43 | 48 |
| Плинтус, пог. м | 16 500 | 16 100 | 21 | 19 |
| Плитка кафельная, шт. | 24 000 | 22 200 | 34 | 32 |

**Задача 23**

Имеется информация о посещаемости занятий по дисциплине "Статистика" семью студентами второго курса и результатах сдачи ими экзамена по данной дисциплине.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Число пропусков занятий | 1 | 2 | 0 | 10 | 2 | 12 | 8 |
| Экзаменационная оценка | 5 | 4 | 5 | 3 | 5 | 2 | 3 |

Определить факторный и результативный признаки. Изобразить поле корреляции. Построить линейное уравнение регрессии, нанести её график на поле корреляции. Установить наличие и характер корреляционной зависимости.

**Задача 24**

Имеются данные Росстата по 10 субъектам РФ о среднегодовой численности занятого в экономике населения и валовом региональном продукте в 2009 г.

| Субъект РФ | Занято в экономике, тыс. ч. | Валовой продукт, млн р. |
|---|---|---|
| 1 Липецкая область | 544,9 | 226 464 |
| 2 Тверская область | 588,8 | 197 892 |
| 3 Сахалинская область | 288,7 | 392 311 |
| 4 Астраханская область | 447,7 | 132 211,8 |
| 5 Карачаево-Черкесская Республика | 170,6 | 38 583 |
| 6 Свердловская область | 2 064,1 | 823 833 |
| 7 Красноярский край | 1 439,3 | 748 512 |
| 8 Кемеровская область | 1 294,7 | 512 421 |
| 9 Приморский край | 980,1 | 367 698 |
| 10 Хабаровский край | 729,4 | 274 983 |

Определить факторный и результативный признаки. Изобразить поле корреляции. Построить линейное уравнение регрессии. Оценить наличие и тесноту связи, значимость линейного коэффициента корреляции.

**Задача 25**

Имеются данные, характеризующие зависимость между объёмом производства некоторого вида продукции и её себестоимостью по восьми предприятиям одной из отраслей промышленности.

| Номер предприятия | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Объём производства продукции, тыс. шт. | 45 | 18 | 74 | 55 | 84 | 13 | 28 | 40 |
| Себестоимость продукции, млн р. | 62 | 34 | 120 | 114 | 154 | 44 | 60 | 76 |

Определить факторный и результативный признаки. Изобразить поле корреляции. Построить логарифмическое уравнение регрессии. Оценить наличие и тесноту связи, значимость индекса корреляции *R*.

## 12. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ РАСЧЁТНО-ГРАФИЧЕСКОГО ЗАДАНИЯ

Расчётно-графическое задание выполняется на тему "Анализ рядов динамики и прогнозирование", при этом используются статистические данные по Хабаровскому краю за 2004 – 2010 гг. Вариант РГЗ определяет преподаватель.

***Задание:***

1) Провести выравнивание ряда динамики с использованием линейной функции и параболы второго порядка.

2) Рассчитать стандартизованные ошибки аппроксимации для линейной функции и параболы второго порядка.

3) На основе стандартизованных ошибок аппроксимации выбрать функцию, более адекватно описывающую динамику анализируемого показателя.

4) Используя найденную функцию, дать точечный прогноз показателя на 2011 г., проиллюстрировать результат графически.

5) Сделать выводы.

**Вариант 1**

Число учреждений начального профессионального образования в Хабаровском крае (ед.).

| 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|
| 48 | 44 | 43 | 29 | 25 | 22 |

**Вариант 2**

Число погибших в дорожно-транспортных происшествиях в Хабаровском крае (на 100 тыс. чел.).

| 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|
| 25,6 | 22,9 | 25,5 | 22,2 | 16,4 | 18,4 |

**Вариант 3**

Производство яиц в хозяйствах всех категорий Хабаровского края (млн шт.).

| 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|
| 272,0 | 271,5 | 268,9 | 269,2 | 275,7 | 298,7 |

**Вариант 4**

Среднедушевые денежные доходы населения в Хабаровском крае (р.).

| 2004 г. | 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|---|
| 7 597 | 9 451 | 11 999 | 14 574 | 15 705 | 19 203 | 22 607 |

**Вариант 5**

Среднемесячная номинальная начисленная заработная плата в Хабаровском крае (р.).

| 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|
| 11 335,6 | 12 887,6 | 15 883,5 | 18 984,5 | 20 455,0 | 22 656,5 |

**Вариант 6**

Средний размер назначенных пенсий в Хабаровском крае (р.).

| 2004 г. | 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|---|
| 2 339,1 | 2 925,4 | 3 290,1 | 4 291,1 | 5 341,9 | 7 239,7 | 8 816,5 |

**Вариант 7**

Число собственных легковых автомобилей на 1000 человек населения в Хабаровском крае.

| 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|
| 156,7 | 160,2 | 158,2 | 167,4 | 183,4 |

**Вариант 8**

Общая площадь жилых помещений, приходящихся в среднем на одного жителя Хабаровского края (на конец года, м$^2$).

| 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|
| 19,9 | 20,1 | 20,1 | 20,4 | 20,7 | 21,8 |

**Вариант 9**

Поголовье крупного рогатого скота в хозяйствах всех категорий Хабаровского края (на конец года, тыс. голов).

| 2004 г. | 2005 г. | 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|---|---|
| 40,3 | 38,0 | 35,0 | 34,7 | 32,7 | 30,0 | 26,9 |

**Вариант 10**

Надой молока на одну корову в сельскохозяйственных организациях Хабаровского края (кг).

| 2006 г. | 2007 г. | 2008 г. | 2009 г. | 2010 г. |
|---|---|---|---|---|
| 2617 | 2743 | 2739 | 2910 | 3185 |

При выполнении расчётно-графического задания следует использовать методику анализа, изложенную в подразделах 8.5 и 8.6 данного пособия. Решение должно иметь необходимые пояснения и расшифровки обозначений используемых формул.

Расчётно-графическое задание оформляется в соответствии с действующими в университете требованиями к оформлению студенческих текстовых работ.

## ЗАКЛЮЧЕНИЕ

Информатик, работающий с информационными системами, содержащими большие массивы статистической информации, обязан знать основные приёмы и методы обработки и анализа массовых данных. Для корректного построения информационной системы необходимо понимать механизм сбора исходных данных, их природу, взаимосвязь между данными, знать методы проверки данных на полноту и достоверность, сопоставимость. Для эффективной работы с массивами данных информатик должен владеть навыками использования различных статистических методов анализа массовых данных при решении типовых задач.

В пособии рассмотрены основные приёмы и методы обработки и анализа статистических данных, позволяющие выявлять структуру совокупностей, рассчитывать средние показатели и вариацию, оценивать результаты выборочных наблюдений, выявлять и измерять связи между явлениями, определять тенденции развития и делать прогнозы. Знание этих методов и владение навыками их практического использования позволяют специалистам в полной мере использовать потенциал существующих информационных систем и создавать новые, более эффективные системы.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1 Гусаров, В. М. Статистика : учеб. пособие для экономических специальностей вузов / В. М. Гусаров. – М. : ЮНИТИ-ДАНА, 2001. – 463 с.

2 Ефимова, М. Р. Практикум по общей теории статистики : учеб. пособие для экономических специальностей вузов / М. Р. Ефимова, О. И. Ганченко, Е. В. Петрова. – М. : Финансы и статистика, 2007. – 280 с.

3 Общая теория статистики. Статистическая методология в изучении коммерческой деятельности : учеб. для вузов / О. Э. Башина, А. А. Спирин, В. Т. Бабурин, И. А. Ионсен, М. П. Федоров, А. И. Харламов. – М. : Финансы и статистика, 2005. – 440 с.

4 Общая теория статистики : учеб. для вузов / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев, Н. К. Агеева, Н. Е. Варакина, Е. А. Саблина. – М. : ИНФРА, 2000. – 416 с.

5 Теория статистики : учеб. для вузов / Г. Л. Громыко [и др.]. – М. : ИНФРА-М, 2006. – 476 с.

6 Теория статистики : учеб. для вузов / Р. А. Шмойлова, Е. Б. Бесфамильная, Н. Ю. Глубокова, А. Б. Бусыгин, В. Г. Минашкин, А. А. Романов, Н. А. Садовникова. – М. : Финансы и статистика, 2009. – 576 с.

7 Шмойлова, Р. А. Практикум по теории статистики : учеб. пособие для экономических специальностей вузов / Р. А. Шмойлова, В. Г. Минашкин, Н. А. Садовникова. – М. : Финансы и статистика, 2006. – 416 с.

*Учебное издание*

**Инзарцев** Алексей Вячеславович

# СТАТИСТИКА В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Учебное пособие

Научный редактор – кандидат технических наук,
доцент А. В. Еськова

Редактор Ю. Н. Осинцева

Подписано в печать 06.11.2013.
Формат 60 × 84 1/16. Бумага 65 г/м$^2$. Ризограф EZ570E.
Усл. печ. л. 6,97. Уч.-изд. л. 6,68. Тираж 75 экз. Заказ 25868.

Редакционно-издательский отдел
Федерального государственного бюджетного образовательного учреждения
высшего профессионального образования
«Комсомольский-на-Амуре государственный технический университет»
681013, Комсомольск-на-Амуре, пр. Ленина, 27.

Полиграфическая лаборатория
Федерального государственного бюджетного образовательного учреждения
высшего профессионального образования
«Комсомольский-на-Амуре государственный технический университет»
681013, Комсомольск-на-Амуре, пр. Ленина, 27.