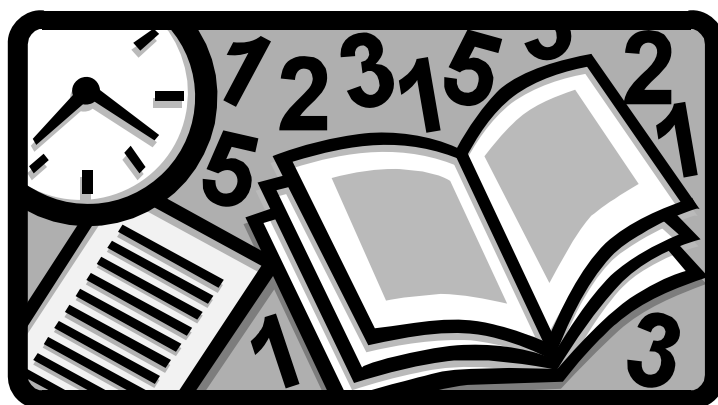


Т.М. Сизова

Статистика



Санкт-Петербург
2005

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

Т.М. СИЗОВА

СТАТИСТИКА

Учебное пособие



Санкт-Петербург
2005

ББК

Сизова Т.М.

Статистика: Учебное пособие. – СПб.: СПб ГУИТМО, 2005. – 80 с.

Учебное пособие разработано в соответствии с программой дисциплины «Статистика» и предназначено для студентов гуманитарного факультета СПб ГУИТМО, обучающихся по специальностям 080507 «Менеджмент организации», 080103 «Национальная экономика», 08081 «Прикладная информатика (в экономике) и по направлению 521600 «Экономика».

Одобрено на заседании Совета гуманитарного факультета 17.02.2005, протокол № 6.

Рецензенты:

***доктор экономических наук, профессор Э.Н. Кроливецкий, СПб ГУ КиТ,
доктор экономических наук, профессор А.М. Колесников, СПб ГУ АКП.***

Рекомендовано Учебно-методическим объединением по образованию в области национальной экономики и экономики труда в качестве учебного пособия для студентов высших учебных заведений, обучающихся по специальности 080103 (060700) «Национальная экономика» и другим экономическим специальностям.

ISBN

© Санкт-Петербургский
государственный университет
информационных технологий,
механики и оптики, 2005

© Т.М. Сизова, 2005

Содержание

1. Статистика как наука	7
1.1. Предмет, метод и задачи статистики	7
1.2. Основные категории статистики	11
2. Статистические показатели	16
2.1. Понятие, формы выражения и виды статистических показателей	16
2.2. Абсолютные статистические показатели	18
2.3. Относительные показатели	20
2.4. Средние показатели	26
2.5. Сопоставимость статистических показателей	33
3. Статистическое наблюдение	35
3.1. Информационная база статистического исследования, статистическое наблюдение и его этапы	35
3.2. Программно-методологические и организационные задачи статистического наблюдения	38
3.3. Формы, виды и способы проведения статистического наблюдения	41
3.4. Ошибки статистического наблюдения и контроль данных	46
4. Сводка и группировка данных	49
4.1. Задачи сводки и группировки	49
4.2. Типы группировок	51
4.3. Сложные группировки	59
4.4. Статистические таблицы	63
5. Обобщающие характеристики статистических совокупностей	66
5.1. Ряды распределения	66
5.2. Частотные характеристики рядов распределения	67
5.3. Графическое представление рядов распределения	69
5.4. Показатели центра распределения	71
5.5. Показатели вариации	76
5.6. Правило сложения дисперсий	81
5.7. Исследование формы рядов распределения	82
5.8. Теоретические функции в анализе рядов распределения	86
5.9. Оценка близости эмпирического и теоретического распределений	92
6. Ряды динамики	95
6.1. Понятие и классификация рядов динамики	95
6.2. Система характеристик рядов динамики	97
6.3. Разложение рядов динамики	102
6.4. Выявление тренда	103
6.5. Метод аналитического выравнивания	106
6.6. Анализ сезонных колебаний	112

6.7.Экстраполяция в рядах динамики и прогнозирование	117
7. Статистическое изучение взаимосвязи социально-экономических явлений.....	120
7.1.Понятие функциональной, статистической корреляционной связи ...	120
7.2.Парная корреляция	123
7.3.Парная регрессия на основе метода наименьших квадратов	127
7.4.Оценка существенности парной корреляционной связи	135
7.5.Множественная корреляция	138
7.6.Анализ взаимосвязанных рядов динамики	146
8. Выборочное исследование	149
8.1.Постановка задачи выборочного исследования	149
8.2.Статистические оценки параметров генеральной совокупности	151
8.3.Ошибки выборки	153
8.4.Способы формирования выборочной совокупности	159
8.5.Численность выборки и способы распространения ее характеристик на генеральную совокупность.....	164
9. Экономические индексы	168
9.1.Индексы и их использование в экономико-статистических исследованиях	
9.2.Виды и формы индексов	170
9.3.Агрегатные индексы количественных показателей	173
9.4.Агрегатные индексы качественных показателей	176
9.5.Индексные системы и факторный анализ	179
9.6.Средние индексы	181
Литература	188

Введение

Учебное пособие предназначено для использования в учебном процессе студентами, обучающимися по специальностям 080507 «Менеджмент организации», 080103 «Национальная экономика» и 080801 «Прикладная информатика (в экономике)», и разработано в соответствии с требованиями действующих учебных планов.

Материалы учебного пособия позволяют студентам получить необходимые знания по одной из важнейших дисциплин подготовки специалистов – статистике.

В учебном пособии рассмотрены основные методы статистического исследования: статистическое наблюдение, сводка и группировка, исследование рядов распределения, анализ рядов динамики, выборочный метод, корреляционно-регрессионный анализ, индексный метод анализа.

Структура учебного пособия соответствует двум основным функциям, выполняемым статистикой как наукой: описательной и объясняющей. Соответственно. Пособие состоит из двух частей - описательной и аналитической статистики.

В первой части (п.п. 1-4) раскрываются:

- сущность статистики как науки, особенности статистической методологии, основные понятия и категории статистики;
- методология исчисления абсолютных, относительных и средних показателей и их использование в экономико-статистическом анализе;
- методы сбора статистической информации (формы, виды и способы статистического наблюдения), программно-методологические и организационные вопросы статистического наблюдения, сущность ошибок наблюдения и контроль данных наблюдения;
- метод статистических группировок и таблиц.

Во второй части (п.п. 5-8) рассматриваются:

- статистические методы и показатели структуры изучаемого явления, такие как построение и анализ вариационных рядов, определение показателей положения центра распределения, уровня вариации признака, формы распределения;
- анализ рядов динамики, построение моделей их разложения, выявление и описание тренда, оценка сезонности, методы статистического прогнозирования на основе экстраполяции тренда;
- статистическое изучение взаимосвязей социально-экономических явлений на основе корреляционно-регрессионного анализа;
- проблемы экономических индексов.

1. Статистика как наука

1.1. Предмет и метод статистики

Термин статистика происходит от латинских “Status”, что означает «определенное состояние явления, положение вещей», и “Stato”, означающее «государство». Он был введен в научный оборот в 1749 году немецким ученым Готфридом Ахенвалем, опубликовавшим книгу под названием «Статистика», в которой приводилось описание политического устройства государств Европы. Как наука статистика возникла в 17-ом веке, однако статистический учет проводился уже в глубокой древности. Так, известно, что в 5-м веке до н. э. в Китае проводились переписи населения; в Древнем Риме велся учет имущества граждан. В античных Афинах проводился учет естественного движения населения: число родившихся и умерших определялось посредством подсчета подношений жрице богини Миневры - за каждого родившегося приносился дар в 1 меру пшеницы, за каждого умершего - 1 мера ячменя. На Руси в 10-12 вв. собирались разнообразные сведения, связанные с налогообложением населения.

Как научная дисциплина статистика оформилась в середине 18 в. в результате слияния двух научных школ: немецкой описательной школы (Г. Ахенваль), которая трактовала статистику как государствоведение и английской школы политических арифметиков (В.Петти). От немецкой описательной школы современная статистика взяла систему словесного описания социально-экономических явлений без цифр, вне динамики; от английской школы политических арифметиков - изучение общественных явлений с помощью числовых характеристик, статистическое обобщение полученных характеристик с целью выявления закономерностей развития изучаемых явлений.

Представители описательной школы пытались систематизировать существующие способы описаний «государственных достопримечательностей» - территорию, государственное устройство, население, религию, внешнюю политику и т.п. Описание таких «достопримечательностей» они проводили в словесной форме, без цифр и вне динамики, только на момент наблюдения.

Политические арифметики своей целью ставили изучение общественных явлений с помощью числовых характеристик, они пытались выявить закономерности развития и взаимосвязи экономических явлений с помощью математических расчетов, осознавали необходимость учета в статистических исследованиях требований закона больших чисел, поскольку закономерность

может проявиться только при большом объеме статистической совокупности.

В первой половине 19в. возникло третье направление статистики – статистико-математическое (А. Кетле), представители которого считали основой статистики теорию вероятностей.

Развитие статистической науки, расширение сферы практической статистической работы привели к изменению содержания термина «статистика».

В настоящее время под статистикой понимается:

- **совокупность итоговых сведений, количественно характеризующих различные стороны общественной жизни:** производство, распределение и обмен товарами, политику, культуру и т.д.;
- **практическая деятельность по сбору, обработке и анализу количественных данных** об общественной жизни и их публикацию;
- **научная дисциплина, отрасль знаний, изучающая количественную сторону массовых явлений и процессов в неразрывной связи с их количественной стороной с целью выявления закономерностей их развития.**

Статистика является важным элементом плана подготовки специалистов высшей экономической квалификации.

Современная статистика – это особая наука, имеющая свой объект, предмет и специфические методы исследования.

Объектом изучения статистики являются **массовые явления и процессы** любой природы, в том числе и в экономике.

Массовыми называются явления, повторяющиеся в пространстве и времени, и отражающие определенную статистическую закономерность. *Массовое явление обладает следующими признаками:*

- | **состоит из множества индивидуальных элементов; каждый элемент – это реально существующие материальные объекты, обладающие некоторым общим свойством, именуемым в статистике – признаком;**
- | **вариация индивидуальных значений признака - каждый элемент должен иметь значение изучаемого статистического признака, отличное от других;**
- | **частичная или полная независимость элементов друг от друга.**

Удалено: -

Удалено: -

Удалено: -

Статистика изучает массовые общественно-экономические явления в конкретных обстоятельствах места и времени.

Предметом статистики выступают **размеры и количественные соотношения массовых общественных явлений в неразрывной связи с их качественной стороной** с целью выявления закономерностей их развития.

Свой предмет статистика изучает на основе особой **методологии, представляющей собой совокупность общих правил (принципов) и**

специальных приемов и методов статистического исследования. Общие правила статистического исследования исходят из *положений социально-экономической теории и принципа диалектического метода познания.* Они составляют теоретическую базу статистики.

Социально-экономическая теория объясняет сущность изучаемых явлений или процессов, законы их развития в конкретных обстоятельствах. Одновременно статистика обогащает социально-экономические науки численными данными, полученными в исследованиях, используется для подтверждения или опровержения выдвигаемых ими теоретических гипотез.

В соответствии с *диалектическим методом познания* статистика изучает массовые явления в их взаимосвязи, в движении и изменении, выявляя их количественные и качественные изменения. Для этого используются специальные статистические методы, такие, например, как сводка и группировка, индексный метод, корреляционно-регрессионный и др.

Опираясь на теоретическую базу, статистика применяет специфический метод, который находит свое выражение в трех стадиях статистического исследования:

- **Статистическое наблюдение**, заключающееся в сборе первичной информации об отдельных фактах изучаемого явления.
- **Группировка и сводка** собранного материала, позволяющие провести их систематизацию и классификацию.
- **Обработка статистических показателей**, полученных в результате группировки и сводки, и анализ полученных результатов с целью получения выводов о состоянии и закономерностях развития изучаемого явления.

Формат: Список

В статистике, как научной дисциплине выделяют *три уровня*:

- **общая теория статистики**, которая занимается разработкой понятийного аппарата и системы категорий статистической науки, общих принципов и правил проведения статистических исследований, универсальных методов обработки информации, то есть разработкой общей методологии статистического исследования массовых общественных явлений;

Формат: Список

- **экономическая и социальная статистика**

Экономическая статистика - занимается исследованием экономических явлений и процессов, их количественной оценкой и разработкой синтетических экономических показателей, таких как валовое национальное богатство, валовой внутренний продукт, национальный доход и др. Социальная статистика – соответственно исследованиями и разработкой обобщающих показателей в различных областях общественной жизни: культура, политика, наука и т.д.

- **отрасли экономической и социальной статистики**

В экономической статистике выделяются такие ее отрасли как статистика промышленности, сельского хозяйства, транспорта, строительства, торговли, связи, природных ресурсов, и т. д., в социальной статистике - статистика науки, права, здравоохранения, политическая статистика, населения и т.д.

Задачами отраслевых статистик является изучение массовых явлений, имеющих место в соответствующих отраслях и сферах общественно-экономической жизни, разработка обобщающих показателей этих отраслей, выявления тенденций и закономерностей развития указанных сфер общественно-экономической жизни.

Статистика развивается как единая наука, и развитие каждой отрасли содействует ее совершенствованию в целом.

Статистика имеет огромное познавательное значение, которое заключается в следующем:

- статистика дает численное и содержательное освещение изучаемых явлений и процессов, служит надежным способом оценки действительности;
- статистика дает доказательную силу экономическим выводам позволяет проверить выдвигаемые гипотезы, отдельные теоретические положения;

- статистика раскрывает взаимосвязи между явлениями, показывает их конкретную форму и силу;

- статистика первой обнаруживает новые явления, процессы и закономерности, дает их количественную и качественную характеристику.

Статистика является важным элементом плана подготовки специалистов высшей экономической квалификации. Знание статистики необходимо современному специалисту для принятия решений в условиях, когда анализируемые явления подвержены влиянию случайностей, для анализа элементов рыночной экономики, прогнозирования и разработки сценариев поведения экономических систем при изменении условий их функционирования.

Теоретическую основу курса статистики, как и любой другой научной дисциплины, составляют категории, через которые выражаются её основные принципы, к важнейшим из которых относятся:

- **статистическая совокупность;**
- **статистическая единица;**
- **статистический признак;**
- **статистическая закономерность;**
- **статистические показатели.**

1.2. Основные категории статистики

Статистика изучает свой предмет при помощи определенных категорий, то есть понятий, отражающих наиболее общие и существенные свойства, признаки, связи и отношения предметов и явлений объективного мира.

В статистике таких основных понятий - пять:

1.2.1. Статистическая совокупность – множество существующих во времени и пространстве варьирующих явлений, однокачественных (однородных) по определенному признаку. Это может быть совокупность жителей Санкт-Петербурга, совокупность студентов ИТМО, и т.д. Статистическая совокупность является объектом статистического изучения.

Предложенное определение статистической совокупности позволяет выделить основные её свойства:

- *Неразложимость* – частичное возникновение или частичное исчезновение элементов статистической совокупности не разрушает её качественной основы, все её качественные характеристики сохраняются. Студенты ИТМО как статистическая совокупность не изменит своей качественной характеристики независимо оттого, что часть студентов (выпускники) ежегодно покидают ИТМО, и часть (первокурсники) вливается в их состав.

- Статистическая совокупность всегда *однородна хотя бы по одному признаку*. Все элементы статистической совокупности обладают хотя бы одним общим свойством. Однако, общее не означает одинакового. Значения общего признака у разных единиц совокупности, как правило, отличаются друг от друга. Однородность статистической совокупности устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями и задачами.

- Важнейшим свойством статистической совокупности является *вариация* - количественное изменение значения статистического признака при переходе от одного ее элемента к другому. Если значения признака у всех элементов одинаково, нет смысла в изучении всей статистической совокупности, достаточно рассмотреть лишь один её элемент, чтобы получить знание обо всем явлении. Вариация возникает под воздействием определённого комплекса условий и причин. Статистика не занимается выявлением и выяснением этих причин. Этим занимаются специальные экономические дисциплины, а статистика количественно оценивает воздействие каждой причины на вариацию конкретного признака, что позволяет учитывать указанное воздействие при принятии управленческих решений различного уровня.

Все статистические совокупности можно разделить на следующие группы:

- созданные самой жизнью и образующие единство вне зависимости от того подвергаются они статистическому исследованию или нет (статистические совокупности работников предприятия, статистические совокупности промышленных предприятий Санкт-Петербурга и т.д.). Это реально существующие статистические совокупности, они имеют определённый конкретный размер.

- образованные специально для целей статистического исследования (совокупность покупателей определённого товара в маркетинговых исследованиях)

- стохастические совокупности (гипотетические множества) – мысленные, нереальные, предполагаемые совокупности (совокупность небесных тел в галактике; совокупность бесконечно большого числа бросаний монеты, падающей “орлом” либо “решкой”).

1.2.2. Статистическая единица – *неразложимый первичный независимый элемент статистической совокупности, являющийся носителем определённого статистического признака.* Она является пределом дробления статистической совокупности, при котором сохраняются все свойства изучаемого явления или процесса, ее частным случаем. Выбор статистической единицы определяется целью и уровнем проводимого статистического исследования. Например, можно изучать производительность труда на уровне отрасли, предприятия, цеха, участка, бригады. В каждом случае статистические единицы будут разными: предприятие отрасли, работник данного предприятия, рабочий данного цеха, бригады.

1.2.3. Статистический признак – *характерное свойство, определённое качество статистической совокупности.* Например, статистическими признаками предприятий могут являться: форма собственности, численность работающих, величина уставного капитала, стоимость активов и т. д. Значение признака отдельной единицы статистической совокупности называется вариантом.

Статистические признаки можно классифицировать по множеству оснований. Классификацию удобно представить в виде таблицы:

Таблица 1.1

Классификация статистических признаков

Основания классификации			
Характер выражения	Характер вариации	Отношение ко времени	Характер взаимосвязи
атрибутивные	альтернативные	моментные	Факторные
количественные	дискретные непрерывные	интервальные	результативные

По характеру выражения различают *атрибутивные* и *количественные* признаки:

- *атрибутивные (описательные)* – выражаются словесно, например, пол, национальность, образование и др. По ним можно получить итоговые сведения о количестве статистических единиц, обладающих данным значением признака;
- *количественные* – выражаются числовой мерой (возраст, стаж работы, объем продаж, размер дохода и т.д.) По ним можно получить итоговые данные о количестве единиц, обладающих конкретным значением признака, и суммарное или среднее значение признака по совокупности.

По характеру вариации признаки делятся на:

- *альтернативные* которые могут принимать только одно из двух возможных значений признака. Это признаки обладания или не обладания чем-либо. Например, пол, семейное положение, в маркетинговых или политологических исследованиях - ответ на вопрос в форме «да или нет»;
- *дискретные* – количественные признаки принимающие только отдельные значения, без промежуточных между ними - как правило целочисленные, например, разряд рабочего, число детей в семье и т.д.);
- *непрерывные* – количественные признаки, принимающие любые значения. На практике они, как правило, округляются в соответствии с принятой точностью (например: бухгалтерская прибыль по балансу в рублях, налоговая по налоговым регистрам – в тыс. руб.

По отношению ко времени различают:

- *моментные* признаки, характеризующие единицы совокупности на критический момент времени, например, стоимость основных производственных фондов (ОПФ) определяется на 01.01. и 31.12.

Удалено: -

Удалено: -

Удалено: -

соответствующего года как стоимость ОПФ на начало и конец отчётного года;

Удалено: -

Удалено: н

- *интервальные* признаки, характеризующие явление за определённый временной период ((год, квартал, месяц и т.д.), например, сменная выработка, дневная выручка, годовой объём продаж и т.д.

По характеру взаимосвязи признаки делятся на:

- *факторные*, вызывающие изменения других признаков, либо создающие возможности для изменений значений других признаков. Факторные признаки подразделяются соответственно на признаки причины и признаки условия;
- *результативные* (признаки следствия), зависящие от вариации других признаков. Например, стоимостной объём выпуска продукции является результативным признаком, величина которого зависит от факторных признаков - численности работников и производительности труда.

1.2.4. Статистическая закономерность

Как философская категория любая **закономерность** есть **форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с высокой степенью вероятности, если причины, порождающие события не меняются или меняются не значительно.**

Статистическая закономерность – это объективная количественная закономерность изменения массовых явлений и процессов, то есть статистическая закономерность является количественной формой проявления причинной связи. Она устанавливается на основе анализа массовых данных и проявляется только на уровне статистической совокупности. Закономерность возникает как результат воздействия большого числа постоянно действующих причин и причин случайных, действующих временами. Постоянно действующие причины придают изменениям в явлениях регулярность и повторяемость, случайные – вызывают отклонения в этой регулярности. На уровне статистических единиц закономерность проявляется не всегда; например, известно, что средняя продолжительность жизни у женщин больше, чем у мужчин. Но это не означает, что каждая женщина живёт дольше, чем каждый мужчина (среди мужчин встречается больше долгожителей).

Так как статистическая закономерность обнаруживается в итоге массовых статистических данных, это обуславливает ее взаимосвязь с *Законом больших чисел* - статистические закономерности являются

следствием действия этого закона. *Закон больших чисел* в самой простой формулировке гласит, что *в массовых явлениях и процессах случайные второстепенные признаки у наблюдаемых единиц взаимопогашаются, в результате чего отчетливо проявляются наиболее существенные признаки, закономерности развития таких явлений*. Таким образом, закон выражает диалектику случайного и необходимого. Например, на 100 девочек рождается 104-106 мальчиков, но в разных семьях и даже в небольших населенных пунктах это соотношение может быть абсолютно иным. В соответствии с природой массовой закономерности тенденции, вскрытые с помощью закона больших чисел, имеют силу только как массовые тенденции, но не как законы, фиксирующие устойчивый, всеобщий характер причинно-следственной связи явлений. При статистической же закономерности эти связи менее устойчивы и не имеют всеобщего характера, а относятся к определенному пространству и времени, справедливы лишь для данных условий существования изучаемого явления.

Сам термин «статистическая закономерность» впервые стал применяться в естественных науках в противовес понятию «динамическая закономерность», выражающему такую форму связи, при которой конкретным значениям каких-либо факторов всегда строго соответствуют определенные значения зависимых от этих факторов признаков. В случае динамической закономерности количественные соотношения между величинами остаются справедливыми для каждой отдельной статистической единицы. Например, площадь круга изменяется с изменением его радиуса, и эта зависимость выражается формулой $S=2\pi r^2$, которая справедлива для любого круга.

2. Статистические показатели

2.1. Понятие, формы выражения и виды статистических показателей

Статистические показатели являются одной из важнейших категорий статистики, Они используются для описания исследуемых массовых явлений и процессов, являются инструментом их познания.

Как философская категория статистический показатель есть мера, т. е. единство качественного и количественного отражения свойств объективных явлений и процессов в научном познании.

Статистический показатель есть количественно-качественная обобщающая характеристика какого-либо свойства статистической совокупности **в условиях конкретного места и времени**. Этим он отличается от индивидуальных значений признака (вариант). Например, средняя заработная плата работников предприятия – статистический показатель, а заработная плата конкретного работника – это индивидуальное значение признака (**варианта**).

В отличие от индивидуального значения признака статистический показатель может быть получен только расчетным путем. Это может быть простой подсчет единиц совокупности, суммирование их значений признаков, или более сложные расчеты.

В соответствии с определением статистический показатель имеет определенную структуру, в нем различают качественную и количественную стороны.

Качественная сторона статистического показателя определяется признаком, который подлежит изучению и отражается в названии показателя, *количественная сторона* - в численном значении показателя.

Еще одной особенностью статистических показателей является то, что они *всегда привязаны к конкретным обстоятельствам места и времени*.

Таким образом, с помощью статистических показателей определяется что, где, когда и каким образом следует измерять или оценивать. Например, численность постоянного населения г. Санкт-Петербурга на 1 января 2004 г. составила 4.631тыс. чел. Качественная сторона показателя – постоянное население; обстоятельство места □ г. Санкт-Петербург; Обстоятельство времени – 1января 2004г.; количественная сторона –4.631тыс. чел.

Важной особенностью статистики является *использование системного подхода* в исследовании социально-экономических явлений, что предполагает использование для их оценки систем статистических показателей.

Система статистических показателей - это совокупность статистических показателей, отражающая объективно существующие

взаимосвязи между явлениями. Система статистических показателей позволяет получить целостную статистическую характеристику социально-экономического явления.

Виды и формы таких систем разнообразны и зависят от решаемых задач и сложности изучаемых явлений.

Например, для характеристики хозяйственной деятельности предприятия используется система, включающая следующие согласованные между собой показатели: прибыль, рентабельность, издержки производства, объем производства, объем реализации, показатели производительности труда, численность ППП, стоимость ОПФ, и т.д. Все эти показатели логически взаимосвязаны. Их наличие в системе вытекает из сущности промышленного предприятия: производство продукции на базе эффективного взаимодействия средств производства и трудовых ресурсов с целью получения прибыли.

Следует иметь в виду, что любая система статистических показателей всегда лишь схематично, с упрощениями, в зависимости от сложившихся представлений отражает изучаемое явление. Поэтому важно постоянно совершенствовать такие системы.

На практике для отражения разнообразных сторон социально-экономических явлений и процессов используются разнообразные статистические показатели, которые можно классифицировать следующим образом:

Таблица 2.1

Классификация статистических показателей

Основания классификации		
Выполняемая функция	Охватываемые единицы совокупности	Форма выражения
плановые учётные прогностические	индивидуальные сводные	абсолютные относительные средние

Плановые показатели - отражают директивную функцию, ориентированы на выполнение поставленных задач, *учётные* – показывают реальное состояние изучаемого явления, а *прогностические* – его возможное состояние в будущем.

Индивидуальные – характеризуют отдельный объект или отдельную единицу совокупности – предприятие, домохозяйство и др. Примером

индивидуальных статистических показателей может быть объем продаж торговой фирмы, численность работающих предприятия и т.д.

Сводные (обобщающие) показатели исчисляются по всей совокупности в целом, являются научными абстракциями и занимают особое место в познании статистических закономерностей.

Удалено: .

Абсолютные – исходная первичная форма выражения статистических показателей.

Относительные – производные, вторичные показатели по отношению к абсолютным, выражающие определённые соотношения между количественными характеристиками статистических совокупностей.

Средние – наиболее распространённая форма статистических показателей, характеризующая наиболее типичный уровень явления. Рассчитываются на единицу статистической совокупности или на единицу признака.

2.2. Абсолютные статистические показатели

Абсолютные показатели характеризуют численность совокупности, либо объём изучаемого явления в конкретных границах пространства и времени, т. е. отражают уровень развития явления, его размер.

Абсолютный показатель можно получить одним из двух способов:

- путём подсчёта единиц совокупности, обладающих конкретным значением признака; например, число транспортных предприятий в Санкт-Петербурге на конкретную дату, численность промышленно-производственного персонала предприятия и т.д.

- путём суммирования значения признака по всей статистической совокупности; например, объём товарооборота предприятий торговли города в 2004г.

Абсолютные показатели всегда являются именованными числами. В зависимости от социально-экономической сущности исследуемых явлений они выражаются в **натуральных, стоимостных и трудовых единицах измерения.**

Натуральные измерители используются в тех случаях, когда единицы измерения соответствуют потребительским свойствам изучаемых явлений, например: производство автомобилей измеряется в штуках, производство стали - в тоннах, урожайность - в центнерах и т.д.

Натуральные единицы измерения могут быть составными (сложными). Такие единицы применяются в тех случаях, когда для характеристики изучаемого явления одной единицы измерения недостаточно, и используется произведение двух единиц. Например,

производство электроэнергии измеряется в киловатт-часах, грузооборот – в тонно-километрах и т. д.

В группу натуральных, включаются также *условно- натуральные единицы измерения*. Они используются, когда какой-либо продукт имеет несколько разновидностей, и общий объем можно получить только исходя из общего для всех разновидностей потребительского свойства. Например, в консервной промышленности объёмы производства определяются в условных консервных банках объёмом 353,4 см³., в топливной – в условном топливе с теплотой сгорания 7000 ккал/кг (29,3 мДж/кг).

Чтобы получить обобщённые итоги одна из разновидностей продукта принимается в качестве единого измерителя, а другие приводятся к нему с помощью соответствующих коэффициентов пересчёта. Например, если месторождение даёт за год 100 тыс. тонн нефти с теплотой сгорания 45,0 МДж/кг, то в условном топливе это будет эквивалентно $100 \times 45,0 / 29,3 = 153,6$ тыс. тонн условного топлива.

Стоимостные измерители позволяют дать денежную оценку изучаемым явлениям и процессам. Эти измерители используются при обобщении данных, начиная с уровня предприятия и до уровня народного хозяйства, при оценке неоднородных статистических совокупностей. В стоимостных единицах измеряется объем выпущенной продукции предприятия, доходы населения и т. д. Схема получения общего объёма статистического признака в стоимостном выражении выглядит следующим образом:

$$Q = \sum_{i=1}^m p_i \cdot q_i ,$$

где p_i – цена (стоимостная оценка единицы признака)

q_i - объём признака в натуральном выражении

m – количество признаков.

Показатели, выраженные в стоимостных единицах, можно суммировать, получать по ним итоговые данные, но при их использовании необходимо учитывать изменение цен с течением времени. Для устранения указанного недостатка стоимостных измерителей следует применять «неизменные» или «сопоставимые» цены одного итого же периода.

Трудовые единицы измерения применяются для оценки общих затрат труда и трудоемкости отдельных операции техпроцесса. К ним относятся

человеко-часы, человеко-дни (оценка затрат рабочего времени), нормо-минуты (оценка трудоёмкости).

Сами по себе абсолютные показатели не дают полного представления об изучаемом явлении, не показывают его структуру, развитие во времени, соотношение между частями явления, на их основе сложно проводить сравнения с другими подобными явлениями. Перечисленные аналитические функции выполняют относительные показатели.

2.3. Относительные показатели

Относительным статистическим показателем называется обобщающая характеристика, выраженная в виде числовой меры соотношения двух сопоставляемых абсолютных величин. Такие показатели используются в различных целях: для выяснения структуры изучаемого явления, для сравнения его уровня развития с уровнем развития другого явления, для оценки происходящих в изучаемом явлении изменений и т. д.

Относительный статистический показатель получают путём деления одного абсолютного показателя на другой. Схема расчета относительного показателя выглядит следующим образом:

$$ОП = \frac{\text{величина_сравнения}}{\text{база_сравнения}};$$

где ОП – относительная величина,
величина сравнения и база сравнения – сравниваемые абсолютные показатели.

Таким образом, по способу получения относительные показатели всегда являются величинами производными, их можно получить только расчетным путем.

Относительные показатели выражаются в разных формах - *коэффициентов, процентов, промилле, продецимилле*. Если база сравнения принимается за единицу, то относительный показатель выражается в форме **коэффициента**. Например, в 2003г в Санкт-Петербурге родилось 40,3тыс. детей, а в 2002г. – 37,2тыс. Коэффициент роста числа родившихся составит 1,083 (40,3/37,2).

Если база сравнения принимается за 100 единиц, то относительный показатель выражается **в процентах**. По предыдущим данным рост количества родившихся составит 108,3%.

Если база сравнения принимается за 1000 единиц, то относительный показатель выражается в **промилле** (десятая часть процента), если – за 10 000, то относительный показатель выражается в **продецимилле** (сотая часть процента). *Промилле* широко применяются в демографической статистике для характеристики рождаемости, смертности населения и других демографических процессов. *Продецимилле* используются для оценки обеспечения населения больничными койками, местами в высших учебных заведениях и т.д.

Следует заметить, что безразмерным по форме **относительным показателям может быть приписана конкретная единица измерения**. Например, показатели естественного движения населения – коэффициент рождаемости, коэффициент смертности, исчисляется в промилле, но показывает число родившихся или умерших за год в расчёте на 1000 чел. Например, показатель рождаемости в Санкт-Петербурге в 2003г. составил 8,9 промилле, что означает 8,9 рождений на тысячу населения.

По содержанию выражаемых количественных соотношений выделяют шесть видов относительных показателей: динамики, плана и выполнения плана, структуры, координации, интенсивности и уровня экономического развития, сравнения.

2.3.1. Относительный показатель динамики характеризует изменение изучаемого явления во времени и представляет собой соотношение показателей, характеризующих явление в текущем периоде и предшествующем (базисном) периоде.

$$ОПД = \frac{\text{текущий_показатель}}{\text{предшествующий_показатель(базисный)}}.$$

Рассчитанный таким образом показатель называется коэффициентом роста (снижения). Он показывает, во сколько раз показатель текущего периода больше (меньше) показателя предшествующего (базисного) периода. Выраженный в %, относительный показатель динамики называется темпом роста (снижения)

Например, если численность населения РФ по данным переписи населения 2002г. составила 145 181,9тыс. чел., а по данным переписи 1989г. –

147 021,9тыс. чел., то коэффициент (темп) снижения численности населения составил: $K=145181,9/147021,9=0,987$ или 98,7%.

2.3.2. Относительный показатель плана (прогноза) и выполнения плана

Относительный показатель плана (ОПП) и относительный показатель выполнения плана (ОПВП) используют все субъекты финансово-хозяйственной деятельности, осуществляющие текущее и стратегическое планирование. Они рассчитываются следующим образом:

$$\text{ОПП} = \frac{\text{Плановый_показатель_данного_периода}}{\text{Фактический_показатель_предшествующего_периода}};$$

$$\text{ОПВП} = \frac{\text{Фактический_показатель_данного_периода}}{\text{Плановый_показатель_данного_периода}}.$$

Относительный показатель выполнения плана характеризует напряженность планового задания, а относительный показатель выполнения плана – степень его выполнения.

Пример расчета ОПП и ОПВП: фактический оборот фирмы в 2003 г. составил 2млрд. руб., анализ рынка показал, что за 2004г. реально довести оборот до 2,6 млрд. руб., фактический же оборот в 2004г. составил 2,5 млн. руб.

$$\text{ОПП} = \frac{2,60}{200} = 1,3;$$

$$\text{ОПВП} = \frac{2,50}{2,60} = 0,96.$$

Расчеты показывают, что плановое задание на 2004г. в 1,3 раза превышает фактический уровень 2003г., но план 2004г. выполнен только на 96%.

2.3.3. Относительные показатели структуры (ОПС)

характеризуют доли (удельные веса) составных частей совокупности в общем ее объеме. Они показывают структуру совокупности, ее строение. Расчет относительных показателей структуры заключается в исчислении удельных весов отдельных частей во всей совокупности:

$$\text{ОПС} = \frac{\text{показатель_части_совокупности}}{\text{показатель_всей_совокупности}}$$

ОПС обычно выражаются в форме коэффициентов или процентах, сумма коэффициентов должна составлять 1, а сумма процентов – 100, так как удельные веса приведены к общему основанию.

Относительные показатели структуры используются при изучении состава сложных явлений, распадающихся на части, например: при изучении состава населения по различным признакам (возрасту, образованию, национальности и др.).

Пример расчёта структуры розничного товарооборота приведен в таблице 2.2.

Таблица 2.2.

Структура розничного товарооборота города в 2004г

Показатели	Сумма, млрд. руб.	В % к итогу
Розничный товарооборот, всего:	83,1	100
в том числе:		
1. Организации розничной торговли	42,4	51
2. Неторговые организации	20,8	5
3. Физические лица на рынках	19,9	21

Совокупность относительных величин структуры показывает строение совокупности.

2.3.4. Относительные показатели координации (ОПК) характеризуют отношение частей данных статистической совокупности к одной из них, взятой за базу сравнения и показывают, во сколько раз одна часть совокупности больше другой, или сколько единиц одной части совокупности приходится на 1,10,100 и т.д. единиц другой части. За базу

сравнения выбирается часть, имеющая наибольший удельный вес или являющаяся приоритетной в совокупности.

Так, приняв за базу в предыдущем примере товарооборот организаций розничной торговли, можно рассчитать ОПК для неторговых организаций:

$ОПК = 25/51 = 0,49$; что означает следующее - на каждый рубль товарооборота розничной торговли приходится 49 копеек товарооборота неторговых организаций.

Относительные показатели координации играют важную роль в экономическом анализе, так как с их помощью существующие в совокупности соотношения представляются более отчетливо и наглядно.

2.3.5. Относительные показатели интенсивности и уровня экономического развития (ОПИ) характеризуют степень распространения или уровень развития изучаемых явлений или процессов в определённой среде и образуются как результат сравнения разноименных, но определенным образом связанных между собой величин. Указанные показатели рассчитываются следующим образом:

$$ОПИ = \frac{\text{Показатель}_\text{характеризующий}_\text{явление}}{\text{Показатель}_\text{характеризующий}_\text{среду}_\text{распространения}_\text{явления}}.$$

ОПИ исчисляются в расчете на 100, 1000, 1000 и т.д. единиц изучаемой совокупности и используются в тех случаях, когда невозможно по значению абсолютного показателя определить масштаб распространения явления. Так, при изучении демографических процессов рассчитываются показатели рождаемости, смертности, естественного прироста (убыли) населения как отношение числа родившихся (умерших) или величины естественного прироста за год к среднегодовой численности населения данной территории на 1000 или 10 000 человек.

Например, по состоянию на 1.01.04г. в Санкт-Петербурге родилось 40,3тыс. новорожденных, в Ленинградской области -13,4тыс. Сопоставление абсолютных показателей не позволяет оценить уровень рождаемости, определить, где этот уровень выше. Это можно сделать через ОПИ – коэффициенты рождаемости в Санкт-Петербурге и Ленинградской области. Население Санкт-Петербурга в 2003г. составило – 4631 тыс. чел., в Ленинградской области - 1662 тыс. чел.

$$ОПИ_{спб} = \frac{40300}{4631} \approx 8,71 \text{ чел} / \text{тыс. чел} ;$$

$$ОПИ_{обл} = \frac{13400}{1662} \approx 8,06 \text{ чел / тыс. чел.}$$

Сравнивая полученные значения показателей рождаемости, можно сделать следующий вывод: рождаемость в Санкт-Петербурге выше, чем в Ленинградской области.

В эту же группу включаются **относительные показатели уровня экономического развития, характеризующие эффективность использования ресурсов и эффективность производства**. Это показатели выработки продукции, затрат на единицу продукции, эффективности использования производственных фондов и т.д., поскольку их получают сопоставлением разноименных величин, относящихся к одному и тому же явлению и одинаковому периоду времени.

2.3.6. Относительные показатели сравнения (ОПСр) характеризуют сравнительные размеры одноименных абсолютных показателей, относящихся к различным объектам или территориям, но за одинаковый период времени. Их получают как частные от деления одноименных абсолютных показателей, характеризующих разные объекты, относящихся к одному и тому же периоду или моменту времени.

$$ОПСр = \frac{\text{Показатель, характеризующий объект } A}{\text{Показатель, характеризующий объект } B}$$

С помощью таких показателей сравнения можно сопоставлять производительность труда в разных странах и определять, где и во сколько раз она выше; сравнивать цены на различные товары, экономические показатели разных предприятий и т. д.

Например, можно сравнить среднюю заработную плату в промышленности Санкт-Петербурга в 2003г. и в образовании, принимая заработную плату в промышленности за базу сравнения. Средняя заработная плата в промышленности составила 7871руб., в образовании – 5403руб. $ОПСр = 5403/7871 = 0,686$; следовательно, средняя заработная плата в образовании составляет 68,6% от заработной платы в промышленности.

Относительные показатели имеют важное значение в практической деятельности, но их нельзя рассматривать в отрыве от абсолютных показателей, через которые они рассчитываются, в противном случае можно прийти к неправильным выводам. Таким образом, только совместное использование абсолютных и относительных показателей позволяет провести качественный анализ различных явлений социально-экономической жизни.

2.4. Средние показатели

Средние показатели являются наиболее распространённой формой статистических показателей, используемых в социально-экономических исследованиях. **Средним называется обобщающий показатель статистической совокупности, характеризующий наиболее типичный уровень явления.** Он выражает величину признака, отнесённую к единице совокупности. **Особенности средних** показателей заключаются в том, что они, во-первых, отражают то общее, что присуще всем единицам совокупности; во-вторых, в них взаимопогашаются те отклонения значений признака, которые возникают под воздействием случайных факторов. Это означает, что средний показатель отражает типичный уровень признака, формирующийся под воздействием основных доминирующих неслучайных факторов. Применение средних величин позволяет охарактеризовать определенный признак совокупности одним числом, несмотря на то, что у разных единиц совокупности значения признака отличны друг от друга.

В социально-экономическом анализе используются два класса средних величин:

- степенные средние;
- структурные средние.

К степенным средним относятся несколько видов средних, построенных по одному общему принципу:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^n x_i^k}{n}},$$

где x_i - варианта,

$n=N$ - объем статистической совокупности,

k - показатель степени.

Показатель степени k может принимать любые значения, но на практике обычно используются несколько его значений: *при $k = 1$ получают среднюю арифметическую; $k = -1$ – среднюю гармоническую; $k = 0$ – среднюю геометрическую; $k = 2$ – среднюю квадратическую.*

Степенные средние в зависимости от формы представления исходных данных могут быть **простыми и взвешенными.**

Если исходные данные представлены простым перечислением значений признака у статистических единиц, то используется формула степенной средней простой:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^k x_i^k}{n}}$$

Если данные предварительно сгруппированы (представлены рядом распределения), то используется формула степенной средней взвешенной:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^m x_i^k n_i}{\sum_{i=1}^m n_i}}$$

где n_i – частота повторения индивидуальных значений признака;

m – количество однородных групп.

Средняя арифметическая является наиболее распространенным видом степенных средних, **используется в случаях, когда объём усредняемого признака является аддитивной величиной**, т.е. образуется как сумма его значений по всем единицам статистической совокупности. При этом если индивидуальные значения признака у статистических единиц заменить средней арифметической, то суммарный объём признака по совокупности в целом сохраняется неизменным. Это означает, что *средняя арифметическая есть среднее слагаемое*.

Средняя арифметическая простая используется при работе с несгруппированными данными и рассчитывается по формуле:

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

Например, известна сменная выработка рабочих бригады токарей:

табельный номер рабочего	1	2	3	4	5
количество изготовленных деталей, шт.	21	19	20	18	21.

Требуется определить среднюю выработку бригады.

Для ее нахождения используется формула средней арифметической простой:

Средняя арифметическая простая рассчитывается как:

$$\bar{x} = \frac{21+19+20+18+21}{5} = 19,8 \sim 20 \text{ шт.}$$

Если в исходных данных отдельные значения усредняемого признака повторяются, то расчет средней проводится по сгруппированным данным или вариационным рядам. В подобных случаях для расчета необходимо применять **среднюю арифметическую взвешенную** – среднюю сгруппированных величин.

$$\bar{x} = \frac{\sum_{i=1}^m x_i \cdot n_i}{\sum_{i=1}^m n_i}, \text{ или } \bar{x}_i = \sum_{i=1}^m x_i \cdot q_i;$$

где $q_i = \frac{n_i}{\sum_{i=1}^m n_i}$ - частота, т. е. удельный вес статистических единиц,

обладающих определенным значением признака в общем объеме совокупности.

Средняя арифметическая обладает рядом полезных свойств, к важнейшим из которых относятся:

1. Средняя арифметическая постоянной величины равна этой величине:

$$\bar{A} = A \text{ при } A = \text{const};$$

2. Алгебраическая сумма отклонений вариантов от их средней арифметической равно нулю:

$$\sum (x_i - \bar{x}) \cdot n_i = 0;$$

3. Если все варианты уменьшить (увеличить) на постоянное число А, то средняя арифметическая из них уменьшится (увеличится) на это же число:

$$\frac{\sum (x_i \pm A) \cdot n_i}{\sum n_i} = \bar{x} \pm A;$$

4. Если все варианты одинаково увеличить (уменьшить) в одно и то же число раз, то средняя арифметическая увеличится (уменьшится) во столько же раз:

$$\frac{\sum (x_i \cdot A) \cdot n_i}{\sum n_i} = \bar{x} \cdot A;$$

5. Если все веса средней одинаково увеличить (уменьшить) в несколько раз, то средняя арифметическая не изменится.

$$\frac{\sum x_i \cdot \left(\frac{n_i}{A}\right)}{\sum \frac{n_i}{A}} = \bar{x}.$$

Пример использования средней арифметической: рассчитать среднюю продажную цену товара по данным, приведенным в таблице 2.3:

Таблица 2.3.

Объём продаж и цена товара А. в магазинах города

Магазины	Продажная цена единицы, руб.	Объём продаж, шт.
Космос	20	25000
Ариадна	18	40000
Вега	19	40000
Итого	?	105000

Использовать среднюю арифметическую простую в данном случае нельзя, так как в разных магазинах продано разное количество товара А. Для расчёта средней продажной цены товара А. следует применить среднюю арифметическую взвешенную:

$$\bar{x} = \frac{\sum_{i=1}^m x_i \cdot n_i}{\sum_{i=1}^m n_i} = \frac{20 \cdot 25000 + 18 \cdot 40000 + 19 \cdot 40000}{105.000} = 18,86 \text{ руб.}$$

При применении средней арифметической простой средняя продажная цена товара составляла бы:

$$\bar{x} = \frac{20 + 18 + 19}{3} = 19 \text{ руб.}, \text{ т.е. оказалась бы завышенной.}$$

Средняя гармоническая имеет более сложную конструкцию, чем средняя арифметическая. *Используется в тех случаях, когда статистическая информация не содержит частот по отдельным значениям признака, а представлена произведением значения признака на частоту.* Средняя гармоническая как вид степенной средней выглядит следующим образом:

$$\bar{x} = \sqrt[n]{\frac{\sum_{i=1}^n x_i^{-1}}{N}};$$

В зависимости от формы представления исходных данных средняя гармоническая может быть рассчитана как простая и как взвешенная.

*Если исходные данные несгруппированны, то применяется **средняя гармоническая простая**:*

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

К ней прибегают в случаях определения, например, средних затрат труда, материалов и т. д. на единицу продукции по нескольким предприятиям.

Рассмотрим пример использования средней гармонической простой:

Три предприятия производят микроволновые печи. Себестоимость их производства на 1-ом предприятии составила 4000руб., на 2-ом - 3000руб., на 3-ем - 5000руб. Необходимо определить среднюю себестоимость производства микроволновой печи при условии, что на каждом предприятии общие затраты на ее изготовление составляют 600тыс. руб.

Применять среднюю арифметическую в данном случае нельзя, так как предприятия выпускают разное количество микроволновых печей: первое – 150шт. (600000/4000); второе – 200шт. (600000/3000); третье – 120шт. (600000/5000).

Среднюю себестоимость микроволновой печи можно получить, если общие затраты трех предприятий разделить на общий выпуск:

$$\bar{x} = \frac{600 + 600 + 600}{150 + 200 + 120} = 3,830 \text{руб.}$$

К аналогичному результату можно прийти, используя *формулу средней гармонической простой*:

$$\bar{x} = \frac{1+1+1}{\frac{1}{4000} + \frac{1}{3000} + \frac{1}{5000}} = 3,830 \text{руб.}$$

*При работе со сгруппированными данными используется **средняя гармоническая взвешенная**:*

$$\bar{x} = \frac{\sum_{i=1}^m w_i}{\sum_{i=1}^m \frac{w_i}{x_i}},$$

где w_i – статистический вес; $w_i = x_i \cdot n_i$.

Если в предыдущем примере принять, что на предприятиях было произведено разное количество печей при разных общих затратах, то для определения средней себестоимости следует использовать формулу средней гармонической взвешенной. Пусть на первом предприятии общие затраты на производство микроволновых печей составили 600тыс. руб., на втором – 660тыс. руб., на третьем – 500тыс. руб.; произведено было соответственно 150, 220 и 100 единиц продукции. Средняя себестоимость одной микроволновой печи составила:

$$\bar{x} = \frac{600 + 660 + 500}{\frac{600}{4} + \frac{660}{3} + \frac{500}{5}} = 4,757 \text{ тыс. руб.}$$

Средняя геометрическая применяется в тех случаях, когда общий объем усредняемого признака является мультипликативной величиной, т.е. определяется не суммированием, а умножением индивидуальных значений признака.

$$\bar{x}_{\text{геом}} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i}.$$

Форма средней геометрической взвешенной в практических расчётах не применяется.

В социально-экономических исследованиях средняя геометрическая применяется в анализе рядов динамики при определении среднего коэффициента роста, когда задана последовательность относительных величин динамики.

Рассмотрим пример:

В результате инфляции за первый год цена товара возросла в 2 раза по сравнению к предыдущему году, а за второй ещё в 1,5 раза по сравнению к предыдущему. Необходимо определить средний коэффициент роста цены.

За два года цена возросла в 3 раза (2·1,5). Если использовать среднюю арифметическую, то средний коэффициент роста составит $\frac{2+1,5}{2} = 1,75$ раза; за два года цена при таком среднем коэффициенте роста должна составить

$1,75 \cdot 1,75 = 3,0625$ раза, что выше реального на 0,625 или на 6,25%; В действительности средний коэффициент роста следует определить по формуле средней геометрической:

$$\bar{x}_{\text{геом}} = \sqrt{2 \times 1,5} = 1,73.$$

Средняя геометрическая используется также для определения равноудаленной величины от максимального и минимального значения признака. Например, страховая фирма заключает договоры страхования имущества граждан. В зависимости от вида имущества, его состояния, категории фирмы, конкретного рискованного случая и т. д. страховая сумма может изменяться от 3 тыс руб. до 1 млн. руб. Средняя сумма по страховке составит:

$$\bar{x}_{\text{геом}} = \sqrt{3 \times 1000} = 54,772 \text{ тыс. руб.}$$

Средняя квадратическая используется в тех случаях, когда при замене индивидуальных значений признака на среднюю величину необходимо сохранить неизменной сумму квадратов исходных величин.

Главная сфера её использования – измерение степени колеблемости индивидуальных значений признака относительно средней арифметической (среднее квадратическое отклонение). Кроме этого, средняя квадратическая применяется в тех случаях, когда необходимо вычислить среднюю величину признака, выраженного в квадратных или кубических единицах измерения (при вычислении средней величины квадратных участков, средних диаметров труб, стволов и т. д.).

Средняя квадратическая рассчитывается в двух формах:

как простая	$\bar{x}_{KB} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{N}};$
как взвешенная	$\bar{x}_{KB} = \sqrt{\frac{\sum_{i=1}^m x_i^2 \cdot n_i}{\sum_{i=1}^m n_i}}.$

Все степенные средние различаются между собой значениями показателя степени. При этом, чем выше показатель степени, тем больше количественное значение среднего показателя:

$$\bar{x}_{\text{гарм}} \leq \bar{x}_{\text{геом}} \leq \bar{x}_{\text{арифм}} \leq \bar{x}_{\text{кв}}.$$

Это свойство степенных средних называется **свойством мажорантности средних**.

Таким образом, *выбор вида среднего показателя оказывает существенное влияние на его численную величину*. Выбор вида средней определяется в каждом отдельном случае путем анализа исследуемой совокупности, изучения содержания явления. **Степенная средняя выбрана правильно, если на всех этапах вычислений не меняется её логическая формула**, т.е. реально сохраняется социально-экономическое содержание усредняемого признака.

Особый вид средних показателей – **структурные средние**. Они используются при изучении внутреннего строения рядов распределения значений признака и будут рассматриваться в главе “Обобщающие характеристики статистических совокупностей”.

2.5. Сопоставимость показателей

Главнейшим требованием статистики является требование обеспечения сопоставимости показателей, так как без сопоставимости нет сравнения, а значит, нет объективных выводов об изучаемом социально-экономическом явлении или процессе.

В статистике выработана определённые правила, обеспечивающие сопоставимость показателей:

- **показатели должны обладать общим содержанием:** еще древние говорили, что абсурдно сравнивать “что длиннее - дерево или ночь” или “чего больше – ума или зерна”.

- **статистические показатели должны выражаться в одинаковых единицах измерения:** расстояние – в километрах, вес – в килограммах, и т.д. Если используются стоимостные измерители, то для обеспечения сопоставимости должны применяться сопоставимые цены - цены базисного, либо отчётного периода.

Например: если сравниваются стоимостные объёмы продаж 1995 и 2004 г., то для обеспечения корректности сравнения необходимо физические (натуральные) объёмы продаж выразить либо в ценах 1995г, либо в ценах 2004г. Кроме этого, сопоставимость разных по содержанию главного компонента разновидностей продукта может быть обеспечена применением условно-натуральных измерителей.

- **сравниваемые показатели должны рассчитываться по единой методике.**

Например, нельзя непосредственно сравнивать показатели безработицы в России и в США, так как они рассчитываются по разным методикам.

Удалено: .

Удалено: .

• **сравниваемые статистические показатели должны быть однородными по времени и территории** – они должны определяться за одинаковые периоды времени, на одни и те же даты, по одной территории.

В соответствии с перечисленными правилами для обеспечения сопоставимости статистических показателей на практике используются следующие статистические приёмы:

- для обеспечения общего содержания – **разделение разнородных совокупностей на однородные части, т.е. группировку;**

- для приведения к одинаковым единицам измерения - **использование единой системы мер и весов, условно-натуральных измерителей, сопоставимых цен или индексов при сравнении стоимостных показателей;**

- **пересчёт несопоставимых показателей по единой методике;**

- **приведение показателей к одинаковым периодам и моментам времени;**

- **приведение показателей к единой территории или кругу охватываемых единиц.**

Например, с изменением границ Санкт-Петербурга для обеспечения сопоставимости показателей промышленного производства или розничного товарооборота за 1970 и 2004г., необходимо показатели 1970г пересчитать по новой территории - к данным по г. Ленинграду за 1970г. приплюсовать данные за тот же год по присоединённой территории.

- **Замена несравнимых абсолютных показателей** относительными или средними показателями: показателями структуры, координации и т.д.

Удалено: :

Все статистические показатели исчисляются на основе первичных данных, собранных и определённым образом обработанных в процессе статистического исследования.

3. Статистическое наблюдение

3.1. Информационная база статистического исследования, статистическое наблюдение и его этапы

Для исследования социально-экономических явлений и процессов необходимо, прежде всего, собрать первичные статистические данные (информацию), под которыми понимается совокупность количественных характеристик массовых явлений и процессов, полученных в результате статистического наблюдения. Эти данные являются исходным материалом для получения обобщающих показателей и выводов о тенденциях его развития. Не всякие собранные факты об изучаемом явлении могут считаться статистической информацией. Они должны отвечать определенным требованиям:

- Быть полной - это означает, что а) она должна охватывать либо все единицы статистической совокупности, либо такую их часть, по которой можно делать выводы о совокупности в целом; б) информация должна охватывать все существенные стороны явления, его свойства, внутренние и внешние связи; в) она должна собираться за максимально длительный срок. Это способствует ослаблению воздействия случайных факторов и выявлению закономерностей развития явления;

- быть достоверной, что означает соответствие данных о явлении, собранных в процессе наблюдения фактическому состоянию явления;

- быть сопоставимой – для этого данные должны собираться в установленные сроки, по единой программе, с использованием одинаковых методов и т. д., - иначе невозможно обеспечить их дальнейшее сопоставление;

- своевременно предоставляться – особенно, если она используется для осуществления управленческих функций.

Основными свойствами статистической информации являются ее массовость и стабильность. Первое свойство связано с особенностями предмета статистики, вторая - с неизменностью собранной информации, ее способностью устаревать и с необходимостью получения новой информации для принятия верных управленческих решений. Практический менеджмент нуждается в постоянно пополняемых статистических данных, достоверная, полная, но запоздалая информация оказывается практически ненужной.

Состав статистической информации определяется потребностями развития общества. В условиях рыночной экономики ее потребителями являются как государственные органы, так и различные негосударственные структуры. Так, данные о состоянии экономики, численности и структуре населения, его покупательной способности, об

Формат: Список

уровне инфляции и т. п. нужны как государственным органам, так и частным структурам для планирования и организации своей деятельности.

Удалено: ¶

• Основными источниками статистической информации являются издания органов государственной статистики. Высшим органом государственной статистики является Государственный Комитет Статистики Российской Федерации. Его ежегодное официальное издание – статистический сборник «Российский статистический ежегодник» содержит наиболее полную информацию о Российской Федерации – макроэкономические, демографические показатели, показатели состояния различных отраслей национального хозяйства, данные о развитии здравоохранения, о доходах и потреблении населения, продолжительности жизни и т. д.

Формат: Список

Кроме этого, данные о социально-экономическом положении страны в каком-либо году можно найти в кратком статистическом сборнике Госкомстата РФ «Россия в цифрах».

Местные статистические органы издают региональные статистические сборники. Например, Петербургкомстат издает ежегодник «Народное хозяйство Санкт-Петербурга и Ленинградской области в 20...году», комитет по статистике г. Москва – сборник «Москва в цифрах» и т. д.

Предоставление статистической информации является основной задачей органов госстатистики, а сама статистическая информация – продукцией их деятельности, имеющей, как и любая другая продукция, свою стоимость. Особенно высокую стоимость имеет та информация, получение которой выходит за пределы программы работы органов государственной статистики.

Основным источником получения первичной статистической информации является статистическое наблюдение.

Статистическое наблюдение представляет собой *планомерный, научно организованный сбор данных или сведений о массовых явлениях и процессах, который заключается в регистрации отобранных признаков у каждой единицы совокупности.*

Не всякий сбор сведений может называться статистическим наблюдением. О статистическом наблюдении можно говорить лишь тогда, когда изучаются статистические закономерности, проявляющиеся в массовых процессах, в большом количестве единиц совокупности. Поэтому наблюдение считается статистическим, если оно соответствует следующим условиям:

- является планомерным;
- является массовым;
- является систематическим.

Формат: Список

Планомерность статистического наблюдения предполагает, что оно готовится и проводится по заранее разработанному плану, являющего частью общего плана проведения статистического исследования; в такой план включаются вопросы методологии, организации, техники сбора информации, контроля ее качества, его достоверности и оформления итоговых результатов.

Массовый характер статистического наблюдения означает, что оно охватывает количество случаев проявления изучаемого явления, достаточное для получения достоверных статистических данных, характеризующих совокупность в целом.

Систематичность наблюдения определяется тем, что оно должно проводиться либо непрерывно, либо систематически, либо регулярно, так как только такой подход позволяет изучать тенденции и закономерности социально-экономических явлений и процессов.

Примером статистического наблюдения являются опросы общественного мнения, проводимые с целью изучения мнения граждан по вопросам, представляющим для них интерес.

Процесс проведения статистического наблюдения состоит из нескольких этапов:

- подготовка наблюдения;
- проведение массового сбора данных;
- подготовка данных наблюдения к обработке;
- разработка предложений по совершенствованию проведения статистического наблюдения.

← --- **Формат:** Список

При проведении первого этапа – подготовке наблюдения необходимо в первую очередь решить программно-методологические вопросы, важнейшими из которых являются определение цели и задач наблюдения, его объекта, выбор единиц наблюдения, состава признаков, подлежащих регистрации, формы, вида и способа наблюдения, разработка документов для сбора информации, программы наблюдения. На этом же этапе решаются и организационные вопросы, такие как: подготовка работников, проводящих наблюдение, тиражирование документов для проведения наблюдения и т. д.

Второй этап связан с непосредственным проведением наблюдения и включает в себя такие работы как рассылка бланков, анкет, форм статистической отчетности, переписных листов, их заполнение и дача в органы, проводящие наблюдение.

При выполнении третьего этапа – подготовке данных к наблюдению собранная информация проверяется на полноту, подвергается арифметическому и логическому контролю с целью выявления и исключения допущенных ошибок.

На последнем этапе проведения статистического наблюдения анализируются причины, которые вызвали ошибки в заполнении статистических формуляров, и разрабатываются предложения по совершенствованию проведения статистического наблюдения.

3.2. Программно-методологические и организационные задачи статистического наблюдения

Статистическое наблюдение должно проводиться по заранее разработанному плану, при разработке которого необходимо решить множество важных задач. Эти задачи можно разделить на программно-методологические и организационные.

К программно-методологическим задачам статистического наблюдения относятся:

- определение цели и задач наблюдения;
- выбор объектов и единиц наблюдения;
- разработка программы наблюдения;
- выбор формы, вида и способа проведения наблюдения.

Цель и задачи наблюдения. Каждое статистическое наблюдение проводится с целью получения достоверных данных об исследуемых процессах и явлениях. Она должна быть конкретной и четко сформулированной, исходить из общих задач, поставленных перед статистическим исследованием явления. В соответствии с принципами системного подхода задачи наблюдения должны соподчиняться поставленной цели, исходить из нее. Цель и задачи определяют программу и форму организации наблюдения. Если они поставлены неясно, неконкретно, то будут собраны излишние сведения или, наоборот, получены неполные статистические данные.

В зависимости от цели и решаемых задач определяются объект и единица наблюдения.

Объект статистического наблюдения - это статистическая совокупность, в которой проистекают исследуемые социально-экономические явления и процессы. Установление объекта наблюдения означает определение точных границ и состава совокупности. Например, при переписи населения необходимо установить, какое население следует регистрировать - наличное, под которым понимается фактически находящееся в момент переписи в данной местности, или постоянное, т. е. живущее на данной территории постоянно.

В ряде случаев для отграничения объекта наблюдения используют понятие ценза. *Ценз* есть пороговое значение признака, которое ограничивает

Удалено: ¶ O
Формат: Список
Удалено: B
Удалено: P
Удалено: B

объект наблюдения. Например, при обследовании промышленности объектом могут быть средние и крупные предприятия, к которым в соответствии с существующим законодательством, относятся предприятия с числом работников более 100 человек.

Любой объект наблюдения состоит из единиц наблюдения.

Единица наблюдения – элемент статистической совокупности, являющийся носителем признаков, подлежащих регистрации, то есть это то первичное звено, от которого должны быть получены необходимые статистические сведения. Например, при проведении демографических обследований это может быть человек, но может быть и семья, при бюджетных обследованиях – домохозяйство или семья.

В соответствии с поставленной целью, задачами, выбранным объектом разрабатывается программа наблюдения.

Программа наблюдения – перечень признаков, подлежащих регистрации (при непосредственном наблюдении), либо это перечень вопросов, по которым собираются сведения (при опросах). Составление программы наблюдения является сложной и ответственной задачей, поскольку от этого зависит качество собранной информации. Состав и содержание программы наблюдения определяются задачами исследования и особенностями изучаемого общественно-экономического явления. Всякое явление обладает множеством признаков. Собирать информацию по всем – нецелесообразно и невозможно. Поэтому необходимо отобрать наиболее важные, отвечающие поставленным задачам и соответствующие цели наблюдения.

Удалено: (микрорепись 1994 г 2 единицы – человек и домохозяйство).

К программе наблюдения предъявляются следующие **требования**:

- в программу включаются только существенные признаки, непосредственно характеризующие изучаемое явление, его тип, основные черты и свойства. Не должны включаться второстепенные вопросы, не связанных с решением представленных задач.
- в программу не включаются вопросы, на которые могут быть получены заведомо неточные ответы, - то есть вопросы, способные вызвать подозрения, что ответы на них могут быть использованы во вред опрашиваемым.
- в программу должны включаться вопросы контрольного характера для проверки собираемой информации, например, это логически связанные вопросы о возрасте, семейном положении, образовании, наличии детей и т.д.
- все вопросы программы должны быть ориентированы на определенную форму ответа: либо цифровую, либо альтернативную, либо многовариантную.

Удалено: (

Удалено: .)

При цифровой – ответ дается в количественной форме (о возрасте, стаже, заработке и т.д.); при альтернативной – в форме «да» или «нет» (пол, семейное положение и т.д.); при многовариантной – выбор одного или нескольких вариантов из предлагаемого меню. Например, на вопрос о состоянии в браке возможны следующие варианты ответов: - состоит в браке;

- никогда не состоял; - вдовец; - разведен.

- ответы в программах для облегчения обработки кодируются числовыми кодами.
- при разработке программы необходимо не только определить состав вопросов, но и их последовательность, так как логика расположения вопросов способствует получению более достоверных данных.

Одновременно с программой разрабатывается инструментарий наблюдения в виде статистических формуляров и инструкций по их заполнению.

Статистический формуляр – это документ единого образца, содержащий программу и результаты наблюдения. Он может иметь разные названия: бланк обследования, переписной лист, анкета, отчет и т. д. Обязательными элементами статистического формуляра являются титульная и адресная части. В титульной части указываются: наименование статистического наблюдения и органа, его проводящего; номер формуляра, а также, кто и когда его утвердил. В адресной – адрес отчетной единицы, ее подчиненность.

Кроме формуляра разрабатывается инструкция по порядку проведения наблюдения, по заполнению формуляра. В зависимости от сложности программы наблюдения это может быть документ в виде отдельной брошюры, либо подсказки в ответах, либо разъяснения на обратной стороне бланка.

При подготовке статистического наблюдения помимо программно-методологических вопросов необходимо решить и организационные вопросы. К ним относятся:

- определение органа (исполнителя) наблюдения. Наблюдение может проводиться собственными силами или организациями, специализирующимися на проведении наблюдений;
- определение времени наблюдения: даты начала, даты окончания наблюдения, критической даты. Срок (период) наблюдения устанавливается исходя из объема работы и численности персонала, занятого сбором данных. Критической датой считается конкретный день года, час дня, по состоянию на который проводится регистрация признаков по каждой единицы статистической совокупности.

Удалено: a
Удалено: я
Удалено: ,
Удалено: ая
Удалено: ¶
Удалено: ¶
Удалено: ¶
Удалено: ¶
Удалено: ¶

Удалено:

Формат: Список

- определение места (территории) проведения наблюдения. Выбор места проведения наблюдения определяется его целью. Например, если определяется стоимость потребительской корзины в Санкт-Петербурге, то местом проведения наблюдения будет территория города.

Важнейшим организационно-методологическим вопросом, определяющим эффективность статистического наблюдения (достоверность полученной информации, затраты на проведение наблюдения и т.д.) является выбор формы, вида и способа проведения наблюдения.

Удалено:

3.3. Формы, виды и способы проведения статистического наблюдения

В статистической практике используют три организационные формы статистического наблюдения:

- отчетность (предприятий, организаций, учреждений и т. п.);
- специально организованное статистическое наблюдение;
- регистры.

← Формат: Список

Статистическая отчетность является основной формой статистического наблюдения в Российской Федерации. Статистические сведения в виде установленных законом отчетных документов в определенные сроки представляются всеми предприятиями, организациями и учреждениями страны.

Статистическая отчетность является составной частью государственной статистики, для нее характерно следующее:

- она утверждается органами Государственной статистики;
- имеет обязательный характер;
- имеет юридическую силу (подписывается должностными лицами);
- имеет документальную обоснованность (базируется на документах первичного учета).

Постановлением Государственного Комитета по Статистике РФ от 16 сентября 1997 года с 1 января 1998 года предприятия и организации (юридические лица) обязаны предоставлять следующие унифицированные формы федерального государственного статистического наблюдения:

Форма № П-1 «Сведения о производстве и отгрузке товаров и услуг»;

Форма № П-2 «Сведения об инвестициях»;

Форма № П-3 «Сведения о финансовом состоянии организации»;

Форма № П-4 «Сведения о численности, заработной плате и движении работников»;

Форма № П-5(м) «Основные сведения о деятельности организации» (введена постановлением Росстата от 27.07.2004 № 34).

Состав представляемых форм является различным для разных организаций в зависимости от специфики деятельности и количества работников. Предприятия с численностью работников более 15 человек с января 2005г обязаны ежемесячно представлять формы № П-1 с приложением 3 «Сведения об объемах платных услуг населению по видам», П-3, П-4, а ежеквартально – форму № П-2. Малые предприятия, штат которых не превышает 15 человек отчитываются ежемесячно по форме № П-4 и ежеквартально – по форме № П-5.

Статистическая отчетность подается в территориальные управления статистики. Региональные подразделения отраслевых ведомств, структуры, регулирующие деятельность естественных монополий.

Специально организованные статистические наблюдения представляют собой сбор сведений посредством переписей, единовременных учетов и обследований. Эта форма наблюдения позволяет более углубленно изучать изменения в составе населения, семейном бюджете, изучать рыночный спрос и т.д. Такие наблюдения проводятся с целью получения данных, отсутствующих в отчетности, или для проверки сведений, взятых из нее.

Специально организованное статистическое наблюдение требует специально подготовленных людей, специально разработанной программы, является трудоемким и дорогостоящим.

Например, перепись населения проводится одновременно по всей территории страны, охватывает всех живущих на критический момент времени.

Регистром называется форма непрерывного статистического наблюдения за долговременными процессами по совокупности показателей.

К наиболее известным относятся два вида регистров: регистры населения и регистры предприятий.

Регистры населения представляет собой поименованный и регулярно обновляемый перечень жителей страны. Наблюдение проводится по следующим признакам: пол, дата и место рождения (постоянные данные), брачное состояние (переменный признак). Информация в регистр заносится на каждого родившегося или прибывшего из-за границы. В случае смерти или отъезда на постоянное место жительства в другую страну данные из регистра убираются. Такие регистры ведутся в государствах с небольшой численностью населения и по ограниченному числу признаков, так как это требует больших затрат.

Удалено: ¶

Удалено: ¶

Регистр предприятий ведется по всем видам экономической деятельности. По каждому предприятию в регистре находятся следующие сведения: время регистрации предприятия; его название; адрес; организационно-правовая форма; организационная структура и структура управления; виды экономической деятельности; численность занятых; и т. д.

Удалено: ¶

Удалено:

В настоящее время в Российской Федерации действует единый государственный регистр предприятий и организаций (ЕГРПО), который содержит данные по всем предприятиям, организациям, учреждениям, любой формы собственности и любого вида деятельности, в том числе и общественным организациям.

В информационном фонде регистра содержатся следующие данные:

- сведения об отраслевой принадлежности и территориальной, вид экономической деятельности;
- вид собственности и организационная форма;
- справочные данные (адрес, фамилии руководителей, телефоны, факсы);
- экономические показатели (среднесписочная численность, средства, направленные на потребление, остаточная стоимость ОПФ, уставный фонд и т.д.)

Регистр ведется в территориальном разрезе.

Если предприятие закрывается, то служба ведения регистра получает информацию об этом в течение 10 дней после решения ликвидации комиссией.

По характеру регистрации фактов во времени выделяют несколько **видов наблюдений:**

- текущее;
- периодическое;
- единовременное.

Удалено: .

Формат: Список

Удалено: .

Текущее наблюдение состоит в том, что факты регистрируются непрерывно по мере их поступления (регистрация рождений, смертей, учет отработанного времени табельным путем, учет посещаемости и т.д.).

При периодическом наблюдении данные, отражающие изменение изучаемых явлений собираются в ходе нескольких наблюдений. Такие наблюдения проводятся по схожей программе через определенные промежутки времени, то есть периодически (переписи населения, регистрация цен на определенные виды товаров, учет успеваемости).

При единовременном наблюдении сведения об изучаемом явлении получают один раз в момент наблюдения. Повторное наблюдение может и не проводиться, как, например, инвентаризация незавершенного производственного строительства в 1990г.

По полноте охвата исследуемой совокупности различают два **вида статистических наблюдений**:

- сплошное наблюдение;
- несплошное наблюдение.

Формат: Список

Удалено: .

Задачей **сплошного наблюдения** является получение информации обо всех единицах исследуемой совокупности. Оно применяется при проведении переписи населения, инвентаризаций разного рода и т. д.

При **несплошном наблюдении** обследованию подлежит часть единиц статистической совокупности. Несплошное наблюдение обладает существенными преимуществами по сравнению со сплошным: требует меньших материальных и трудовых затрат, позволяет применять более совершенные способы учета, повышает оперативную значимость статистических данных, поскольку они могут быть получены в более короткие сроки.

Несплошное наблюдение подразделяется на:

- *выборочное*;
- *монографическое*;
- *обследование основного массива (метод основного массива)*.

Выборочное наблюдение основано на принципе случайного отбора части единиц статистической совокупности, подлежащих исследованию. Случайный отбор гарантирует независимость результатов наблюдения.

При правильной организации выборочное наблюдение дает достаточно точные результаты, позволяющие оценить всю исследуемую совокупность. При этом значительно сокращаются затраты на проведение наблюдения. Выборочное наблюдение широко используется в промышленности – для контроля качества продукции, в маркетинге – для изучения спроса населения на различные товары и т. п.

Монографическое наблюдение (описание) заключается в подробном описании отдельных единиц статистической совокупности. Наблюдение проводится по широкой программе, используется для глубокого всестороннего изучения особенностей наблюдаемых объектов. При этом не ставится цель получить характеристику всей совокупности. Такие наблюдения позволяют уловить пропорции и связи, которые ускользают из поля зрения при массовых наблюдениях. Часто монографические наблюдения проводятся для составления программы массового наблюдения.

Обследование основного массива характеризуется тем, что наблюдение проводится за наиболее крупными единицами совокупности – основным массивом. Например, наблюдение за финансовым положением в строительстве, транспорте, промышленности ведется по данным предприятий с численностью работников не менее 500 человек.

По способам получения информации различают следующие **типы наблюдений**:

- непосредственно;
- документальное наблюдение (учет);
- опрос.

← Формат: Список

При непосредственном наблюдении информацию собирают специально подготовленные люди, в задачу которых входит получение сведений путем личного учета единиц совокупности: подсчета, взвешивания, измерения и т. д. значения признака. Это наиболее точный и надежный способ получения данных, но и наиболее трудоемкий и дорогостоящий.

При документальном наблюдении (учете) в качестве источника информации используются различного рода учетные документы. Это способ наблюдения используется предприятиями, организациями, учреждениями при составлении отчетности на основе документов первичного учета и является достаточно достоверным и надежным источником информации.

Опрос – способ наблюдения, при котором сведения получают со слов респондента (опрашиваемого). Опрос используется для получения информации о процессах и явлениях, неподдающихся непосредственному прямому наблюдению. Он может быть организован по-разному. В практике статистики применяются следующие основные способы опроса: - экспедиционный (устный);

← Формат: Список

- саморегистрации;

- корреспондентский;

- анкетный. При устном (экспедиционном) опросе

Удалено: ¶

Удалено: ¶

специально подготовленные регистраторы на основе опроса обследуемого лица заполняют переписные листы (фиксируют факты). При этом регистратор одновременно контролирует достоверность полученных сведений. Устный опрос обеспечивает достаточно точные результаты, но является весьма дорогостоящим способом получения данных.

Способ саморегистрации состоит в том, что формуляры заполняются самими опрашиваемыми лицами (респондентами), а регистраторы раздают опросные бланки, инструктируют респондентов, собирают заполненные формуляры, контролируя при этом правильность заполненных сведений.

Удалено: .

Корреспондентский способ заключается в том, что формуляры заполняются и отсылаются опрашиваемыми лицами в адрес организации, проводящей исследование, без участия регистраторов на основе инструкций по их заполнению. Этот вид опроса требует наименьших затрат, но не обеспечивает высокого качества полученных данных, так как проверить точность сообщаемых сведений не всегда возможно.

Анкетный способ предполагает сбор информации в виде анкет. Анкеты распространяются разными способами, заполняются на

добровольных началах, как правило, анонимно. При этом количество анкет, полученных организацией, проводящей обследование, всегда значительно меньше количества разосланных анкет. Данный вид опроса дает приблизительные, ориентировочные данные и используется обычно при изучении общественного мнения по различным вопросам.

3.4. Ошибки статистического наблюдения и контроль данных

Любое статистическое наблюдение, как бы тщательно оно не готовилось, допускает наличие в собранной информации ошибок, которые необходимо своевременно устранить.

Ошибками наблюдения называются расхождения между данными наблюдения и фактическими значениями признаков исследуемого явления. Ошибки наблюдения разнообразны по происхождению и своему содержанию.

В зависимости от причин возникновения различают следующие виды ошибок:

- *методические ошибки;*
- *ошибки регистрации;*
- *ошибки репрезентативности (представительности).*

Методические ошибки *возникают в результате использования несовершенных методик, неправильных теоретических концепций, лежащих в основе исследования.* Ошибки регистрации возникают при получении данных об отдельных единицах совокупности вследствие неправильного установления фактов в процессе наблюдения или неправильной их записи. Они подразделяются на:

- объективные (непреднамеренные) причиной появления которых является неправильное восприятие наблюдаемых фактов, неисправность измерительных приборов и неправильная регистрация. Такие ошибки являются результатом добросовестного заблуждения регистратора;

- субъективные (преднамеренные) ошибки, возникающие по причине сознательного искажения фактов. К ним относятся всевозможные преднамеренные ошибки и приписки, при которых опрашиваемый преднамеренно сообщает неправильные сведения; регистратор преднамеренно воздействует на респондента с целью получения нужного ответа; регистратор преднамеренно искажает в формулярах результаты наблюдения.

Ошибки репрезентативности (представительности), характерны только для сплошного наблюдения. Они возникают в результате того, что состав отобранной для обследования части единиц совокупности (выборки) не полностью отражает состав и свойства

Удалено: ¶

Формат: Список

Удалено: –

всей изучаемой совокупности, несмотря на то, что регистрация сведений по каждой отобранной единице была проведена точно.

По форме проявления (по влиянию на результат) ошибки делятся на:

- систематически;
- случайные.

← Формат: Список

Систематические ошибки возникают по какой-то определенной причине и вызывают одностороннее искажение значений признака у наблюдаемых единиц (увеличение или уменьшение). Они очень опасны, так как величина показателя, рассчитанная в целом по всей совокупности будет включать накопленную ошибку. Случайные ошибки являются результатом действия различных случайных факторов. Они не имеют какой-либо направленности. В больших совокупностях в результате действия закона больших чисел эти ошибки взаимно погашаются и не оказывают существенного влияния на точность наблюдения.

Оба вида ошибок в любом исследовании выступают совместно и составляют совокупную ошибку наблюдения Δ :

Удалено: ¶

$$\Delta = \sigma + \varepsilon;$$

где σ - систематическая ошибка наблюдения,

ε - случайная ошибка наблюдения.

Для выявления и исправления ошибок, данные наблюдения необходимо тщательно контролировать.

Процедура контроля сводится к следующему:

- Проверка материалов наблюдения на полноту и правильность оформления. Проверяется полнота охвата статистических единиц наблюдения, правильность заполнения каждого формуляра.

- Арифметический (счетный) контроль. Этот вид контроля основан на использовании количественных связей между показателями, которые могут быть проверены арифметическими действиями. Такие связи обычно отражаются в заголовках граф или строк формуляров. Например, графа $x =$ графа $y -$ графа z и т.д. Арифметический контроль используется для проверки итоговых данных, с его помощью устанавливается наличие ошибки.

Удалено: же

Удалено: (

Удалено: равна

Удалено: минус

Удалено:).

- Логический контроль основан на использовании логической взаимосвязи показателей, установлении логического соответствия между ними. Он не выявляет ошибки наблюдения, а лишь ставит под сомнение правильность полученных данных. Логический контроль заключается в проверке ответов на вопросы программы наблюдения путем их логического осмысления или сравнения полученных данных с другими источниками по данному вопросу. Классическим примером логического

контроля является соответствие данных при переписи населения о возрасте, образовании и семейном положении.

Для проверки данных наблюдения обычно составляется **схема контроля**, в которую включаются различные виды контроля.

При обнаружении ошибок нельзя самостоятельно их исправлять. Для этого необходимо получить дополнительную информацию путем повторного наблюдения. Данные наблюдения считаются принятыми, если они прошли контроль, и в них внесены все необходимые исправления.

Удалено: ¶

Удалено:

Проверкой собранных данных заканчивается начальная стадия статистического исследования.

После этого можно переходить ко второй стадии исследования обработке данных наблюдения. Обработка заключается в классификации и систематизации полученного статистического материала, осуществляемых через **сводку и группировку**.

4. Сводка и группировка статистических данных

4.1. Задачи сводки и группировки

Удалено: ¶

Статистические данные, собранные в процессе наблюдения не позволяют получить обобщающие характеристики изучаемой совокупности, выявить закономерности ее развития, так как в процессе наблюдения фиксируются характеристики только отдельных единиц совокупности.

Для получения обобщающих характеристик собранную информацию необходимо систематизировать, превратить ее в упорядоченную систему статистических показателей. Систематизация полученной информации и обобщение наблюдаемых факторов является содержанием второй стадии статистического исследования, **называемой сводкой и группировкой**.

Удалено: ¶

Статистическая сводка представляет собой комплекс последовательных операций по обобщению конкретных единичных фактов, образующих совокупность, для выявления типичных черт и закономерностей, присущих изучаемому явлению.

Таким образом, **целью сводки** является получение итоговых данных путем подсчета единичных сведений.

По глубине проработки материала различают *простые* и *сложные* сводки.

Простой сводкой называется операция по подсчету общих итогов по совокупности единиц наблюдения, то есть определение размера исследуемого явления.

Сложной сводкой называется комплекс операций, включающих группировку единиц наблюдения, подсчет итогов по каждой группе и совокупности в целом, а также представление результатов группировки в табличной форме.

По форме обработки материала сводки делятся на *централизованные* и *децентрализованные*.

При **централизованной сводке** весь первичный материал поступает в одну организацию, где и подвергается обработке по принятой программе, по единой методике (например, в Государственном комитете по статистике РФ или территориальных управлениях статистики).

При **децентрализованной сводке** разработка статистического материала осуществляется по иерархической системе управления, подвергаясь соответствующей обработке на каждом уровне. Например, предприятия сдают отчеты в районные отделы статистики, которые делают сводку по своему району, и отправляют обобщенную информацию в региональные управления или комитеты, которые свои сводки отправляют в Государственный комитет по статистике РФ, где и определяются показатели в целом по народному хозяйству страны.

Статистическая сводка осуществляется по специальной программе, которая должна составляться одновременно с разработкой плана и программы проведения наблюдения.

Программа статистической сводки включает в себя:

- выбор группировочных признаков;
- определение порядка формирования групп;
- разработка системы статистических показателей для характеристики выделенных групп и совокупности в целом;
- разработка макетов таблиц для представления результатов сводки.

План статистической сводки содержит указания о сроках и последовательности выполнения отдельных этапов сводки, ее исполнителях, о порядке представления ее результатов.

Научной основой сводки является статистическая группировка, которая представляет собой процесс образования однородных групп на основе расчленения (разделения) статистической совокупности на части или объединение изучаемых статистических единиц в частные совокупности по существенным для них признакам.

Группировка является методом исследования содержания изучаемого явления. На ее основе рассчитываются обобщающие показатели по группам, выявляется строение совокупности, взаимосвязи между изучаемыми признаками, а затем проводится анализ полученных результатов.

Основными категориями метода группировок являются группировочный признак (основание группировки) и интервал.

Группировочным признаком (основанием группировки) называется признак, по которому происходит выделение однородных групп. В качестве группировочного обычно выбирается один из существенных легко распознаваемых признаков, носящих как атрибутивный, так и количественный характер.

Интервал – это совокупность варьирующих значений признака в группе, он определяет количественные границы групп, а его ширина представляет собой промежуток между максимальным и минимальным значениями признака в группе.

При выполнении группировок используются следующие типы интервалов:

- равные – во всех выделенных группах ширина интервала является одинаковой;
- неравные – в каждой группе ширина интервала различна; при этом ее изменение может изменяться закономерно (например, равномерно возрастать), или произвольно, то есть быть свободной;
- закрытые если известны верхняя и нижняя границы интервалов (максимальное и минимальное значения признака в группах);

Формат: Список

- открытые - если известна только одна граница интервала, верхняя или нижняя.

При выполнении группировок необходимо исходить из следующего принципа: различия между единицами, отнесенными к одной группе должны быть меньше, чем между единицами, отнесенными к разным группам.

Удалено: ¶

С помощью группировок в статистике решают следующие задачи:

- изучение состава статистических совокупностей;
- выделение отдельных типов явлений внутри совокупности;
- выявление причинно-следственных связей разных признаков внутри совокупности;
- классификация единиц совокупности по множеству признаков.

Удалено:

Для решения указанных задач применяют разные **типы статистических группировок**.

4.2. Типы группировок

В зависимости от степени сложности изучаемого явления и от поставленных задач статистические группировки могут выполняться по одному или нескольким группировочным признакам. **Группировка называется простой (одномерной), если однородные группы формируются по одному признаку одновременно.**

Удалено: ¶

Если однородные группы образуются по двум и более признакам, то группировка называется сложной.

В классе одномерных группировок выделяют следующие типы:

- *структурные* – предназначены для выявления состава изучаемого явления;
- *типологические* – предназначены для выделения в статистической совокупности различных социально-экономических типов явлений;
- *аналитические (факторные)* – используются для изучения связей и зависимости между варьирующими признаками.

4.2.1. Структурные группировки

Структурные группировки используются для изучения внутреннего строения статистической совокупности и характеристики структурных сдвигов.

Удалено:

Они дают информацию о текущем состоянии массовых явлений и применяются в целях оперативного управления.

Структурная группировка выполняется в несколько этапов:

- выбор группировочного признака;
- определение необходимого числа групп;
- определение параметров групп;

← Формат: Список

- распределение единиц наблюдения по выделенным группам;
- расчет структурных характеристик;
- формулировка выводов.

Выбор группировочного признака осуществляется в соответствии с целями статистического исследования. В качестве группировочного обычно выступает существенный признак. Обязательным условием выполнения любой группировки, в том числе и структурной является упорядочение статистической совокупности по значениям группировочного признака.

Определение необходимого числа групп Число групп должно быть достаточным для объективного представления изучаемой совокупности. При большом числе групп различия между ними становятся малозаметными, а в самих группах в виду их малой наполняемости перестает действовать закон больших чисел и возможно проявления случайности. При малом же их числе в одну группу могут попасть статистические единицы с существенно различающимися значениями признака.

На количество выделяемых групп влияют следующие факторы:

- уровень колеблемости группировочного признака - чем значительнее вариация признака, тем большее количество групп необходимо выделять при прочих равных условиях;
- размер изучаемой статистической совокупности - чем больше размер исследуемой совокупности, тем большее количество групп необходимо выделять.

Удалено: ¶

Выделенные группы должны быть достаточно заполненными. Наличие пустых групп или малое число статистических единиц в них свидетельствуют о неправильном определении их числа.

Ориентировочно число групп можно определить используя эмпирическую зависимость, называемую формулой Стерджесса:

$$m \approx 1 + 3,322 \times \lg N ,$$

где m – количество групп;

N - численность единиц статистической совокупности.

Удалено: ¶

В практических расчетах можно использовать следующие соотношения, полученные на основании формулы Стерджесса:

N	15-24	25-44	45-89	90-179	180-359	360 и более
m	5	6	7	8	9	10

Зависимость Стерджесса дает хорошие результаты, если совокупность состоит из большого числа единиц, распределение близко к нормальному, и при этом используются равные интервалы.

Существует еще один способ определения количества выделяемых групп, он связан с применением среднеквадратичного отклонения равными и неравными σ : если ширина интервала равна $0,5\sigma$, то выделяется 12 групп, если $2/3\sigma$, - то 9 групп, если σ – то 6 групп.

Определение параметров группы

В каждой выделенной группе рассчитываются следующие параметры:

- верхняя граница интервала x_i^H , нижняя граница интервала x_i^G ;
- ширина интервала a_i ;
- середина интервала b_i .

Нижней границей интервала x_i^H называется наименьшее значение признака в группе. Верхней границей интервала x_i^G называется наибольшее значение признака в группе.

Интервалы группировки бывают равными и неравными (прогрессивно возрастающими, прогрессивно убывающими, произвольными, специализированными).

Если вариация признака проявляется в сравнительно узких границах, и распределение статистических единиц носит достаточно равномерный характер, то строят группировку с равными интервалами.

Для равноинтервальной группировки ширина интервала a_i определяется по формуле:

$$a_i = \frac{x_{\max} - x_{\min}}{m} = \frac{R}{m},$$

где R – размах вариации, $R = x_{\max} - x_{\min}$.

При определении размаха вариации R из наблюдения исключаются аномальные значения признака. Полученное значение ширины интервала a_i округляется в большую сторону.

На основе рассчитанной ширины интервала a_i последовательно определяется границы интервалов x_i^H и x_i^G . Определение границ начинается с первой группы. Ее нижняя граница принимается равной минимальному значению признака в совокупности, т. е. $x_1^H = x_{\min}$, верхняя граница определяется как $x_1^G = x_1^H + a_i$. Для второй группы нижняя граница принимается равной верхней границе первой группы, т. е.

$x_2^H = x_1^G$, верхняя определяется как $x_2^G = x_2^H + a_i$ и т. д.

В целом границы интервалов определяются формулами:

$$x_i^H = x_{i-1}^G ; \quad x_i^G = x_i^H + a_i .$$

Удалено :

Удалено :

Середина интервала (центральная варианта) b_i определяется как полусумма верхней и нижней границ, т.е. по формуле: $b_i = \frac{x_i^g + x_i^h}{2}$.

Параметр середина интервала используется при расчете обобщающих характеристик изучаемой совокупности.

Достаточно часто при выполнении группировки используются открытые интервалы. В открытых интервалах указывается только одна граница: верхняя - у последнего интервала, нижняя - у первого. Для закрытия таких интервалов необходимо предварительно определить их ширину. Проблема ширины открытых интервалов решается следующим образом: при равноинтервальной группировке она есть величина постоянная; при неравноинтервальной - предварительно определяется закономерность изменения ширины интервала для некрайних групп, выявленная закономерность позволяет определить ширину соответствующего интервала и рассчитать недостающую границу.

Удалено: .

Распределение единиц совокупности по группам

Основной задачей данного этапа является подсчет числа единиц, попавших в каждую из выделенных групп n_i . При распределении единиц наблюдения по выделенным группам, особенно если группировочный признак является непрерывным, имеет место неопределенность: к какой группе относить единицы со значениями признака, совпадающими с границами интервалов? Для устранения неопределенности используют принцип единообразия – такие единицы включаются в группу, в которой нижняя граница совпадает со значением признака. Например, имеются группы предприятий по объему производства, млн. руб.:

400 – 450; 450 – 500; 500 – 550; 550 – 600; 600 – 650.

К какой группе следует отнести предприятия с объемом производства 500млн. руб.? В соответствии с принципом единообразия - ко второй группе.

Расчет структурных характеристик

Расчет заключается в определении для каждой группы **удельного веса (доли) ее единиц в общем объеме статистической совокупности.** Как и любая относительная величина этот показатель может быть определен в виде коэффициентов: $d_i = \frac{n_i}{N}$;

Удалено: ¶

или в виде процентов:

$$d_i = \frac{n_i}{N} \times 100 \% .$$

Расчитав такие доли для всех групп, мы получаем структуру изучаемой статистической совокупности, равную полному набору долей:

Удалено: ¶

$$\sum_i^m d_i = 1 \text{ или } \sum_i^m d_i = 100\% .$$

На основе анализа показателей структуры делаются соответствующие выводы.

Формулировка выводов о составе совокупности. Для структурных группировок в выводах отражаются два положения:

- Какие значения признака встречаются в совокупности наиболее часто, какие наиболее редко.
- Каков характер изменения структуры в зависимости от изменения значения признака. С увеличением x доля может увеличиваться, либо уменьшаться. Это довольно типично для экономических показателей.

Выводы должны быть сделаны обязательно, иначе пропадает смысл группировки.

Данные структурных группировок обычно представляются в форме соответствующей таблицы.

Пример выполнения структурной группировки: требуется произвести группировку с равными интервалами по данным о стаже работы рабочих участка. На участке работает 50 рабочих, со стажем работы от 0 до 26 лет.

По формуле Стерджесса рассчитаем оптимальное количество групп:

$$m = 1 + 3,332 \lg 50 \approx 7.$$

Далее рассчитывается постоянная ширина интервалов:

$$a = \frac{26-0}{7} \approx 4 \text{ года.}$$

Структурная группировка, выполненная по рассчитанным параметрам приведена в таблице 4.1:

Таблица 4.1.

Группировка рабочих участка по стажу работы

№ группы	Параметры групп, лет				Число рабочих в группе	Удельный вес рабочих в группах, %
	Границы интервалов		Ширина интервала	Середина интервала		
	нижняя	верхняя				
1	0	4	4	2	6	12
2	4	8	4	6	8	16
3	8	12	4	10	11	22
4	12	16	4	14	13	26
5	16	20	4	18	6	12
6	20	24	4	22	4	8
7	24	28	4	26	2	4
Итого	0	28	28	14	50	100

По проведенной группировке можно сделать следующие выводы:

1. Больше всего на участке имеется рабочих со стажем работы от 12 до 16 лет, они составляют 26% от общей численности рабочих участка; меньше всего - со стажем от 24 до 28 лет, их доля составляет 4%.

Удалено: ¶

Удалено: .

2. С увеличением стажа работы число рабочих вначале растет, достигает максимума в 4-ой группе, а затем снижается.

Удалено:

4.2.2. Типологическая группировка

Ее цель состоит в изучении распространенности различных типов экономических явлений в статистической совокупности. Типологические группировки применяются, как правило, к неоднородной совокупности и осуществляются посредством сложных неравноинтервальных группировок. Результатом типологических группировок является разделение совокупности на классы, социально-экономические типы, однородные группы единиц.

По своей сути типологическая группировка представляет собой группировку-классификатор. Такие группировки часто основываются на устойчивом перечне групп, не меняющихся или меняющихся незначительно во времени.

Примером такой группировки является группировка предприятий по форме собственности (государственная, муниципальная, частная, смешанная) или группировка секторов экономики.

При выполнении типологических группировок важно правильно выбрать основание группировки. Для этого необходимо предварительно выявить возможные типы явления на основе анализа сущности и закономерностей его развития. Число групп и их параметры устанавливаются неформально на основе выделенных качественных закономерностей, часто с привлечением количественных признаков.

Удалено: ¶

Например, при выполнении группировки населения по возрастному составу выделяются 4 возрастные категории: дошкольный возраст - до 7 лет; школьный - с 7 до 17 лет; рабочий - с 17 до 55(60) лет; пенсионный - с 55 (60) лет.

Удалено: ¶

По технике выполнения типологическая группировка похожа на структурную группировку, за исключением первых этапов – группировочный признак, количество групп, их параметры определяются на основе качественного анализа. В таких группировках очень часто применяются специализированные интервалы. Типологические группировки представляются в табличной форме, объектом анализа в них являются показатели структуры. Пример типологической группировки приведен в таблице 4.2:

Удалено: .

Таблица 4.2

Группировка населения поселка городского типа по возрастным категориям.

Возрастные категории населения	Границы интервалов, лет		Численность в группе, чел.	Показатель структуры в %
	x_i^h	x_i^e		
Дошкольный возраст	до	7	192	14,3
Школьный	7	17	218	16,3
рабочий	17	55(60)	574	42,8
Пенсионный	55(60)	и более	357	26,6
Итого	-	-	1340	100,0

Удалено:

4.2.3. Аналитические группировки

Аналитические группировки предназначены для выявления связи между изучаемыми признаками. Они позволяют выявить наличие и направление связи, а также измерить ее тесноту и силу.

Все исследуемые признаки в этом случае делятся на две группы: факторные и результативные. *Взаимосвязь между ними проявляется в том, что с изменением среднего значения факторного признака систематически изменяется среднее значение результативного признака.*

Аналитические группировки отличаются от структурных и типологических по технике выполнения, которая заключается в следующем:

Удалено:

1. Производится группировка единиц совокупности по факторному признаку, она выполняется как структурная.

2. В каждой выделенной группе отбираются соответствующие значения результативного признака, и на их основе рассчитывается некоторый обобщающий показатель, обычно, среднее значение.

3. Анализируются изменения обобщающего показателя - среднего значения результативного признака по группам, и делается вывод о наличии или отсутствии взаимосвязи и ее направлении. Если при изменении значений факторного признака, положенного в основу группировки изменяется величина результативного, то признается наличие связи между признаками; при этом, если с увеличением значений факторного признака увеличивается значение результативного, то связь относится к прямой; в противном случае – к обратной.

Пример выполнения аналитической группировки: необходимо установить зависимость между стажем работы и величиной заработной платы рабочих участка. По каждому рабочему известны стаж работы и месячная заработная плата.

В рассматриваемом примере факторным признаком является стаж работы, результативным – величина заработной платы. По факторному признаку ранее была проведена структурная группировка. Рабочие были разбиты на 7 однородных групп. Результат структурной группировки представлен в таблице 3.1. Дополнительно в каждой группе была рассчитана суммарная заработная плата рабочих за месяц и средняя ее величина по формуле средней арифметической простой:

$$y_i = \frac{\sum_{i=1}^{n_i} y_i}{n_i}$$

где y_i - заработная плата i-ого рабочего.

Результаты расчета приведены в таблице 4.3;

Удалено: .

Удалено: e

Удалено: .

Таблица 4.3.

Исследование зависимости заработной платы рабочих участка от стажа работы.

Удалено: ¶

№ группы	Группы рабочих по стажу, лет			Число рабочих в группе, n_i	Общая зарплата рабочих в группе, руб.	Средняя зарплата в группе, руб.
	Границы интервала		Середина интервала, b_i			
	x_i^H	x_i^B				
1	0	4	2	6	18000	3000
2	4	8	6	8	28000	3500
3	8	12	10	11	41800	3800
4	12	16	14	13	59800	4600
5	16	20	18	6	34800	5800
6	20	24	22	4	27200	6800
7	24	28	26	2	14800	7400
Итого	0	28	14	50	224400	4488

Объектом анализа в этой таблице является среднее значение результативного признака – среднемесячная заработная плата рабочих в группах и середина интервала – средний стаж работы. Если среднее

значение резульативного признака, установленное по группам имеет некоторое различие, то связь между признаками можно считать установленной. Если средний результат при переходе от одной группе к другой практически не меняется, то связь между признаками отсутствует.

В рассматриваемом примере изменение стажа работы приводит к изменению заработной платы.

Таким образом, с помощью аналитической группировки можно установить наличие связи между признаками, но описать ее нельзя. Для этого необходимо использовать аппарат корреляционно-регрессионного анализа.

4.3. Сложные группировки

К сложным группировкам относятся группировки, выполняемые по двум и более основаниям. Сложные группировки делятся на комбинационные и многомерные.

Комбинационные группировки выполняются по нескольким признакам последовательно. Последовательность устанавливается исходя из логики взаимосвязи показателей. Как правило, группировку начинают с атрибутивного признака. При комбинационной группировке совокупность логически последовательно разбивается на однородные части по отдельным признакам: на группы - по одному признаку, затем внутри каждой группы по второму признаку - на подгруппы и т.д.

Такие группировки предназначены для более глубокого анализа изучаемого явления, позволяют выявить и сравнить различия и связи между исследуемыми признаками, которые невозможно установить на основе изолированных группировок по каждому из исследуемых признаков. Однако следует иметь в виду, что при изучении влияния большого числа признаков применение комбинационных группировок невозможно, так как это приводит к дроблению информации, а значит, к затуханию проявлений закономерности. Даже при наличии больших объемов информации приходится ограничиваться двумя – четырьмя признаками.

Комбинационная группировка по двум признакам (X , Y) оформляется в виде шахматной таблицы, в которой значения одного признака X откладываются по строкам, а значения второго признака Y – по столбцам. На пересечении j -ого столбца и i -ой строки (в теле таблицы) находятся частоты совместного проявления значения признака Y в j -ом столбце и значения признака X в i -ой строке.

4.3.1. Многомерные группировки

К ним относятся группировки, выполненные по нескольким группировочным признакам одновременно.

Цель многомерных группировок – классификация данных на основе множества признаков, то есть выделение групп статистических единиц,

однородных по нескольким признакам одновременно. В процессе такой группировки решаются, например, задачи типизации – выделяются самостоятельные экономические или социальные типы явлений. Так, приемами многомерной классификации можно всю совокупность промышленных предприятий разбить на «мелкие», «средние» и «крупные», используя следующие признаки: численность промышленно-производственного персонала, объем продукции, стоимость ОПФ, потребление материальных ресурсов и т.д.

Можно выделить типы предприятий по финансовому положению на основе таких показателей как размер прибыли, уровень рентабельности производства, уровень капитализации, уровень ликвидности ценных бумаг и т.д.

В психологии многомерные группировки используются для выделения типов людей по степени их профессиональной пригодности, в медицине – для диагностики болезней на основе множества симптомов.

При выполнении многомерных группировок могут быть использованы два основных подхода:

- Первый заключается в том, что рассчитывается обобщающий показатель по совокупности группировочных признаков и проводится простая группировка по этому обобщающему показателю.

- Второй подход состоит в использовании методом кластерного анализа.

Представителем первого подхода является **метод многомерной средней**, алгоритм которого заключается в следующем:

1. **Составляется матрица абсолютных значений признаков** по всем статистическим единицам - x_{ij} , $i=\overline{1,n}$ – статистические единицы, $j=\overline{1,k}$ – признаки.

2. **Абсолютные значения признаков заменяются их нормированными по среднему значению уровнями:**

$$P_{ij} = \frac{x_{ij}}{\bar{x}_j},$$

где P_{ij} - нормированное значение j -ого признака у i -ой статистической единицы;

\bar{x}_j - среднее значение j -того признака, $\bar{x}_j = \frac{\sum_{i=1}^n P_{ij}}{n}$;

3. Для каждой статистической единицы **рассчитывается многомерная средняя:**

$$\bar{P}_i = \frac{\sum_{j=1}^k P_{ij}}{k},$$

k – число оснований группировки;

4. В соответствии со значениями многомерной средней совокупность разделяется на однородные группы, то есть **выполняется простая группировка по многомерной средней**.

Рассмотрим пример выполнения группировки на основе многомерной средней: необходимо выделить однородные группы статистических единиц по трем признакам; объем статистической совокупности составляет 10 объектов, каждый из которых характеризуется условными значениями признаков.

Исходные данные и расчет многомерной средней представлен в таблице 4.4.

Таблица 4.4.

Расчет многомерной средней

Номер объекта	Абсолютные значения признаков			Нормированные значения признаков			Расчет многомерной средней	
	x_{i1}	x_{i2}	x_{i3}	P_{i1}	P_{i2}	P_{i3}	$\sum_{j=1}^k P_{ij}$	\bar{P}_i
1	2	18	62	0,57	1,67	1,0	3,24	1,08
2	1	5	40	0,29	0,46	0,64	1,39	0,46
3	2	7	40	0,57	0,65	0,64	1,86	0,62
4	6	15	77	1,71	1,39	1,24	4,34	1,4
5	1	9	43	0,29	0,83	0,69	1,81	0,60
6	6	20	95	1,71	1,85	1,53	5,1	1,7
7	5	9	62	1,42	0,83	1,0	3,25	1,08
8	1	1	46	0,29	0,10	0,74	1,13	0,38
9	8	15	84	2,29	1,39	1,35	5,03	1,68
10	3	9	72	0,86	0,83	1,16	2,85	0,95
Итого	35	108	621	10,00	10,00	10,00	10,00	10,00
\bar{X}_j	3,5	10,8	62,1	1,0	1,0	1,0	1,0	1,0

Таким образом, для каждого из 10 объектов рассчитана условная характеристика – многомерная средняя, заменяющая три первоначальных признака. По многомерной средней как группировочному признаку необходимо выполнить одномерную структурную группировку. Для этого определяется:

- Количество однородных групп $m = 1 + 3.322 \times \lg 10 \approx 4$;
- Ширина интервалов $a = \frac{x_{\max} - x_{\min}}{m} = \frac{1,7 - 0,3}{4} = 0,33$.

Структурная группировка представлена в таблице 4.5.

Таблица 4.5.

Группировка объектов по многомерной средней

№ групп	Параметры групп			Номера объектов	Количество объектов в группе, n_i
	x_i^H	x_i^G	a_i		
1	0,38	0,71	0,33	2,3,5,8	4
2	0,71	1,04	0,33	10	1
3	1,04	1,37	0,33	1,7	2
4	1,37	1,07	0,33	4,6,9	3
Итого	0,38	1,07	1,32	-	10

Таким образом, 10 объектов были распределены по 4-м однородным группам.

При втором подходе к выполнению многомерных группировок каждая единица совокупности, обладающая набором из k признаков рассматривается как точка в k -мерном пространстве – пространстве признаков, а каждому признаку придается смысл координаты. Задача классификации в этом случае сводится к выделению сгущений объектов в этом пространстве. Для этого используются различные алгоритмы, но всегда однородные группы выделяются на основании близости объектов по совокупности признаков. Мерой близости объектов, то есть мерой сходства единиц совокупности, могут служить различные критерии.

Выделяют три типа мер сходства:

- Коэффициенты подобия;
- коэффициенты связи;
- показатели расстояния.

Коэффициенты подобия используются для измерения степени близости между парой объектов, каждый из признаков которых принимает значения 0 или 1.

Наиболее простой коэффициент подобия рассчитывается по формуле:

$$S_{ij} = \frac{P_{ij}}{m};$$

где P_{ij} - число совпадений признаков у объектов i и j ;

m – общее число признаков, по которым осуществляется сравнение.

$$0 \leq S_{ij} \leq 1.$$

Коэффициенты корреляции используются как измерители силы связи между статистическими единицами или между признаками. Для измерения тесноты связи количественных признаков применяют коэффициенты линейной корреляции.

В кластерном анализе мерой сходства является *мера расстояния между двумя объектами I и j* . Для количественных признаков используется *Евклидово расстояние*:

$$R_{ij} = \sqrt{\sum_{j=1}^m (P_{i1} - P_{j1})^2},$$

где P_{i1} , P_{j1} - стандартизованные значения 1-ого признака i -ого и j -ого объекта наблюдения.

4.4. Статистические таблицы

Результаты сводок и группировок заносятся в статистические таблицы, являющиеся средством наглядного выражения результатов исследования. **Статистическая таблица** представляет собой рационального и наглядного изложения статистических данных об исследуемых явлениях. Основу статистической таблицы составляет графленая сетка, вертикальные столбцы которой называются графами, а горизонтальные – строками. Если строки и графы имеют название, то это будет *макет таблицы* (рис.3.1).

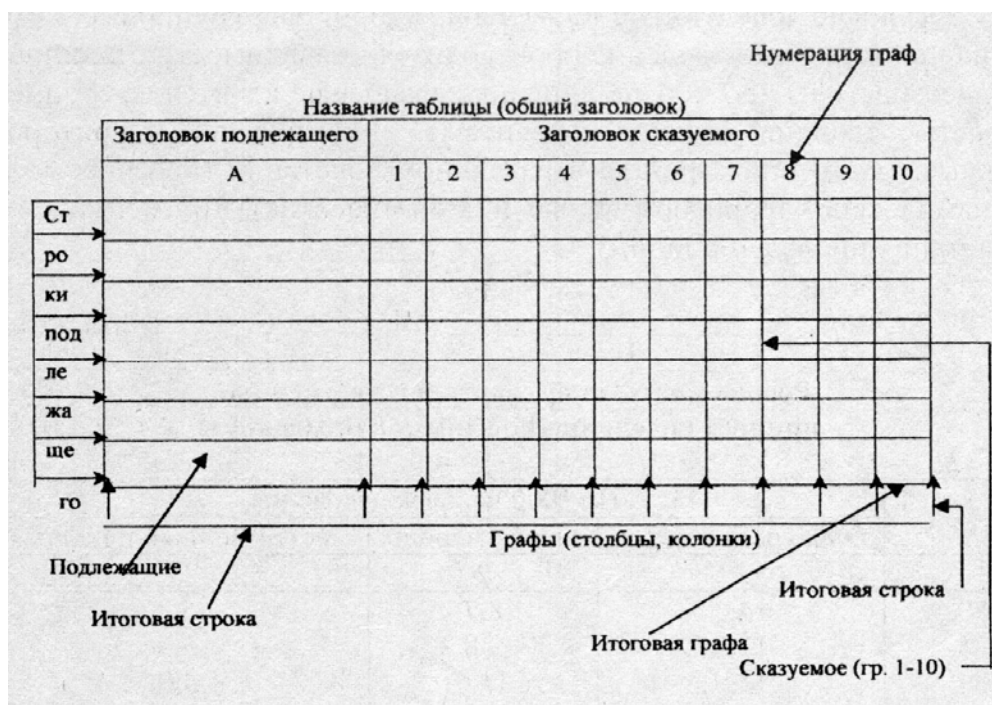


Рис. 4.1. Макет статистической таблицы

В таблице различают: заголовок, подлежащее и сказуемое. *Заголовок таблицы* отражает содержание таблицы, место и время, к которому относятся ее данные, единицы измерения, если они являются общими для приведенных данных. *Подлежащим* таблицы является перечень единиц совокупности или группы, т. е. объект изучения, *сказуемым* – цифровые данные, характеризующие подлежащее, т. е. результаты сводки. Обычно подлежащее располагается слева в виде названий строк, а сказуемое – сверху в виде названий граф.

По содержанию подлежащего все статистические таблицы можно разделить на следующие группы:

- **Простые** таблицы, в подлежащем которых отсутствуют группировки. Они содержат обобщающие показатели, относящиеся к перечню единиц совокупности (перечневые таблицы), к перечню хронологических дат (хронологические таблицы) или к перечню территорий (территориальные таблицы).

- **Групповые** таблицы, в подлежащем которых изучаемый объект разделен на группы по определенному признаку. При этом каждая группа может быть охарактеризована рядом показателей.

- **Комбинационные** таблицы, в подлежащем которых дана группировка единиц совокупности по двум и более признакам, взятым в комбинации.

Сказуемого таблицы может быть *простым или сложным*.

Простое сказуемое предусматривает параллельное расположение показателей, (графы 1,2 в макете таблицы) а **сложное** – комбинированное (графы 3, 4, 5).

При построении статистических таблиц необходимо соблюдать определенные **правила по их оформлению**:

1. Таблица должна быть компактной, легко обозримой. Ее не следует загружать излишними подробностями, затрудняющими анализ.

2. Заголовок таблицы должен ясно и кратко выражать ее содержание. Заголовки строк подлежащего и граф сказуемого также должны быть сформулированы точно и кратко.

3. В таблице желательно давать нумерацию граф. Это облегчает пользование таблицей, показывает способ расчета чисел в графах. Графы, содержащие подлежащее, обозначаются заглавными буквами алфавита; графы, содержащие сказуемое, нумеруются арабскими числами. Не допускается в заголовках подлежащего и сказуемого сокращение слов.

4. Если единицы измерения различны, то они указываются в названиях строк и граф.

5. Приводимые в подлежащем и сказуемом признаки должны располагаться в логическом порядке с учетом необходимости их совместного рассмотрения. Информация размещается от частного к общему, т. е. сначала показывают слагаемые, а в конце подводят итоги.

6. Если в таблице приводятся не все данные, а только наиболее значимые из них то сначала показывают итог, а затем выделяют наиболее важные части с помощью оборотов «в том числе», «из них».

7. Следует различать «Итого» и «Всего». «Итого» является итогом для определенной части совокупности, а «Всего» - итог для всей совокупности.

8. При оформлении таблицы применяются следующие обозначения:

- прочерк (-) – когда явление отсутствует;
- символ «x» - если явление не имеет осмысленного содержания;
- многоточие (...) – если отсутствуют сведения (или делается запись «нет сведений»).
- если сведения имеются, но числовое их значение меньше принятой в таблице точности, оно выражается дробным числом 0,0.

9. Округление чисел, приводимых в таблице, должно проводиться с одинаковой степенью точности.

10. Если одна величина превосходит другую многократно, то полученные относительные показатели лучше выражать не в процентах, а в количестве раз.

Соблюдение приведенных правил построения и оформления статистических таблиц делает их основным средством представления, обработки и обобщения статистической информации.

5. Обобщающие характеристики статических совокупностей

5.1. Ряды распределения

Результаты статистических сводок и группировок могут быть представлены в виде **статистических рядов** – упорядоченных совокупностей значений показателей (статистического признака). По своему содержанию *статистические ряды подразделяются на ряды динамики и ряды распределения.*

Рядом динамики называют систематизированную совокупность числовых данных, характеризующих изменения изучаемых явлений во времени.

Ряд распределения, *представляет собой систематизированную последовательность статистических единиц, сгруппированных по конкретному признаку.* Он характеризует состав изучаемого явления, позволяет судить об однородности совокупности, закономерности распределения статистических единиц. Обычно ряд распределения представляет собой результат структурной группировки.

Ряд распределения считается построенным, если известно, каким образом меняются в совокупности значения признака и как часто встречаются отдельные значения признака.

Для различных статистических признаков строятся ряды распределения разного типа:

- **атрибутивные** – строятся по описательным признакам в порядке возрастания или убывания наблюдаемых значений признака; примером атрибутивных рядов могут служить распределения населения по национальности, по профессиям, по полу; распределение предприятий по формам собственности;

- **вариационные** - строятся по количественным признакам, например, распределение рабочих по уровню квалификации, по заработной плате, распределение студентов по успеваемости.

Вариационные ряды делятся на дискретные и интервальные.

В дискретных рядах признак принимает только целые значения, например, размер семьи, тарифный разряд.

Интервальные ряды основаны на непрерывных признаках, принимающих любые, в том числе и дробные значения. В зависимости от того, какая структурная группировка лежит в основе интервального ряда, различают *равноинтервальные и неравноинтервальные* ряды.

В *равноинтервальных рядах* ширина интервала является величиной постоянной, в *неравноинтервальных* – она различна для разных групп.

Основными элементами рядов распределения являются:

- 1) значения признака (варианты):
 - χ_i - дискретное в дискретных рядах;

← - - - - - **Формат:** Список

- $\chi_i^h - \chi_i^e$ - интервал для интервальных рядов, где $i = \overline{1, n_i}$, n_i - частота;
- 2) частота n_i - число единиц совокупности, обладающих данным значением признака. Частота показывает, сколько раз данное значение признака встречается в совокупности; сумма всех частот всегда равна объему статистической совокупности, т. е. $\sum_1^m n_i = N$.

Исследование рядов распределения осуществляется в два этапа:

- *эмпирическое исследование*, целью которого является получение обобщающих характеристик изучаемой совокупности;
 - *теоретическое исследование* с целью выявления закономерности данного распределения и его теоретического описания.
- Эмпирическое исследование начинается с определения частотных характеристик ряда распределения.

5.2. Частотные характеристики рядов распределения

Исходной частотной характеристикой любого ряда распределения является **частота** n_i . На ее основе можно рассчитать следующие характеристики:

- **Частость** – *удельный вес (доля) единиц совокупности, имеющих определенное значение признака*, т. е. это частота, выраженная в виде относительной величины (доли единицы или процента):

$$q_i = \frac{n_i}{N}, \quad i = \overline{1, m}, \quad \sum_{i=1}^m q_i = 1.$$

Эта характеристика имеет важное значение при исследовании рядов распределения, так как позволяет связать показатели рядов распределения с соответствующими показателями и аппаратом теории вероятностей. В теории вероятностей q_i есть вероятность того, что данное значение признака встретится в совокупности. Частость используется для сопоставления рядов распределения, содержащих равное число статистических единиц.

- **Накопленная частота** – число единиц совокупности, у которых значение признака не превышает данного x^* , т. е. это частота нарастающим итогом:

$$N_{x^*} = \sum_{i=1}^{m^*} n_i, \quad N_{x_m} = N.$$

x^* - данное значение признака в \square - ой группе, для которой рассчитывается накопленная частота.

По накопленным частотам можно построить *кумулятивный ряд распределения* – ряд значений числа единиц совокупности с меньшими и

Формат: Список

равными верхней границе соответствующего интервала значениями признака.

• **Накопленная частость** – удельный вес (доля) единиц, у которых значение признака не превосходит данное x^* , т. е. это частость нарастающим итогом:

$$Q_{x^*} = \sum_{i=1}^{m^*} q_i, \quad Q_{x_m} = 1;$$

• **Плотность распределения** – универсальная частотная характеристика, позволяющая перейти от эмпирического к теоретическому распределению. Для рядов с неравными интервалами только эта характеристика дает правильное представление о характере распределения. Плотность распределения рассчитывается в 2-х вариантах:

- как *абсолютная плотность распределения* φ_i , показывающая число единиц совокупности, приходящихся на единицу ширины интервала значения признака:

$$\varphi_i = \frac{n_i}{a_i}.$$

- как *относительная плотность распределения* φ'_i , показывающая удельный вес единиц совокупности, приходящихся на единицу ширины интервала:

$$\varphi'_i = \frac{q_i}{a_i}.$$

Плотность распределения обеспечивает сопоставимость различных рядов распределения.

Разные ряды распределения характеризуются разным набором частотных характеристик: минимальным – атрибутивные ряды (частота n_i , и частость q_i), для дискретных используются четыре характеристики (частота n_i , частость q_i , накопленная частота N_i , накопленная частость Q_i), для интервальных – все пять (частота n_i , частость q_i , накопленная частота N_i , накопленная частость Q_i , абсолютная φ_i и относительная φ'_i плотности распределения).

Расчет частотных характеристик рассмотрим на следующем примере: имеется распределение рабочих участка по стажу работы. $N=50$ человек, стаж измеряется числом полностью отработанных лет. На основании структурной группировки, выполненной ранее, построен равноинтервальный вариационный ряд, $m=7$, $a_i=4$ года. Для такого ряда рассчитываются все частотные характеристики, результаты расчета приведены в таблице 5.1.

Таблица 5.1.

Расчет характеристик распределения рабочих участка
по стажу работы

№ п/п	Стаж работы, лет		Частота , n_i , чел	Часто сть q_i	Накоп ленна я частот а N_i , чел.	Накоп ленная частот ь Q_i	Абсо лютна я плот ность распре - делен ия	Относ и тельна я плотн ость распре де ления	
	интервал								a_i
	x_i^H	x_i^6					φ_i	φ_i'	
1	0	4	4	6	0,12	6	0,12	1,5	0,03
2	4	8	4	8	0,16	14	0,28	2,0	0,04
3	8	12	4	11	0,22	25	0,50	2,75	0,055
4	12	16	4	13	0,26	38	0,76	3,25	0,065
5	16	20	4	6	0,12	44	0,88	1,5	0,03
6	20	24	4	4	0,08	48	0,96	1,0	0,02
7	24	28	4	2	0,04	50	1,00	0,5	0,01
всего	0	28	28	50	1,00	-	-	1,78	0,036

5.3. Графическое представление рядов распределения

Графики являются наглядной формой отображения рядов распределения. Для изображения рядов применяются линейные графики и плоскостные диаграммы, построенные в прямоугольной системе координат.

Для графического представления атрибутивных рядов распределения используются различные диаграммы: столбиковые, линейные, круговые, фигурные, секторные и т. д.

Для дискретных вариационных рядов графиком является полигон распределения.

Полигоном распределения называется ломаная линия, соединяющая точки с координатами $\{x_i; n_i\}$ или $\{x_i; q_i\}$, где x_i - дискретное значение признака, n_i - частота, q_i - частость.

График строится в принятом масштабе. Вид полигона распределения приведен на рис. 5.1.

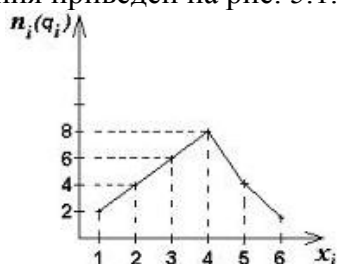


Рис.5.1. Полигон распределения

Для изображения интервальных вариационных рядов применяют **гистограммы**, представляющие собой ступенчатые фигуры, состоящие из прямоугольников, основания которых равны ширине интервала a_i , а высота - частоте n_i (частости q_i) равноинтервального ряда или плотности распределения неравноинтервального φ_i, φ'_i . Построение диаграммы аналогично построению столбиковой диаграммы. Общий вид гистограммы приведен на рис. 5.2.

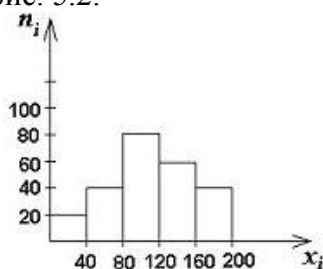


Рис.5.2 Гистограмма распределения равноинтервального ряда

Для графического представления вариационных рядов может использоваться также **кумулята** – ломаная линия, составленная по накопленным частотам (частостям). Накопленные частоты наносятся в виде ординат; соединяя вершины отдельных ординат отрезками прямой, получаем ломаную линию, имеющую неубывающий вид. Координатами точек на графике для дискретного ряда являются $\{x_i; N_i\}$; для интервального ряда - $\{x_i^e; N_i\}$. Начальная точка графика имеет координаты $\{x_1^n; 0\}$, самая высокая точка - $\{x_m^e; N\}$. Общий вид кумуляты приведен на рис.5.3. Использование кумуляты особенно удобно при проведении сравнений вариационных рядов.

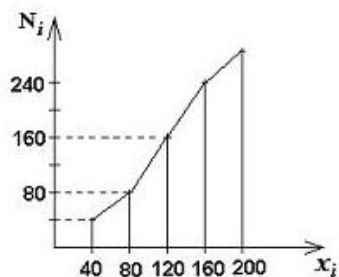


Рис. 5.3. Кумулята распределения

При построении графиков рядов распределения *большое значение имеет соотношение масштабов по оси абсцисс и оси ординат*. В этом случае и необходимо руководствоваться «правилом золотого сечения», в соответствии с которым *высота графика должна быть примерно в два раза меньше его основания*.

5.4. Показатели центра распределения

При проведении эмпирического исследования ряда распределения рассчитываются и анализируются следующие группы показателей:

- показатели положения центра распределения;
- показатели степени его однородности;
- показатели формы распределения.

← - - - - - **Формат:** Список

Показатели положения центра распределения. К ним относятся *степенная средняя в виде средней арифметической и структурные средние – мода и медиана*.

Средняя арифметическая для дискретного ряда распределения рассчитывается по формуле:

$$\bar{x} = \frac{\sum_{i=1}^m x_i n_i}{\sum_{i=1}^m n_i},$$

где x_i - варианты значений признака,
 n_i - частота повторения данного признака.

В интервальном вариационном ряду *средняя арифметическая определяется по формуле:*

$$\bar{x} = \frac{\sum_{i=1}^m b_i n_i}{\sum_{i=1}^m n_i},$$

где b_i - середина соответствующего интервала.

В отличие от средней арифметической, рассчитываемой на основе всех вариантов, мода и медиана характеризует значение признака у статистической единицы, занимающей определенное положение в вариационном ряду.

Медиана (Me) - значение признака у статистической единицы, стоящей в середине ранжированного ряда и делящей совокупность на две равные по численности части.

Мода (Mo) - наиболее часто встречаемое значение признаков в совокупности. Мода широко используется в статистической практике при изучении покупательского спроса, регистрации цен и др.

Для дискретных вариационных рядов Mo и Me выбираются в соответствии с определениями: мода - как значение признака с наибольшей частотой n_i ; положение медианы при нечетном объеме совокупности определяется ее номером $n_{Me} = \frac{N+1}{2}$, где N - объем статистической совокупности. При четном объеме ряда медиана равна средней из двух вариантов, находящихся в середине ряда.

Медиану используют как наиболее надежный показатель *типичного* значения неоднородной совокупности, так как она нечувствительна к крайним значениям признака, которые могут значительно отличаться от основного массива его значений. Кроме этого, медиана находит практическое применение вследствие особого математического свойства:

$$\sum |x_i - Me| \rightarrow \min .$$

Рассмотрим определение моды и медианы на следующем примере: имеется ряд распределения рабочих участка по уровню квалификации. Данные приведены в таблице 5.2.

Таблица 5.2

Распределения рабочих участка по уровню квалификации

№ группы	Разряд рабочих	Число рабочих	Накопленная частота
1	1	3	3
2	2	5	8
3	3	9	17
4	4	14	31
5	5	10	41
6	6	9	50
Всего	-	50	-

Мода выбирается по максимальному значению частоты: при $n_{\max} = 14$ $Mo=4$, т.е. чаще всего встречается 4-ый разряд. Для нахождения медианы

Me определяются центральные единицы $\frac{N+1}{2}$. Это 25 и 26-ая единицы.

По накопленным частотам определяется группа, в которую попадают эти единицы. Это 4-ая группа, в которой значение признака равно 4. Таким образом, $Me = 4$, это означает, что у половины рабочих разряд ниже 4-го, а у другой – выше четвертого.

В интервальном ряду значения Mo и Me вычисляются более сложным путем.

Мода определяется следующим образом:

● По максимальному значению частоты определяется интервал, в котором находится значение моды. Он называется модальным.

Формат: Список

● Внутри модального интервала значение моды вычисляется по формуле:

$$Mo = x_{Mo}^n + a_{Mo} \times \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} + n_{Mo+1})},$$

где x_{Mo}^n - нижняя граница модального интервала,

a_{Mo} - ширина модального интервала,

n_{Mo} , n_{Mo-1} , n_{Mo+1} - соответственно частоты модального, предмодального (предшествующего модальному) и постмодального (следующего за модальным) интервалов.

Для расчета медианы в интервальных рядах используется следующий подход:

● По накопленным частотам находится медианный интервал. Медианным называется интервал, содержащий центральную единицу.

Формат: Список

● Внутри медианного интервала значение Me определяется по формуле:

$$Me = x_{Me}^n + a_{Me} \cdot \frac{\frac{N}{2} - N_{Me-1}}{n_{Me}},$$

где x_{Me}^n - нижняя граница медианного интервала,

a_{Me} - ширина медианного интервала,

N – объем статистической совокупности,

N_{Me-1} - накопленная частота предмедианного интервала,

n_{Me} - частота медианного интервала.

В неравноинтервальных рядах при вычислении Mo используется другая частотная характеристика – абсолютная плотность распределения:

$$Mo = x_{Mo}^n + a_{Mo} \cdot \frac{\varphi_{Mo} - \varphi_{Mo-1}}{(\varphi_{Mo} - \varphi_{Mo-1}) + (\varphi_{Mo} - \varphi_{Mo+1})},$$

где φ_{Mo} - абсолютная плотность распределения модального интервала,

φ_{Mo-1} - абсолютная плотность распределения предмодального интервала,

φ_{Mo+1} - абсолютная плотность распределения постмодального интервала.

Расчет моды и медианы для интервального ряда распределения рассмотрим на примере ряда распределения рабочих по стажу по стажу, приведенного в таблице 5.3.

Таблица 5.3

Распределение рабочих участка по стажу

№ группы	Интервал		a_i	n_i	N_i
	x_i^H	x_i^S			
1	0	4	4	6	6
2	4	8	4	8	14
3	8	12	4	11	25
4	12	16	4	13	28
5	16	20	4	6	44
6	20	24	4	4	48
7	24	28	4	2	50
Всего	0	28	28	50	-

Расчет M_o :

• Максимальная частота $n_{\max} = 13$, она соответствует четвертой группе, следовательно, модальным является интервал с границами 12 – 16 лет.

Формат: Список

• Моду рассчитаем по формуле:

$$M_o = x_{Mo}^H + a_{Mo} \times \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} + n_{Mo+1})} =$$

$$12 + 4 \cdot \frac{13 - 11}{(13 - 11) + (13 - 6)} = 12 + 4 \cdot \frac{2}{2 + 7} = 12 + 4 \cdot 0,22 \approx 13 \text{ лет.}$$

Чаще всего встречаются рабочие со стажем работы около 13 лет. Мода не находится в середине модального интервала, она смещена к его нижней границе, связано это со структурой данного ряда распределения (частота предмодального интервала значительно больше частоты постмодального интервала).

Расчет медианы:

• По графе накопленных частот определяется медианный интервал. Он содержит 25 и 26-ую статистические единицы, которые

находятся в разных группах – в 3-ей и 4-ой. Для нахождения Me можно использовать любую из них. Расчет проведем по 3-ей группе:

$$Me = 8 + 4 \cdot \frac{\frac{50}{2} - 14}{11} = 8 + 4 \cdot \frac{25 - 14}{11} = 12 \text{ лет.}$$

Такое же значение Me можно получить при её расчете по 4-ой группе:

$$Me = 12 + 4 \cdot \frac{\frac{50}{2} - 25}{13} = 8 + 4 \cdot \frac{25 - 14}{11} = 12 \text{ лет.}$$

При сдвоенном центре Me всегда находится на стыке интервалов, содержащих центральные единицы. Вычисленное значение Me показывает, что у первых 25 рабочих стаж работы - менее 12 лет, а у оставшихся 25-ти, следовательно, - более 12 лет.

Моду можно определить графически по полигону распределения в дискретных рядах, по гистограмме распределения – в интервальных, а медиану - по кумуляте.

Для нахождения моды в интервальном ряду правую вершину модального прямоугольника нужно соединить с правым верхним углом предыдущего прямоугольника, а левую вершину – с левым верхним углом последующего прямоугольника. Абсцисса точки пересечения этих прямых и будет модой распределения.

Для определения медианы высоту наибольшей ординаты кумуляты, соответствующей общей численности совокупности, делят пополам. Через полученную точку проводят прямую, параллельную оси абсцисс, до пересечения ее с кумулятой. Абсцисса точки пересечения является медианой.

Кроме Mo и Me в вариантных рядах могут быть определены и другие структурные характеристики – квантили. Квантили предназначены для более глубокого изучения структуры ряда распределения. **Квантиль** – это значение признака, занимающее определенное место в упорядоченной по данному признаку совокупности. Различают следующие виды квантилей:

- **квартили** ($Q_{1/4}, Q_{2/4} = Me, Q_{3/4}$) – значения признака, делящие упорядоченную совокупность на 4 равные части;
- **децили** (d_1, d_2, \dots, d_9) – значения признака, делящие совокупность на 10 равных частей;
- **перцентели** - значения признака, делящие совокупность на 100 равных частей.

Если данные сгруппированы, то значение квартиля определяется по накопленным частотам: номер группы, которая содержит i -ый квантиль. Определяется как номер первой группы от начала ряда, в котором сумма накопленных частот равна или превышает $i \cdot N$, где i – индекс квантиля.

Если ряд интервальный, то значение квантиля определяется по формуле:

$$Q_i = x_{Q_i}^n + a_{Q_i} \cdot \frac{i \cdot N - N_{Q_{i-1}}}{n_{Q_i}},$$

где $x_{Q_i}^n$ - нижняя граница интервала, в котором находится i -ый квантиль;

$N_{Q_{i-1}}$ - сумма накопленных частот интервалов, предшествующих интервалу, в котором находится i -ый квантиль;

n_{Q_i} - частота интервала, в котором находится i -ый квантиль.

Рассчитаем квартили для ряда распределения рабочих участка по стажу работы:

• нижний квартиль $Q_{1/4}$ – соответствует 13-ой единице, верхний квартиль $Q_{3/4}$ – 38-ой. Это соответственно 2-ая и 4-ая группы.

$$Q_{1/4} = 4 + 4 \cdot \frac{\frac{50}{4} - 6}{8} = 4 + 4 \cdot 0,8 \approx 7,2 \text{ лет};$$

$$Q_{3/4} = 12 + 4 \cdot \frac{\frac{3 \cdot 50}{4} - 25}{13} = 12 + 4 \approx 16 \text{ лет}.$$

Следовательно, у четверти рабочих стаж менее 7 лет и у четверти – более 16 лет.

Таким образом, для характеристики положения центра ряда распределения можно использовать 3 показателя: **среднее значение признака, мода, медиана**.

При выборе вида и формы конкретного показателя центра распределения необходимо исходить из следующих рекомендаций:

• для устойчивых социально-экономических процессов в качестве показателя центра используют среднюю арифметическую. Такие процессы характеризуются симметричными распределениями, в которых $\bar{x} = Me = Mo$;

• для неустойчивых процессов положение центра распределения характеризуется с помощью Mo или Me . Для асимметричных процессов предпочтительной характеристикой центра распределения является медиана, поскольку занимает положение между средней арифметической и модой.

5.5. Показатели вариации

Вторая важная задача при определении общего характера распределения – это оценка степени его однородности. Однородность статистических совокупностей характеризуется величиной вариации

(рассеяния) признака, т.е. несовпадением его значений у разных статистических единиц.

Для измерения вариации в статистике используются абсолютные и относительные показатели.

К абсолютным показателям вариации относятся:

- размах вариации R ,
- среднее линейное отклонение \bar{d} ,
- средний квадрат отклонений (дисперсия) δ^2 ,
- среднее квадратическое отклонение δ ,

← Формат: Список

Размах вариации R является наиболее простым показателем вариации, рассчитывается по формуле:

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

Этот показатель представляет собой разность между максимальным и минимальным значениями признаков и характеризует разброс элементов совокупности. Размах улавливает только крайние значения признака в совокупности, не учитывает повторяемость его промежуточных значений, а также не отражает отклонений всех вариантов значений признака.

Размах часто используется в практической деятельности, например, различие между \max и \min пенсией, заработной платой в различных отраслях и т.д.

Среднее линейное отклонение \bar{d} является более строгой характеристикой вариации признака, учитывающей различия всех единиц изучаемой совокупности. *Среднее линейное отклонение* представляет собой *среднюю арифметическую абсолютных значений отклонений* отдельных вариантов от их средней арифметической. Этот показатель рассчитывается по формулам простой и взвешенной средней арифметической:

$$\bar{d} = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N} - \text{для несгруппированных данных};$$

$$\bar{d} = \frac{\sum_{i=1}^m |x_i - \bar{x}| \cdot n_i}{\sum_{i=1}^m n_i} - \text{для сгруппированных данных}.$$

В практических расчетах среднее линейное отклонение используется для оценки ритмичности производства, равномерности поставок.

Так как модули обладают плохими математическими свойствами, то на практике часто применяют другие показатели среднего отклонения от средней – дисперсию и среднее квадратическое отклонение.

Дисперсия признака σ^2 представляет собой средний квадрат отклонений вариантов от их средней величины, является общепринятой мерой вариации. В зависимости от исходных данных дисперсия вычисляется по формулам простой и взвешенной средней арифметической:

$$\text{для несгруппированных данных } \sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N};$$

$$\text{для сгруппированных данных } \sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^m n_i}.$$

При использовании взвешенной средней для расчета дисперсии в интервальных рядах распределения в качестве вариантов значений признака используются срединные значения b (середины интервалов), не являющиеся средним значением в группе. В результате получают приближенное значение дисперсии.

Существуют более простые подходы в вычислении дисперсии. Наиболее часто используется *сокращенный способ расчета дисперсии (метод моментов)*, в соответствии с которым дисперсия σ^2 есть разница между средним из квадратов значений признака $\overline{x^2}$ и квадратом их средней $(\bar{x})^2$:

$$\sigma^2 = \overline{x^2} - (\bar{x})^2,$$

где $\overline{x^2} = \frac{\sum x_i^2}{N}$ - для несгруппированных данных;

$$\overline{x^2} = \frac{\sum x_i^2 \cdot n_i}{\sum n_i} - \text{для сгруппированных данных.}$$

Этот способ позволяет вести расчет дисперсии по исходным данным без предварительного расчета отклонений.

Дисперсия как базовый показатель вариации обладает рядом вычислительных свойств, позволяющих упростить её расчет. К ним относятся:

- дисперсия постоянной величины равна 0;
- дисперсия не меняется, если все варианты увеличить или уменьшить на одно и то же число A ;
- если все варианты умножить (разделить) на число A , то дисперсия увеличится (уменьшится) в A^2 раз.

← **Формат:** Список

Размерность дисперсии соответствует квадрату размерности исследуемого признака, поэтому данный показатель не имеет экономической интерпретации. Для сохранения экономического смысла рассчитывается ещё один показатель вариации – среднее квадратическое отклонение.

Среднее квадратическое отклонение представляет собой среднюю квадратическую из отклонений отдельных значений признака от их средней арифметической:

для несгруппированных да

для сгруппированных данных
$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^m n_i}} .$$

Среднее квадратическое отклонение является именованной величиной, имеет размерность усредняемого признака, экономически хорошо интерпретируется. Она также используется для оценки надежности средней: чем меньше среднее квадратическое отклонение σ , тем надежнее среднее значение признака \bar{x} , тем лучше средняя представляет исследуемую совокупность.

Для распределений, близких к нормальным между средним квадратическим отклонением и средним линейным отклонением существует следующая зависимость:

$$\sigma \approx 1,25 \cdot \bar{d} .$$

- **Относительные показатели вариации** предназначены для оценки и сравнения вариации нескольких признаков по одной совокупности или же вариации одного и того же признака по нескольким совокупностям. Базой для их исчисления является средняя арифметическая.

Самым распространенным относительным показателем вариации является **коэффициент вариации** V_σ . Он представляет собой отношение среднего квадратического отклонения к средней арифметической, выраженное в процентах:

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

Коэффициент вариации используется для характеристики однородности исследуемой совокупности. *Статистическая совокупность считается количественно однородной, если коэффициент вариации не превышает 33%.*

Расчет показателей вариации рассмотрим на примере ряда распределения рабочих участка по стажу работы. Для этого составим вспомогательную таблицу:

Таблица 5.4.

Расчет показателей вариации для распределения рабочих по стажу работы

№ группы	Стаж работы, лет			n_i	$n_i \cdot x_i$	Расчет среднего линейного отклонения		Расчет дисперсии	
	x_i^H	x_i^6	b_i			$ x_i - \bar{x} $	$ x_i - \bar{x} \cdot n_i$	x_i^2	$x_i^2 \cdot n_i$
1	0	4	2	6	12	10	60	4	24
2	4	8	6	8	48	6	48	36	288
3	8	12	10	11	110	2	22	100	1100
4	12	16	14	13	182	2	26	196	2548
5	16	20	18	6	108	6	36	324	1944
6	20	24	22	4	88	10	40	484	1936
7	24	28	26	2	52	14	28	676	1352
Итого	0	28	14	50	600	-	260	-	9192

- Определение среднего стажа работы:

$$\bar{x} = \frac{\sum x_i \cdot n_i}{\sum n_i} = \frac{600}{50} = 12 \text{ лет.}$$

Таким образом, наиболее типичным для рабочих участка является стаж работы, равный 12 годам.

- Определение размаха:

$$R = 28 - 0 = 28 \text{ лет.}$$

Размах показывает общий диапазон изменения стажа, он составляет 28 лет.

- Среднее линейное отклонение составляет

$$\bar{d} = \frac{\sum |x_i - \bar{x}| \cdot n_i}{\sum n_i} = \frac{260}{50} = 5,2 \text{ года.}$$

- Дисперсия для данного ряда составляет

$$\sigma^2 = \bar{x}^2 - (\bar{x})^2 = \frac{\sum x_i^2 \cdot n_i}{\sum n_i} - \left(\frac{\sum x_i \cdot n_i}{\sum n_i} \right)^2 = \frac{9192}{50} - 12^2 = 183,84 - 144 = 39,84 \text{ лет}^2.$$

Показатель с такой размерностью невозможно интерпретировать, поэтому рассчитаем среднее квадратическое отклонение

- Среднее квадратическое отклонение составляет $\sigma = \sqrt{39,84} = 6,3$ года.

Проверим соотношение между средним линейным отклонением и средним квадратическим отклонением: $\sigma \approx 1,25 \cdot \bar{d} \approx 6,5$. Можно сделать вывод, что распределение рабочих по стажу близко к нормальному.

- Коэффициент вариации составляет $V_\delta = \frac{6,3}{12} \cdot 100\% \approx 53\%$, что свидетельствует о высокой колеблемости признака в совокупности.

5.6. Правило сложения дисперсий

Если изучаемая совокупность состоит из нескольких частей, то для каждой из них можно рассчитать среднее значение признака и дисперсию. Кроме этого можно рассчитать дисперсию, измеряющую вариацию признака между выделенными частями совокупности.

Таким образом, с помощью разных видов дисперсии можно более глубоко изучить вариацию признака в совокупности. Различают следующие *виды дисперсий*: *общая дисперсия, межгрупповая и внутригрупповая.*

Общая дисперсия σ^2 *измеряет вариацию признака во всей статистической совокупности под влиянием всех факторов, вызывающих эту вариацию.* Она рассчитывается по формуле:

$$\sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot n_i}{\sum n_i}.$$

Межгрупповая дисперсия δ^2 *характеризует изменение признака обусловленное факторами, положенными в основу группировки.* Таким образом, межгрупповая дисперсия есть дисперсия локальных средних. Ее расчет проводится по формуле:

$$\delta^2 = \frac{\sum_{i=1}^m (\tilde{x}_i - \bar{x})^2}{m}, \text{ где}$$

\tilde{x}_i - локальная средняя (среднее значение признака) в каждой группе,
 m – количество групп (частей) в совокупности.

Внутригрупповая дисперсия σ_i^2 *характеризует случайную вариацию, т.е. колебания признака, возникающие под воздействием неучтенных факторов и независящую от вариации признака – фактора, положенного в основу группировки.* Внутригрупповая дисперсия σ_i^2 рассчитывается для каждой однородной группы:

$$\sigma_i^2 = \frac{\sum_{i=1}^{n_i} (x_i - \tilde{x}_i)^2}{n_i}.$$

На основании внутригрупповой дисперсии рассчитывается **средняя из внутригрупповых дисперсий (остаточная)** $\overline{\sigma_i^2}$

$$\overline{\sigma_i^2} = \frac{\sum_{i=1}^m \sigma_i^2}{m}.$$

Перечисленные виды дисперсий связаны между собой следующим отношением:

$$\sigma_0^2 = \delta^2 + \overline{\sigma_i^2}.$$

Указанное соотношение называется **правилом сложения дисперсий**. Очевидно, что, чем больше величина межгрупповой дисперсии, тем более качественно проведена группировка, тем сильнее факторный признак влияет на общую вариацию. Кроме этого, пользуясь указанным правилом, можно по двум известным дисперсиям рассчитать неизвестную третью дисперсию.

5.7. Исследование формы распределения

Выяснение общего характера распределения предполагает не только оценку степени его однородности, но и исследование формы распределения, т.е. оценку симметричности и эксцесса.

Из математической статистики известно, что при увеличении объема статистической совокупности ($N \rightarrow \infty$) и одновременного уменьшении интервала группировки ($x_i \rightarrow 0$) полигон либо гистограмма распределения все более и более приближается к некоторой плавной кривой, являющейся для указанных графиков пределом. Эта кривая называется **эмпирической кривой распределения** и представляет собой **графическое изображение в виде непрерывной линии изменения частот, функционально связанного с изменением вариант**.

В статистике различают следующие **виды кривых распределения**:

- одновершинные кривые;
- многовершинные кривые.

← Формат: Список

Однородные совокупности описываются одновершинными распределениями. Многовершинность распределения свидетельствует о неоднородности изучаемой совокупности или о некачественном выполнении группировки.

Одновершинные кривые распределения делятся на симметричные, умеренно асимметричные и крайне асимметричные.

Распределение называется симметричным, если частоты любых 2-х вариантов, равноотстоящих в обе стороны от центра распределения, равны между собой. В таких распределениях $\bar{x} = Mo = Me$.

Для характеристики асимметрии используют коэффициенты асимметрии.

Наиболее часто используются следующие из них:

- **Коэффициент асимметрии Пирсона** $As = \frac{\bar{x} - Mo}{\sigma}$.

В одновершинных распределениях величина этого показателя изменяется от -1 до +1.

в симметричных распределениях $As=0$.

При $As>0$ наблюдается правосторонняя асимметрия (рис.5.4). В распределениях с правосторонней асимметрией $Mo \leq Me \leq \bar{x}$.

При $As<0$ – асимметрия отрицательная левосторонняя, $Mo>Me>\bar{x}$.

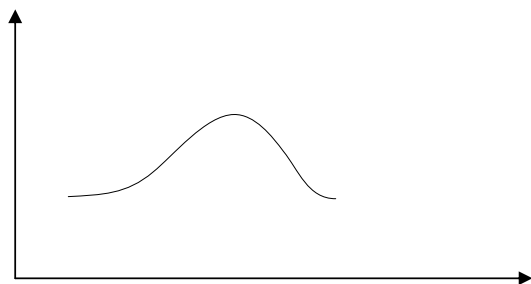


Рис. 5.4. Правосторонняя асимметрия

При $As<0$ имеет место левосторонняя асимметрия (Рис. 5.4.)
 $Mo>Me>\bar{x}$.

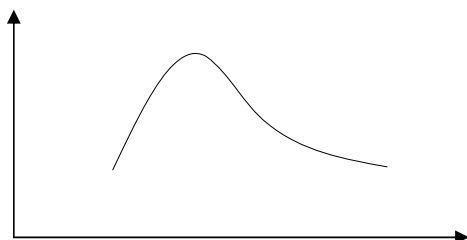


Рис. 5.5. Левосторонняя асимметрия

Чем ближе по модулю As к 1, тем асимметрия существеннее:

- если $|As|<0,25$, то асимметрия считается незначительной;
- если $0,25 < |As| < 0,5$ то асимметрия считается умеренной;
- если $|As|>0,5$ – асимметрия значительна.

Коэффициент асимметрии Пирсона характеризует асимметрию только в центральной части распределения, поэтому более

распространенным и более точным является **коэффициент асимметрии**, рассчитанный на основе центрального момента 3-его порядка:

$$As = \frac{\mu_3}{\sigma^3},$$

где μ_3 - центральный момент третьего порядка;

σ^3 - среднее квадратическое отклонение в третьей степени.

Центральным моментом в статистике называется среднее отклонение индивидуальных значений признака от его среднеарифметической величины.

Центральный момент k-ого порядка рассчитывается как:

$$\mu_k = \frac{\sum_{i=1}^N (x_i - \bar{x})^k}{n} - \text{для несгруппированных данных};$$

$$\mu_k = \frac{\sum_{i=1}^m (x_i - \bar{x})^k}{i = 1 \sum_{i=1}^m n_i} - \text{для сгруппированных данных}.$$

Соответственно формулы для определения центрального момента третьего порядка имеют следующий вид:

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3}{n} - \text{для несгруппированных данных};$$

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3 \cdot n_i}{\sum n_i} - \text{для сгруппированных данных}.$$

Для оценки существенности рассчитанного вторым способом коэффициента асимметрии определяется его средняя квадратическая ошибка:

$$\sigma_{As} = \sqrt{\frac{6 \cdot (N-1)}{(N+1) \cdot (N+3)}}.$$

Если $\frac{|AS|}{\sigma_{AS}} > 3$, асимметрия является существенной.

Для одновершинных распределений рассчитывается еще один показатель оценки его формы – *эксцесс*. *Эксцесс является показателем островершинности распределения*. Он рассчитывается для симметричных распределений на основе центрального момента 4-ого порядка μ_4 :

$$Ex = \frac{\mu_4}{\sigma^4} - 3,$$

где μ_4 - центральный момент 4-го порядка.

$$\mu_4 = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{N} - \text{для несгруппированных данных};$$

$$\mu_4 = \frac{\sum_{i=1}^m (x_i - \bar{x})^4 \cdot n_i}{\sum_{i=1}^m n_i} - \text{ для сгруппированных данных.}$$

При симметричных распределениях $E_x=0$. если $E_x>0$, то распределение относится к **островершинным**, если $E_x<0$ – к **плосковершинным**.

Рассчитаем показатели асимметрии и эксцесса для ряда распределения рабочих по стажу работы. Ранее для данного ряда были получены следующие характеристики:

$$\bar{x} = 12 \text{ лет,} \quad Mo = 12,9 \text{ лет,} \quad \delta = 6,3 \text{ года.}$$

Коэффициент асимметрии Пирсона получается равным:

$$As = \frac{\bar{x} - Mo}{\sigma} = \frac{12 - 12,9}{6,3} \approx -0,14 < 0, \text{ что говорит о наличии незначительной}$$

левосторонней асимметрии в центральной части распределения.

Коэффициент асимметрии, рассчитанный через центральный момент 3-его порядка:

$$As = \frac{\mu_3}{\sigma^3} = \frac{\sum (x_i - \bar{x})^3 \cdot n_i}{\sum n_i} := \frac{61,44}{6,3^3} = \frac{61,44}{250} = 0,24 > 0.$$

Это означает, что в целом по всему ряду наблюдается правосторонняя асимметрия.

Расчет центрального момента 3- его порядка μ_3 приведен во вспомогательной таблице 5.6.

Таблица 5.6.

Расчет центральных моментов 3- его и 4-ого порядка

№	x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^3 \cdot n_i$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})^4 \cdot n_i$
1	2	6	-10	-1000	-6000	10000	60000
2	6	8	-6	-216	-1728	1296	10368
3	10	11	-2	-8	-88	16	176
4	14	13	2	8	104	16	208
5	18	6	6	216	1296	1296	7776
6	22	4	10	1000	4000	10000	40000
7	26	2	14	2744	5488	38416	76832
Итого	14	50	-	-	3072	-	195360

Показатель эксцесса:

$$E_x = \frac{\mu_4}{\delta^4} - 3 = \frac{195360}{50} \div 6,3^4 - 3 = \frac{3907,2}{1575,3} - 3 = 2,5 - 3 = -0,5, \quad \text{что}$$

свидетельствует о том, что распределение плосковершинное.

5.8. Теоретические распределения в анализе вариационных рядов

Эмпирические кривые распределения, построенные на основе, как правило, небольшого числа наблюдений очень трудно описать аналитически, поэтому для выявления статистических закономерностей, сравнения и обобщения различных совокупностей аналогичных данных используются теоретические распределения.

Теоретические распределения – это хорошо изученные в теории распределения, представляющие собой зависимости между плотностями распределения и значениями признака, отражающие закономерности распределения. Они описываются статистическими функциями, параметры которых вычисляются по статистическим характеристикам изучаемой совокупности.

Исследование формы распределения предполагает замену эмпирического распределения известным теоретическим, близким ему по форме. При этом необходимо соблюдать условие: различия между эмпирическим и теоретическим распределениями должны быть минимальными. Это означает, что **сумма частот эмпирического распределения должна соответствовать сумме частот теоретического распределения**, т.е. $\sum_{i=1}^m n_i \approx \sum_{i=1}^m n_{i_r}$, где n_{i_r} – частота теоретического распределения.

Теоретическое распределение в этом случае является некоторой идеализированной моделью эмпирического распределения, и анализ вариационного ряда сводится к сопоставлению эмпирического и теоретического распределений и определению различий между ними.

В статистической практике наиболее широко используют следующие теоретические распределения:

- **Биномиальное распределение** – для описания распределения дискретного альтернативного признака. Оно представляет собой распределение вероятности исходов события, которые можно оценить как положительные или отрицательные.

- **Распределение Пуассона** – для изучения маловероятных событий в большой серии независимых испытаний (объем совокупностей $n \geq 100$, доля единиц, обладающих данным признаком $q \leq 0,1$). Например, количество бракованных деталей в массовом производстве, число отказов автоматических линий – т.е. в статистическом контроле.

Вероятность появления таких событий подчиняется P_n закону Пуассона – «закону редких событий»:

$$P_n = \frac{\lambda^n \cdot e^{-\lambda}}{n!},$$

где P_n - вероятность события при одном испытании;

n - частота данного события

$\lambda = n \cdot p$ - среднее число появления события в одинаковых условиях;

$e = 2,72$ - основание натурального логарифма.

Распределение Пуассона обычно применяется в статистическом контроле качества в массовом производстве.

- **Распределение Максвелла** применяется при исследовании признака, для которого заранее известно, что распределение имеет положительную асимметрию. Чаще всего Распределение Максвелла используется при описании технологических характеристик производственных процессов.

- **Распределение «Стюдента»** применяют для описания распределения

ошибок в малых выборках ($n < 30$).

Плотность распределения ошибок малой выборки определяется как:

$$\varphi_t = A \cdot \left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}},$$

где $t = \frac{\tilde{x} - \bar{x}}{S \cdot \sqrt{n-1}}$ - отношение Стюдента,

S - выборочное среднее квадратическое отклонение,

\tilde{x} - выборочная средняя;

$K = n - 1$ - число степеней свободы при определении выборочной дисперсии,

$$A = \frac{\gamma \cdot \left(\frac{k+1}{2}\right)}{\gamma \cdot \left(\frac{k}{2}\right) \cdot \sqrt{\pi \cdot k}},$$

γ - значение γ функции.

Распределение Стюдента используется только при оценке ошибок выборок, взятых из генеральной совокупности с нормальным распределением признака.

- **Нормальное распределение (распределение Гаусса)** применяется для описания распределения признаков, на которые действуют множество независимых факторов, среди которых нет доминирующих.

Функция нормального распределения выглядит следующим образом:

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

где $\varphi'(x)$ - относительная плотность распределения (ордината кривой нормального распределения);

$\pi=3,14$, $e=2,72$ - математические константы;

\bar{x} - среднее значение признака в распределении;

σ - среднее квадратическое отклонение.

Для конкретного распределения среднее значение признака \bar{x} и среднее квадратическое отклонение σ являются постоянными величинами.

Графически нормальное распределение может быть представлено в виде симметричной колоколообразной кривой (рис. 5.6):

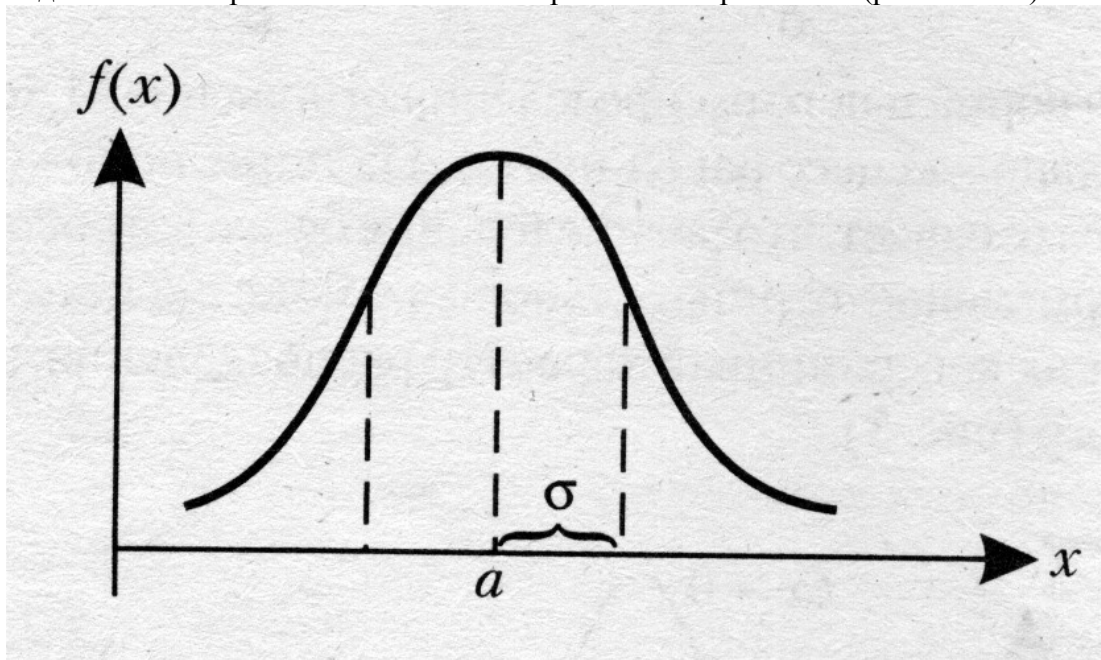


Рис. 5.6. Нормальное распределение

К **основным свойствам** кривой нормального распределения относятся:

• кривая распределения является **одновершинной**; координаты вершины - $\{\bar{x}; \frac{1}{\sigma \cdot \sqrt{2\pi}}\}$;

• кривая распределения симметрична относительно оси, проходящей через центр распределения $\bar{x} = Mo = Me$;

• кривая имеет **три точки перегиба**: в вершине, на левой ветви $\{\bar{x} - \sigma; \frac{1}{\sigma \cdot \sqrt{2\pi} \cdot e}\}$, и на правой - $\{\bar{x} + \sigma; \frac{1}{\sigma \cdot \sqrt{2\pi} \cdot e}\}$;

Формат: Список

• кривая имеет две ветви, асимптотически приближающиеся к оси абсцисс, продолжаясь до бесконечности;

• если меняется значение \bar{x} , кривая перемещается вдоль оси ординат, при этом форма кривой не меняется;

• если меняется значение σ , меняется форма распределения при неизменном положении центра распределения: при уменьшении σ - уменьшается вариация, кривая становится более пологой, увеличивается эксцесс; при увеличении σ - увеличивается вариация, эксцесс уменьшается;

• площадь, ограниченная кривой сверху и осью абсцисс снизу, характеризует вероятность появления определенных значений признака: если всю её принять за 100%, то в пределах $\bar{x} \pm \sigma$ находится 68,3% всех значений признака, в пределах $\bar{x} \pm 2\sigma$ - 95,44% значений, в пределах $\bar{x} \pm 3\sigma$ - 99,73% значений признака.

Этот вывод называется правилом “трех сигм”, в соответствии, с которым можно считать, что все возможные значения нормально распределенного признака укладываются в интервал $\bar{x} \pm 3\sigma$.

Пользоваться функцией нормального распределения в её первоначальном виде сложно, так как для каждой пары значений \bar{x} и σ необходимо создавать свои таблицы значений. Поэтому функцию стандартизируют и затем используют для обработки рядов распределения, для чего вводится понятие *стандартного отклонения* t_i :

$$t_i = \frac{x_i - \bar{x}}{\sigma}.$$

тогда:

$$\varphi'(x) = \frac{1}{\sigma} \cdot \left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \right).$$

Выражение $\varphi'(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$ состоит из констант, не содержит параметров, **называется стандартизованной функцией нормального распределения**. Для неё разработаны специальные таблицы, позволяющие находить конкретные значения $\varphi'(t)$ при различных значениях аргумента t_i (Приложение 1).

Исходная функция нормального распределения связана со стандартизированной соотношением:

$$\varphi'(x) = \frac{1}{\sigma} \cdot \varphi'(t).$$

Стандартизованная функция является четной, т.е. $\varphi'(-t) = \varphi'(t)$.

Для примера рассмотрим подбор теоретического распределения к ряду распределения рабочих участка по стажу.

Данный ряд распределения характеризуется следующими параметрами:

$$\bar{x} = 12 \text{ лет}, \quad \sigma = 6,3 \text{ года.}$$

Для того чтобы оценить близость указанного ряда распределения к нормальному, необходимо рассчитать частоты теоретического ряда распределения n_{i_r} .

Для их расчета определяются стандартные отклонения $t = \frac{x - \bar{x}}{\sigma}$, затем по таблицам значений функции Лапласа (Приложение 1) находят значения $\varphi'(t)$.

Для получения частот теоретического распределения n_{i_r} необходимо иметь в виду, как относительная плотность распределения $\varphi'(x)$ связана с одной стороны с частотой n_i , а с другой - со стандартизованной функцией нормального распределения $\varphi'(t)$. Эти связи выражаются следующими зависимостями:

$$\varphi'(x) = \frac{q_{i_r}}{a_i}, \quad q_{i_r} = \frac{n_{i_r}}{N}, \text{ следовательно, } \varphi'(x) = \frac{n_{i_r}}{N \cdot a_i}.$$

С другой стороны, $\varphi'(x) = \frac{1}{\sigma} \cdot \varphi'(t)$, таким образом, имеет место равенство:

$$\frac{n_{i_r}}{N \cdot a_i} = \frac{1}{\sigma} \cdot \varphi'(t), \text{ отсюда } n_{i_r} = \frac{a_i \cdot N}{\sigma} \cdot \varphi'(t);$$

где a_i - ширина интервала,

N - объем статистической совокупности,

σ - среднее квадратическое отклонение,

$\varphi'(t)$ - стандартизованная функция нормального распределения.

Полученные значения n_{i_r} округляются до целых значений в соответствии со смыслом характеристики частоты.

Расчеты теоретических частот распределения рабочих по стажу приведены в таблице 5.6.

Таблица 5.6.

Вспомогательные расчеты для построения теоретического распределения по данным о стаже работы рабочих участка.

№ п/п	Стаж, лет		n_i	$x - \bar{x}$	$t = \frac{x - \bar{x}}{\delta}$	$\varphi'(t)$	n	Расчет χ^2 -критерия		Расчет λ -критерия		
	интервал	b_i						$n_i - n_{i_r}$	$\frac{(n_i - n_{i_r})^2}{n_{i_r}}$	N_i	N_{i_r}	$N_i - N_{i_r}$
1	0 - 4	2	6	-10	-1,59	0,1127	4	2	1,00	6	4	2
2	4 - 8	6	8	-6	-0,95	0,2541	8	0	0,00	14	12	2
3	8 - 12	10	11	-2	-0,32	0,3790	12	-1	0,08	25	24	1
4	12 - 16	14	13	+2	+0,32	0,3790	12	+1	0,08	38	36	2
5	16 - 20	18	6	+6	+0,95	0,2541	8	-2	0,50	44	44	0
6	20 - 24	22	4	+10	+1,59	0,1127	4	0	0,00	48	48	0
7	24 - 28	26	2	+14	+2,22	0,0339	2	0	0,00	50	50	0
Все го	0 - 28	14	50	-	-	-	50	0	1,66	-	-	-

Для определения близости эмпирического и теоретического распределений, можно построить эмпирическую и теоретическую кривые распределения. Их сопоставление позволяет оценить степень расхождения между ними.

Эмпирическую кривую строим по точкам с координатами $\{b_i, n_i\}$, теоретическую – по точкам с координатами $\{b_i, n_{i_r}\}$.

Визуальное сопоставление эмпирической и теоретической кривых распределения позволяет получить субъективную оценку их близости. Сравнивая графики, можно утверждать, что наблюдается довольно большая близость фактических и теоретических частот распределения. Следовательно, можно сделать вывод о том, что исследуемый ряд подчиняется закону нормального распределения. Для

получения объективной оценки расхождения между эмпирической и теоретической кривыми распределения используются специальные статистические показатели – критерии согласия.

5.9. Оценка близости эмпирического и теоретического распределений

Эмпирическое распределение отличается от теоретического тем, что на значения признака в нем влияют случайные факторы. С увеличением объема статистической совокупности влияние случайных факторов ослабевает, и эмпирическое распределение все менее отличается от теоретического.

Для оценки близости распределений используются особые показатели – **критерии согласия**. Они основаны на использовании различных мер расстояний между эмпирическим и теоретическим распределением.

Наиболее часто на практике используются следующие критерия согласия:

- **«хи-квадрат»- критерий (критерий Пирсона);**
- **«лямбда»- критерий» (критерий Колмогорова).**

← **Формат:** Список

5.9.1. «Хи-квадрат» - критерий является случайной величиной, имеющей распределение, близкое к распределению «хи-квадрат». Его величина определяется по формуле:

$$\chi_p^2 = \sum_{i=1}^m \frac{(n_i - n_{i_t})^2}{n_{i_t}}.$$

Чем меньше эмпирические и теоретические частоты в отдельных группах отличаются друг от друга, тем меньше эмпирическое распределение отличается от теоретического, то есть тем в большей степени эмпирическое и теоретическое распределения согласуются между собой.

Для оценки существенности расчетной величины «хи-квадрат» - критерия оно сравнивается с табличным (критическим) значением χ_k^2 , определяемым по статистическим таблицам значений χ^2 -критерия. χ_k^2 определяют в зависимости от уровня значимости α и параметра $k=m- m_1 -1$, где α - вероятность ошибки, m_1 - число оцененных параметров теоретического распределения по наблюдаемым значениям признака.

Уровень значимости τ выбирается таким образом, что $P(\chi_p^2 > \chi_k^2) = \alpha$. Обычно α принимается равным 0,05 или 0,01, что соответствует вероятности 95% или 99%.

Если $\chi_p^2 \leq \chi_k^2$, то считают, что распределения близки друг другу, различия между ними несущественны.

Критерий Пирсона можно использовать можно при соблюдении следующих условий:

- в совокупности не менее 50 единиц наблюдения ($N \geq 50$),
- теоретические частоты $n_{i_r} \geq 5$, - если это условие не соблюдается, то следует объединить интервалы.

Формат: Список

Рассчитаем в таблице 4.6. значения отклонений $(n_i - n_{i_r})$ и фактическое значение χ^2 - критерия. По расчету $\chi_p^2 = 1,66$. Это значение сравнивается с табличным, определенном при числе степеней свободы $k=4$ и уровне значимости $= 0,05$. Оно равно $\chi_k^2 = 9,49$.

Таким образом $\chi_p^2 < \chi_k^2$; эмпирическое и теоретическое распределения признаются близкими друг другу с вероятностью 95%, расхождения между ними - несущественными, вызываемыми случайной вариацией признака в совокупности.

На основе χ^2 - критерия может быть рассчитан ещё один критерий согласия – **критерий Романовского:**

$$C = \frac{\chi_p^2 - (m - 3)}{\sqrt{2 \cdot (m - 3)}}.$$

Эмпирическое и теоретическое распределения признаются близкими друг другу, если $C < 3$.

5.9.2. Критерий согласия Колмогорова основан на другой мере близости распределений. Для оценки близости эмпирического распределения к нормальному используется максимальная разница между накопленными эмпирическими и накопленными теоретическими частотами. Расчетное значение «лямбда»- критерия» определяется по формуле:

$$\lambda_p = \frac{D}{\sqrt{\sum_{i=1}^m n_i}} = D : \sqrt{N},$$

$$\text{где } D = \max_{i=1, m} \{N_i - N_{i_r}\}$$

N_i - накопленная эмпирическая частота,

N_{i_r} - накопленная теоретическая частота.

По рассчитанному значению λ_p по специальной таблице вероятностей «лямбда»- критерия» определяется вероятность того, что рассматриваемое эмпирическое распределение подчиняется закону нормального распределения.

Для рассматриваемого примера $D=2$ - в соответствии с расчетом, приведенным в таблице 4.6.

$$\text{Тогда } \lambda_p = \frac{2}{\sqrt{50}} = \frac{2}{7,07} = 0,283.$$

По таблице вероятностей $P(\lambda)$ определяем, что $\lambda=0,283$ соответствует вероятность $P(\lambda)$, близкая к 1.

Полученное значение вероятности свидетельствует о том, что расхождение между эмпирическим и теоретическим распределениями несут существенны, вызваны случайной вариацией признака в статистической совокупности. В основе эмпирического распределения рабочих по стажу лежит закон нормального распределения.

6. Ряды динамики

6.1. Понятие и классификация рядов динамики

В статистике динамикой принято называть процесс развития, движения социально-экономических явлений во времени. Для отображения таких процессов строятся **ряды динамики** (хронологические, временные, динамические ряды), **представляющие собой последовательность упорядоченных во времени значений статистического показателя**. Любой ряд динамики состоит из двух элементов:

1. *показатель времени t_i - это моменты или периоды времени, к которым относятся числовые значения показателей;*

2. *уровень ряда y_i , под которым понимается значение статистического показателя, относящееся к определенному моменту или периоду времени.*

Каждый ряд динамики может быть представлен в табличной форме - в виде пар значений t_i и y_i (таблица 6.1); и в графической форме - в виде линейной диаграммы.

Таблица 6.1.

Ряд динамики

t_1	t_2	t_i	t_n
y_1	y_2	y_i	y_n

При обработке статистических данных *используются ряды динамики, различающиеся по следующим признакам: по времени, форме представления уровней, числу показателей, по расстоянию между датами или интервалами.*

*По времени различают **моментные и интервальные ряды динамики.***

В моментных рядах уровни выражают состояние явления на критический момент времени – начало месяца, квартала, года и т.д. Например, численность населения, численность работающих и т.д. В таких рядах каждый последующий уровень полностью или частично содержит значение предыдущего уровня, поэтому суммировать уровни нельзя, так как это приводит к повторному счету.

В интервальных – уровни отражают состояние явления за определенный период времени – сутки, месяц, год и т.д. Это ряды показателей объема производства, объема продаж по месяцам года, количества отработанных человеко-дней и т.д.

По форме представления уровней различают **ряды абсолютных, относительных и средних величин.**

По числу показателей выделяют **изолированные и комплексные ряды** динамики (многомерные).

Изолированный ряд строится по отдельному показателю, комплексный – по системе взаимосвязанных показателей.

По расстоянию между датами или интервалами ряды динамики делятся на ряды с **равноотстоящими и неравноотстоящими уровнями.**

В рядах с равноотстоящими уровнями расстояние между датами или периодами одинаково, в рядах с неравноотстоящими уровнями – оно различно.

Например, рассмотрим ряд динамики числа построенных квартир и их среднего размера, приведенный в таблице 6.2.

Таблица 6.2.

Динамика численности построенных квартир
и их среднего размера за 2000 – 2004гг.

Наименование показателя	1991	1992	1993	1994	1995
Число квартир, тыс.	1190	1151	682	682	630
Средний размер квартир, m^2 общей площади	49,9	54,4	63,8	67,3	71,9
Удельный вес жилой площади в общей площади квартир, %	62,7	60,7	59	56,1	50,1

Ряд относится к интервальным, многомерным, состоит из рядов абсолютных, средних, относительных величин с равноотстоящими уровнями.

Чтобы ряды динамики давали правильное представление о процессах, которые они представляют, при их составлении *необходимо соблюдать следующие требования:*

1. **Периодизация развития** – расчленение процесса во времени на однородные этапы, в пределах которых показатель подчиняется одному закону. Например, весь советский этап развития России является особым однородным периодом, кардинально отличающимся от предыдущих периодов и того, что мы наблюдаем сейчас. Внутри него можно выделить более короткие и более однородные периоды:

- Довоенные годы (индустриализация и коллективизация),
- Великая Отечественная Война,
- Послевоенное восстановление народного хозяйства и т.д.

2. **Обеспечение сопоставимости уровней** – использование единых методик расчета показателей, одинаковых единиц измерения,

единого круга объектов наблюдения, единых территориальных границ, единого содержания показателей.

3. **Систематизация уровней** в хронологическом порядке - в рядах динамики не должно быть пропусков отдельных уровней. Если данных не хватает, то их восполняют условными расчетными значениями уровней.

С помощью рядов динамики в статистике решают следующие задачи:

- **Получение характеристик интенсивности изменения явления во времени и характеристик отдельных уровней;**
- **Выявление и количественная оценка основной долговременной тенденции развития явления;**
- **Изучение периодических и сезонных колебаний явления;**
- **Экстраполяция и прогнозирование.**

Обработка рядов динамики осуществляется в 3 этапа:

1. *Определение системы характеристики динамического ряда;*
2. *Разложение ряда на отдельные компоненты;*
3. *Прогнозирование на основе экстраполяции.*

6.2. Система характеристик динамического ряда

Система характеристик динамического ряда включает в себя:

- **индивидуальные (частные) характеристики;**
- **сводные (обобщающие) характеристики.**

← Формат: Список

К индивидуальным показателям интенсивности изменения явления относятся:

- *абсолютный прирост Δy_i ;*
- *темп роста T_i (коэффициент роста K_i);*
- *темп прироста T'_i (коэффициент прироста K'_i);*
- *абсолютное значение одного процента прироста A_i ;*
- *пункт роста P_i .*

← Формат: Список

Первые три из перечисленных характеристик можно рассчитать двумя способами в зависимости от применяемой базы сравнения. База сравнения может быть постоянной или переменной. Соответственно, можно рассчитать *базисные или цепные характеристики динамического ряда.*

Абсолютный прирост Δy_i характеризует размер увеличения (уменьшения) уровня ряда по сравнению с выбранной базой:

- **цепной абсолютный прирост** показывает, на сколько изменилось значение данного уровня по сравнению с предыдущим, то есть приращение уровня по сравнению с предыдущим:

$$y_{\bar{u}_i} = y_i - y_{i-1}, \quad i = \overline{2, n}.$$

- **базисный абсолютный прирост** показывает, на сколько изменилось значение данного уровня по сравнению с исходным (начальным) уровнем:

$$y_{\bar{\sigma}_i} = y_i - y_1, \quad i = \overline{2, n}.$$

y_1 - начальный уровень ряда.

Между базисными и цепными абсолютными приростами существует взаимосвязь: сумма всех цепных абсолютных приростов равна базисному приросту конечного уровня:

$$\sum_{i=2}^n \Delta y_{\bar{u}_i} = \Delta y_{\bar{\sigma}_n},$$

где y_n - конечный уровень ряда.

Коэффициент роста (относительный прирост) характеризует интенсивность изменения уровней ряда (скорость изменения уровней). Он показывает, во сколько раз уровень данного периода выше или ниже базисного уровня. Этот показатель как относительная величина, выраженная в долях единицы, называется **коэффициентом (индексом) роста**; выраженная в процентах, называется **темпом роста**.

- **Цепной коэффициент роста** показывает, во сколько раз текущий уровень выше или ниже предыдущего:

$$K_{\bar{u}_i} = \frac{y_i}{y_{i-1}}, \quad i = \overline{2, n};$$

- **Базисный коэффициент роста** показывает, во сколько раз текущий уровень выше или ниже начального уровня:

$$K_{\bar{\sigma}_i} = \frac{y_i}{y_1}, \quad i = \overline{2, n}.$$

Между базисными и цепными темпами (коэффициентами) роста имеется зависимость: *произведения последовательных цепных коэффициентов роста равно базисному коэффициенту роста за весь промежуток времени:*

$$K_{\bar{u}_2} \cdot K_{\bar{u}_3} \cdot \dots \cdot K_{\bar{u}_n} = K_{\bar{\sigma}_n};$$

а частное от деления текущего базисного коэффициента роста на предыдущий базисный коэффициент роста равно текущему цепному коэффициенту роста:

$$K_{\bar{u}_i} = \frac{K_{\bar{\sigma}_i}}{K_{\bar{\sigma}_{i-1}}}, \quad i = \overline{2, n}.$$

Коэффициент роста всегда есть положительная величина, область его допустимых значений- $(0 - + \infty)$.

Коэффициент прироста характеризует относительную скорость изменения уровня ряда в единицу времени. Показывает, на какую долю единицы (или процент) уровень данного периода или момента времени выше или ниже базисного уровня.

- **Цепной коэффициент прироста** рассчитывается по формуле:

$$K'_{u_i} = \frac{\Delta y_{u_i}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}};$$

Цепной темп прироста равен: $T_{u_i} = K_{u_i} \cdot 100\%$. Он показывает, на сколько процентов уровень текущего периода выше или ниже предыдущего уровня.

- **Базисный коэффициент прироста** равен:

$$K'_{\sigma_i} = \frac{\Delta y_{\sigma_i}}{y_1} = \frac{y_i - y_1}{y_1};$$

а базисный темп прироста - $T'_{\sigma_i} = \frac{\Delta y_{\sigma_i}}{y_1} \cdot 100\%$. T'_{σ_i} показывает, на сколько процентов уровень текущего периода выше или ниже начального уровня ряда.

Между коэффициентом (темпом) роста и коэффициентом (темпом) прироста существует зависимость:

$$K'_i = K_i - 1 \text{ или } T'_i = T_i - 100\%.$$

Абсолютное значение одного процента прироста используется для оценки значения полученного темпа прироста. Он показывает, какое абсолютное значение соответствует одному проценту прироста. Показатель считается по цепным характеристикам:

$$A_i = \frac{\Delta y_{u_i}}{T_{u_i}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100} = \frac{y_{i-1}}{100}.$$

Пункты роста используется в тех случаях, когда сравнение производится с отдалением периода времени, принятого за базу. Они представляют собой разность базисных темпов роста двух смежных периодов:

$$P_i = T_{\sigma_i} - T_{\sigma_{i-1}} = \frac{y_i}{y_1} - \frac{y_{i-1}}{y_1} = \frac{y_i - y_{i-1}}{y_1} = \frac{\Delta y_{u_i}}{y_1}.$$

Пункты роста можно суммировать, в результате получаем базовый темп прироста последнего периода:

$$\sum_{i=2}^n P_i = T_{\sigma_n}.$$

Вторая часть системы характеристик динамического ряда состоит из обобщающих характеристик, к которым относятся его средние показатели и характеристики вариации уровней:

- *средний уровень ряда \bar{y}* ;
- *общий абсолютный прирост Δ* ;
- *средний абсолютный прирост $\bar{\Delta}$* ;
- *средний темп роста \bar{T} (\bar{K})*;
- *средний темп прироста \bar{T}' (\bar{K}')*;
- *дисперсия и среднее квадратическое отклонение уровней ряда σ_y^2 ,*

σ_y ;

- *коэффициент вариации уровней ряда V_y* .

Расчет среднего уровня ряда динамики определяется видом ряда и величиной интервала, соответствующего каждому уровню. **Средний уровень** характеризует наиболее типичную величину уровней, центр ряда.

В интервальных рядах с равноотстоящими интервалами средний уровень ряда определяется по формуле средней арифметической простой:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

В интервальных рядах с неравноотстоящими уровнями используется формула средней арифметической взвешенной:

$$\bar{y} = \frac{\sum_{i=1}^n y_i \cdot t_i}{\sum_{i=1}^n t_i};$$

t_i - длительность интервала.

Пример: известна динамика среднемесячной численности промышленно-производственного персонала предприятия: с 1 января по 1 мая она не менялась и составляла – 500 человек за каждый месяц, с 1 мая по 1 октября – 450 человек, с 1 октября по 31 декабря – 440 человек. Необходимо определить среднемесячную численность ППП по году.

Данный ряд относится к рядам с неравноотстоящими уровнями. Для расчета среднемесячной численности промышленно-производственного персонала предприятия используем формулу средней арифметической взвешенной:

$$\bar{y} = \frac{500 \cdot 4 + 450 \cdot 5 + 440 \cdot 3}{12} = \frac{5570}{12} \approx 464 \text{ чел.}$$

В моментных рядах при определении среднего уровня ряда используется формула средней хронологической:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}.$$

Например, имеются данные о валютном курсе шведской кроны на московской межбанковской валютной бирже:

Дата	13.12	15.12	16.12	17.12
Курс	4,13	4,14	4,15	4,15

Ряд динамики относится к моментным, поэтому средний курс рассчитаем по формуле средней хронологической:

$$\bar{y} = \frac{4,13/2 + 4,14 + 4,14 + 4,15 + 4,15/2}{5-1} = 4,142 \text{ руб.}$$

Средний абсолютный прирост является обобщающим показателем изменения явления во времени. Он **показывает, на сколько в среднем за единицу времени изменяется уровень ряда** и рассчитывается как простая средняя арифметическая из показателей абсолютных цепных приростов:

$$\bar{\Delta} = \frac{\sum_{i=2}^n \Delta y_{t_i}}{n-1} = \frac{\Delta y_{\bar{t}_n}}{n-1}.$$

Средний коэффициент роста (средний относительный прирост) *показывает, во сколько раз в среднем за единицу времени изменился уровень динамического ряда.* Эта характеристика имеет важное значение при выявлении и описании основной долговременной тенденции развития, используется в качестве обобщенного показателя интенсивности развития явления за длительный период времени.

Средний коэффициент роста вычисляется по формуле простой средней геометрической:

$$\bar{K} = \sqrt[n-1]{K_{t_2} \cdot K_{t_3} \cdot \dots \cdot K_{t_n}} = \sqrt[n-1]{K_{\bar{t}_n}};$$

Средний коэффициент прироста *характеризует среднюю относительную скорость изменения уровней в единицу времени.* Он определяется на основе среднего темпа роста:

$$\bar{K}' = \bar{K} - 1;$$

Средний коэффициент прироста показывает, на какую долю единицы в среднем изменяется уровень ряда за единичный промежуток времени.

Средний темп прироста показывает, на сколько процентов в среднем за единицу времени изменяется уровень ряда. Он рассчитывается на основе среднего темпа роста:

$$\bar{T}' = \bar{T} - 100\%$$

Дисперсия уровней динамического ряда σ_y^2 , среднее квадратическое отклонение σ_y и коэффициент вариации V_y используются для оценки уровня вариации уровней.

Дисперсия уровней динамического ряда рассчитывается по формуле:

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}.$$

Среднее квадратическое отклонение как абсолютный показатель колеблемости уровней ряда равно: $\delta_y = \sqrt{\sigma_y^2}$, а *коэффициент вариации* как относительный показатель уровней ряда -

$$V_y = \frac{\delta_y}{\bar{y}}.$$

6.3. Разложение рядов динамики

Уровни любого ряда динамики формируются под совместным влиянием факторов, различных как по характеру, так и силе воздействия. В первую очередь необходимо выделить **факторы эволюционного характера**, оказывающие постоянное воздействие и определяющие **общее направление развития явления**, его долговременную эволюцию. Такие изменения динамического ряда называют основной тенденцией развития или трендом.

Вторую группу факторов составляют **факторы осциллятивного характера**, оказывающие периодическое воздействие. Они вызывают *циклические и сезонные колебания* уровней динамического ряда.

Циклические (или периодические) долговременные колебания – это регулярные колебания, вызываемые постоянно действующими причинами, например, циклы экономической конъюнктуры, циклы гарвардской школы. Схематично циклические колебания можно представить в виде синусоиды $y_i = \sin t_i$ (значение признака вначале возрастает, достигает определенного max, затем снижается, достигает своего min, вновь возрастает и т.д.).

Сезонные колебания – колебания, периодически повторяющиеся в некоторое определенное время каждого года, в определенные дни каждого месяца или в определенные часы суток. Они могут вызываться природно-климатическими условиями, действием экономических, культурных и иных факторов.

Последней группой факторов, влияющих на ряд динамики являются **факторы, вызывающие нерегулярные колебания уровней**. Эти факторы подразделяются в свою очередь на:

- **вызывающие спорадические изменения уровней** (война, экологические катастрофы, эпидемии и т.д.),
- **случайные, слабо воздействующие, второстепенные факторы** вызывающие случайные разнонаправленные изменения уровней.

Таким образом, уровни ряда динамики подвержены разным воздействиям, и теоретически ряд динамики может быть представлен как функция следующих компонент:

$$y = f(T, R, S, E),$$

где T – тренд;

R – циклические колебания;

S - сезонные колебания;

E – случайные колебания.

Так как каждый фактор вызывает повышение или понижение уровней, то каждую компоненту и исходный динамический ряд можно представить в векторной форме:

$$\vec{y} = f(\vec{T}, \vec{R}, \vec{S}, \vec{E}).$$

В зависимости от связи компонент между собой можно построить две модели ряда динамики:

- **аддитивная модель:** $\vec{y} = \vec{T} + \vec{R} + \vec{S} + \vec{E}$ - характеризуется тем, что характер циклических и сезонных колебаний остается постоянными,
- **мультипликативная модель:** $\vec{y} = \vec{T} \cdot \vec{R} \cdot \vec{S} \cdot \vec{E}$ - если характер циклических и сезонных колебаний остается постоянным только по отношению к тренду.

6.4. Выявление тренда

Первая задача, которая возникает при анализе рядов динамики, заключается в выявлении и описании основной тенденции развития изучаемого явления (тренда).

Трендом называется плавное и устойчивое изменение уровней явления во времени, свободное от случайных колебаний.

Изучение тренда включает в себя два этапа:

1. Проверка ряда на наличие тренда;
2. Выравнивание ряда динамики и непосредственное выделение тренда.

Проверка ряда на наличие тренда проводится разными методами, самым простым из которых является **метод средних**. Суть его заключается в следующем: изучаемый ряд динамики разбивается на несколько интервалов (чаще всего на два), для каждого из которых определяется средняя величина - \bar{y}_1 и \bar{y}_2 . Выдвигается гипотеза о существенном различии средних. Если выдвинутая гипотеза принимается, то признается наличие тренда.

Для непосредственного выявления тренда используют следующие методы:

- **метод укрупнения интервалов;**
- **метод скользящей средней;**
- **метод аналитического выравнивания.**

Все перечисленные методы относятся к группе методов сглаживания, предполагающих наличие в исходном ряду динамики только одной компоненты – тренда.

Метод укрупнения интервалов является одним из наиболее простых методов непосредственного выявления основной тенденции. При использовании этого метода *ряд динамики, состоящий из мелких интервалов, заменяется рядом, состоящим из более крупных интервалов*. Так как на каждый уровень исходного ряда влияют факторы, вызывающие их разнонаправленное изменение, то это мешает видеть основную тенденцию. При укрупнении интервалов влияние факторов нивелируется, и основная тенденция проявляется более отчетливо. *Расчет среднего значения уровня по укрупненному интервалу осуществляется по формуле простой средней арифметической*.

Недостатком этого способа является то, что сокращается число уровней ряда, а это не позволяет учитывать изменения внутри укрупненного интервала. К его *преимуществам* можно отнести сохранение природы явления.

Применение метода укрупнения интервалов рассмотрим на примере данных об объеме выпуска продукции предприятия (таблица 6.3).

Таблица 6.3.
Объем выпуска продукции в 2003 году

t, мес.	Объем выпуска, млн. руб.	t, квартал	Итого, за квартал	\bar{y}
Январь	190	I	600	200
Февраль	210			
Март	200			
Апрель	220	II	690	230
Май	240			
Июнь	230			
Июль	220	III	720	240
Август	240			
Сентябрь	260			
Октябрь	260	IV	810	270
Ноябрь	280			
Декабрь	270			
Всего	2790		2790	-

Исходный ряд не показывает последовательного роста или снижения объемов выпуска. Изменение уровней не имеет общего направления, они то растут, то снижаются. Заменяем месячные интервалы квартальными, соответственно изменив и уровни показателя. Для этого рассчитаем среднемесячные уровни по данным кварталов. Новый ряд состоит из 4-х уровней, каждый из которых является среднемесячным объемом выпуска, рассчитанным по данным соответствующего квартала. В полученном ряду отчетливо просматривается последовательный рост объемов производства в течение года.

Метод скользящей средней предполагает замену исходный ряда теоретическим, уровни которого рассчитываются по формуле скользящей средней. Скользящая средняя относится к подвижным динамическим средним, вычисляемым по ряду при последовательном перемещении на один интервал. При этом, как и в предыдущем методе, происходит укрупнение интервалов. Число уровней, по которым укрупняется интервал, называется диапазоном укрупнения, интервалом или периодом сглаживания α . Период сглаживания может быть нечетным ($\alpha=3; 5$; и т.д.) и четным ($\alpha=2; 4$; и т.д.).

При нечетном периоде сглаживания полученное среднее значение уровня \hat{y}_i .

закрепляется за серединой расчетного интервала. При $\alpha=3$ формула имеет вид:

$$\hat{y}_i = \frac{y_{i-1} + y_i + y_{i+1}}{3}, \quad i = 2, n-1.$$

При четном периоде сглаживания возникает проблема центрирования, для решения которой необходимо осуществить сдвиг сглаженных уровней.

При использовании этого метода получают укороченный теоретический ряд, при этом при $\alpha=3$ ряд укорачивается на 2 уровня (крайних), при $\alpha=5$ соответственно - на 4 и т.д., а это приводит к потере информации.

Применение метода скользящей средней рассмотрим на данных предыдущего примера в таблице 6.4. Период скольжения принят равным $\alpha=3$ -м месяцам.

Таблица 6.4.

Расчет скользящей средней

t, мес.	Объем выпуска, y_i	Скользящая сумма $y_{i-1} + y_i + y_{i+1}$	Скользящая средняя \hat{y}_i
Январь	190	-	-
Февраль	210	600	200
Март	200	630	210
Апрель	220	660	220
Май	240	690	230
Июнь	230	690	230
Июль	220	690	230
Август	240	720	240
Сентябрь	260	760	253
Октябрь	260	800	257
Ноябрь	280	810	270
Декабрь	270	-	-
Всего	2790	-	

Последняя графа таблицы 5.4. показывает *последовательный рост объемов выпуска на протяжении отчетного года.*

Рассмотренные методы дают возможность определить общую тенденцию развития явления, освобожденную от случайных и волнообразных колебаний, но не позволяют получить количественного описания тренда исследуемого ряда. Для получения обобщенной статистической модели тренда применяют метод аналитического выравнивания.

6.5. Метод аналитического выравнивания

Основная тенденция развития рассчитывается как временная функция $\hat{y}_i = f(t_i)$, где \hat{y}_i - теоретические уровни (уровни динамического

ряда, вычисленные по соответствующему аналитическому уравнению на момент времени t_i) т.е. развитие явления рассматривается в зависимости только от течения времени. Отклонения эмпирических уровней ряда y_i от уровней, соответствующих общей тенденции \hat{y}_i объясняются действием случайных или циклических факторов. В результате получаем *трендовую модель* вида:

$$\hat{y}_i = f(t_i) + \varepsilon_i,$$

где ε_i - случайное и циклическое отклонение от тенденции.

Целью аналитического выравнивания динамического ряда является определение аналитической или графической зависимости $f(t_i)$. Функция $f(t_i)$ выбирается таким образом, чтобы она давала содержательное объяснение изучаемого процесса.

Подбор функции обычно осуществляется методом наименьших квадратов (МНК), в соответствии с которым наилучшим образом тренд описывает временная функция, обеспечивающая минимальную величину суммы квадратов отклонений эмпирических уровней ряда от соответствующих уровней теоретического ряда:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min,$$

где y_i - фактические уровни;

\hat{y}_i - выровненные по временной функции уровни ряда.

Наиболее часто в анализе рядов динамики при выравнивании используются следующие зависимости:

- *линейная* $\hat{y} = a + b \cdot t$;
- *параболическая* $\hat{y} = a + b \cdot t + c \cdot t^2$;
- *показательная функция* $\hat{y} = a \cdot b^t$.

Линейная зависимость $\hat{y} = a + b \cdot t$ выбирается в тех случаях, когда в исходном ряду наблюдаются в среднем *постоянные абсолютные цепные приросты* $\Delta_{ц_i} \approx \text{const}$.

Параметры уравнения a и b находятся по методу наименьших квадратов, в соответствии с которым получают систему нормальных уравнений:

$$\begin{cases} n \cdot a + b \sum t = \sum y, \\ a \cdot \sum t + b \sum t^2 = \sum yt \end{cases}$$

где y – фактические (эмпирические) уровни ряда;

t – хронологические показатели времени (порядковый номер периода или момента времени).

Для решения системы можно использовать любой известный метод, но предварительно необходимо решить проблему замены показателей

времени, что позволит значительно упростить расчет параметров. Хронологические показатели заменяются числовыми аналогами таким образом, чтобы сумма новых показателей времени по ряду $\sum_{i=1}^n t_i = 0$:

- при нечетном числе уровней (например, - 7) за начало отсчета времени ($t=0$) принимается центральный интервал (момент) – 4-ый:

1998г. 1999г. 2000г. 2001г. 2002г. 2003г. 2004г.
 -3 -2 -1 0 +1 +2

+3;

- при четном числе уровней (например, - 6) значения условных показателей времени будут выглядеть следующим образом:

1999г. . 2000г. 2001г. 2002г. 2003г. 2004г.
 -3 -2 -1 +1 +2 +3.

Применение условных показателей времени позволяет привести систему нормальных уравнений к виду:

$$\begin{cases} n \cdot a = \sum y \\ b \sum t^2 = \sum yt \end{cases}$$

Из первого уравнения $a = \frac{\sum y}{n}$.

Из второго уравнения $b = \frac{\sum yt}{\sum t^2}$.

Параметр a в линейной трендовой модели обычно интерпретации не имеет, но иногда его рассматривают как обобщенный начальный уровень ряда.

Параметр b в трендовом уравнении *называется коэффициентом регрессии*. Он определяет направление развития явления: при $b > 0$ – уровни ряда динамики равномерно возрастают, при $b < 0$ – равномерно снижаются. *Коэффициент регрессии показывает, насколько в среднем изменится уровень ряда при изменении времени на единицу*. Это означает, что параметр b можно рассматривать как средний абсолютный прирост с учетом тенденции к равномерному росту (росту в арифметической прогрессии).

Парабола второго порядка используется для описания рядов динамики, в которых меняется направление развития: со снижения показателей на их рост и наоборот.

Трендовое уравнение имеет вид: $\hat{y} = a + b \cdot t + c \cdot t^2$.

Параметр c называется коэффициентом регрессии и характеризует изменение интенсивности развития в единицу времени. При $c > 0$ наблюдается ускоренное развитие, при $c < 0$ – замедленное.

Система уравнений, полученная по МНК имеет вид:

$$\begin{cases} n \cdot a + b \cdot \sum t + c \cdot \sum t^2 = \sum y, \\ a \cdot \sum t + b \cdot \sum t^2 + c \cdot \sum t^3 = \sum yt, \\ a \cdot \sum t^2 + b \cdot \sum t^3 + c \cdot \sum t^4 = \sum yt^2. \end{cases}$$

Так как $\sum t_i = 0$, то система упрощается:

$$\begin{cases} n \cdot a + c \cdot \sum t^2 = \sum y, \\ b \cdot \sum t^2 = \sum yt, \\ a \cdot \sum t^2 + c \cdot \sum t^4 = \sum yt^2. \end{cases}$$

Отсюда:

$$b = \frac{\sum yt}{\sum t^2};$$

$$a = \frac{\sum t^4 \cdot \sum y - \sum t^2 \cdot \sum yt^2}{n \cdot \sum t^4 - (\sum t^2)^2};$$

$$c = \frac{n \cdot \sum yt^2 - \sum t^2 \cdot \sum y}{n \cdot \sum t^4 - (\sum t^2)^2}.$$

Показательная функция $\hat{y} = a \cdot b^t$ применяется для описания динамических рядов со стабильными ценными темпами роста: $T_{qi} = const$. Такие динамические ряды отражают развитие в геометрической прогрессии.

Параметр b называется коэффициентом регрессии, интерпретируется как средний темп роста изучаемого явления в единицу времени.

Для нахождения параметров модели функцию предварительно логарифмируют:

$$\ln y = \ln a + t \cdot \ln b.$$

Система нормальных уравнений для нахождения параметров трендового уравнения имеет вид:

$$\begin{cases} \sum \lg y = n \lg a + \lg b \cdot \sum t \\ \sum \lg y \cdot t = \lg a \cdot \sum t + \lg b \cdot \sum t^2 \end{cases}$$

На практике выбор формы кривой может быть основан на анализе графического изображения уровней ряда динамики (линейной диаграммы). При этом целесообразно использовать графическое изображение сглаженных уровней, в которых погашены случайные колебания.

Для оценки близости трендового уравнения эмпирическому ряду динамики применяется критерий Фишера (F). Фактический (расчетный) уровень F-критерия сравнивается с теоретическим (табличным) значением:

$$F_{\text{факт}}^{\text{(расч)}} = \frac{\eta_T^2}{1 - \eta_T^2} \cdot \frac{n - m}{m - 1} = \frac{V_1}{V_2},$$

где m - число параметров;

η_T^2 - теоретический коэффициент детерминации.

$$\eta_T^2 = 1 - \frac{\sigma_{y-\hat{y}}^2}{\sigma_y^2},$$

$\sigma_{y-\hat{y}}^2$ - остаточная дисперсия,

σ_y^2 - общая дисперсия.

Остаточная дисперсия рассчитывается по

формуле: $\sigma_{y-\hat{y}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n};$

общая дисперсия - $\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}.$

Для признания модели надежной необходимо соблюдение условия: $F_{\text{факт}} > F_{\text{крит}}.$

$F_{\text{крит}}$ подбирается по специальным таблицам распределения Фишера при $v_1 = m - 1$, $v_2 = n - m$ и заданном уровне значимости α .

Для динамических рядов, имеющих небольшую длину и подверженных значительным колебаниям, использовать метод аналитического выравнивания с помощью временной функции не рекомендуется, так как аппроксимация практически не адаптируется к изменяющимся условиям формирования уровней, при появлении новых данных нужно строить новые модели.

Для сглаживания таких рядов динамики используются методы адаптивного моделирования и прогнозирования. В основе указанных методов лежит модель экспоненциального сглаживания. Временной ряд сглаживается с помощью взвешенной скользящей средней, в которой веса распределяются по экспоненциальному закону.

В качестве примера применения метода аналитического выравнивания рассмотрим ряд динамики, приведенный в таблице 6.5.

Таблица 6.5.

Выравнивание ряда динамики выпуска продукции

Показатель времени и t_i	Условный показатель времени t_i	y_i млн руб.	Расчет параметров			Расчет F-критерия			
			t^2	$y \cdot t$	\hat{y}	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
Январь	-6	190	36	-1140	196,0	-6	36	-45,0	2025
Февраль	-5	210	25	-1050	202,5	-7,5	56,25	-25,0	625
Март	-4	200	16	-800	209,0	-9	81,0	-35,0	1225
Апрель	-3	220	9	-660	215,5	4,5	20,25	-15,0	225
Май	-2	240	4	-480	222,0	18	324	5,0	25
Июнь	-1	230	1	-230	228,5	1,5	2,25	-5,0	25
Июль	+1	220	1	220	241,5	-21,5	462,25	-15,0	225
Август	+2	240	4	480	248,0	-8	640	5,0	25
Сентябрь	+3	260	9	780	254,5	5,5	30,25	25,0	625
Октябрь	+4	260	16	1040	261,0	-1,0	1,0	25,0	625
Ноябрь	+5	280	25	1400	267,0	12,5	156,25	45,0	2025
Декабрь	+6	270	36	1620	274,0	-4,0	16,0	35,0	1225
Всего	0	2820	182	1180	2820	-	1249,5	-	8900

Выравнивание проводится по линейной модели $\hat{y} = a + b \cdot t$. Оценка параметров уравнения регрессии выполнена методом наименьших квадратов:

$$a = \frac{\sum y}{n} = \frac{2820}{12} = 235,0 \text{ млн. руб.};$$

$$b = \frac{\sum y^t}{\sum t^2} = \frac{1180}{182} = 6,48 \approx 6,5 \text{ млн. руб.};$$

Трендовое уравнение имеет вид:

$$\hat{y} = 235,0 + 6,5t.$$

Для оценки надежности модели определим расчетное значение F-критерия. Для этого предварительно рассчитаем на основе данных таблицы 5.5:

среднее значение уровней ряда $\bar{y} = 232,5$ млн. руб.;

$$\text{остаточную дисперсию } \sigma_{y-\hat{y}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} = \frac{1249,5}{12} \approx 104,1;$$

$$\text{общую дисперсию уровней ряда } \sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{8900}{12} \approx 741,7;$$

$$\text{коэффициент детерминации } \eta^2 = 1 - \frac{104,1}{741,7} = 1 - 0,14 = 0,86.$$

$$F_{\text{расч}} = \frac{0,86}{1 - 0,86^2} \cdot \frac{12 - 2}{2 - 1} = \frac{0,74}{0,26} \cdot 10 = 2,84 \cdot 10 = 28,4.$$

$$F_{\text{крит}} = 19,39;$$

$$\alpha=0,95 \\ V_1=10, V_2=1$$

$F_{\text{факт}} > F_{\text{крит}}$ - уравнение прямой адекватно отражает сложившуюся в исследуемом ряду динамики основную тенденцию.

Параметры модели можно интерпретировать следующим образом:

Коэффициент регрессии $b=6,5$ – показатель силы связи, означающий, что ежемесячно объем выпуска продукции возрастал на 6,5 млн. рублей.

6.6. Анализ сезонных колебаний

Сезонными называют периодические колебания, возникающие под влиянием смены времени года и других причин природного или социально-культурного порядка. Они имеют устойчивый характер, повторяются регулярно с интервалом в один год.

Их роль велика в агропромышленном комплексе, строительстве, транспорте, здравоохранении, торговле и т.д. При этом сезонные колебания в одних отраслях экономики вызывает соответствующие колебания в других. Таким образом, проблема сезонности носит общий характер для экономики страны. Как правило, сезонность отрицательно влияет на результаты работы, поскольку приводит к неравномерному использованию рабочей силы, производственных мощностей, материальных ресурсов. Поэтому хозяйственные организации принимают

меры для смягчения сезонности или стараются учитывать её влияние на свою деятельность.

Для выявления и измерения сезонных колебаний используются различные статистические методы, такие как, например, *построение модели сезонной волны и гармонический анализ*.

Метод построения «сезонной волны» заключается в расчете специальных показателей, которые называются **индексами сезонности** I_s^i .

Совокупность индексов сезонности отражают сезонную волну.

Индексами сезонности называется процентные отношения фактических (эмпирических) внутригрупповых уровней к теоретическим уровням, рассчитанным по трендовому уравнению, либо к средним уровням.

Для выявления устойчивой сезонной волны, на которой не отражаются случайные условия одного года, индексы сезонности рассчитываются за период не менее чем 3 года распределенный по месяцам или кварталам.

Расчет индексов сезонности выполняют двумя методами в зависимости от характера динамики:

- **если тренд неявно выражен**, то есть годовой уровень явления из года в год остается относительно неизменным, то *индексы сезонности рассчитываются методом постоянной средней*. Они рассчитываются по формуле:

$$\bar{I}_s^i = \frac{\bar{y}_s^i}{\bar{y}} \cdot 100\%$$

где i – номер одноименного периода (сезона);

\bar{y}_s^i - средняя из фактических уровней одноименных периодов (месяцев или кварталов), вычисляется по формуле:

$$\bar{y}_s^i = \frac{\sum_{t=1}^n y_s^i}{n};$$

y_s^i - фактический уровень одноименного периода;

\bar{y} - средний уровень ряда за исследуемый период.

Индексы сезонности рассчитываются в такой последовательности:

- Рассчитываются средние уровни для каждого одноименного периода по данным за все - годы наблюдения \bar{y}_s^i .

- Определяется общая средняя \bar{y} за весь период наблюдения.

- Вычисляется индекс сезонности по приведенной выше формуле.

- **Если тренд явно выражен**, то для исчисления индексов сезонности *используется метод переменной средней*, в соответствии с которым их расчет проводится по формуле:

$$\bar{I}_s^i = \frac{\sum_{i=1}^n i_s^i}{n} \cdot 100\% ;$$

где $i_s^i = \frac{y_i}{\hat{y}_i} \cdot 100\%$ - индивидуальный индекс сезонности одноименных периодов,
 n – число лет наблюдения.

Совокупность средних индексов сезонности одноименных периодов составляет модель сезонной волны.

Если при построении модели сезонной волны случайные колебания гасятся полностью, то сумма средних индексов сезонности одноименных периодов = 1200%, если уровни брались за месяц, и 400%, если уровни были квартальными. Если это условие не выполняется, то проводится корректировка модели. Для этого рассчитывается поправочный коэффициент:

$$П = \frac{1200(400)}{\sum \bar{I}_s^i} .$$

На величину этого коэффициента корректируются все рассчитанные средние индексы сезонности $\bar{Y}_{s,кор}^i = \bar{Y}_s^i \cdot П$.

Построение модели сезонности рассмотрим на следующем примере: известна динамика реализации продовольственных товаров в магазинах города за 2001- 2004 гг., (данные приведены в таблице 5.6). Необходимо выявить и измерить сезонные колебания.

Для рассматриваемого примера n=4; число одноименных периодов (кварталов) i =4, число элементов ряда – 16.

Чтобы выявить наличие тенденции в изучаемом ряду, рассчитаем его индивидуальные характеристики (по году): годовой абсолютный прирост, темп роста и темп прироста (таблица 5.5.). Расчеты показывают наличие явно выраженного тренда в виде ежегодного устойчивого роста объемов реализации. Это означает, что средние индексы сезонности следует рассчитывать способом переменной средней:

$$\bar{I}_s^i = \frac{\sum_{i=1}^n i_s^i}{n} \cdot 100\% .$$

Для определения теоретических уровней тренда \hat{y}_i используем прямолинейную функцию $\hat{y} = a + b \cdot t$. О возможности ее применения говорит графическое представление ряда - линейная диаграмма (рис 6.1).

Таблица 6.6.

Среднедневная реализация продовольственных товаров
В 2001 - 2004 гг., млн. руб.

Кварталы	Годы			
	2001	2002	2003	2004
I	40	42	43	49
II	64	70	60	75
III	62	72	80	90
IV	50	44	53	64
Среднеквартальная реализация	54,0	57,0	59	69,5
Годовой абсолютный прирост	-	+3	+2	+10,5
Темп роста в % к 2001 г.	-	105,6	103,5	116,1
Темп прироста в % к 2001 г.	-	5,6	3,5	16,1

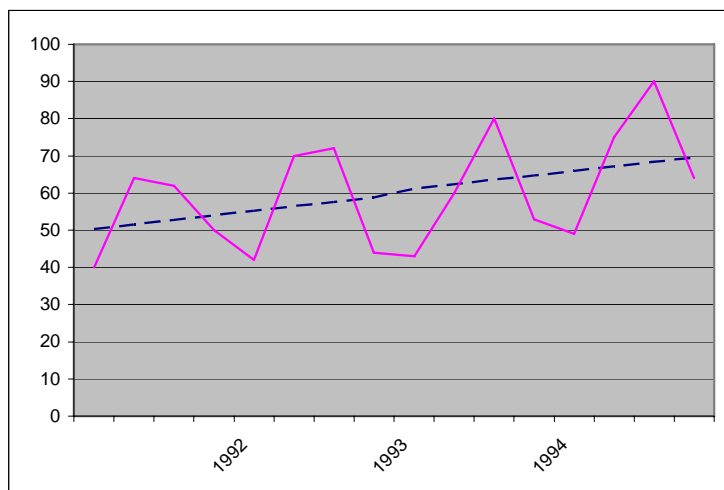


Рис. 6.1. Динамика среднедневной реализации продовольственных товаров

Для определения теоретических уровней тренда \hat{y}_t используем прямолинейную функцию $\hat{y} = a + b \cdot t$.

Расчет параметров линейного уравнения регрессии по методу наименьших квадратов дает следующие их значения:

$$a = 59,9, \quad b = 1,2.$$

Трендовая модель имеет вид: $\hat{y} = 59,9 + 1,2t$.

Расчет индивидуальных индексов сезонности проведем в таблице 6.7.

Таблица 6.7.

Расчет индивидуальных индексов сезонности

Год, квартал	y_i	\hat{y}_i	$\frac{y_i}{\hat{y}_i} \cdot 100$	Год, квартал	y_i	\hat{y}_i	$\frac{y_i}{\hat{y}_i} \cdot 100$
2001, I	40	50,3	79,5	2003, I	43	61,1	70,4
II	64	51,5	124,3	II	60	62,3	96,3
III	62	52,7	117,6	III	80	63,5	126,0
IV	50	53,9	92,8	IV	53	64,7	81,9
2002, I	42	55,2	76,1	2004, I	49	65,9	74,4
II	70	56,4	124,1	II	75	67,1	111,8
III	72	57,5	125,2	III	90	68,3	131,8
IV	44	58,7	75,0	IV	64	69,5	92,1

Для устранения воздействия случайных факторов проведем усреднение индивидуальных индексов сезонности по кварталам. Используем формулу переменной средней.

$$\text{По I кварталу } \bar{I}_s^I = \frac{79,5 + 76,1 + 70,4 + 74,4}{4} = 75,1\% .$$

$$\text{По II кварталу } \bar{I}_s^{II} = \frac{124,3 + 124,1 + 96,3 + 111,8}{4} = 114,1\% .$$

$$\text{По III кварталу } \bar{I}_s^{III} = \frac{117,6 + 125,2 + 126,0 + 131,8}{4} = 125,2\% .$$

$$\text{По IV кварталу } \bar{I}_s^{IV} = \frac{92,8 + 75,0 + 81,9 + 92,1}{4} = 85,5\% .$$

$$\text{Рассчитаем поправочный коэффициент } \Pi = \frac{400}{71,5 + 114,1 + 125,2 + 85,5} \approx 1 .$$

Вычисленные и откорректированные средние индексы сезонности составляют модель сезонности волны реализации продовольственных товаров во внутригодовом цикле. Модель отражает квартальные колебания уровней.

Наибольшие объемы реализации ежегодно приходятся на II и III кварталы, снижение объемов – в I и IV квартале. Графически модель представлена на рис. 6.2.

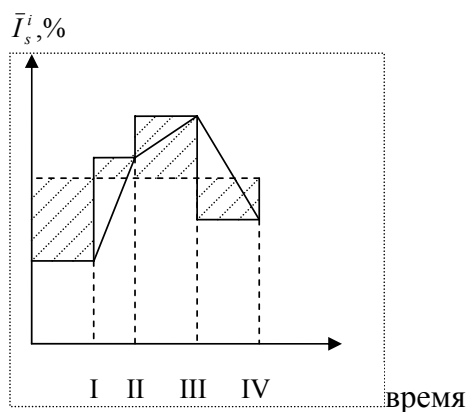


Рис. 6.2. Графическое представление модели сезонной волны

Кроме указанного способа для выявления сезонных колебаний можно использовать метод скользящих средних, гармонический анализ.

При применении гармонического анализа ряд динамики представляется как совокупность колебательных процессов, описываемых с помощью гармонического ряда Фурье.

Модель сезонных колебаний на основе гармоник Фурье имеет вид:

$$\hat{y}_i = a_0 + \sum_{i=1}^m (a_i \cdot \cos kt_i + b \cdot \sin kt_i),$$

k – номер гармоники, определяющий степень адекватности модели ($k = 1 \div 4$),

a_0, a, b – параметры уравнения, определяются по МНК:

$$a_0 = \frac{\sum y}{n}; \quad a_1 = \frac{2}{n} \cdot \sum y \cos kt; \quad b = \frac{2}{n} \cdot \sum y \sin kt.$$

При $k=1$ модель принимает вид: $\hat{y}_i = a_0 + a_1 \cdot \cos kt + b \cdot \sin kt$.

При изучении внутригодовой динамики n равным 12, а показатели времени нужно перевести в условные, как части окружности. Для этого можно использовать данные таблицы 6.8.

Таблица 6.8.

Перевод хронологических показателей времени в условные

t_i	1	2	3	4	5	6	7	8	9	10	11	12
$t_i^{усл}$	0	$1/6\pi$	$1/3\pi$	$1/2\pi$	$2/3\pi$	$5/6\pi$	π	$7/6\pi$	$4/3\pi$	$3/2\pi$	$5/3\pi$	$11/6\pi$

6.7. Экстраполяция в рядах динамики и прогнозирование

Полученные при анализе динамических рядов характеристики используются для получения **статистических прогнозов**, под которыми

понимаются **статистические оценки состояния явления в будущих периодах.**

Статистическое прогнозирование основано на предположении, что закономерность развития, основная тенденция, действующая в прошлом (внутри ряда динамики), сохранится и в будущем. Такое предположение называется **экстраполяцией**. Теоретической основой распространения тенденции на будущее является инерционность социально-экономических явлений.

Следует иметь в виду, что экстраполяция в рядах динамики носит приближенный характер. Точность прогноза зависит от сроков прогнозирования: чем они короче, тем надежнее результат экстраполяции, так как за короткий период времени не успевают значительно измениться условия развития явления и характер его динамики. Обычно рекомендуется, чтобы срок прогноза не превышал 1/3 длительности базы расчета тренда.

С помощью метода экстраполяции получают два вида прогноза: *точечные и интервальные.* **Точечный прогноз** представляет собой конкретное численное значение уровня в прогнозируемый период (момент) времени. **Интервальный прогноз** – диапазон численных значений, предположительно содержащий прогнозируемое значение уровня.

В зависимости от того, какие принципы и исходные данные положены в основу прогноза, выделяют следующие **методы экстраполяции (прогнозирования):**

- на основе среднего абсолютного прироста $\bar{\Delta}$,
- на основе среднего коэффициента роста \bar{K} ,
- на основе аналитического выравнивания ряда.

Метод прогнозирования на основе среднего абсолютного прироста $\bar{\Delta}$ применяется в том случае, если уровни изменяются равномерно (линейно).

Прогнозируемое значение уровня определяется по формуле:

$$\hat{y}_{n+l} = y_n + \bar{\Delta} \cdot l;$$

где y_{n+l} - экстраполируемый уровень;

y_n - конечный уровень ряда динамики;

l - период упреждения прогноза (срок экстраполяции).

Прогнозирование по среднему коэффициенту роста \bar{K} применяется, если общая тенденция характеризуется экспоненциальной кривой. В этом случае экстраполируемый уровень определяется по формуле:

$$\hat{y}_{n+l} = y_n \cdot (\bar{K})^l.$$

Прогнозирование на основе аналитического выравнивания является наиболее распространенным методом прогнозирования. Для

получения прогноза используется аналитическое выражение тренда. Чтобы получить прогноз, достаточно в модели продолжить значение условного показателя времени t_i до t_{n+l} .

Интервальные прогнозы имеют значительные преимущества перед точечными – они учитывают вероятность свершения прогноза.

Величина доверительного интервала определяется в общем виде так:

$$\hat{y}_{n+l} \pm t_{\alpha} \cdot \sigma_{y_i - \hat{y}_i},$$

где t_{α} - коэффициент доверия по распределению Стьюдента;

$\sigma_{y_i - \hat{y}_i}$ - средняя квадратическая ошибка тренда, рассчитываемая по формуле:

$$\sigma_{y_i - \hat{y}_i} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m}};$$

n – число уровней исходного ряда,

m – число параметров трендового уравнения.

Коэффициент доверия t_{α} выбирается по таблице распределения Стьюдента

Таким образом, при использовании интервального прогноза:

$$\hat{y}_{n+l} - t_{\alpha} \cdot \sigma_{y_i - \hat{y}_i} \leq \hat{y}_{n+l} \leq \hat{y}_{n+l} + t_{\alpha} \cdot \sigma_{y_i - \hat{y}_i}.$$

Для ряда динамики объемов выпуска продукции:

$$l = 2; \quad \hat{y} = 235,0 + 6,5t; \quad \sigma_{y_i - \hat{y}_i} = \sqrt{111} = 10,5;$$

$$y_{12\phi} = 280; \quad t = +8; \quad \hat{y}_{02} = 235,0 + 6,5 \cdot 8 = 287 \text{ млн. руб.}$$

$$t_{\alpha=0,05} = 2;$$

Интервальный прогноз: $287 \pm 2,0 \cdot 10,5 = 287 \pm 21,0$.

С вероятностью в 95% следует, что $266 \leq \hat{y}_{14} \leq 308$.

7. Статистическое изучение взаимосвязей социально-экономических явлений

7.1. Понятие статистической и корреляционной связи

Одной из важнейших задач статистики является изучение объективно существующих связей между явлениями. При исследовании таких связей выясняются причинно-следственные отношения между явлениями, а это, в свою очередь, позволяет выявить факторы, оказывающие основное влияние на вариацию изучаемых явлений и процессов. *Причинно-следственные отношения представляют собой такую связь явлений, при которой изменение одного из них – причины, ведёт к изменению другого – следствия.* Причинно-следственная форма связи определяет все другие формы, носит всеобщий и многообразный характер.

Для описания причинно-следственной связи между явлениями и процессами используется деление статистических признаков, отражающих отдельные стороны взаимосвязанных явлений, на факторные и результативные. *Факторными считаются признаки, обуславливающие изменение других, связанных с ними признаков, являющихся причинами и условиями таких изменений. Результативными являются признаки, изменяющимися под воздействием факторных.*

Формы проявления существующих взаимосвязей весьма разнообразны. В качестве самых общих их видов выделяют функциональную и статистическую связи.

Функциональной называют такую связь, при которой определённому значению факторного признака соответствует одно и только одно значение результативного. Такая связь возможна при условии, что на поведение одного признака (результативного) влияет только второй признак (факторный) и никакие другие.

Такие связи являются абстракциями, в реальной жизни они встречаются редко, но находят широкое применение в точных науках и в первую очередь, в математике. Например: зависимость площади круга от радиуса: $S = \pi \cdot r^2$

Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы изучаемой совокупности.

В массовых явлениях проявляются *статистические связи, при которых строго определённому значению факторного признака ставится в соответствие множество значений результативного.* Такие связи имеют место, если на результативный признак действуют несколько факторных, а для описания связи используется один или несколько определяющих (учтённых) факторов.

Строгое различие между функциональной и статистической связью можно получить при их математической формулировке.

Функциональную связь можно представить уравнением:

$$y_i = f(x_i),$$

где y_i - результирующий признак ($i=1, \dots, n$);

$f(x_i)$ - функция связи результирующего и факторного признаков.

x_i - факторный признак

Статистическая связь может быть представлена уравнением следующего вида:

$$\tilde{y}_i = f(x_i) + \varepsilon_i,$$

где \tilde{y}_i - расчётное значение результирующего признака.

$f(x_i)$ - часть значения результирующего признака, сформировавшаяся под воздействием учтённых факторов.

ε_i - часть значения результирующего признака, возникающая вследствие действия неконтролируемых факторов или ошибок измерения.

Примером статистической связи может служить зависимость себестоимости единицы продукции от уровня производительности труда: чем выше производительность труда, тем ниже себестоимость. Но на себестоимость единицы продукции помимо производительности труда влияют и другие факторы: стоимость сырья, материалов, топлива, общепроизводственные и общехозяйственные расходы и т.д. Поэтому нельзя утверждать, что изменение производительности труда на 5% (повышение) приведет к аналогичному снижению себестоимости. Может наблюдаться и обратная картина, если на себестоимость будут влиять в большей степени другие факторы, - например, резко возрастут цены на сырье и материалы.

Любую статистическую связь можно представить в виде набора локальных распределений результирующего признака при фиксированных значениях факторного:

$$x_1: \quad y_{1,1}, y_{1,2} \dots y_{1,j} \dots y_{1,m}$$

$$x_2: \quad y_{2,1}, y_{2,2} \dots y_{2,j} \dots y_{2,m}$$

.....

$$x_n: \quad y_{n,1}, y_{n,2} \dots y_{n,j} \dots y_{n,m}$$

где $i = \overline{1, n}$, $j = \overline{1, m}$.

Каждое локальное распределение результирующего признака можно описать на эмпирическом уровне, рассчитав такие его характеристики как локальная средняя результирующего признака \tilde{y}_i , характеризующая положение центра распределения, и среднеквадратическое отклонение результирующего признака σ_i^2 , характеризующее форму локального распределения.

Если при изменении значений факторного признака x_i будут смещаться центры локальных распределений (меняться значение локальных средних \tilde{y}_i), но не будет меняться форма локальных распределений (значения внутригрупповых средних квадратических отклонений), то можно говорить о наличии между признаками **корреляционной связи**.

Корреляционная связь является частным случаем статистической связи. При корреляционной связи с изменением значения факторного признака x_i закономерно изменяется среднее значение результативного признака \tilde{y}_i , в то время как в каждом отдельном случае факторный признак может принимать множество различных значений.

Корреляционная связь может быть представлена уравнением:
 $\tilde{y}_i = F(x_i)$,

где $F(x_i)$ – функция связи среднего значения результативного признака с факторным.

Корреляционная связь проявляется только на всей статистической совокупности, а не в каждом отдельном случае, так как только при достаточно большом числе случаев каждому случайному значению факторного признака будет соответствовать распределение средних значений случайного признака y .

По направлению корреляционные связи делятся на прямые и обратные. При прямой связи результативный признак растёт с увеличением факторного, при обратной – рост факторного признака приводит к снижению значений результативного признака. Например, чем больше стаж работы, тем выше производительность труда – прямая связь, а чем выше производительность труда, тем ниже себестоимость единицы продукции – обратная связь.

По форме (аналитическому выражению) **связи делятся на линейные** (прямолинейные) **и нелинейные** (криволинейные) связи. Линейные связи выражаются уравнением прямой, а нелинейные – уравнением параболы, гиперболы, степенной и т. п.

По количеству взаимодействующих факторов связи делятся на **парную** (однофакторную) и **множественную** (многофакторную) связи. При парной связи на результативный признак действует один факторный, при множественной – несколько факторных признаков.

Исследование статистической связи проводится в следующем порядке:

- качественный анализ связи - определение состава признаков, предварительный анализ формы связи;
- сбор данных на основе статистического наблюдения;
- количественная оценка тесноты связи по эмпирическим данным;
- регрессионный анализ (аналитическое описание связи):

Формат: Список

- выбор формы связи,
- оценка параметров модели,
- оценка качества модели.

Основным методом изучения статистической взаимосвязи является статистическое моделирование связи на основе корреляционного и регрессионного анализа.

Задачей корреляционного анализа является количественное определение тесноты связи между двумя признаками при парной связи или между результативным и несколькими факторными при множественной связи.

Регрессионный анализ заключается в определении аналитического выражения связи в виде уравнения регрессии. **Регрессией** называется зависимость среднего значения случайной величины результативного признака от величины факторного, а **уравнением регрессии** – уравнение описывающее корреляционную зависимость между результативным признаком и одним или несколькими факторными.

7.2. Парная корреляция

Наиболее полно в статистике разработана методология парной корреляции, рассматривающей влияние вариации одного факторного признака на результативный.

Исследование парной корреляции осуществляется на основе корреляционного анализа, который предполагает последовательное решение ряда задач:

- Выявление связи;
- Описание связи в табличной и графической формах;
- Измерение тесноты связи;
- Формулировка выводов о характере существующей связи.

← Формат: Список

7.2.1. Задача выявления связи между факторным и результативным признаками может быть решена при помощи следующих приёмов:

- *визуализации связи* (построение и визуальный анализ корреляционного поля);
- *использования результатов аналитической группировки* и др.

Корреляционное поле представляет собой точечный график в системе координат $\{x,y\}$. Каждая точка соответствует единице совокупности. Положение точек на графике определяется величиной двух признаков – факторного и результативного. Точки корреляционного поля могут располагаться на графике хаотично, без всякой закономерности - тогда делается вывод об отсутствии связи между признаками (рис.1); или

определённым образом вдоль некоторой гипотетической линии (рис.2) – тогда делается вывод о существовании связи между признаками.

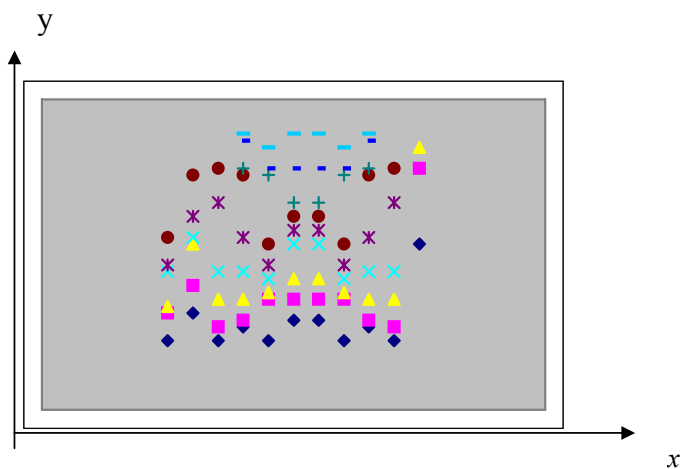


Рис. 7.1. Корреляционное поле при отсутствии связи между признаками

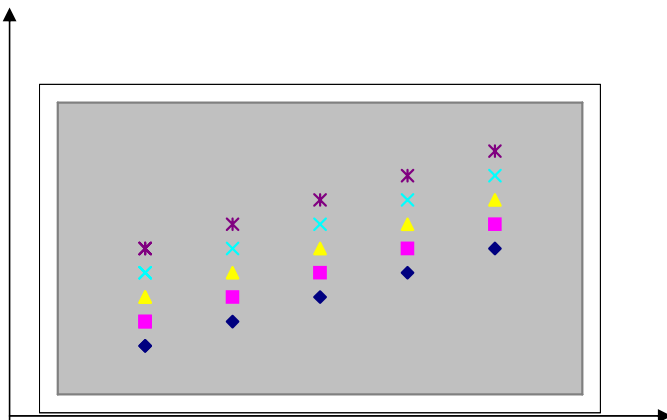


Рис.7.2. Корреляционное поле при наличии связи между признаками

При втором способе – *использовании результатов аналитической группировки* связь считается установленной, если группировка показывает изменение среднего значения результативного признака в группах при изменении факторного признака (основания группировки).

7.2.2. Описание выявленной связи при проведении корреляционного анализа *проводится в двух формах – табличной и графической.*

При табличном описании связи статистические единицы *группируются по значению факторного признака (располагаются в порядке его возрастания или убывания):*

x_i			
\tilde{y}_i			

Графическое описание связи заключается в построении **линии эмпирической регрессии** – ломаной линии, соединяющей на корреляционном поле точки, абсциссами которых являются значения факторного признака (индивидуальные значения или групповые значения), а ординатами – средние значения результативного признака.

Эмпирическая линия регрессии отражает основную тенденцию рассматриваемой зависимости. Если по своему виду она приближается к прямой линии, то можно предположить наличие прямолинейной связи между признаками.

7.2.3. Теснота связи показывает меру влияния факторного признака на общую вариацию результативного признака.

Для описания корреляционной связи используется зависимость $\tilde{y} = F(x)$, которая проявляется только на всей статистической совокупности. Так как на результат всегда действует множество факторов, то для каждой отдельной единицы наблюдения значение результативного признака состоит из двух частей:

$$y_i = \tilde{y}_i + \varepsilon_i,$$

где \tilde{y}_i – локальная средняя, характеризующая значение результативного признака, сформированное под воздействием только данного фактора x_i ;

$\varepsilon_i = (y_i - \tilde{y}_i)$ – отклонение, характеризующее вариацию результативного признака под влиянием неучтённых факторов.

Таким образом, **теснота связи** – это характеристика соотношения между локальной средней \tilde{y}_i и отклонением ε_i . Через тесноту связи определяется, в какой степени влияют на результат учтённые и неучтённые факторы.

На эмпирическом уровне, при проведении корреляционного анализа теснота связи измеряется с помощью интегральных показателей, построенных на правиле сложения дисперсии. В соответствии с ним

общая дисперсия результативного признака разлагается на внутригрупповую и межгрупповую:

$$\sigma_y^2 = \overline{\sigma_i^2} + \delta^2,$$

где $\overline{\sigma_i^2}$ - средняя из внутригрупповых дисперсий;

δ^2 - межгрупповая дисперсия.

Через соотношение дисперсий определяются показатели, измеряющие степень тесноты связи между результативными и факторными признаками: коэффициент детерминации η^2 и эмпирическое корреляционное отношение η .

- **Коэффициент детерминации** рассчитывается по формуле:

$$\eta^2 = \frac{\delta^2}{\sigma_y^2} = 1 - \frac{\overline{\sigma_i^2}}{\sigma_y^2}.$$

Приведенное отношение определяет удельный вес вариации, объясняемой влиянием учтенного фактора на результат, в общей вариации результативного признака. Показатель изменяется в диапазоне от 0 до 1. При $\eta^2 = 0$ межгрупповая дисперсия $\delta^2 = 0$, - это означает, что локальные средние во всех распределениях результативного признака строго одинаковы, центры распределений не смещаются; **связь между признаками отсутствует**. При $\eta^2 = 1$ межгрупповая дисперсия δ^2 равна общей дисперсии результативного признака $\delta^2 = \sigma_y^2$; следовательно, $\overline{\sigma_i^2} = 0$, а внутригрупповые значения результативного признака не варьируют, то есть $y_i = \tilde{y}_i$. Это означает, что на значения результативного признака влияют только учтенные факторы, и **связь между признаками является функциональной**: значению факторного признака соответствует единственное значение результативного.

Коэффициент детерминации сложно интерпретируется, поэтому на его основе рассчитывается ещё один показатель тесноты связи – эмпирическое корреляционное отношение η .

• **Эмпирическое корреляционное отношение** рассчитывается по формуле:

$$\eta = \sqrt{\eta^2} = \sqrt{1 - \frac{\overline{\sigma_i^2}}{\sigma_y^2}}.$$

Диапазон изменения этого показателя: $\eta = \{0 \div 1\}$. Нулевое значение эмпирического корреляционного отношения означает отсутствие связи между результативным и факторным признаками, при $\eta = |1|$ связь классифицируется как функциональная.

Формат: Список

Используя численное значение эмпирического корреляционного отношения η , связь можно классифицировать по шкале Чеддока, таблица 7.1.:

Таблица 7.1.

Шкала Чеддока

η	0 ÷ 0,1	0,11 ÷ 0,3	0,31 ÷ 0,5	0,51 ÷ 0,7	0,71 ÷ 0,9	0,91 ÷ 0,99	0,991 ÷ 1
Характеристика связи	отсутствует	Слабая	умеренная	заметная	тесная	сильная	функциональная

Если известно, что между результативным и факторным признаком существует линейная связь, то для оценки её тесноты используется **линейный коэффициент корреляции** $r_{y,x}$, рассчитываемый по формуле:

$$r_{y,x} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{[\sum x^2 - \frac{(\sum x)^2}{n}] \cdot [\sum y^2 - \frac{(\sum y)^2}{n}]}} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sigma_x \cdot \sigma_y}.$$

Значения линейного коэффициента корреляции важно для исследований, в которых распределение признака близко к нормальному. Он принимает значение в интервале $-1 \leq r_{y,x} \leq +1$. Отрицательные значения $r_{y,x}$ свидетельствуют о наличии обратной связи между признаками, положительные – о прямой связи. При $r_{y,x} = 0$ связь между признаками отсутствует. Для классификации связи по значению линейного коэффициента корреляции используется шкала Чеддока.

7.2.4. Выводы по результатам корреляционного анализа включают в себя констатацию факта наличия связи, определение её направления, предварительную оценку формы связи по линии эмпирической регрессии и классификацию связи по степени её тесноты.

7.3. Парная регрессия на основе метода наименьших квадратов

Парная регрессия характеризует связь между двумя признаками: факторным и результативным.

Задача построения уравнения регрессии для одного факторного и одного результативного признака формулируется следующим образом:

Пусть имеется набор значений двух переменных: результативного признака y_i и факторного признака x_i . Между этими переменными существует объективная связь вида: $y_i = f(x_i) + \varepsilon_i$. Необходимо по данным наблюдения $(y_i, x_i, i=1, n)$ подобрать функцию $\hat{y} = F(x)$, наилучшим образом описывающую существующую связь.

При подборе функции последовательно решаются две задачи:

• Определяется вид функциональной зависимости, то есть проводится спецификация модели.

Формат: Список

• Рассчитываются значения параметров уравнения регрессии.

В парной регрессии выбор вида математической функции может быть осуществлён разными методами:

- аналитическим, исходя из материальной природы связи;
- графическим, на основе линии эмпирической регрессии;
- на основе показателей качества уравнения регрессии.

Показателем качества уравнения регрессии является величина остаточной дисперсии:

$$\sigma_{y-\hat{y}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}.$$

Этот показатель рассчитывается для уравнений регрессии, построенных по разным математическим функциям. Лучшим по качеству является уравнение, для которого $\sigma_{y-\hat{y}}^2 \rightarrow \min$.

При построении уравнений парной регрессии чаще всего используют следующие уравнения:

1. прямой $\hat{y} = a + bx$,
2. параболы второго порядка $\hat{y} = a + bx + cx^2$,
3. гиперболы $\hat{y} = a + \frac{b}{x}$,
4. степенной $\hat{y} = a \cdot x^b$,
5. показательной $\hat{y} = a \cdot b^x$,
6. логистической кривой $\hat{y} = \frac{a}{1 + bc^{-cx}}$ и т.д.

Оценка параметров уравнений регрессии может быть проведена разными методами.

Классический подход к оцениванию параметров основан на методе наименьших квадратов (МНК).

Метод наименьших квадратов позволяет получить такие оценки параметров уравнения регрессии, которые минимизируют функционал вида:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min ;$$

Применение метода наименьших квадратов для расчёта параметров уравнения регрессии рассмотрим на примере прямолинейной зависимости $\hat{y} = a + bx$.

Подставим аналитическое выражение функции $\hat{y} = a + bx$ в функционал S:

$$S = \sum (y - a - bx)^2 \rightarrow \min .$$

Для нахождения минимума функции двух переменных a и b необходимо взять частные производные по каждому параметру и приравнять их к нулю:

$$\frac{dS}{da} = 0; \quad \frac{dS}{db} = 0 .$$

В результате получаем систему нормальных уравнений:

$$\begin{cases} na + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum xy. \end{cases}$$

Решение системы уравнений даёт оценки параметров a и b :

$$a = \frac{\sum y - b \cdot \sum x}{n}; \quad b = \frac{\sum x \cdot \sum y - \sum xy}{(\sum x)^2 - \sum x^2};$$

В линейном уравнении регрессии параметр a показывает усреднённое влияние на результативный признак неучтённых факторов. Формально $a = \bar{y}$ при $x=0$. **Интерпретация параметра a** как среднего значения результативного признака возможно лишь при условии, что среди наблюдаемых значений факторного признака есть значения, равные или близкие к 0. **Параметр b** в уравнении линейной регрессии называется **коэффициентом регрессии**. Коэффициент регрессии показывает, на сколько в среднем изменится значение результативного признака при увеличении факторного на единицу собственного измерения.

Для получения качественного уравнения регрессии необходимо чтобы данные наблюдения соответствовали следующим требованиям:

1. **число наблюдений должно в 6-7 раз превышать число рассчитываемых параметров при переменных x** . Таким образом, искать линейную регрессию имея менее 7 наблюдений не имеет смысла;

2. **распределение единиц наблюдения по факторному признаку должно быть однородным и подчиняться нормальному закону распределения.**

Построение линейного уравнения регрессии рассмотрим на следующем примере: имеются экспериментальные данные исследования влияния времени вулканизации на сопротивление резины разрыву. Данные наблюдения приведены в таблице 7.3. На основе приведённых данных провести исследование влияния времени вулканизации на качество резины.

Таблица 7.3.
Данные о времени вулканизации и сопротивлению резины

№ анализа	Время вулканизации, мин	Сопротивление разрыву, кг/см ²	№ анализа	Время вулканизации, мин	Сопротивление разрыву, кг/см ²
2	35	162	8	33	160
2	40	174	9	36	167
3	30	155	10	31	153
4	42	172	11	36	163
5	37	179	12	43	173
6	38	166	13	39	168
7	34	162	14	44	176

Результативный признак y – сопротивление резины, факторный x – время вулканизации.

Прежде чем подбирать соответствующую математическую функцию и строить уравнение регрессии, необходимо проверить качество исходной информации, уровень её вариации, нормальность распределения.

1. Для проверки совокупности на однородность по факторному признаку, используется коэффициент вариации ($V_x \geq 33\%$):

$$V_x = \frac{\sigma_x}{\bar{x}} \cdot 100\%$$

Среднее время вулканизации по данным таблицы 7.4. составляет

$$\bar{x} = \frac{518}{14} = 37 \text{ мин.}$$

Таблица 7.4.

Вспомогательная таблица для расчета \bar{x} и σ_x

№ п/п	x	$x - \bar{x}$	$(x - \bar{x})^2$	№ п/п	x	$x - \bar{x}$	$(x - \bar{x})^2$
1	35	-2	4	8	33	-4	16
2	40	+3	9	9	36	-1	1
3	30	-7	49	10	31	-6	36
4	42	+5	25	11	36	-1	1
5	37	0	0	12	43	+6	36
6	38	+1	1	13	39	+2	4
7	34	-3	9	14	44	+7	49
	-	-	-	Итого	518	-	240

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{240}{14}} = \sqrt{17.14} \approx 4.1_{мин}; \quad V_x = \frac{4,1}{37} \cdot 100 = 11,1\% < 33\%,-$$

это означает, что изучаемая совокупность однородна.

2. Проверка первичной информации на нормальность распределение проводится на основе «правила 3-х сигм»:

$$\bar{x} \pm 1\sigma = 37 \pm 4.1 = 39.2 \div 41.1;$$

$$\bar{x} \pm 2\sigma = 37 \pm 2 \cdot 4.1 = 3 \pm 8.2 = 28.8 \div 45.2;$$

$$\bar{x} \pm 3\sigma = 37 \pm 3 \cdot 4.1 = 37 \pm 12.3 = 24.7 \div 49.3.$$

Для проверки составим следующую таблицу:

Таблица 7.5.

Проверка данных наблюдения на нормальность распределения

Интервалы значений признака	Число единиц попадающих в интервал	Удельный вес единиц, попавших в интервал, %	Удельный вес единиц, входящих в интервал при нормальном распределении, %
39,2 - 41,1	9	64,3	68,3
28,8 - 45,2	14	100,0	95,4
24,7 - 49,3	14	100,0	99,73

Первичная информация по факторному признаку не полностью подчиняется закону нормального распределения, однако это не является основанием для отказа использования корреляционно-регрессионного анализа для описания связи между признаками.

3. Исключение из первичной информации аномальных значений факторного признака, то есть значений, не попадающих в интервал $\bar{x} \pm 3\sigma$ ($24,7 < x_i < 49,3$). Таких аномальных значений в первичной информации нет.

4.

5. Для установления факта наличия связи проведём аналитическую группировку. Она выполняется как равноинтервальная группировка. При $m=4$, $a = \frac{x_{\max} - x_{\min}}{m} = \frac{44 - 30}{4} = 3.5$, принимаем $a = 4$ мин.

Результаты группировки приведены в таблице 7.6:

Таблица 7.6.
Исследование зависимости сопротивления резины от времени вулканизации

№ группы	Время вулканизации, мин	Номера анализов, попавших в группу	Число анализов	Сопротивление резины в группе, y_i	$\sum y_i$	Среднее сопротивление резины, кг/см ²
1	30-34	3,8,10	3	155,160,153	468	156,0
2	34-38	1,5,7,9,11	5	162,173,162,167,163	827	165,4
3	38-42	2,6,13	3	174,166,168	508	169,3
4	42-44	4,12,14	3	172,173,176	521	173,7
Итого	30-46	-	14	-	2324	-

Анализ таблицы 7.6. позволяет сделать следующий вывод: связь между признаками существует, так как при увеличении времени вулканизации возрастает сопротивление резины. Графически это выглядит следующим образом:

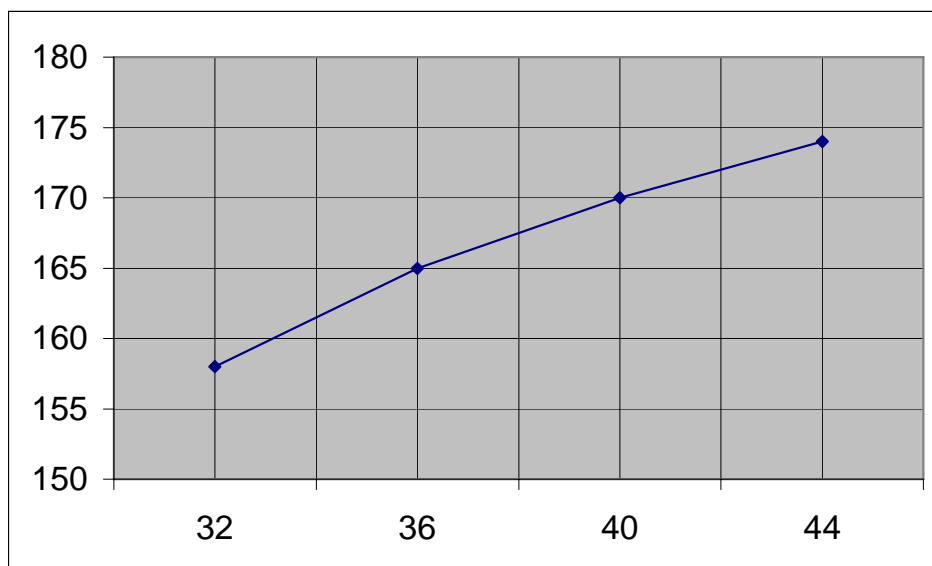


Рис. 7.3. Эмпирическая линия регрессии сопротивления резины на время вулканизации

Эмпирическая линия регрессии приближается к прямой. Следовательно, можно считать, что между признаками имеется прямолинейная связь вида $\hat{y} = a + bx$.

Коэффициент регрессии b рассчитаем по формуле:

$$b = \frac{\sum x \cdot \sum y - \sum xy}{(\sum x)^2 - \sum x^2};$$

параметр $a = \frac{\sum y - b \sum x}{n}$.

Для расчёта параметров составим вспомогательную таблицу:

Таблица 7.7.

Расчёт параметров уравнения регрессии

№ анализа	x_i	y_i	x^2	$x \cdot y$	y^2	\hat{y}_i	$y_i - \hat{y}_i$	$(y - \hat{y})^2$
1	35	162	1225	5670	26244	163.0	-1.0	1.0
2	40	174	1600	6960	30276	170.5	+3.5	12.25

3	30	155	900	4650	24025	155.5	-0.5	0.25
4	42	172	1764	7224	29584	173.5	-1.5	2.25
5	37	173	1369	6401	29929	166.0	+7.0	49.0
6	38	166	1444	6308	27556	167.5	-1.5	2.25
7	34	162	1156	5508	26244	161.5	+0.5	0.25
8	33	160	1089	5280	25600	160.0	0.0	0.00
9	36	167	1296	6012	27889	164.5	+2.5	6.25
10	31	153	961	4743	23409	157.0	-4.0	16.0
11	36	163	1296	5868	26569	164.5	-1.5	2.25
12	43	73	1849	7439	29929	175.0	-2.0	4.0
13	39	168	1521	6552	28224	169.0	-1.0	1.0
14	44	176	1936	7744	30976	176.5	-0.5	0.25
Итого	518	2324	19406	86359	386.454	2364,0	-	97.0

$$b = \frac{518 \cdot \frac{2324}{14} - 86359}{518^2 - 19 \cdot 406} = 1.5;$$

$$a = \frac{2324 - 1.5 \cdot 518}{14} = 110.5.$$

Уравнение регрессии имеет вид $\hat{y} = 110.5 + 1.5x$. Параметры модели могут быть интерпретированы следующим образом: коэффициент регрессии $b=1,5$ показывает, что при увеличении времени вулканизации на 1 минуту сопротивление резины увеличивается на $1,5 \text{ кг/см}^2$. Параметр $a = 110,5$ интерпретировать невозможно, так как среди наблюдаемых значений факторного признака – времени вулканизации отсутствуют значения равные или близкие к нулю.

Аналогичным образом на основе МНК рассчитываются параметры нелинейной регрессии.

Для параболы второго порядка: $\hat{y} = a + bx + cx^2$ получаем систему нормальных уравнений следующего вида:

$$\begin{cases} \sum y = n \cdot a + b \cdot \sum x + c \cdot \sum x^2 \\ \sum x \cdot y = a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 \\ \sum x^2 \cdot y = a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 \end{cases}$$

Для показательной функции $\hat{y} = a \cdot b^x$ предварительно необходимо выполнить процедуру минимизации, то есть привести функцию к линейному виду. Это можно сделать, прологарифмировав обе части уравнения:

$$\ln \hat{y} = \ln a + x \ln b.$$

Введём следующие обозначения: $\ln \hat{y} = Y$; $\ln a = A$; $\ln b = B$. Тогда уравнение регрессии принимает вид: $Y = A + B \cdot x$, то есть приводится к линейному уравнению регрессии.

7.4. Оценка существенности парной корреляционной связи

Для проверки существенности парной корреляционной связи, то есть соответствия полученной модели данным наблюдения используется следующий подход: **модель признаётся значимой, если таковыми являются параметры модели или показатели тесноты связи**. При этом выясняется, не являются ли вычисленные значения параметров регрессии случайными величинами? **Значимость параметров линейной модели определяется с помощью t-критерия Стьюдента**. Для каждого из параметров уравнения регрессии вычисляются расчетные (фактические) значения t-критерия:

$$\text{для параметра } a: t_{a=0} = a \cdot \frac{\sqrt{n-2}}{\sigma_{y-\hat{y}}};$$

$$\text{для параметра } b: t_b = b \cdot \frac{\sqrt{n-2}}{\sigma_{y-\hat{y}}} \cdot \sigma_x$$

где n – число наблюдений;

$$\sigma_{y-\hat{y}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n}} - \text{остаточное среднее квадратическое отклонение}$$

результативного признака y от выровненных значений \hat{y} , рассчитанных по модели;

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} - \text{среднее квадратическое отклонение факторного}$$

признака x_i от общей средней \bar{x} .

Вычисленные значения t -критериев сравниваются с критическими значениями $t_{\alpha, \nu}$, определёнными по таблице распределения Стюдента с учётом принятого уровня значимости α и числа степеней свободы вариации $\nu = n - 2$.

Параметр признаётся значимым, если выполняется неравенство: $t_{\text{расч}} > t_{\alpha, \nu}$.

В этом случае найдённые значения параметров не являются случайными, а уравнение регрессии признаётся существенным.

Значимость линейной регрессии можно оценить по линейному коэффициенту корреляции. **Модель признаётся значимой, если расчётное значение t -критерия для линейного коэффициента корреляции превышает табличное, то есть выполняется неравенство:**

$$t_{r_{yx}} > t_{\alpha, \nu}.$$

Расчётное значение t -критерия для линейного коэффициента корреляции определяется по формуле:

$$t_{r_{yx}} = r_{yx} \cdot \sqrt{\frac{n-2}{1-r_{yx}^2}}.$$

Для нелинейных моделей их существенность проверяется с помощью F -критерия Фишера.

Модель признаётся значимой, если выполняется следующее неравенство:

$$F_p > F_{\nu_1, \nu_2};$$

где F_p - расчётное значение критерия Фишера,

F_{ν_1, ν_2} - критическое значение критерия Фишера, выбираемое по специальной таблице распределения F -критерия.

Расчётное значение F -критерия определяется по формуле:

$$F_{\text{расч}} = \frac{\eta^2}{1-\eta^2} \cdot \frac{n-m}{m-1},$$

где η^2 - теоретический коэффициент детерминации,

m - число параметров уравнения регрессии.

Теоретический коэффициент детерминации η^2 является показателем тесноты связи результативного и факторного признака в

уравнении регрессии. Рассчитывается η^2 на основе правила сложения дисперсий.

При наличии уравнения регрессии, описывающего существующую связь, степень влияния факторного признака на результативный может быть выражена следующим образом:

$$y_i = \hat{y}_i + \varepsilon_i,$$

где \hat{y}_i - теоретическое (сглаженное) значение результативного признака, просчитанное по уравнению регрессии.

Соответственно, **дисперсия результативного признака σ_y^2 должна включить в себя дисперсию теоретических значений результативного признака (объясняемую) $\sigma_{\hat{y}}^2$ и дисперсию отклонений эмпирических (наблюдаемых) значений результативного признака от теоретических $\sigma_{y-\hat{y}}^2$.**

$$\text{Таким образом, } \sigma_y^2 = \sigma_{\hat{y}}^2 + \overline{\sigma_{y-\hat{y}}^2},$$

где $\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}$ - общая дисперсия результативного признака,

$\sigma_{\hat{y}}^2 = \frac{\sum (\hat{y} - \bar{y})^2}{n}$ - объяснённая дисперсия результативного признака,

$\sigma_{y-\hat{y}}^2 = \frac{\sum (y - \hat{y})^2}{n}$ - остаточная дисперсия результативного признака.

Объяснённая дисперсия $\sigma_{\hat{y}}^2$ характеризует влияние фактора, включённого в модель, на общую вариацию результативного признака.

Остаточная дисперсия $\sigma_{y-\hat{y}}^2$ характеризует влияние факторов, не включённых в уравнение регрессии, на вариацию результативного признака.

Теоретический коэффициент детерминации определяется через соотношение объясняемой и общей дисперсии результативного признака.

$$\eta^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}, \text{ так как } \sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_{y-\hat{y}}^2, \text{ то } \eta^2 = 1 - \frac{\sigma_{y-\hat{y}}^2}{\sigma_y^2}.$$

Оценим качество линейного уравнения регрессии $\hat{y} = 110.5 + 1.5x$, выражающего влияние времени вулканизации на качество резины (сопротивление на разрыв)

Оценку проведём по линейному коэффициенту корреляции по данным таблицы 7.7:



$$r_{yx} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} = \frac{86389 - \frac{518 \cdot 2324}{14}}{\sqrt{\left(19406 - \frac{518^2}{14}\right) \cdot \left(386454 - \frac{2324^2}{14}\right)}} = +0.925$$

Расчётное значение критерия Стьюдента составит:

$$t_{yx} = r_{yx} \cdot \sqrt{\frac{n-2}{1-r_{yx}^2}} = 0.925 \cdot \sqrt{\frac{14-2}{1-0.925^2}} = 83,3$$

Табличное значение t-критерия составляет $t_{\alpha, \nu} = 2.179$ при $\alpha = 0,05$ и $\nu = 14-2 = 12$.

Таким образом, с вероятностью в 95% можно утверждать численные значения линейного коэффициента корреляции не является случайной величиной, а уравнение регрессии $\hat{y} = 110.5 + 1.5x$ является статистически значимым.

7.5. Множественная корреляция

Двухмерные корреляционные модели (парная корреляция) используются в случаях, когда среди факторов, влияющих на результативный признак, есть доминирующий. Таких связей немного, чаще встречаются зависимости результативного признака от нескольких факторных, так как экономические явления находятся под влиянием значительного числа одновременно и совокупно действующих факторов. Для описания совместного влияния одновременно действующих факторов на результат используют множественные корреляционные модели вида: $\hat{y} = f(x_1, x_2, \dots, x_k)$

Модели подобного класса используются при изучении спроса, функции потребления, доходности акций и т.д.

Задача множественного корреляционно-регрессионного анализа в общем виде формулируется следующим образом:

Пусть некоторая статистическая совокупность, состоящая из n единиц наблюдения обладает определённым набором признаков, один из которых играет роль результативного y , а остальные – факторных (x_1, x_2, \dots, x_k) . На основе наблюдаемых значений всех признаков требуется

выявить и описать связь между ними в виде множественной корреляционной модели вида: $\hat{y} = f(x_1, x_2, \dots, x_k)$.

Решение данной задачи требует последовательного выполнения следующих этапов исследования множественной корреляционной связи:

- предварительный отбор факторов, включаемых в модель;
- предварительное описание связи;
 - уточнение модели на основе анализа корреляционной матрицы;
 - определение тесноты связи;
- оценка надёжности множественной корреляционной модели;
- интерпретация модели.

← Формат: Список

← Формат: Список

7.5.1. Предварительный отбор факторов

Изучение множественной регрессии (корреляции) требует измерения не только прямого воздействия каждого фактора на результат, но и учёта влияния факторов друг на друга, то есть учёта наличия межфакторных связей. **Общее число связей** всегда значительно больше числа факторов, включаемых в модель. Оно определяется выражением:

$$l = \frac{k(k+1)}{2};$$

где k – количество факторных признаков, включенных в модель.

Например, если в модель включаются четыре фактора, то количество связей равно 10, если в модель включено 10 факторов, то общее число связей составит 55.

В общем случае, при большом числе учитываемых факторов необходимо строить сложные модели, требующие проведения сложных расчётов; модели получаются громоздкими. С другой стороны, чем большее количество факторов учитывается, тем адекватнее построенная модель.

Для разрешения указанного противоречия предварительно ограничивается число учитываемых факторов. **Целесообразность их включения в модель определяется следующими соображениями:**

• они должны быть соизмеримы, иметь количественное выражение;

← Формат: Список

• факторы не должны быть интеркоррелированы, то есть тесно связанными между собой;

• они должны объяснять вариацию результативного признака.

При включении в модель интеркоррелированных факторов невозможно определить изолированное влияние таких факторов на результативный показатель, а оценки параметров уравнения

множественной регрессии будут ненадёжными, зависимыми от наблюдений.

7.5.2. Предварительное описание множественной корреляционной связи (МКС) осуществляется через построение соответствующего уравнения регрессии. Практика показывает, что можно использовать следующие пять функций, так как они описывают все реально существующие зависимости между социально-экономическими явлениями:

1. линейная $\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_k b_k$;
2. степенная $\hat{y} = a_0 \cdot x_1 a_1 \cdot x_2 a_2 \cdot \dots \cdot x_k a_k$;
3. показательная (экспотенциальная) $\hat{y} = e^{a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k}$;
4. параболическая $\hat{y} = a_0 + a_1 x_1^2 + a_2 x_2^2 + \dots + a_k b_k^2$;
5. гиперболическая $\hat{y} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_k}{x_k}$.

Работать с нелинейными функциями сложно, поэтому **основное значение имеют линейные модели** в силу их простоты и логичности экономической интерпретации. Нелинейные формы всегда можно привести к линейной, используя известный в математике приём линеаризации функций.

Величина каждого параметра в уравнении прямой может быть определена по методу наименьших квадратов. Для этого в функционал $S = \sum (y - \hat{y})^2 \rightarrow \min$ подставим выражение прямолинейной функции. В результате подстановки получаем:

$$\sum (y - a_0 - a_1 x_1 - a_2 x_2 - \dots - a_k x_k)^2 \rightarrow \min$$

Продифференцировав полученное выражение по каждому из параметров, получаем систему нормальных уравнений вида:

$$\begin{cases} \sum y = a_0 \cdot n + a_1 \cdot \sum x_1 + a_2 \cdot \sum x_2 + \dots + a_k \cdot \sum x_k^2 \\ \sum y \cdot x_1 = a_0 \cdot \sum x_1 + a_1 \cdot \sum x_1^2 + a_2 \cdot \sum x_1 \cdot x_2 + \dots + a_k \cdot \sum x_k^2 \\ \sum y \cdot x_2 = a_0 \cdot \sum x_2 + a_1 \cdot \sum x_1 x_2 + a_2 \cdot \sum x_2 \cdot x_k + \dots + a_k \cdot \sum x_k^2 \\ \dots \\ \sum y \cdot x_k = a_0 \cdot \sum x_k + a_1 \cdot \sum x_1 \cdot x_k + a_2 \cdot \sum x_2 \cdot x_k + \dots + a_k \cdot \sum x_k^2 \end{cases}$$

Число уравнений в системе равно числу параметров, поэтому задача их нахождения является разрешимой.

При выборе формы уравнения множественной регрессии необходимо иметь в виду:

1. Чем сложнее функция, тем хуже интерпретируются параметры модели.
2. Сложные функции (полиномы) с большим количеством факторов требуют большого числа наблюдений (на каждый параметр не менее 6 наблюдений)

7.5.3. Окончательный отбор факторов

← Формат: Список

Окончательный отбор факторов, то есть уточнение корреляционной модели проводится на основе анализа корреляционной матрицы. **Корреляционная матрица** состоит из *парных линейных коэффициентов корреляции* r_{yx} , отражающих тесноту связи результативного и факторного признака и коэффициентов интеркорреляции r_{xixj} , отражающих тесноту связи между *i*-м и *j*-м факторными признаками, рассчитываемых по известной формуле:

$$r_{ij} = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[x^2 - \frac{(\sum x)^2}{n} \right] \left[y^2 - \frac{(\sum y)^2}{n} \right]}} ; \quad r_{ij} = r_{ji}$$

Общий вид корреляционной матрицы приведен в таблице 7.8.

Таблица 7.8.

Общий вид корреляционной матрицы

	y	x_1	x_2	x_j	x_k
y	1	r_{yx_1}	r_{yx_2}		r_{yx_j}		r_{yx_k}
x_1	r_{yx_1}	1	$r_{x_1x_2}$		$r_{x_1x_j}$		$r_{x_1x_k}$
x_2	r_{yx_2}	$r_{x_1x_2}$	1		$r_{x_2x_j}$		$r_{x_2x_k}$
.....							
x_j	r_{yx_j}	$r_{x_jx_1}$	$r_{x_jx_2}$		1		$r_{x_jx_k}$
.....							
x_k	r_{yx_k}	$r_{x_kx_1}$	$r_{x_kx_2}$		$r_{x_kx_j}$		1

Факторы, теснота связи между которыми оценивается как высокая, считаются коллинеарными. **Окончательный отбор факторов заключается во включении в модель независимых (неколлинеарных) факторов.** Процедура отбора осуществляется способом шаговой регрессии.

Применение шаговой регрессии рассмотрим на следующем примере: в процессе предварительного отбора были выявлены 5 факторов, влияющих на результат. Значения линейных коэффициентов корреляции приведено в таблице 7.9. Необходимо построить двухфакторную корреляционную модель.

Корреляционная матрица имеет вид:

Таблица 7.9.

	y	x_1	x_2	x_3	x_4	x_5
y	1	-0,43	0,50	0,55	0,7	0,31
x_1	-0,3	1	0,65	-0,44	0,5	0,38
x_2	0,50	0,62	1	0,85	0,79	0,7
x_3	0,55	-0,44	0,85	1	-0,95	0,50
x_4	0,70	0,5	0,79	-0,95	1	0,78
x_5	0,1	0,38	0,70	0,50	0,78	1

1. Для обоснования включения факторов в модель оценивается первая строка матрицы, отражающая связь факторов с результатом. В модель включаются факторы, оказывающие наибольшее влияние на результат (с максимальными линейными коэффициентами корреляции). Такими являются факторы x_3 и x_4 .

2. Оценивается теснота межфакторной связи. Если она высока, то между данными факторами существует тесная зависимость, то есть факторы коллинеарны, а коллинеарность (тесная зависимость между факторами) существенно искажает результаты исследования. **Связь относится к коллинеарной, если:** $|r_{ij}| \geq 0.8$. $r_{x_3x_5} = -0.95$, - факторы коллинеарны. Один из них необходимо исключить из модели. Исключается фактор с меньшим значением линейного коэффициента корреляции - x_3 .

3. Для включения недостающего вошедшие в модель на первом этапе. Это факторы x_1 , x_2 , x_5 . Выбирается фактор с максимальным значением линейного коэффициента корреляции. Это x_2 с $r_{yx_2} = 0.50$. Проверяем тесноту межфакторной связи x_2, x_4 , У нее $r_{x_2x_4} = 0.79$ - факторы не коллинеарны. Таким образом, в модель включаются факторы x_2 и x_4 . Она имеет вид: $\hat{y} = F(x_2, x_4)$.

7.5.4. Оценка тесноты связи

Оценка тесноты множественной корреляционной связи проводится на основе двух показателей: множественного коэффициента детерминации $R_{yx1...xk}^2$ и множественного коэффициента корреляции $R_{yx1...xk}$.

Сложность расчёта этих показателей связана с необходимостью учёта межфакторных связей. Гипотетически данные показатели рассчитываются по формулам:

$$R_{yx1...xk}^2 = 1 - \frac{\sigma_{y-\hat{y}}^2}{\sigma_y^2};$$

$$R_{yx1...xk} = \sqrt{R_{yx1...xk}^2}.$$

На практике множественный коэффициент корреляции R рассчитывается на основе определителей корреляционной матрицы:

$$R_{yx1...xk} = \sqrt{1 - \frac{\Delta r}{\Delta r_{xixj}}};$$

где Δr - общий определитель корреляционной матрицы;

Δr_{xixj} - определитель матрицы межфакторной корреляции.

Для двухфакторной модели множественный коэффициент корреляции определяется по формуле:

$$R_{yx1x2} = \sqrt{\frac{r_{yx1}^2 + r_{yx2}^2 - 2 \cdot r_{yx1} \cdot r_{yx2} \cdot r_{x1x2}}{1 - r_{x1x2}^2}}.$$

Диапазон изменения множественного коэффициента корреляции $R_{yx1...xk} = |0 \div 1|$.

«0» означает отсутствие связи, «1» - наличие функциональной множественной связи между признаками. Для классификации тесноты связи используется шкала Чеддока.

7.5.5. Оценка надёжности модели

Для оценки надёжности выявленной связи сравнивается множественный коэффициент корреляции с линейными корреляционными коэффициентами корреляции между результатом и факторными признаками, включёнными в модель.

Связь признаётся надёжной, если $R_{yx1...xk} \geq \max\{r_{yxj}\}$.

7.5.6. Интерпретация параметров модели

Завершающим этапом множественной корреляции является интерпретация параметров построенной корреляционной модели. **Чем больше величина этих параметров** (коэффициентов регрессии), **тем большее влияние данных факторов на результат**. Важное значение имеют знак перед коэффициентами регрессии. Знак “+” свидетельствует о росте результата при увеличении факторного признака, знак “-” – об уменьшении результата при росте факторного.

Пример построения многофакторной корреляционной модели: В результате выборочного исследования была собрана следующая информация об акционерных обществах таблица 7.10:

Таблица 7.10.

Финансовые показатели акционерных обществ

№ п/п	Размер дивидендов, %	Сумма кредитов	Уставной капитал, млн.руб.
1	2	10	4
2	3	14	14
3	5	16	10
4	6	8	8
5	7	12	20
6	11	10	10
7	12	2	36
8	18	8	26
Итого	64	80	128

Требуется описать связь между размером получаемых дивидендов, суммой взятых кредитов и размером УК.

Считаем, что связь между результативным и факторными признаками линейная, описываемая уравнением: $\hat{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2$.

Для нахождения параметров используется метод наименьших квадратов, получаем систему уравнений:

$$\begin{cases} \sum y = a_0 \cdot n + a_1 \cdot \sum x_1 + a_2 \cdot \sum x_2 \\ \sum y \cdot x_1 = a_0 \cdot \sum x_1^2 + a_2 \cdot \sum x_1 \cdot x_2 \\ \sum y \cdot x_2 = a_0 \cdot \sum x_2^2 + a_1 \cdot \sum x_1 \cdot x_2 + a_2 \cdot \sum x_2^2 \end{cases}$$

Для расчёта численных значений параметров составим вспомогательную таблицу:

Таблица 7.11.

Таблица вспомогательных расчетов

№ п/п	y	x_1	x_2	x_1^2	x_2^2	yx_1	yx_2	x_1x_2	y^2
1	2	10	4	100	16	20	8	40	4
2	3	14	14	196	196	42	42	196	9
3	5	16	10	256	100	80	50	160	25
4	6	8	8	64	64	48	48	64	36
5	7	12	20	144	400	84	140	240	49
6	11	10	10	100	100	110	110	100	121
7	12	2	36	4	1296	24	432	72	144
8	18	8	26	64	676	144	468	208	324
Итого	64	80	128	928	2848	552	1298	1080	712

$$\begin{cases} 64 = 8a_0 + 80a_1 + 128a_2; \\ 552 = 80a_1 + 928a_2 + 1080a_2; \\ 1298 = 128a_0 + 1080a_1 + 2848a_2. \end{cases}$$

Решение системы даёт следующие значения параметров:

$$a_0 = 5.14; \quad a_1 = -0.21; \quad a_2 = 0.31.$$

Модель имеет вид: $\hat{y} = 5.14 - 0.21x_1 + 0.31x_2$.

Определим тесноту связи и надежность данной модели. Для этого предварительно рассчитаем линейные коэффициенты корреляции:

$$r_{yx1} = \frac{\sum x_1 y - \frac{\sum x_1 \sum y}{n}}{\sqrt{\left[\sum x_1^2 - \frac{(\sum x_1)^2}{n} \right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}} = \frac{552 - \frac{80 \cdot 64}{8}}{\sqrt{\left(928 - \frac{80^2}{8} \right) \left(712 - \frac{64^2}{8} \right)}} =$$

$$= \frac{552 - 640}{\sqrt{128 \cdot 200}} = -\frac{88}{160} = -0.55,$$

$$r_{yx2} = +0.685, \quad r_{x1x2} = -0.625.$$

Такие значения линейных коэффициентов корреляции означают, что факторы не являются коллинеарными, оба должны быть включены в модель.

Множественный коэффициент корреляции получается равным:

$$R_{yx1x2} = \sqrt{\frac{(-0.55)^2 + (0.685)^2 - 2 \cdot (-0.55) \cdot 0.685 \cdot (-0.625)}{1 - (-0.625)^2}} = 0.703.$$

По шкале Чеддока связь классифицируется как тесная. Поскольку $R_{yx1x2} > \max\{r_{yx1}, r_{yx2}\}$, ($0,703 > 0,685$), модель надёжна, связь статистически значима. Параметры модели интерпретируются следующим образом:

$a_1 = -0,25$ показывает, что при неизменности уставного капитала дополнительная сумма кредитов на 1 млн. руб. приводит к снижению размеров дивидендов на 0,25%; соответственно

$a_2 = 0,28$ показывает, что при неизменности суммы взятых кредитов прирост уставного капитала на 1 млн. руб. приводит к росту дивидендов на 0,28%.

7.6. Анализ взаимосвязанных рядов динамики

При изучении развития явления во времени часто возникает необходимость оценивать степень зависимости изменения уровней двух рядов динамики различного содержания, но связанных между собой. Например, динамики урожайности и внесения удобрений; динамики средней выработки и энерговооружённости труда.

Ряды динамики, в которых уровни одного ряда определяют уровни другого называются взаимосвязанными. Их анализ является наиболее сложным элементом при изучении экономических процессов. Это связано со следующими обстоятельствами:

• **Наличие зависимости между последующими и предыдущими уровнями ряда:** $y_{t+1} = f(y_t)$. Такая зависимость называется **автокорреляцией**. Обычно автокорреляция имеет место в моментных рядах динамики. Каждый из взаимосвязанных рядов необходимо проверить на наличие автокорреляции; и, если она имеет место, устранить её.

• **Существование лага**, то есть смещения во времени одного показателя по сравнению с изменением другого (например, при рассмотрении взаимосвязанных рядов динамики, числа заключённых браков и числа родившихся). При наличии такого смещения нужно сдвинуть уровни одного ряда относительно другого на некоторый промежуток времени. Это даёт возможность получить более правильную тесноты корреляционной связи между рядами динамики.

• **На разных отрезках времени изучаемого периода условия формирования могут меняться (непериодизированные ряды динамики).** В этом случае имеет место **переменная корреляция** – степень тесноты связи меняется во времени.

Оценка степени взаимосвязи уровней двух рядов динамики проводится в такой последовательности:

Формат: Список

1. Проверка взаимосвязанных рядов динамики на наличие автокорреляции

Для этого можно использовать один из двух способов:

- Проверка по коэффициенту автокорреляции (коэффициенту

корреляции первого порядка)
$$r_{ap} = \frac{y_t \cdot y_{t+1} - y_t \cdot y_{t+1}}{\sigma_{y_t} \cdot \sigma_{y_{t+1}}}$$
,

где y_t - уровни исходного ряда динамики,

y_{t+1} - уровни ряда, сдвинутые на единицу времени.

y_{t+1} рассматривается как результативный признак; y_t - как факторный. Полученное расчётное значение коэффициента r_{ap} автокорреляции сравнивается с табличным, выбранным по таблице Р.Андерсена r_{akp} .

При $r_{ap} > r_{akp}$ - автокорреляция имеет место.

- Проверка по критерию Дарбина-Уотсона (DW)

$$DW = \frac{\sum (\varepsilon_{t+1} - \varepsilon_t)^2}{\sum \varepsilon_t^2},$$

где $\varepsilon_t = y_t - \hat{y}_t$ - отклонение уровней исходного ряда от теоретических значений, просчитанных по трендовой модели,

$\varepsilon_{t+1} = y_{t+1} - \hat{y}_{t+1}$ - отклонение уровней ряда, сдвинутых на единицу от соответствующих теоретических, также сдвинутых на единицу.

Если рассчитанное значение критерия Дарбина-Уотсона $DW=2$ – автокорреляция отсутствует, при $DW=0$ наблюдается полная положительная автокорреляция, при $DW=4$ имеет место полная отрицательная автокорреляция.

2. Определение взаимосвязи рядов динамики

- Если автокорреляция не обнаружена, для измерения тесноты связи между взаимосвязанными рядами динамики рассчитывается коэффициент корреляции:

$$r = \frac{\sum_{i=1}^n \varepsilon_{x_t} \cdot \varepsilon_{y_t}}{\sqrt{\sum_{t=1}^n \varepsilon_{x_t}^2 \sum_{t=1}^n \varepsilon_{y_t}^2}};$$

где x_t, y_t - уровни взаимосвязанных рядов динамики;

$\varepsilon_{x_t} = x_t - \hat{x}_t$; $\varepsilon_{y_t} = y_t - \hat{y}_t$ - отклонение эмпирических уровней рядов динамики от теоретических, рассчитанных по трендовым моделям.

• **Если автокорреляция уровней имеет место, то её необходимо исключить.**

Для этого можно использовать несколько способов, самым простым из которых является **способ коррелирования первых разностей**.

Суть способа заключается в *переходе от исходных рядов динамики X и Y к новым, построенным по цепным абсолютным приростам (по первым разностям)*:

$$\Delta x_t = x_t - x_{t-1};$$

$$\Delta y_t = y_t - y_{t-1}.$$

Для новых рядов рассчитывается коэффициент корреляции разностей:

$$r_{\Delta X_t, \Delta Y_t} = \frac{\sum_{t=2}^n \Delta x_t \cdot \Delta y_t}{\sqrt{\sum_{t=2}^n \Delta x_t^2 \cdot \sum_{t=2}^n \Delta y_t^2}}.$$

Рассматриваемый способ исключает автокорреляцию только в тех рядах динамики, которые характеризуются прямолинейным трендом.

8. Выборочное исследование

8. 1. Постановка задачи выборочного исследования

Выборочным называется такое статистическое исследование, при котором обобщающие показатели изучаемой совокупности устанавливаются по некоторой ее части, сформированной на основе положений случайного отбора.

В основе выборочного исследования лежит *несплошное наблюдение*, при котором обследуются не все единицы совокупности, а лишь определенная их часть.

Выборочное исследование широко применяется на практике, поскольку обладает *существенными преимуществами* по сравнению с другими методами получения статистических данных. К ним относятся:

• Достаточно высокая точность результатов обследования благодаря использованию более квалифицированных кадров, что приводит к сокращению ошибок регистрации;

• Экономия времени и средств в результате сокращения объема работы, большая оперативность в получении данных о результатах обследования;

• Возможность исследования очень больших статистических совокупностей;

• Выборочный метод является единственно возможным, если сбор информации связан с разрушением или потерей единиц наблюдения, например, при органолептическом контроле качества продукции;

• Возможность исследования полностью недоступных совокупностей.

При выборочном исследовании изучается сравнительно небольшая часть статистической совокупности (5-10%, реже 20-25% объема ее единиц).

Проведение выборочного исследования является достаточно сложным процессом, выполнение которого включает в себя:

- *обоснование* целесообразности применения выборочного метода в данном исследовании;

- *составление* программы исследования;

- *установление* объема выборки – n ;

- *обоснование* способа формирования выборки;

- *отбор единиц* из Генеральной совокупности (формирование выборки);

- *измерение* изучаемых признаков у отдельных единиц;

- *обработка* полученной информации и расчет характеристик выборки;

- *определение* ошибки выборки;

Формат: Список

- *распространение* выборочных характеристик на Генеральную совокупность.

Для постановки задачи выборочного исследования необходимо ввести следующие понятия:

- **Генеральная совокупность** – изучаемая совокупность, из которой производится отбор единиц, подлежащих изучению, она может быть конечной (N) или бесконечной (∞).

- **Выборочная совокупность (выборка)** – часть единиц генеральной совокупности, отобранная для изучения (n).

Качество результатов выборочного исследования зависит от того, насколько состав выборки представляет генеральную совокупность, иначе говоря, *насколько выборка репрезентативна*.

Под репрезентативностью выборки понимается *соответствие ее свойств и структуры свойствам и структуре генеральной совокупности*.

Репрезентативность выборки может быть обеспечена только при объективности отбора данных, гарантируемую принципами случайности отбора единиц.

Принцип случайности предполагает, что на включение или исключение статистической единицы из выборки не может повлиять никакой другой фактор, кроме случая.

Этот принцип лежит в основе методов случайного отбора, с помощью которых формируется выборка.

Использование методов случайного отбора при формировании выборки позволяет в дальнейшем при обработке использовать аппарат теории вероятности.

Чаще всего с помощью выборочного исследования определяются следующие характеристики генеральной совокупности.

- **Среднее значение признака в совокупности** - \bar{X} , рассчитывается как средняя арифметическая.

- **Доля альтернативного признака в совокупности** - \bar{d} . *Альтернативным* считается признак, принимающий два значения. Если одно из них изменяется как заданное, то *доля альтернативного признака будет характеризовать удельный вес статистических единиц, обладающих заданным значением альтернативного признака*, например, доля брака в изготовленной партии продукции;

- **Дисперсия признака в совокупности** - σ^2 , как показатель вариации.

В общем виде *задача выборочного исследования* формулируется следующим образом:

Пусть имеется *некоторая генеральная совокупность* известного объема (N единиц), обладающая *неизвестными статистическими характеристиками*:

$$\bar{d} = \frac{P}{N} - \text{генеральная доля (удельный вес статистических единиц}$$

генеральной совокупности, обладающих данным значением признака), где P- число единиц генеральной совокупности, обладающих данным значением признака.

\bar{X} - **генеральная средняя** (среднее арифметическое значение признака в генеральной совокупности).

σ^2 - **генеральная дисперсия** (дисперсия исследуемого признака в генеральной совокупности).

σ - **генеральное среднеквадратического отклонения** (среднее квадратическое отклонение исследуемого признака в генеральной совокупности).

Для их определения *сформирована выборочная совокупность* объемом n статистических единиц ($n \ll N$), обладающая аналогичными характеристиками:

ω - **выборочная доля** (удельный вес статистических единиц, обладающих данным значением признака в выборочной совокупности).

\tilde{X} - **выборочная средняя** (среднее арифметическое значение признака в выборочной совокупности).

S^2 - **выборочная дисперсия** (дисперсия исследуемого признака в выборочной совокупности).

S - **выборочное среднее квадратическое отклонение** (среднее квадратическое отклонение изучаемого признака в выборке).

Необходимо на основе известных характеристик выборки получить статистические оценки характеристик генеральной совокупности.

8.2. Статистические оценки параметров (характеристик) генеральной совокупности

Статистической оценкой или статистикой характеристики (параметра) генеральной совокупности *называют приближенное значение искомой характеристики (параметра), полученное по данным выборки.*

В статистике используются два вида оценок - *точечные и интервальные.*

Точечной статистической оценкой параметра генеральной совокупности называется конкретное числовое значение искомой характеристики.

Интервальная оценка представляет собой числовые интервалы, предположительно содержащие значение параметра генеральной совокупности.

Качество статистических оценок **определяется следующими их свойствами:**

- **Состоятельность**

Оценка считается состоятельной, если при неограниченном увеличении объема выборки $[n \rightarrow \infty (N)]$, ее ошибка стремится к 0:

$$\lim_{n \rightarrow \infty} (\tilde{\alpha} - \alpha) = 0, \text{ т. к. при } n \uparrow \lim_{n \rightarrow \infty} \tilde{\alpha} = \alpha;$$

где α - значение характеристики генеральной совокупности;

$\tilde{\alpha}$ - значение характеристики выборки;

$\tilde{\alpha} - \alpha$ - ошибка выборки.

- **Несмещенность**

Оценка считается несмещенной, если при данном объеме выборки n математическое ожидание ошибки равно 0. Для несмещенной оценки ее математическое ожидание точно равно математическому ожиданию характеристики выборки:

$$M[\tilde{\alpha} - \alpha] = 0 \text{ или } M[\tilde{\alpha}] = M[\alpha].$$

Несмещенная оценка не всегда дает хорошее приближение оцениваемого параметра, так как возможные значения получаемой оценки могут быть сильно рассеяны вокруг своего среднего значения. Поэтому оценка должна соответствовать еще одному требованию – эффективности.

- **Эффективность**

Оценка считается эффективной, если ее ошибка, называемая ошибкой выборки, является величиной минимальной. В математической статистике доказывается, что ошибка выборки определяется как:

$$\mu(\tilde{\alpha}) = \sqrt{M^2[\tilde{\alpha} - \alpha] + S^2};$$

где $M^2[\tilde{\alpha} - \alpha]$ - квадрат математического ожидания ошибки выборки;

S^2 - выборочная дисперсия.

Оценка эффективна, если выполняется условие: $\mu(\tilde{\alpha}) \rightarrow \min$.

Для точечных оценок справедливы следующие утверждения:

- Точечной оценкой генеральной доли является выборочная доля, то есть $d \sim \omega$.

- Точечной оценкой генеральной средней является выборочная средняя, то есть $\bar{x} \sim \bar{\mu}$.

Таким образом, заранее известно, что оценки для указанных параметров являются состоятельными и несмещенными.

Для остальных параметров генеральной совокупности это утверждение не является справедливым, то есть $\sigma^2 \neq S^2$, а $\sigma \neq S$.

В математической статистике доказывается, что точечной оценкой генеральной дисперсии является выборочная дисперсия, откорректированная на отношение $\frac{n}{n-1}$, то есть $\sigma^2 = S^2 \times \frac{n}{n-1}$; при увеличении n $\frac{n}{n-1} \rightarrow 1$, поэтому в выборках, объемом больше 30 единиц наблюдения, указанным отношением можно пренебречь.

Аналогично, точечной оценкой генерального среднеекватрического отклонения является выборочное среднеекватрическое отклонение, откорректированное на $\frac{n}{n-1}$, то есть $\sigma = S \times \frac{n}{n-1}$.

В этом случае точечные оценки генеральной дисперсии и генерального среднеекватрического отклонения являются состоятельными и несмещенными.

Основным недостатком точечных оценок является то, что они не учитывают ошибки выборки, то есть не являются эффективными. Поэтому более предпочтительными являются интервальные оценки параметров генеральной совокупности, в которых эти ошибки учитываются. Интервальные оценки соответствуют всем трем требованиям качества статистической оценки.

В математической статистике доказывается, что:

- **Интервальной оценкой генеральной доли** является ее выборочная доля с учетом ошибки выборочной доли, то есть $\bar{d} \approx \omega \pm \mu_{\omega}$, где μ_{ω} - ошибка выборочной доли.

- **Интервальной оценкой генеральной средней** является выборочная средняя с учетом ошибки выборочной средней, то есть $\bar{x} \approx \tilde{x} \pm \mu_{\tilde{x}}$, где $\mu_{\tilde{x}}$ - ошибка выборочной средней.

Применение интервальных оценок означает, что характеристики генеральной совокупности укладываются в определенный диапазон значений. Чтобы их получить, необходимо рассчитать соответствующие ошибки выборки.

8. 3. Ошибки выборки

При правильном формировании выборки величину ее ошибки можно рассчитать заранее. В общем случае **под ошибкой выборки** понимают объективно возникающее расхождение между характеристиками выборки и генеральной совокупности.

Ошибки выборки подразделяются на *ошибки регистрации и ошибки репрезентативности*.

Ошибки регистрации возникают из-за неправильных или неточных сведений. Их источником является невнимательность регистратора, неправильное заполнение формуляров, описки или же непонимание существа исследуемого вопроса.

Ошибки репрезентативности возникают вследствие несоответствия структуры выборки структуре генеральной совокупности. Источником их существования является разная вариация признака у статистических единиц, в результате которой распределение единиц в выборочной

совокупности отличается от распределения единиц в генеральной совокупности.

Ошибки репрезентативности делятся на систематические и случайные.

Систематические ошибки репрезентативности возникают из-за неправильного формирования выборки, при котором нарушается основной принцип научно организационной выборки – принцип случайности.

Случайные ошибки репрезентативности означают, что даже при соблюдении принципа случайности отбора единиц, расхождения между характеристиками выборки и генеральной совокупности все же имеют место.

Сущность случайной ошибки репрезентативности рассмотрим на следующем примере: сравниваются данные об успеваемости студентов по факультету в целом (генеральная совокупность) и по двум 10%-ным случайным выборкам. Данные приведены в таблице 8.1.

Таблица 8.1.

Данные об успеваемости студентов

Оценка	Количество студентов, чел.		
	Генеральная совокупность	Первая выборка	Вторая выборка
2 (неудовлетворительно)	100	9	13
3 (удовлетворительно)	500	50	49
4 (хорошо)	380	30	32
5 (отлично)	120	21	15
всего	1000	100	100

Объем генеральной совокупности N составил 1000 студентов, объем каждой выборки составляет 10% от N , то есть $n=0,1 \cdot 1000=100$ человек.

Так как исходные данные представлены дискретными рядами распределения студентов по уровню успеваемости, то *средний балл* рассчитаем по формуле средней арифметической взвешенной. Он составляет:

$$\text{по генеральной совокупности } \bar{x} = \frac{100 \times 2 + 3 \times 500 + 4 \times 380 + 5 \times 120}{1000} = 3,82;$$

$$\text{по первой выборке } \tilde{x}_1 = \frac{2 \times 9 + 3 \times 50 + 4 \times 30 + 5 \times 21}{100} = 3,93;$$

$$\text{по второй выборке } \tilde{x}_2 = \frac{2 \times 13 + 3 \times 49 + 4 \times 32 + 5 \times 15}{100} = 3,76.$$

Доля студентов, получивших оценки «хорошо» и «отлично», равна:

по генеральной совокупности $\bar{d} = \frac{380+120}{1000} = 0,5$;

по первой выборке $\omega_1 = \frac{30+21}{100} = 0,51$;

по второй выборке $\omega_2 = \frac{32+15}{100} = 0,47$.

Разность между показателями выборочной и генеральной совокупности является случайной ошибкой репрезентативности.

Ошибки репрезентативности для средней составляет:

$$\tilde{x}_1 - \bar{x} = 3,93 - 3,82 = +0,11;$$

$$\tilde{x}_2 - \bar{x} = 3,76 - 3,82 = -0,06.$$

Соответственно, ошибки репрезентативности для доли равна:

$$\omega_1 - \bar{d} = 0,51 - 0,5 = +0,01;$$

$$\omega_2 - \bar{d} = 0,47 - 0,5 = -0,03.$$

Из приведенных расчетов следует, что выборочная средняя и выборочная доля являются случайными величинами, которые могут принимать различные значения в зависимости от единиц, попавших в выборку. *Ошибки выборки также можно считать случайными величинами.* Они могут принимать разные значения, поэтому определяют среднюю из возможных ошибок (стандартную).

Величина ошибки выборки зависит от следующих факторов.

- Степень колеблемости признака в генеральной совокупности

← Формат: Список

Чем однороднее исследуемая совокупность, тем меньше величина средней ошибки при той же самой численности выборки.

- Объем (численность) выборки

← Формат: Список

Увеличивая или уменьшая объем выборки n , можно регулировать величину средней ошибки. Чем больше единиц будет включено в выборку, тем меньше будет величина ошибки, так как тем точнее в выборке будет представлена генеральная совокупность.

- Способ отбора единиц в выборочную совокупность

← Формат: Список

Для каждого способа формирования выборки величина ее ошибки определяется по-разному. В практической деятельности используются различные способы формирования выборочной совокупности, но принципиальное значение имеет их деление на *способы случайного повторного и бесповторного отбора.*

При собственно случайном повторном отборе общее число единиц генеральной совокупности в процессе выборке не меняется. *Статистическая единица, попавшая в выборку, после регистрации изучаемого признака возвращается в генеральную совокупность* и может вновь попасть в выборку. Таким образом, для **всех единиц генеральной совокупности обеспечивается равная вероятность отбора.**

В математической статистике доказывается, что средняя ошибка выборки $\mu_{\tilde{\alpha}}$ определяется по формуле:

$$\mu_{\tilde{\alpha}} = \sqrt{\frac{\sigma_{\tilde{\alpha}}^2}{n}},$$

где $\sigma_{\tilde{\alpha}}^2$ - генеральная дисперсия.

Генеральная дисперсия, также как и остальные параметры генеральной совокупности является неизвестной величиной, но известно соотношение между генеральной и выборочной дисперсией: $\sigma_{\tilde{\alpha}}^2 \sim S_{\tilde{\alpha}}^2 \times \frac{n}{n-1}$;

тогда при достаточно большом объеме выборки ($n > 30$), $\frac{n}{n-1}$ является величиной близкой к 1, и можно считать, что $\sigma_{\tilde{\alpha}}^2 \sim S_{\tilde{\alpha}}^2$.

В случаях малой выборки при $n < 30$ необходимо учитывать отношение $\frac{n}{n-1}$ и рассчитывать среднюю ошибку малой выборки по формуле:

$$\mu_{\tilde{\alpha} \text{ м.в.}} = \sqrt{\frac{S_{\tilde{\alpha}}^2}{n-1}}.$$

Таким образом, для **средней количественного признака средняя ошибка выборки $\mu_{\tilde{x}}$** равна:

$$\mu_{\tilde{x}} = \sqrt{\frac{S_{\tilde{x}}^2}{n}};$$

где $S_{\tilde{x}}^2 = \frac{\sum_{i=1}^n (x_i - \tilde{x})^2}{n}$ - выборочная дисперсия количественного признака.

Средняя ошибка выборки для доли μ_{ω} определяется по формуле:

$$\mu_{\omega} = \sqrt{\frac{S_{\omega}^2}{n}};$$

где $S_{\omega}^2 = \omega \times (1 - \omega)$ - выборочная дисперсия доли альтернативного признака.

Применение простой случайной повторной выборки на практике весьма ограничено. Это связано с тем, что практически нецелесообразно, а иногда и невозможно повторное наблюдение одних и тех же единиц, и поэтому однажды обследованная единица повторному учету не подвергается. Поэтому чаще на практике применяется бесповторный отбор.

При бесповторном собственно случайном отборе общее количество статистических единиц в генеральной совокупности в процессе формирования выборки меняется, уменьшаясь каждый раз на единицу, попавшую в выборку, поскольку отобранные единицы в генеральную совокупность не возвращаются. Таким образом, *вероятность попадания отдельных единиц в выборку при бесповторном случайном отборе также меняется* (для оставшихся единиц она возрастает). В целом *вероятность попадания любой статистической единицы в выборку при бесповторном отборе может быть определена как $1 - \frac{n}{N}$* . На эту величину должна быть скорректирована и средняя ошибка выборки при бесповторном отборе.

Таким образом, расчетные **формулы средней ошибки выборки при бесповторном отборе принимают вид:**

- для **средней** количественного признака

$$\mu_{\bar{x}} = \sqrt{\frac{S_{\bar{x}}^2}{n} \times \left(1 - \frac{n}{N}\right)};$$

- для **доли** альтернативного признака

$$\mu_{\omega} = \sqrt{\frac{\omega \times (1 - \omega)}{n} \times \left(1 - \frac{n}{N}\right)}.$$

На практике при применении выборочного метода *определяются пределы*, за которые не выйдет величина конкретной ошибки выборочного исследования. Величина пределов конкретной ошибки определяется степенью вероятности, с которой измеряется ошибка выборки.

Ошибка выборки, исчисленная с заданной степенью вероятности, называется предельной ошибкой выборки.

Предельная ошибка выборки является максимально возможной при данной вероятности ошибки. Это означает, что с заданной вероятностью гарантируется, что ошибка любой выборки не превысит предельную ошибку. Такая вероятность называется доверительной.

Предельная ошибка Δ выборки рассчитывается по формуле:

$$\Delta_{\bar{\alpha}} = t \times \mu_{\bar{\alpha}};$$

где t – коэффициент доверия, значения которого определяются доверительной вероятностью $P(t)$.

Значения коэффициента доверия t задаются в таблицах нормального распределения вероятностей. Чаще всего используются следующие сочетания:

t	P(t)
1	0,683
1,5	0,866
2,0	0,954
2,5	0,988
3,0	0,997
3,5	0,999

Так, если $t = 1$, то с вероятностью 0,683 можно утверждать, что расхождение между выборочными характеристиками и параметрами генеральной совокупности не превысит одной средней ошибки.

Предельные ошибки выборки Δ для разных параметров при разных методах отбора статистических единиц рассчитываются по формулам, приведенным в таблице 8.2.

Таблица 8.2.

Предельные ошибки выборки

Метод отбора	Предельные ошибки выборки	
	Для средней	Для доли
Повторный	$\Delta_{\bar{x}} = t \times \sqrt{\frac{S_{\bar{x}}^2}{n}}$	$\Delta_{\omega} = t \times \sqrt{\frac{\omega \times (1 - \omega)}{n}}$
Бесповторный	$\Delta_{\bar{x}} = t \times \sqrt{\frac{S_{\bar{x}}^2}{n} \times (1 - \frac{n}{N})}$	$\Delta_{\omega} = t \times \sqrt{\frac{\omega \times (1 - \omega)}{n} \times (1 - \frac{n}{N})}$

Зная величину предельной ошибки выборки, можно рассчитать интервалы для характеристик генеральной совокупности:

Доверительный интервал для генеральной средней равен $(\bar{x} \pm \Delta_{\bar{x}})$; **для генеральной доли** - $(\omega \pm \Delta_{\omega})$.

Рассмотрим нахождение средних и предельных ошибок выборок, определение доверительных интервалов для *средней и доли* на следующем примере:

При оценке спроса на товар А. было проведено пятипроцентное бесповторное обследование регионального рынка. При этом было выяснено, что в 90 из 100 обследованных семей данный товар потребляется. В среднем каждая из обследованных семей потребляла 5 единиц товара ($\bar{x} = 5$) при стандартном отклонении 0,5 единицы ($S=0,5$ ед.). С вероятностью $p=0,954$ установить *долю семей*, потребляющих данный товар и среднее его потребление (спрос).

Для получения статистических оценок параметров генеральной совокупности выполним следующие процедуры:

1. Определим характеристики выборочной совокупности:

- *выборочную долю* ω (удельный вес семей в выборке, потребляющих товар А.): $\omega = \frac{m}{n} = \frac{90}{100} = 0,9$.

- *выборочную среднюю* (средний объем потребления товара А. одной семьей в выборке): $\bar{x} = 5$ единиц.

2. Определим предельные ошибки выборки:

для

доли

$$\Delta\omega = t \times \sqrt{\frac{\omega \times (1-\omega)}{n} \times \left(1 - \frac{n}{N}\right)} = 2 \times \sqrt{\frac{0,9 \times (1-0,9)}{100} \times \left(1 - \frac{100}{2000}\right)} = 0,0585,$$

где $N = \frac{n \times 100}{5} = 2000$ семей;

$$\text{для средней } \Delta\bar{x} = t \times \sqrt{\frac{S_{\bar{x}}^2}{n} \times \left(1 - \frac{n}{N}\right)} = 2 \times \sqrt{\frac{0,5^2}{100} \times \left(1 - \frac{100}{2000}\right)} \approx 0,1.$$

3. Рассчитаем доверительные интервалы характеристик генеральной совокупности:

для доли - $\Delta\omega \leq \bar{d} \leq \omega + \Delta\omega$,

$$0,9 - 0,059 \leq \bar{d} \leq 0,9 + 0,059,$$

$$0,841 < \bar{d} < 0,959;$$

для средней - $\Delta\bar{x} \leq \bar{x} \leq \bar{x} + \Delta\bar{x}$,

$$5 - 0,1 \leq \bar{x} \leq 5 + 0,1,$$

$$4,9 < \bar{x} < 5,1.$$

Таким образом, с вероятностью 0,954 можно утверждать, что доля семей потребляющих данный товар не меньше 84,1%, но не более 95,9%, а среднее потребление товара в семьях находится в пределах от 4,9 до 5,1 единиц. На основании проведенных расчетов можно определить границы потребления (спроса) товара А. на данном рынке:

$$4,9 \cdot 0,841 \cdot 2000 < Q < 5,1 \cdot 0,959 \cdot 2000,$$

$$8240 < Q < 9780.$$

Таким образом, с вероятностью в 95% можно утверждать, что спрос на товар А. не будет ниже 8240 единиц, но и не превысит 9780 единиц.

8.4. Способы формирования выборочной совокупности

Способы формирования выборки (отбора) влияют на результат выборочного исследования, в частности, на точность статистических оценок параметров генеральной совокупности.

Основное требование к отбору заключается в том, что он должен быть по возможности простым.

Различают два способа формирования выборки:

- простой собственно-случайный,
- отбор с предварительным разделением генеральной совокупности на части.

8.4.1. При простом собственно-случайном отборе на включение или исключение какой-либо статистической единицы в выборку влияет только случай. Это обеспечивает равную вероятность каждой единице попасть в выборку.

Технически собственно-случайный отбор проводят методом жеребьевки или по таблице случайных чисел. При этом можно ожидать, что среди отобранных единиц имеются представители разных состояний, которыми характеризуется признак в общей совокупности. В таком случае среднее значение изучаемого признака окажется представленным достаточно точно.

Собственно-случайный отбор в «чистом виде» применяется редко, но он является исходным для всех других видов отбора.

Случайный отбор может быть повторным или бесповторным.

- **При повторном отборе** статистические единицы, отобранные ранее, возвращаются в генеральную совокупность и могут вновь попасть в выборку.

При этом численность генеральной совокупности при проведении отбора остается постоянной, тем самым обеспечивается каждой статистической единице равная возможность попасть в выборку

- **При бесповторном отборе** единицы не возвращаются обратно в генеральную совокупность, ее численность с каждой единицей сокращается, абсолютно равная возможность каждой статистической единице попасть в выборку полностью не обеспечивается. Но при этом при одном и том же объеме выборки наблюдение охватывает больше единиц генеральной совокупности, что обеспечивает более точные результаты по сравнению с повторным отбором (меньшую ошибку выборки).

Бесповторный отбор находит более широкое применение на практике. Повторный отбор используется в тех случаях, когда нельзя применить бесповторную выборку; например, при обследовании потребительского спроса, изучении общественного мнения по какому-либо вопросу и т. п.

Отбор с предварительным разделением генеральной совокупности на части может быть организован различными способами, которым соответствуют свои виды отбора.

В практике выборочных исследований наибольшее распространение получили следующие виды выборки:

- *механическая;*
- *типическая;*

- *серийная;*
- *комбинированная.*

Указанные виды выборки являются дальнейшим развитием и видоизменением собственно-случайного отбора. Их применение вызывается соображениями удешевления или облегчения процесса наблюдения, особым характером объектов наблюдения.

8.4.2. Механический отбор является одним из наиболее применяемых способов формирования выборки. При механическом отборе *генеральная совокупность предварительно упорядочивается по несущественному для цели исследования признаку* (списки избирателей, табельные номера работников, различные другие базы данных). Отбор осуществляется *бесповторным способом через равные интервалы*. Из каждого интервала в выборку попадает только одна единица.

При проведении механической выборки необходимо установить *шаг отсчета* (расстояние между отбираемыми единицами) и *начало отсчета* (номер единицы, которая должна быть обследована первой). Шаг отсчета устанавливается, исходя из предполагаемого процента отбора. Например, при 10%-ой выборке отбирается каждая десятая единица, при 20%-ой – каждая двадцатая.

Особенностью механического отбора является то, что при его применении *возможно появление систематических ошибок*, связанное со случайным совпадением выбранного интервала и циклических закономерностей в расположении единиц генеральной совокупности. Чтобы избежать систематических ошибок, следует отбирать статистическую единицу, находящуюся в середине каждого интервала.

Этот способ очень удобен в тех случаях, когда нельзя заранее составить список единиц генеральной совокупности (выборка берется из постоянно формирующейся во времени совокупности). В таком случае, например, при изучении спроса на определенный товар, удобно наблюдать каждого десятого или каждого двадцатого входящего в магазин покупателя; или же при контроле качества продукции – проверять каждое пятое или каждое десятое изделие, сходящее с конвейера.

При определении средней ошибки механической выборки используются формулы средней ошибки при собственно-случайном бесповторном отборе:

$$\text{для выборочной средней } \mu_{\bar{x}} = \sqrt{\frac{S_{\bar{x}}^2}{n} \times \left(1 - \frac{n}{N}\right)};$$

$$\text{для выборочной доли } \mu_{\omega} = \sqrt{\frac{\omega \times (1 - \omega)}{n} \times \left(1 - \frac{n}{N}\right)}.$$

8.4.3. Расслоенный (стратифицированный) отбор *используется при изучении сложных совокупностей*, которые можно разбить на

несколько качественно однородных групп по существенным для целей исследования признакам. *Внутри каждой группы проводится собственно-случайный или механический отбор.* Полученные группы по численности единиц, как правило, не равны между собой, поэтому *отбор единиц осуществляется пропорционально объему группы*, т. е. количество отбираемых в выборку единиц пропорционально удельному весу данной группы по числу единиц в генеральной совокупности. Таким образом, число наблюдений по каждой группе определяется по формуле:

$$n_{ig} = N \cdot \frac{n_i}{N},$$

где n_{ig} - число наблюдений из i -ой группы генеральной совокупности,

N – объем генеральной совокупности,

n_i - объем i -ой группы генеральной совокупности.

Если пропорции между группами в выборке совпадают с пропорциями между группами в генеральной совокупности, то отбор называется **типическим**.

Типическая выборка обеспечивает более точные результаты по сравнению с другими способами отбора единиц в выборочную совокупность, так как позволяет исключить влияние межгрупповой дисперсии δ^2 на среднюю ошибку выборки.

На величину средней ошибки типической выборки влияет только величина средней из внутригрупповых дисперсий.

Типическую выборку можно получить повторным или бесповторным отбором:

Среднюю ошибку типической выборки при повторном отборе определяют по формулам:

- для средней количественного признака

$$\mu_{\bar{x}} = \sqrt{\frac{\bar{S}_i^2}{n}},$$

где \bar{S}_i^2 - средняя из внутригрупповых выборочных дисперсий;

- для доли альтернативного признака

$$\mu_{\omega} = \sqrt{\frac{\omega_i \times (1 - \omega_i)}{n}},$$

где $\omega_i \times (1 - \omega_i)$ - средняя из внутригрупповых дисперсий доли альтернативного признака по выборке.

При бесповторном отборе среднюю ошибку типической выборки рассчитывают по следующим формулам:

- для средней количественного признака

$$\mu_{\bar{x}} = \sqrt{\frac{S_i^2}{n} \times \left(1 - \frac{n}{N}\right)},$$

- для доли альтернативного признака

$$\mu_{\omega} = \sqrt{\frac{\omega_i \times (1 - \omega_i)}{n} \times \left(1 - \frac{n}{N}\right)}.$$

8.4.4. Серийная выборка применяется в тех случаях, когда *единицы статистической совокупности объединены в небольшие группы или серии*. В качестве таких серий могут рассматриваться, например, упаковки с определенным количеством готовой продукции.

Для отбора серий применяют либо собственно-случайную, либо механическую выборку. Наблюдению подвергаются все единицы отобранной серии.

Серийный отбор имеет большое практическое значение, так как обследуется незначительное число серий, и это сокращает расходы на проведение наблюдения; однако при серийном отборе случайная ошибка получается несколько большей, чем при других способах отбора.

При серийном отборе, поскольку внутри серий обследуются все без исключения статистические единицы, величина средней ошибки зависит только от межгрупповой (межсерийной) дисперсии.

Средняя ошибка серийной выборки при повторном отборе определяется следующим образом:

- для средней количественного признака:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta_{\bar{x}}^2}{r}},$$

$$\text{где } \delta_{\bar{x}}^2 = \frac{\sum (\bar{x}_i - \bar{x})^2}{r},$$

\bar{x}_i - среднее i -той серии,

\bar{x} - средняя по всей выборке,

r – число отобранных серий;

- для доли альтернативного признака:

$$\mu_{\omega} = \sqrt{\frac{\delta_{\omega}^2}{r}},$$

$$\text{где } \delta_{\omega}^2 = \frac{\sum (\omega_i - \bar{\omega})^2}{r} \text{ - межгрупповая дисперсия доли серийной}$$

выборки;

ω_i - доля признака в i -той доли;

$\bar{\omega}$ - общая доля признака во всей выборке.

При бесповторном отборе средняя ошибка серийной выборки может быть определена:

- для средней количественного признака:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta_{\bar{x}}^2}{r} \times \left(1 - \frac{r}{R}\right)},$$

где R – общее число серий в генеральной совокупности.

- для доли альтернативного признака:

$$\mu_{\omega} = \sqrt{\frac{\delta_{\omega}^2}{r} \times \left(1 - \frac{r}{R}\right)}.$$

Рассмотренные способы формирования выборки могут применяться в «чистом виде», а могут комбинироваться в различных сочетаниях и последовательности. Использование нескольких методов формирования выборки в одном выборочном исследовании называется **комбинированной выборкой** (отбором).

Такая выборка проводится в несколько этапов, и на каждом из них применяется свой способ отбора.

Например, при обследовании семейных доходов выборочное обследование проводится в такой последовательности:

- устанавливаются населенные пункты, попадающие под обследование. Используется расслоенный отбор. С его помощью отбираются крупные города, средние города, и другие населенные пункты;
- в каждом населенном пункте устанавливаются места, где проживают семьи - улицы, дома. Для этого используется механический отбор (по списку улиц и нумерации домов);
- в каждом месте проживания семей отбираются конкретные семьи, для чего применяется собственно-случайный бесповторный или механический отбор. Для отбора используют перечень номеров квартир или списки семей.

Формат: Список

Методы формирования выборки влияют на точность статистических оценок через ошибки выборки), а также на объем выборочной совокупности, на ее численность.

8.5. Численность выборки и способы распространения ее характеристик на Генеральную совокупность

Численность выборки – один из факторов, влияющих на величину ее ошибки: *чем она больше, тем меньше ошибка*. С другой стороны, с объемом выборки связаны затраты на проведение исследования: *чем она больше, тем больше затраты*.

Таким образом, **выборка должна быть оптимальной по численности**, чтобы обеспечить достоверность результатов исследования и не вызвать дополнительных затрат труда и денежных средств.

Численность выборки может быть определена исходя из допустимой ошибки при выборочном наблюдении, способа отбора статистических единиц.

Для определения необходимой численности выборки необходимо задаться предельной ошибкой выборки.

В общем случае предельная ошибка выборки связана с ее численностью следующим соотношением:

$$\Delta = t \times \mu_{\alpha} = t \times \sqrt{\frac{S_{\alpha}^2}{n}}, \text{ откуда } n = \frac{t^2 \times S_{\alpha}^2}{\Delta^2}.$$

Эта формула показывает, что с увеличением предполагаемой ошибки значительно уменьшается необходимый объем выборки и наоборот.

Для разных характеристик и разных методов формирования выборок формулы для определения необходимой численности выборки приведены в таблице 8.5.

Таблица 8.5.

Численность выборки при разных методах отбора

Метод отбора	Формулы определения объема выборки	
	Для средней	Для доли
Повторный	$n = \frac{t^2 \times S_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \times \omega \times (1-\omega)}{\Delta_{\omega}^2}$
Бесповторный	$n = \frac{t^2 \times S_{\bar{x}}^2 \times N}{N \times \Delta_{\bar{x}}^2 + t^2 \times S_{\bar{x}}^2}$	$n = \frac{t^2 \times \omega \times (1-\omega) \times N}{N \times \Delta_{\omega}^2 + t^2 \times \omega \times (1-\omega)}$

Н

а практике определение необходимого объема выборки часто составляет серьезную проблему, связанную с определением показателя вариации изучаемого признака. К началу проведения выборочного наблюдения показатели вариации неизвестны.

Приблизительно показатель вариации определяют одним из следующих способов:

- берут из предыдущих исследований;
- по правилу «трех сигм» общий размах вариации R при нормальном распределении укладывается в 6 среднеквадратических отклонений σ : $R \cong 6\sigma$, откуда $\sigma \approx \frac{R}{6}$; для бóльшей точности R делят на 5;

← Формат: Список

• если хотя бы приблизительно известна средняя величина изучаемого признака \bar{x} , то среднее квадратическое отклонение $\sigma \approx \frac{\bar{x}}{3}$;

• при изучении альтернативного признака, если нет других данных можно брать максимальную величину дисперсии, равную 0,25, то есть $\omega \times (1 - \omega) = 0,25$.

• Проводят «пробную» выборку, по которой рассчитывают показатель вариации, используемый в качестве оценки генеральной совокупности.

Рассмотрим следующий пример:

Исходя из технических условий установить оптимальный объем выборки для партии стальных листов, равной $N=2000$ шт., чтобы с вероятностью $P=0,954$ предельная ошибка не превышала 10% толщины листа. По техническим условиям толщина листа составляет 5 мм.

При этом среднее квадратическое отклонение толщины листа $S_x = \pm 1$ мм. Формирование выборки проведено методом бесповторного отбора.

Относительную предельную ошибку толщины листа переведем в абсолютную ошибку: $\Delta\tilde{x} = \frac{10 \times 5 \text{ мм}}{100} = 0,5 \text{ мм}$.

Рассчитаем оптимальный объем выборки для средней при бесповторном отборе на основе следующих данных: $N=2000$ шт., $S_x = \pm 1$ мм., $\Delta\tilde{x} = 0,5 \text{ мм}$., $t=2$ при $P=0,954$.

$$n_x = \frac{N \times t^2 \times S_x^2}{\Delta_{\tilde{x}}^2 \times N + t^2 \times S_x^2} = \frac{2000 \times 2^2 \times 1^2}{0,5^2 \times 2000 + 2^2 \times 1^2} = 148 \text{ листов}.$$

Таким образом, выборка численностью 148 листов обеспечивает заданную точность при бесповторном отборе.

Характеристики выборки могут быть распространены на генеральную совокупность с помощью одного из двух способов распространения выборочных данных:

- 1) способа прямого пересчета;
- 2) способа поправочных коэффициентов.

При первом способе средние величины и доли, полученные по выборке, переносятся на генеральную совокупность. При этом генеральная средняя определяется как $\bar{x} \cong \tilde{x} \pm \Delta\tilde{x}$, а генеральная доля – как $\bar{d} \cong \omega \pm \Delta\omega$.

Способ поправочных коэффициентов применяется, когда целью выборочного исследования является уточнение результатов сплошного наблюдения. Для этого после обобщения данных сплошного наблюдения практикуется 10%-ное выборочное наблюдение с установлением

поправочного коэффициента γ , который устанавливает процент расхождений между данными сплошного и выборочного наблюдения.

Например, при проведении сплошного учета уличных торговых мест в городе их было зарегистрировано $N=1000$ шт. С целью уточнения данных через полгода был проведен контрольный обход части города и зарегистрировано 210 уличных торговых мест. По данным сплошного учета их было 200 шт.

Необходимо уточнить число уличных торговых мест на новую дату.

$$\gamma = \frac{210}{200} = 1,05.$$

На новую дату число торговых мест составят
 $N' = N \times \gamma = 1000 \times 1,05 = 1050$.

9. Экономические индексы

9.1. Индексы и их использование в экономико-статистических исследованиях

Индексы используются в качестве обобщающих характеристик изучаемых явлений. В переводе с латинского “index” означает указатель, показатель.

Индексы являются относительными величинами, характеризующими изменение уровней простых или сложных социально-экономических явлений во времени, пространстве или по сравнению с планом, то есть это соответственно относительные показатели динамики (индексы динамики), относительные показатели сравнения (территориальные индексы) и относительные показатели плана и выполнения плана.

От обычных относительных показателей индексы отличаются тем, что характеризуют изменение не только простых, но и сложных явлений. Сложные явления состоят из непосредственно несоизмеримых элементов, а простые – только из однородных элементов.

Показатель, для которого рассчитывается индекс, называется *индексируемой величиной*. Так, в индексе себестоимости индексируемой величиной является себестоимость, в индексе физического объема – объем выпуска в натуральном выражении.

С помощью индексов решаются следующие задачи:

1. Оценка изменений сложных явлений и отдельных их частей (например, на сколько в текущем периоде изменился объем продаж по сравнению с предыдущим).
2. Определение влияния отдельных факторов на общую динамику сложного явления (например, влияние изменения цен на объем продаж), для чего используется индексный анализ.

В практической деятельности используются разнообразные индексы, которые можно классифицировать по следующим основаниям:

- *содержание изучаемых объектов (характер);*
- *степень охвата элементов совокупности;*
- *методы расчета.*

← - - - - - **Формат:** Список

По содержанию и характеру изучаемых показателей различают два вида индексов:

- *индексы количественных показателей (объемных)*

← - - - - - **Формат:** Список

К ним относятся индексы физического объема произведенной продукции, физического объема потребления и т.д. *Индексируемой величиной* в таких индексах является объемный показатель, измеряемый в натуральных единицах.

- *индексы качественных показателей*

← - - - - - **Формат:** Список

Эти индексы используются для измерения изменения показателя, рассчитываемого на единицу совокупности. Такие показатели называются качественными и характеризуют интенсивность изучаемого явления или процесса. *Индексируемой величиной* в индексах качественных показателей является уровень явления в расчете на единицу совокупности.

К индексам качественных показателей относятся индекс цен, себестоимости единицы продукции, трудоемкости, производительности труда и т.д.

По степени охвата элементов совокупности выделяют три формы индексов:

- индивидуальные,
- сводные (общие),
- групповые (субиндексы).

← - - - - **Формат:** Список

Индивидуальные индексы характеризуют изменение отдельных элементов, входящих в состав сложного явления. Это простая форма индексов (например, индивидуальный индекс цен отдельного вида товара).

Сводные индексы характеризуют изменение всего сложного явления, выражаемого сложным показателем. В таком явлении его элементы являются величинами несопоставимыми. Для решения проблемы несопоставимости индексируемых величин используются специальные показатели, называемые *соизмерителями индексируемых величин (статическими весами)*.

Групповые индексы (субиндексы) рассчитываются для определенной части элементов совокупности. Например, индекс физического объема по отдельным отраслям или территориям.

По методам расчета классифицируются только общие индексы. Они делятся на *агрегатные и средние*.

В *агрегатных индексах* числитель и знаменатель (величина и база сравнения) представляют собой набор или агрегат разнородных элементов («aggregatus» - складываемый, суммируемый).

Средние индексы используются в тех случаях, когда данных для построения агрегатных индексов недостаточно. Они рассчитываются на основе индивидуальных индексов и делятся на средние арифметические и средние гармонические индексы.

Для удобства применения индексов используется определенная символика и специальная терминология.

Каждая *индексируемая величина имеет свое обозначение:*

- q – количество продукции одного вида в натуральном выражении,
- p – цена единицы продукции,
- z – себестоимость единицы продукции,
- w – выработка продукции на 1-ого работника или в единицу времени,
- t – трудоемкость единицы продукции.

Индивидуальные индексы обозначаются следующими символами:

i_q - индивидуальный индекс физического объема,

i_p - индивидуальный индекс цен,

i_z - индивидуальный индекс себестоимости и т.д.

Общие (сводные) индексы имеют обозначения:

I_q - общий индекс физического объема,

I_p - общий индекс цен,

I_z - общий индекс себестоимости и т.д.

При расчете индексов используются **два вида данных**:

- **данные базисного уровня** – уровня, с которым производится сравнение; для их обозначения к символу соответствующего показателя добавляется «0».

- **данные текущего уровня** – уровня, который сравнивается - обозначаются добавлением «1» к символу соответствующего показателя.

В соответствии с принятыми обозначениями *индивидуальный индекс физического объема рассчитывается как $i_q = \frac{q_1}{q_0}$* , а *сводный индекс*

физического объема в агрегатной форме как $I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$ или

$$I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1}.$$

Индексы могут рассчитываться в виде коэффициентов или процентов.

9.2. Виды и формы индексов

Виды индексов выделяются *по виду индексируемой величины*. Различают **индексы физического показателя (объемные) индексы и индексы качественного показателя**.

Индексы физического показателя применяют для измерения изменения объемных показателей (объема продаж, численности работающих и т. п.).

Индексы качественного показателя используются для измерения изменений качественных показателей (цены, себестоимости единицы продукции и т. п.).

Формы индексов выделяются по степени охвата элементов совокупности. Элементами совокупности считаются её разнородные части. Например, предприятие выпускает несколько видов продукции. Каждый вид продукции – это отдельный элемент совокупности.

В практической деятельности применяют *три формы индексов: индивидуальные, общие (свободные) и групповые (субиндексы).*

Самая простая форма индексов – **индивидуальные**, так как они являются **обычными относительными величинами и представляют собой соотношение двух уровней индексируемой величины.**

Например, индивидуальный индекс физического объема $i_q = \frac{q_1}{q_0}$, где q_1, q_0 - количество произведенной одноименной продукции в текущем (отчетном) периоде и базисном. Этот индекс показывает, во сколько раз больше (меньше) в текущем периоде было произведено продукции по сравнению с базисным.

Индивидуальный индекс цен $i_p = \frac{p_1}{p_0}$, где p_1, p_0 - цена единицы продукции отчетного и базисного периодов, показывает, во сколько раз цена единицы продукции отчетного периода выше (ниже) соответствующей цены базисного периода.

Индивидуальный индекс стоимости $i_{pq} = \frac{p_1 \cdot q_1}{p_0 \cdot q_0}$, где p_1, q_1 - стоимость одноименной продукции отчетного периода, p_0, q_0 - стоимость одноименной продукции базисного периода, показывает, во сколько раз стоимость продукции отчетного периода больше (меньше) стоимости этой же продукции в базисном периоде.

Таким образом, *индивидуальный индекс показывает, во сколько раз индексируемая величина изменилась в отчетном (текущем) периоде по сравнению с базисным периодом.*

Сводные (общие) индексы характеризуют изменение всех элементов сложного явления.

Методика их расчета зависит от характера индексируемого показателя, качества исходных данных и целей исследования.

Сводные индексы рассчитываются двумя способами:

- как агрегатные;
- как средние из индивидуальных.

Средние индексы, в свою очередь, рассчитываются как средние арифметические и средние гармонические.

Из 2-х форм сводных индексов **основной является агрегатная форма.**

В числителе и знаменателе агрегатных индексов представлены несопоставимые элементы индексируемой величины. Для обеспечения сопоставимости при расчете используются специальные показатели – *соизмерители или веса индексов.*

← **Формат:** Список

Таким образом, агрегатный индекс строится как отношение сумм произведений индексируемой величины и показателя – соизмерителя, то есть по формуле:

$$I_x = \frac{\sum_{j=1}^n x_j^1 \cdot \omega_j}{\sum_{j=1}^n x_j^0 \cdot \omega_j},$$

где x_j^1, x_j^0 - текущее и базисное значение индексируемой величины j-ого элемента,

ω_j - показатель-соизмеритель явления j-ого элемента,

n – число элементов явления,

$x_j \cdot \omega_j$ - результирующий показатель для j-ого элемента.

Показатель-соизмеритель может относиться либо к текущему периоду, либо к базисному.

Если в качестве соизмерителя используется показатель текущего периода (отчетного), то формула для расчета агрегатного индекса выглядит следующим образом:

$$I_x = \frac{\sum_{j=1}^n x_j^1 \cdot \omega_1}{\sum_{j=1}^n x_j^0 \cdot \omega_1}.$$

Такая формула расчета была предложена в 1874 году Г.Пааше.

Если в качестве соизмерителя выступает показатель базисного периода, то формула для расчета принимает вид:

$$I_x = \frac{\sum_{j=1}^n x_j^1 \cdot \omega_0}{\sum_{j=1}^n x_j^0 \cdot \omega_0}.$$

Эту форму называют агрегатной формой индекса Э. Ласпейреса. Она была предложена в 1864 году.

При выборе формы агрегатного индекса необходимо решить три вопроса:

- *Выбрать индексируемую величину.*
- *Определить состав разнородных элементов, по которым рассчитывается индекс.*
- *Выбрать показатель – соизмеритель индексируемой величины (её вес).*

← **Формат:** Список

Выбор соизмерителя индексируемой величины определяется её характером (содержанием).

При построении агрегатного индекса количественного (объемного) показателя соизмерителем выступает качественный показатель; при

построении агрегатного индекса качественного показателя соизмерителем является количественный (объемный) показатель. Это означает, что агрегатные индексы качественных и количественных показателей рассчитываются по-разному.

9.3. Агрегатные индексы количественных показателей

К агрегатным индексам количественных показателей относятся **агрегатные индексы стоимости продукции** или товарооборота I_{pq} и **агрегатные индексы физического объема** I_q .

Агрегатный индекс стоимости продукции рассчитывается по формуле:

$$I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0},$$

то есть как отношение стоимости продукции текущего периода к стоимости продукции базисного периода.

Агрегатный индекс стоимости I_{pq} показывает, во сколько раз изменилась (возросла или уменьшилась) стоимость продукции или товарооборота отчетного периода по сравнению с базисным периодом.

Разность ($I_{pq} - 100$) показывает, на сколько % изменилась стоимость продукции отчетного периода по сравнению с базисным.

Разность числителя и знаменателя, т.е. $\Delta_{pq} = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_0$, показывает абсолютный прирост результативного показателя, т.е. на сколько денежных единиц изменилась стоимость продукции текущего периода по сравнению с базисным.

Расчет агрегатного индекса стоимости проведем по данным, приведенным в таблице 9.1. Таблица 9.1.

Объемы продаж и цена товаров в 2001-2002 гг.

Наименование товара	Объем продаж, шт.		Оптовая цена 1 шт., тыс. руб.		Расчет индексов			
	2001 г.	2002 г.	2001 г.	2002 г.	$p_0 \cdot q_0$	$p_1 \cdot q_1$	$p_0 \cdot q_1$	$p_1 \cdot q_0$
А	1000	1200	18	20	18000	24000	21600	20000
В	2000	2500	15	16	30000	40000	37500	32000
С	2000	1500	10	10	20000	15000	15000	20000
Итого	-	-	-	-	68000	79000	74100	72000

$p_0 \cdot q_0$ - стоимость продаж в 2001 году,

$p_1 \cdot q_1$ - стоимость продаж в 2002 году.

$$I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0} = \frac{79000}{68000} \cdot 100 = 116,2\%.$$

За счет изменения цен и физических объемов в 2002 году стоимость продаж по сравнению с 2001 годом увеличилась приблизительно в 1,16 раза или на 16,2%, что в стоимостном измерении составляет $\Delta_{pq} = 79000 - 68000 = 11000$ тыс. руб.

Численное значение индекса стоимости определяется двумя факторами:

- изменением количества (объема) товара;
- изменением цен.

← Формат: Список

Для того, чтобы оценить изменение стоимости только за счет одного фактора, необходимо устранить влияние другого фактора. Это можно сделать, если зафиксировать в формуле данный фактор неизменным, т.е. на уровне одного и того же периода.

Так, если объем продаж оценивать по одним и тем же ценам, то можно получить индекс, отражающий изменение только одного фактора – количества товара.

В этом случае **индексируемой величиной является количество товара** или его физический объем, а сам индекс называется **агрегатным индексом физического объема I_q** .

При его расчете в качестве статистических весов можно использовать цены базисного или отчетного периодов. Если выбираются цены базисного периода, то получают **агрегатный индекс физического объема в сопоставимых (базисных) ценах – индекс физического объема Ласпейреса**:

$$I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$$

Индекс Ласпейреса показывает, во сколько раз изменился физический объем продукции (товара) в текущем периоде по сравнению с базисным.

Числитель формулы $\sum q_1 \cdot p_0$ означает расчетную стоимость объема продаж текущего периода в неизменных базисных ценах; знаменатель $\sum q_0 \cdot p_0$ – фактическую стоимость продаж базисного периода.

Разность $(I_q - 1)$ показывает, на сколько % изменилась стоимость объема продаж за счет изменения его физического объема в текущем периоде по сравнению с базисным.

Разница между числителем и знаменателем $\Delta_{pq(a)} = \sum q_1 \cdot p_0 - \sum q_0 \cdot p_0$ есть абсолютное изменение стоимости продаж за счет изменения её физических (натуральных) объемов.

При исчислении агрегатного индекса физического объема в качестве соизмерителя индексируемой величины можно использовать цены текущего периода. В этом случае формула принимает вид:

$$I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1},$$

где $\sum q_1 \cdot p_1$ - стоимость объема продукции текущего периода в ценах текущего периода;

$\sum q_0 \cdot p_1$ - расчетная стоимость объема продаж базисного периода в ценах текущего периода.

Индекс, рассчитанный по приведенной формуле называется **агрегатным индексом физического объема Пааше**, и *показывает, во сколько раз изменился физический объем в текущем периоде по сравнению с базисным, если в базисном периоде цены были бы равны текущим.*

Рассчитаем индексы физического объема по данным, приведенным в таблице 9.1.:

- индекс Ласпейреса:

$$I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0} \cdot 100 = \frac{74100}{68000} \cdot 100 = 108,9\%;$$

Полученное значение индекса свидетельствует о том, что объем продаж за счет изменения физических объемов продаж увеличился на: $108,9 - 100 = 8,9\%$ или в 1,089 раза, что составляет $\Delta q \cdot p_{(q)} = 74100 - 68000 = 6100$ тыс. руб;

- индекс Пааше:

$$I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1} \cdot 100 = \frac{79000}{72000} \cdot 100 = 109,7\%;$$

объем продаж увеличился на $108,9 - 100 = 8,9\%$ или $\Delta q = 79000 - 72000 = 7000$ тыс. руб.

Индекс физического объема, рассчитанный по формулам Пааше и Ласпейреса, имеет разное значение. Численное значение индекса, рассчитанное по формуле Пааше всегда выше, чем рассчитанное по формуле Ласпейреса. Это связано с тем, что в формуле Ласпейреса при использовании в качестве соизмерителя неизменных цен базисного периода полностью устраняется влияние изменения цен на динамику объема продукции.

В формуле Пааше, т.е. при использовании в качестве соизмерителя нефиксированных цен текущего периода, устранить влияние изменения цен на динамику объема продукции не удастся. В связи с этим *использовать формулу Пааше для расчета агрегатного индекса физического объема продукции не рекомендуется.*

Помимо цен в качестве показателя-соизмерителя при построении индекса физического объема можно использовать трудоемкость и себестоимость единицы продукции.

Индекс, построенный с применением в качестве соизмерителя себестоимости, имеет следующий вид:

$$I_q = \frac{\sum q_1 \cdot z_0}{\sum q_0 \cdot z_0},$$

где $\sum q_1 \cdot z_0$ - расчетные издержки производства текущего периода по себестоимости базисного;

$\sum q_0 \cdot z_0$ - издержки производства базисного периода.

Индекс характеризует изменение издержек производства в результате изменения физического объема, а разность между числителем и знаменателем

$\Delta q \cdot z_{(q)} = \sum q_1 \cdot z_0 - \sum q_0 \cdot z_0$ - абсолютное изменение затрат (издержек производства) за счет изменения физического объема производства.

Аналогичным образом строится индекс физического объема с применением в качестве показателя-соизмерителя трудоемкости единицы продукции.

9.4. Агрегатные индексы качественных показателей

К агрегатным индексам качественных показателей относятся:

I_p - агрегатный индекс цен;

I_z - агрегатный индекс себестоимости;

I_t - агрегатный индекс трудоемкости;

I_w - агрегатный индекс производительности труда (выработки).

При построении перечисленных индексов в качестве показателя-соизмерителя используется связанный с индексируемой величиной количественный показатель.

Агрегатный индекс цен I_p характеризует изменение результирующего показателя (общей стоимости) за счет изменения цен в текущем периоде по сравнению с базисным. При его построении важно устранить влияние изменения количества товара, т.е. физического объема. Для этого в качестве соизмерителя индексируемой величины – цены используется неизменный физический объем либо отчетного, либо базисного периода. Таким образом, агрегатный индекс цен можно рассчитать по формуле Пааше и по формуле Ласпейреса:

$$I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} - \text{агрегатный индекс цен Пааше};$$

$$I_p = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} - \text{агрегатный индекс цен Ласпейреса}.$$

Индексы позволяют определить относительное изменение цен, но оно не будет одинаковым, так как имеет различное экономическое содержание.

Индекс Пааше показывает, во сколько раз изменился уровень цен на продукцию текущего периода, а разность между числителем и знаменателем $\Delta p \cdot q_{(p)} = \sum q_1 \cdot p_1 - \sum q_1 \cdot p_0$ - на сколько изменилась стоимость продукции в текущем периоде за счет изменения цен.

Индекс Пааше, рассчитанный по данным таблицы 8.1., равен $I_p = \frac{79000}{74100} = 1,066 = 106,6\%$.

Таким образом, за счет изменения цен объем продаж в текущем периоде увеличился в 1,066 раза или на 6,6%, что составляет в абсолютном исчислении:

$$\Delta p \cdot q_{(p)} = 79000 - 74100 = 4900 \text{ тыс. руб.}$$

Индекс Ласпейреса показывает, во сколько раз подорожала бы или подешевела бы продукция базисного периода из-за изменения цен на нее в отчетном периоде.

По данным предыдущего примера $I_p = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} = \frac{72000}{68000} = 1,059 = 105,9\%$.

Из-за изменения цен в текущем периоде стоимость продукции (объем продаж) 2001 года в 2002 году увеличилась бы в 1,059 раза, т.е. на 5,9% или на $\Delta p \cdot q_{(p)} = 72000 - 68000 = 4000$ тыс. руб.

Согласно практике индекс цен, исчисленный по формуле Пааше, всегда больше индекса, исчисленного по формуле Ласпейреса. Применение того или иного индекса зависит от цели исследования.

Если целью анализа является определение экономического эффекта (прибыль или убыток) от изменения цен в отчетном периоде по сравнению с базисными, то **используется индекс Пааше**.

Если целью анализа является прогнозирование объема продаж в связи с возможным изменением цен в предстоящем периоде, то **используется индекс Ласпейреса**, так как он позволяет определить стоимость продаж одного и того же физического объема базисного периода по новым ценам.

Достаточно часто в экономическом анализе используется ещё один вид общего индекса цен - **индекс Лоу (общий индекс на средних весах)**. В его формуле в качестве соизмерителя используется средний физический объем продаж \bar{q} , рассчитанный как простая средняя арифметическая

$$\bar{q}_i = \frac{q_{0i} + q_{1i}}{2};$$

$$I_{\bar{p}} = \frac{\sum p_1 \cdot \bar{q}}{\sum p_0 \cdot \bar{q}}.$$

Индекс Лоу используется в расчетах, связанных с закупкой или реализацией товаров в течение длительного периода (по долгосрочным

контрактам). Он показывает, во сколько раз в среднем изменился бы объем продаж за счет изменения цен.

По данным таблицы 9.1. средний объем продаж товара А составит $\bar{q}_A = \frac{1000+1200}{2} = 1100$ единиц; товара В - $\bar{q}_B = \frac{2000+2500}{2} = 2250$ единиц; а товара С - $\bar{q}_C = \frac{2000+1500}{2} = 1750$ единиц соответственно.

Индекс Лоу по данным таблицы 9.1. получается равным:

$$I_{\bar{p}} = \frac{20 \cdot 1100 + 16 \cdot 2250 + 10 \cdot 1750}{18 \cdot 1100 + 15 \cdot 2250 + 10 \cdot 1750} = \frac{22000 + 36000 + 17500}{19800 + 33750 + 17500} = \frac{75500}{71050} = 1,063, \quad -$$

это означает, что за счет изменения цен в 2002 году объем продаж в среднем увеличился бы в 1. 063 раза по сравнению с 2001 годом.

Достоинством индекса Лоу является то, что при его использовании устраняются недостатки индекса Пааше и Ласпейреса.

Кроме перечисленных индексов можно использовать «идеальный индекс» Фишера.

Идеальный индекс Фишера рассчитывается как средняя геометрическая из индексов цен Ласпейреса и Пааше:

$$I_{\bar{p}} = \sqrt{\frac{\sum p_1 \cdot q_1 \cdot \sum p_1 \cdot q_0}{\sum p_0 \cdot q_1 \cdot \sum p_0 \cdot q_0}} .$$

Идеальный индекс Фишера используется при исчислении индексов цен на длительный период времени для сглаживания тенденции в структуре и составе объема продукции, в которых происходят значительные изменения. Его недостатком является то, что он не имеет экономической интерпретации.

По данным таблицы 9.1. величина идеального индекса Фишера составит:

$$I_{\bar{p}} = \sqrt{1,066 \cdot 1,059} = \sqrt{1,129} = 1,062 .$$

Аналогично строятся индексы других качественных показателей. Например, агрегатный индекс себестоимости продукции рассчитывается следующим образом:

$$I_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1} ,$$

где $\sum z_1 \cdot q_1$ - затраты на производство продукции отчетного периода,

$\sum z_0 \cdot q_1$ - расчетные затраты на производство продукции текущего периода по себестоимости базисного.

Агрегатный индекс себестоимости продукции показывает, во сколько раз изменился уровень стоимости на продукцию отчетного периода, а разность между числителем и знаменателем $\Delta zq_z = \sum z_1 q_1 - \sum z_0 q_1$ показывает увеличение или снижение затрат за счет изменения себестоимости единицы продукции.

Таким образом, индексы качественных и количественных показателей показывают, как меняется результирующий показатель при изменении либо физического объема проданных (реализованных) товаров, либо цен (себестоимости) единицы товара.

Следует учитывать, что изменение цен и изменение физических объемов действуют на результирующий показатель одновременно. При этом направление действия указанных факторов и их интенсивность могут быть различными. Для оценки совместного их влияния на изменение результирующего показателя используются системы взаимосвязанных индексов, называемые индексными системами.

9. 5. Индексные системы и факторный анализ

В индексных системах отражается взаимосвязь экономических показателей: если экономические показатели связаны между собой определенным образом, то таким же образом связаны между собой и характеризующие их индексы, т.е. если $z = x \cdot y$, то $y_z = I_x \cdot I_y$.

Индексные системы дают возможность использовать индексный метод для изучения взаимосвязи показателей и проведения факторного анализа с целью определения влияния каждого фактора на результирующий показатель.

Построение индексной системы рассмотрим на примере индекса стоимости, индекса цен и индекса физического объема:

I_p и I_q являются факторными по отношению к индексу стоимости продукции.

Индекс стоимости рассчитывается по формуле: $I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0}$; индекс

цен рассчитаем по формуле Пааше: $I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}$; а индекс физического

объема – по формуле Ласпейреса: $I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$.

Перемножение индекса цен и индекса физического объема дает следующий результат:

$$I_p \cdot I_q = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0} = I_p \cdot I_q.$$

Таким образом: $I_p \cdot I_q = I_{pq}$.

Из предыдущих расчетов индекс стоимости $I_{pq} = 116,2\%$; индекс цен, рассчитанный по формуле Пааше, составляет $I_p = 106,6\%$; индекс физического объема, рассчитанный по формуле Ласпейреса, равен $I_q = 108,9\%$.

$$106,6 \cdot 108,9 = 116,2\% .$$

Таким образом, увеличение цен в текущем периоде на 6,6% и физических объемов на 8,9% привело в 2002 году к увеличению стоимости продукции на 16,2% по сравнению с 2001 годом.

Аналогична взаимосвязь других результативных признаков с факторными. Например, индекс объема продукции с индексом численности работающих и индексом производительности труда (выработки) связан таким же образом, как объем производства Q связан с выработкой одного работающего w и численностью работающих r .

Если $Q = w \cdot r$ то $I_{wr} = I_w \cdot I_r$.

$$I_{wr} = \frac{\sum w_1 \cdot r_1}{\sum w_0 \cdot r_0} = \frac{\sum w_1 \cdot r_1}{\sum w_0 \cdot r_1} \cdot \frac{\sum r_1 \cdot w_0}{\sum r_0 \cdot w_0} = I_w \cdot I_r ;$$

где I_w - индекс производительности труда, рассчитываемый по формуле Ласпейреса;

I_r - индекс численности работающих, рассчитываемый по формуле Пааше.

Индексные системы используются для определения влияния отдельных факторов на формирование уровня результативного показателя, позволяют по 2-м известным значениям индексов определить значение неизвестного.

Например, если известно, что затраты на производство возросли на 10%, а себестоимость продукции в среднем – на 5%, то можно определить, как изменится физический объем:

$$I_{zq} = 1,1, \quad I_z = 1,05 ,$$

$$I_q = \frac{I_{zq}}{I_z} = \frac{1,1}{1,05} = 1,05 = 105\% .$$

Рассмотренные индексные системы относятся к двухфакторным, но результативный признак можно разложить и на большее число факторов и соответственно получить многофакторные индексные системы, которые могут разложить изменение результативного показателя на элементы, вызванные влиянием отдельных факторов.

Индексные системы позволяют разложить и абсолютное изменение результативного показателя на составляющие, вызванные влиянием разных факторов, т.е. разложить абсолютное изменение по факторам. Это можно сделать, если результативный показатель представляет собой произведение количественного фактора на качественный.

Абсолютное изменение результативного показателя определяется как разница между числителем и знаменателем формулы расчета индекса стоимости $\Delta pq = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_0$.

Абсолютное изменение результативного показателя за счет изменения цен рассчитывается как $\Delta pq_p = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_1$.

Абсолютное изменение результативного показателя за счет изменения физического объема составит $\Delta pq_q = \sum p_0 \cdot q_1 - \sum p_0 \cdot q_0$.

Сложение абсолютного изменения результативного показателя за счет изменения цен и абсолютного изменения результативного показателя за счет изменения физического объема дает следующий результат:
 $\Delta pq_p + \Delta pq_q = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_1 + \sum p_0 \cdot q_1 - \sum p_0 \cdot q_0 = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_0 = \Delta pq$.

Следовательно $\Delta pq = \Delta pq_p + \Delta pq_q$.

Из предыдущих расчетов известно, что $\Delta pq = 11000$ тыс. руб.;
 $\Delta pq_p = 4900$ тыс. руб.;

$\Delta pq_q = 6100$ тыс. руб. $4900 + 6100 = 11000$ тыс. руб.

В отчетном периоде стоимость продукции увеличилась на 11 млн. руб., из которых на 4900 тыс. руб. – за счет изменения цен и на 6100 тыс. руб. – за счет изменения физического объема.

9.6. Средние индексы

Средние индексы – это вторая форма исчисления общих индексов, применяемая в случаях, когда невозможно вести учет показателей в натуральных измерителях (коммерческие организации, торговля, где в основном ведется стоимостной учет), или в плановых расчетах.

Во всех случаях, когда информация о физических объемах в натуральном исчислении отсутствует, для определения изменения показателей используется средняя форма индексов.

В практических расчетах используются два вида средних индексов:

- **средний индекс качественного показателя,**
- **средний индекс физического объема.**

Каждый из средних индексов может быть рассчитан по формулам средней арифметической взвешенной или средней гармонической взвешенной.

Средний индекс физического объема используется в тех случаях, когда отсутствует информация об объемах выпуска в натуральных измерителях.

Средняя арифметическая форма индекса физического объема применяется, когда имеется информация о стоимости реализованной продукции в базисном периоде, и об индивидуальных индексах физического объема i_q .

Формулу среднеарифметического индекса физического объема можно получить на основе агрегатного индекса физического объема Ласпейреса

$$I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0},$$

заменив физические объемы текущего периода на их выражение через индивидуальный индекс физического объема i_q .

$$i_q = \frac{q_1}{q_0}, \text{ следовательно } q_1 = i_q \cdot q_0;$$

$$\text{тогда } \bar{I}_q = \frac{\sum i_q \cdot q_0 \cdot p_0}{\sum q_0 \cdot p_0},$$

где i_q - усредняемая величина, а $p_0 \cdot q_0$ - статистический вес.

Полученная формула является формулой среднеарифметического индекса.

Среднеарифметический индекс показывает, во сколько раз в среднем изменится физический объем в планируемом (предстоящем) периоде. Таким образом, среднеарифметический индекс физического объема есть средний из индивидуальных индексов физического объема.

Разница между числителем и знаменателем характеризует изменение стоимости продукции в планируемом периоде.

$$\Delta p q_q = \sum i_q \cdot p_0 \cdot q_0 - \sum q_0 \cdot p_0.$$

Пример расчета среднеарифметического индекса физического объема.

Известна стоимость реализованной продукции в текущем периоде (товаров А, В и С) и планируемое изменение физических объемов в предстоящем периоде, в % (таблица 9.2.). Необходимо определить среднее изменение общего объема продаж в планируемом периоде.

Таблица 9.2.

Данные для расчета среднего индекса

Вид продукции	Продукция текущего периода, руб.	к. тыс.	Изменение физического объема планируемом периоде, %	Вспомогательные расчеты	
				$i_q = \frac{q_1}{q_0}$	$i_q \cdot p_0 \cdot q_0$
А	24000		+10	1,1	26400
В	40000		+5	1,05	42000
С	15000		-20	0,8	12000
Σ	79000		-	-	80400

По данным таблицы 9.2 рассчитаем для каждого товара индивидуальный индекс: $i_{q_a} = \frac{100+10}{100} = 1,1$;

$$i_{q_b} = \frac{100+5}{100} = 1,05;$$

$$i_{q_c} = \frac{100-20}{100} = 0,8.$$

Тогда величина среднеарифметического индекса физического объема составит $\bar{I}_q = \frac{\sum i_q \cdot q_0 \cdot p_0}{\sum q_0 \cdot p_0} = \frac{80400}{79000} = 1,018$, - в планируемом периоде

прирост объема продукции в среднем по данной группе товаров составит приблизительно 1,8%, что соответствует общей сумме прироста объема продаж $\Delta p q_q = 80400 - 79000 = 1400$ тыс. руб. Этот прирост произойдет за счет изменения физических объемов продаж.

Средний индекс физического объема можно рассчитать по формуле средней гармонической взвешенной. Она применяется в случае, если исходная информация представлена индивидуальными индексами физического объема i_q (или их легко рассчитать), или фактической стоимостью продукции текущего периода $q_1 \cdot p_1$.

Формула среднего геометрического индекса физического объема может быть получена из агрегатной формы общего индекса физического объема Пааше:

$$I_q = \frac{\sum q_1 \cdot p_1}{\sum q_0 \cdot p_1},$$

которая показывает, во сколько раз изменяется стоимость продукции за счет изменения физических объемов.

В указанной формуле физические объемы базисного периода q_0 заменяются их выражением через индивидуальный индекс физического объема:

$$\text{Если } i_q = \frac{q_1}{q_0} \text{ то } q_0 = \frac{q_1}{i_q}.$$

Таким образом, средний гармонический индекс физического объема рассчитывается по формуле:

$$\bar{i}_q = \frac{\sum q_1 \cdot p_1}{\sum \frac{1}{i_q} \cdot q_1 \cdot p_1},$$

где i_q - усредняемая величина;

$p_1 \cdot q_1$ - статистический вес.

Разница между числителем и знаменателем дает показатель среднего изменения стоимости в текущем периоде за счет изменения физического объема:

$$\Delta pq_p = \sum q_1 \cdot p_1 - \sum \frac{1}{i_q} \cdot q_1 \cdot p_1.$$

Пример расчета: необходимо определить среднее изменение стоимости продукции за счет изменения физических объемов продаж в текущем периоде по данным, приведенным в таблице 9.3.

Таблица 9.3.

Данные для расчета среднего геометрического индекса физического объема

Вид продукции	Продукция текущего периода, $p_1 \cdot q_1$, тыс. руб.	Изменение в текущем периоде, %	i_q	$\frac{1}{i_q}$	$\frac{1}{i_q} \cdot q_1 \cdot p_1$
А	24000	+20%	1,2	0,833	20000
В	40000	+12,5%	1,25	0,80	32000
С	15000	-25%	0,75	1,33	20000
Итого	79000	-	-	-	72000

$$I_q = \frac{\sum q_1 \cdot p_1}{\sum \frac{1}{i_q} \cdot q_1 \cdot p_1} = \frac{79000}{72000} = 1,097 = 109,7\%.$$

Общий прирост стоимости продукции в текущем периоде за счет изменения физических объемов составил $\Delta pq_p = 79000 - 72000 = 5000$ тыс. руб. Такое же значение прироста стоимости продукции получаем по формуле агрегатного индекса физического объема Пааше.

Содержание и расчет **среднего индекса качественного показателя** рассмотрим на примере цен.

Общий индекс цен в средней арифметической форме используется в плановых расчетах (при прогнозировании). Информация для расчета должна быть представлена в виде индивидуальных индексов цен или планируемых изменений цен и стоимости продукции базисного периода (отчетного).

Формулу для расчета общего индекса цен в средней арифметической форме легко получить преобразованием формулы агрегатного индекса цен Ласпейреса

$$I_p = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0},$$

выразив цены отчетного периода p_1 через

индивидуальные индексы цен i_p и цены базисного периода p_0 : $i_p = \frac{p_1}{p_0}$,

следовательно $p_1 = i_p \cdot p_0$. Тогда **формула для расчета среднего арифметического индекса цен имеет вид:**

$$\bar{I}_p = \frac{\sum i_p \cdot p_0 \cdot q_0}{\sum p_0 \cdot q_0},$$

где i_p - усредняемая величина,

$p_0 \cdot q_0$ - статистический вес усредняемой величины.

Средний арифметический индекс цен показывает, во сколько раз в среднем изменится стоимость продукции предстоящего периода за счет изменения цен. Разность числителя и знаменателя $\Delta pq_p = \sum i_p q_0 \cdot p_0 - \sum q_0 \cdot p_0$ определяет общее изменение стоимости продукции предстоящего (планового) периода за счет изменения цен.

Пример расчета: известны объемы продаж и цены на продукцию по 3-м видам товаров, данные приведены в таблице 9.4.

Таблица 9.4.

Данные для расчета среднего арифметического индекса цен

Вид товара	Объем продаж в текущем периоде, шт.	Оптовые цены		Вспомогательные расчеты	
		Текущие p_0	Плановые p_1	i_p	$i_p \cdot p_0 \cdot q_0$
А	24000	20	22	1,1	26400
В	40000	16	20	1,25	50000
С	15000	10	10	1,0	15000
Итого	79000	-	-	-	91400

$$\bar{I}_p = \frac{91400}{79000} = 1,157 = 115,7\%.$$

В предстоящем периоде средний рост стоимости продукции за счет роста цен составит 115,7% или $\Delta pq_p = 91400 - 79000 = 12400$ тыс. руб.

Средняя гармоническая форма общего индекса цен используется, когда информация представлена в виде индивидуальных индексов цен или их изменений и стоимости продукции текущего периода $p_1 \cdot q_1$.

Формулу для расчета среднего гармонического индекса цен можно получить преобразованием агрегатного индекса цен Пааше, заменив цены базисного периода p_0 его выражением через индивидуальный индекс цен

i_p . Так как $i_p = \frac{p_1}{p_0}$, то $p_0 = \frac{p_1}{i_p}$; следовательно

$$\bar{I}_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} = \frac{\sum p_1 \cdot q_1}{\sum \frac{1}{i_p} \cdot p_1 \cdot q_1}.$$

Средний гармонический индекс цен \bar{I}_p показывает, во сколько раз в среднем изменилась стоимость продукции текущего периода за счет изменения цен:

Разность между числителем и знаменателем формулы показывает абсолютное изменение стоимости продукции за счет изменения цен

$$\Delta p = \sum q_1 \cdot p_1 - \sum \frac{1}{i_p} \cdot q_1 \cdot p_1.$$

Пример расчета индекса цен в гармонической форме приведен в таблице 9.5.

Таблица 9.5.

Расчет среднего гармонического индекса цен

Вид продукции	Стоимость продукции, тыс. руб.		Изменение цен в текущем периоде, %	Расчет индексов	
	$p_0 \cdot q_0$	$p_1 \cdot q_1$		$i_p = \frac{p_1}{p_0}$	$\frac{1}{i_p} \cdot p_1 \cdot q_1$
А	18000	24000	11	1,11	21600
В	30000	40000	7	1,07	37500
С	20000	15000	Не изменились	1,00	15000
Итого	68000	79000	-	-	74100

По данным таблицы 9.5. $\bar{i}_p = \frac{79000}{74100} = 1,066 = 106,6\%$, следовательно, в текущем периоде изменение цен привело к росту стоимости продукции в 1,066 раза или на $\Delta p = 79000 - 74100 = 4900$ тыс. руб.

Литература:

1. Алан Р. Экономические индексы. М.; Статистика, 1980.
2. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование. М.; Финансы и статистика, 2001.
3. Гусаров В.М. Статистика: Учебное пособие для вузов. М.; ЮНИТИ-ДАНА, 2001.
4. Елисеева И.И., Юзбашев М.М. Общая теория статистика: Учебник/ Под ред. И.И. Елисеевой. М.; Финансы и статистика, 1995.
5. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. и др. Общая теория статистики: Учебник. М.; Инфра –М, 1998.
6. Общая теория статистики: Учебник/ Под ред. О.Э. Башиной, А.А. Спирина. М.; Финансы и статистика, 1999.
7. Практикум по теории статистики / Под ред. Р.А. Шмойловой. М.; Финансы и статистика, 1998.
8. Статистический словарь. М.; Финансы и статистика, 1989.
9. Теория статистики: Учебно-практическое пособие/ Под ред. В. Г. Минашкина. М.; Изд-во МЭСИ, 1998.
10. Шмойлова Р.А. Теория статистики: Учебник. М.; Финансы и статистика, 2001.

Тамара Митрофановна Сизова

Статистика

Учебное пособие

В авторской редакции
Санкт-Петербургский государственный университет информационных
технологий, механики и оптики
Зав. редакционно-издательским отделом Н.Ф. Гусарова
Лицензия ИД № 00408 от 05.11.99
Подписано к печати «___»___2005.
Отпечатано на ризографе Тираж 100 экз. Заказ № _____

