Министерство образования и науки Российской Федерации

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

ФАКУЛЬТЕТ ДИСТАНЦИОННОГО ОБУЧЕНИЯ (ФДО)

И. В. Подопригора

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебное пособие

Томск
2015

УДК 311(075.8)
ББК 60.60я73
П 444

Рецензенты:
Ярушкина Н. А., канд. экон. наук, доцент кафедры экономики Томского государственного архитектурно-строительного университета;
Золотарева Г. А., канд. экон. наук, доцент кафедры экономики ТУСУРа.

Подопригора И. В.
П 444 Общая теория статистики : учебное пособие / И. В. Подопригора. — Томск : факультет дистанционного обучения ТУСУРа, 2015. — 110 с.

В учебном пособии «Общая теория статистики» рассмотрены основные процедуры сбора, обработки и анализа массовых данных. Оно дает представление об основных статистических методах, их возможностях и границах применения.

Рассматриваются вопросы организации статистики, ее особенности в нашей стране. Приведены методы многомерной классификации, раскрыта основополагающая роль вариационного анализа. Описание теоретических основ выборочного метода и методов проверки статистических гипотез сочетается с примерами их использования в исследованиях и в работе органов государственной статистики. Особое внимание уделено регрессионному анализу, анализу временных рядов.

Для студентов высших учебных заведений, изучающих дисциплину «Статистика».

УДК 311(075.8)
ББК 60.60я73

ОГЛАВЛЕНИЕ

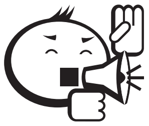
ВВЕДЕНИЕ

Термин «статистика»[1] был введен в науку немецким ученым Готфридом Ахенвалем в 1746 году, и заменил собой название курса «Государствоведение», который преподавался в университетах Германии, положив тем самым начало развитию статистики как учебной дисциплины и науки. Несмотря на это, статистический учет велся намного раньше: проводились переписи населения в Древнем Китае, осуществлялось сравнение военного потенциала государств, велся учет имущества граждан в Древнем Риме и пр. [1].

У истоков статистической науки стояли две школы: *немецкая описательная* и *английская школа политических арифметиков*. Представители описательной школы (Конринг, Ахенваль, Шленцер) своей задачей считали описание достопримечательностей государства: территории, населения, климата, политического устройства, вероисповедания, торговли и т. п. — без анализа закономерностей и связей между явлениями. Представители школы политических арифметиков (Уильям Петти, Граунт, Галлей) своей главной задачей считали выявление на основе большого числа наблюдений различных закономерностей и взаимосвязей в изучаемых явлениях. Каждая школа развивалась своим путем, используя свои методы в исследованиях, но предмет изучения у них был общий — государство, общество и, в частности, массовые явления и процессы, происходящие в нем. Статистика сформировалась как наука в результате синтеза государствоведения и политической арифметики, причем от последней она взяла больше, поскольку статистика и в настоящее время призвана выявлять прежде всего различного рода закономерности в исследуемых явлениях.

Однако представители этих двух школ не дошли до теоретического обобщения практики учетно-статистических работ, до создания теории статистики. Эта задача была решена позднее, в XIX веке, бельгийским ученым Адольфом Кетле, который дал определение предмета статистики, раскрыл суть ее методов. Под влиянием идей Кетле возникло третье направление статистической науки — математико-статистическое, которое получило свое развитие в работах таких ученых, как англичане Гальтон, Пирсон, Госсет, Фишер, русские — Чебышёв, Марков, Ляпунов, Чупров и пр.

[1]От лат. *status* — состояние, положение вещей; первоначально термин употреблялся в значении «политическое состояние».



В настоящее время данный термин употребляется в четырех значениях [1–4]:

1) наука, которая изучает количественную сторону массовых явлений и процессов в неразрывной связи с их качественным содержанием — учебный предмет в высших и средних специальных учебных заведениях;
2) совокупность цифровых сведений, которые характеризуют состояние массовых явлений и процессов общественной жизни; статистические данные, представляемые в отчетности предприятий, организаций, отраслей экономики, а также публикуемые в сборниках, справочниках, периодической печати и в сети Интернет, которые являются результатом статистической работы;
3) отрасль практической деятельности («статистический учет») по сбору, обработке, анализу и публикации массовых цифровых данных о самых различных явлениях и процессах общественной жизни[1];
4) некий параметр ряда случайных величин, получаемый по определенному алгоритму из результатов наблюдений, например статистические критерии (критические статистики), применяющиеся при проверке различных гипотез (предположительных утверждений) относительно природы или значений отдельных показателей исследуемых данных, особенностей их распределения и пр.[2]

Как и любая другая наука, статистика имеет свой предмет и метод исследования.

Статистика изучает количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной или содержанием, а также исследует количественное выражение закономерностей общественного развития в конкретных условиях места и времени.

Такое изучение основывается на системе категорий и понятий, отражающих наиболее общие и существенные свойства, признаки, связи и отношения предметов и явлений объективного мира.

Предметом статистического изучения выступают совокупности — множества однокачественных варьирующих явлений.

[1]Эту деятельность на профессиональном уровне осуществляет государственная статистика — Федеральная служба государственной статистики (ФСГС) и система ее учреждений, организованных по административно-территориальному признаку, а также ведомственная статистика (на предприятиях, ведомствах, министерствах и т. д.). Информация ФСГС публикуется в специальных печатных изданиях, а также в сети Интернет (www.gks.ru) (дата обращения: 16.02.2015).

[2]Термин «статистика» как параметр, как статистический критерий употребляется преимущественно в математической статистике, некоторые из них (χ^2, t и др.) рассмотрены в соответствующих темах данного учебного пособия.

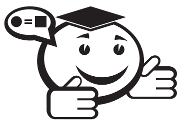
Цель (назначение) статистики — на основе регистрации фактов об отдельных единицах вывести общие закономерности, которые характеризуют некоторый объект или явление.

Задачи статистики состоят в следующем:

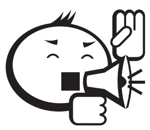
1) определение размеров явлений;
2) определение структуры совокупности;
3) определение динамики явлений;
4) определение степени выполнения плана;
5) определение взаимосвязи между отдельными явлениями;
6) прогнозирование социально-экономических явлений и т. д.

Соглашения, принятые в книге

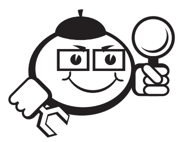
Для улучшения восприятия материала в данной книге используются пиктограммы и специальное выделение важной информации.



Эта пиктограмма означает определение или новое понятие.



Эта пиктограмма означает внимание. Здесь выделена важная информация, требующая акцента на ней. Автор здесь может поделиться с читателем опытом, чтобы помочь избежать некоторых ошибок.



Пример

Эта пиктограмма означает пример. В данном блоке автор может привести практический пример для пояснения и разбора основных моментов, отраженных в теоретическом материале.

Контрольные вопросы по главе

Глава 1

ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

Рассмотрим основные понятия, используемые в статистике.

1. *Статистическая совокупность* — множество социально-экономических объектов или явлений общественной жизни, объединенных качественной основой, но отличающихся друг от друга отдельными признаками. Таковы, например, совокупность домохозяйств, семей, предприятий и т. п.

2. *Единица совокупности* — первичный элемент статистической совокупности, являющийся носителем признаков и основой ведущегося при обследовании счета.

3. *Признак единицы совокупности* — свойства единицы совокупности, которые различаются способами их измерения и другими особенностями, что дает основание для их классификации (табл. 1.1).

Любой социально-экономический объект может быть охарактеризован с четырех сторон [5]:

- качественной;
- количественной;
- пространственной;
- временной.

Например: население Российской Федерации на 01.01.2014 составляло 146,1 млн чел.

Так, «население» — это качественная характеристика изучаемого явления, «146,1 млн чел.» — количественная, «Российская Федерация» — пространственная и «01.01.2014» — временная.

4. *Статистический показатель* — это понятие, отображающее количественные характеристики (размеры) или соотношения признаков общественных явлений.

5. *Система статистических показателей* — совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями.

Таблица 1.1 – Основная классификация признаков в статистике

Параметр классификации	**Вид признака**	**Пример признака**
По характеру выражения	Описательные (атрибутивные)	Цвет волос человека
	Количественные (числовые)	Рост человека
По способу измерения	Первичные (объемные)	Вес человека
	Вторичные (расчетные)	Производительность труда
По характеру вариации	Альтернативные	Пол человека
	Дискретные	Возраст человека
	Интервальные	Возраст группы людей
По отношению ко времени	Моментные	Количество денег в кармане человека
	Периодные	Заработная плата человека за месяц

Совокупность приемов, пользуясь которыми статистика исследует свой предмет, составляет *метод* статистики. Можно выделить четыре группы статистических методов (этапов статистического исследования):

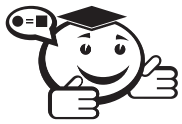
1) статистическое наблюдение;
2) обработка и сводка (группировка) полученных данных;
3) научный анализ исследуемых явлений;
4) интерпретация (представление) результатов исследования.

Обработка и сводка собранных первичных данных и их группировка, обобщение и оформление в таблицах составляет второй этап статистического исследования. Выделяют три основных формы представления обработанных статистических данных: текстовую, табличную и графическую.

На третьем этапе статистического исследования на основе данных сводки осуществляется *научный анализ исследуемых явлений*: рассчитываются различные обобщающие показатели в виде средних и относительных величин, выявляются определенные закономерности в распределениях, динамике показателей и т. п. На основе выявленных закономерностей делаются прогнозы на будущее.

И в заключение анализа производится интерпретация полученных данных, т. е. представление их в форме, доступной для понимания широкого круга потребителей информации.

1.1 Статистическое наблюдение



Статистическое наблюдение *— это научно-организованный сбор данных об изучаемой совокупности путем регистрации заранее намеченных признаков с целью получения обобщающих характеристик.*

Основные формы, виды и способы статистического наблюдения представлены на рисунке 1.1 [6].

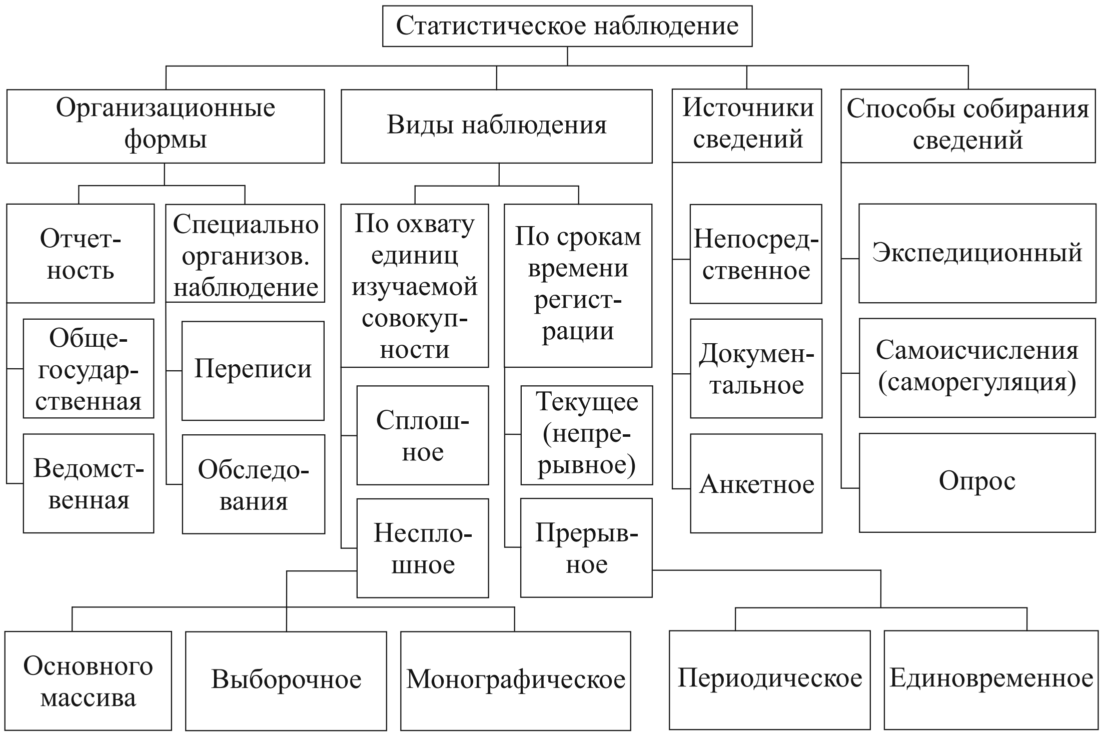


Рис. 1.1 – Классификация форм, видов и способов статистического наблюдения

Отчетность — это форма статистического наблюдения, при которой предприятия или иные хозяйствующие субъекты в установленном виде и в определенные сроки предоставляют в статистические органы необходимую информацию.

Существуют следующие виды сбора информации:

- *Сплошное наблюдение* — все единицы изучаемой совокупности подлежат обследованию.
- *Несплошное наблюдение* — обследуется часть совокупности, отобранная определенным способом.
- *Наблюдение основного массива* — статистическое наблюдение, при котором изучению подвергаются наиболее крупные единицы совокупности, которые вместе взятые имеют преобладающий удельный вес по обследуемому признаку.

- *Выборочное наблюдение* — основано на принципе случайного отбора единиц совокупности и дальнейшем их исследовании.
- *Монографическое наблюдение* — это глубокое детальное изучение отдельных единиц совокупности, имеющих значительные отличия по сравнению со всеми ее элементами.
- *Непосредственное наблюдение* — необходимые сведения получают путем подсчета, измерения и взвешивания единиц совокупности.
- *Экспедиционный способ* наблюдения состоит в том, что обученные и специально привлеченные работники посещают каждую единицу и собирают сведения.
- *Самоисчисление (саморегуляция)* сведения заполняются самими единицами совокупности в соответствии с указаниями.

Статистическое наблюдение проводится по *программе*, в которой указывают цель и задачи наблюдения; определяют объект и единицу наблюдения; разрабатывают формуляры и инструкции; определяют организационные вопросы: место и время наблюдения.

Объектом наблюдения называется совокупность, о которой должны быть собраны сведения.

Определение объекта наблюдения связано с нахождением границ совокупности в пространстве, во времени и по материальной сущности.

Граница по сущности устанавливается так называемым «цензом».

Ценз — это определенная количественная характеристика, которая служит для ограничения объекта наблюдения.

Единица наблюдения — это часть объекта наблюдения, которая является носителем признаков, подлежащих изучению.

При наблюдении заполняется статистический бланк, называемый *формуляром*, который заполняется на основании *инструкции*.

Время наблюдения предусматривает:

- *срок наблюдения* — это время, в течение которого осуществляется сбор информации;
- *критический момент* — момент времени, на который производится регистрация собираемых данных.

Сбор информации осуществляется *органами наблюдения*, которые состоят из конкретных исполнителей, осуществляющих подготовку наблюдения, проведение, и лиц, несущих ответственность за эту работу.

Точность статистического наблюдения определяется степенью соответствия значения какого-либо признака явления, найденного с помощью наблюдения, действительному его значению.

При сборе статистических сведений возможны ошибки при заполнении формуляров, которые разделяются на *случайные* и *систематические*. Систематические ошибки имеют тенденциозный характер и бывают преднамеренными и непреднамеренными.

Устраняются ошибки наблюдения посредством проведения *логического* или *арифметического* контроля собранных данных.

1.2 Статистическая сводка и группировка

*Под **статистической сводкой** понимают научное обобщение первичного статистического материала, полученного в ходе наблюдения с помощью подсчетов, выполняемых по определенной схеме.*

Для более наглядного представления полученной информации ее обычно группируют [7].

***Группировка данных** — это разделение элементов изучаемой совокупности на однородные группы по определенным признакам.*

В зависимости от решаемых задач выделяют следующие виды группировок:

- *типологическая группировка* — это расчленение совокупности на однородные группы по типам экономических явлений;
- *структурная группировка* предназначена для изучения состава однородной совокупности по определенному варьирующему показателю;
- *аналитическая группировка* выявляет взаимосвязи между изучаемыми явлениями и их признаками. Особенность аналитической группировки состоит в том, что ее единицы сгруппированы по факторному признаку и каждая группа охарактеризована средними величинами результативного признака.

  Факторный признак — это признак-причина, под воздействием которого изменяется *результативный* признак — признак-следствие;
- *простая группировка* — группировка, в которой группы выделены по одному признаку;
- *комбинированная группировка* — группировка, в которой расчленение совокупности на группы производится по нескольким признакам.

В основе группировки могут лежать как количественные, так и качественные признаки.

Если группировка строится по качественному признаку, производится лишь подсчет единиц совокупности, которые имеют значение признака. Данная группировка называется *атрибутивным рядом распределения*.

Количественный группировочный признак может быть:

- *дискретным* — принимать только определенные значения. Группировка по дискретному признаку проводится так же как и группировка по качественному признаку, с образованием *дискретного вариационного ряда распределения*;
- *непрерывным*, т. е. принимать любые значения в заданном интервале. В этом случае образуется *интервальный вариационный ряд* распределения.

При построении интервальной группировки определяется число групп и величина интервала.

Число групп совокупности (n) может быть задано:

- *логическим способом* — определяется исследователем субъективно;
- *механическим способом* — с учетом заданной величины интервала;
- *аналитическим способом* — с использованием формулы Стерджесса:

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

где N — число единиц совокупности.

Величина интервала (i) в случае равномерного ряда распределения (i = const) определяется по формуле:

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

где $x_{\max}$, $x_{\min}$ — соответственно максимальное и минимальное значения признаков совокупности.

Для того чтобы построить неравномерный интервальный ряд, используется механический метод увеличения интервала с использованием методики арифметической или геометрической прогрессии, или логистический подход.

Интервалы группировок могут быть закрытыми и открытыми.

Закрытый интервал — это интервал, у которого имеются нижняя и верхняя границы.

Открытый интервал — это тот интервал, у которого указана только одна граница: верхняя — для начального; нижняя — для последнего.

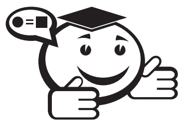
В зависимости от используемых при построении группировки данных их можно разделить на два типа:

- *первичная группировка*, построенная на данных наблюдений;
- *вторичная группировка*, полученная вследствие перегруппировки данных первичной группировки, которая базируется на изменении интервалов первичной группировки.

Для того чтобы сгруппировать единицы совокупности, характеризуемые несколькими признаками, используется метод кластерного анализа. Значения каждого из признаков служат координатами каждой единицы изучаемой совокупности в многомерном пространстве признаков. Каждое наблюдение, характеризующееся значениями нескольких признаков, можно представить как точку в пространстве этих признаков, значение которых рассматривается как координаты в многомерном пространстве. Основным критерием группировки (кластеризации) является то, что различия между группами (кластерами) должны быть более существенными, чем между единицами, отнесенными к одной группе.

Вследствие высокой трудоемкости кластерного анализа его целесообразно выполнять с применением ЭВМ [8].

1.3 Статистические таблицы и графики



***Статистическая таблица** — таблица, содержащая сводную характеристику исследуемой совокупности по одному или нескольким существенным признакам, взаимосвязанным логикой статистического анализа.*

Макет таблицы — заполненный заголовками скелет таблицы. Схема макета таблицы представлена на рисунке 1.2.

Название таблицы (общий заголовок)

<table>
<tr><td>Содержание строк</td><td colspan="3">Наименование граф
(верхние заголовки)</td><td></td></tr>
<tr><td>А</td><td>1</td><td>2</td><td>3 и т. д.</td><td></td></tr>
<tr><td rowspan="2">Наименование строк
(боковые заголовки)</td><td></td><td></td><td>Клетка
таблицы</td><td></td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>Итоговая строка</td><td></td><td></td><td></td><td>Итоговая
графа</td></tr>
</table>

Рис. 1.2 – Схема макета статистической таблицы

Подлежащим таблицы называется объект, характеристика которого представлена в таблице. Обычно, подлежащее таблицы дается в левой части, в наименовании строк.

Сказуемое таблицы — система показателей, которыми характеризуется объект изучения, т. е. подлежащее таблицы, сказуемое, в большинстве случаев, формирует верхние заголовки и составляет содержание граф.

Шапка таблицы — совокупность подлежащего и сказуемого таблицы.

В зависимости от структуры подлежащего таблицы могут быть *простые* и *сложные*. Сложные таблицы в свою очередь подразделяются на *групповые* и *комбинационные* [9].

Простой называется таблица, в подлежащем которой дается перечень каких-либо объектов или территориальных единиц.

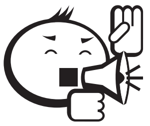
В *групповых* таблицах подлежащее содержит группировку единиц совокупности по одному количественному или атрибутивному признаку.

Комбинационная таблица позволяет охарактеризовать типические группы, выделенные по нескольким признакам, и связь между ними.

По структурному строению сказуемого различают таблицы с простой и сложной его разбивкой.

При *простой разработке сказуемого* показатель, определяющий его, не подразделяется на подгруппы и итоговые значения получаются путем простого суммирования значений по каждому признаку отдельно, независимо друг от друга.

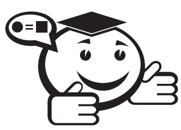
Сложная разработка сказуемого предполагает деление признака, формирующего его на подгруппы. При этом получается более полная и подробная характеристика объекта.



Правила построения таблиц.

1. Таблица должна быть компактной и содержать только те данные, которые непосредственно отражают исследуемое явление в статике или динамике и необходимы для познания его сущности.
2. Заголовки таблицы должны быть четкими, краткими, лаконичными; представлять собой законченное целое, органично вписывающееся в содержание текста.
3. Графики и строки следует нумеровать. Графы, заполненные наименованием строк, принято обозначать буквами алфавита, а все последующие графы — цифрами в порядке возрастания.
4. При изложении цифрового материала должна соблюдаться одинаковая степень точности для всех чисел (одинаковое округление).
5. При заполнении таблиц используются следующие условные обозначения: при отсутствии явления «–», если нет информации о явлении «. . .» или «нет сведений», если изучаемое значение признака не имеет осмысленного содержания «*x*».

При наличии информации по изучаемому явлению, числовое значение которого составляет величину меньше принятой в таблице точности, принято записывать «0,00».



Статистический график *— это чертеж, на котором статистические совокупности, характеризуемые определенными показателями, описываются с помощью условных геометрических образов или знаков.*

Основные элементы графика включают в себя:

- *графический образ* — геометрические знаки, с помощью которых изображаются статистические показатели;
- *поле графика* — часть плоскости, где расположены графические образы;
- *пространственные ориентиры* графика задаются в виде системы координатных сеток. Система координат необходима для размещения геометрических знаков в поле графика;
- *масштабные ориентиры* графика определяются масштабом и системой масштабных шкал;

- *масштаб* — мера перевода числовой величины в графическую;
- *масштабной шкалой* называется линия, отдельные точки которой могут быть прочитаны как определенные числа;
- *экспликация* включает в себя название графика, подписи вдоль масштабных шкал, пояснения и условные обозначения графика.

По способу построения статистические графики делят на *диаграммы* и *статистические карты*.

В зависимости от круга решаемых задач выделяют диаграммы сравнения, структурные диаграммы, диаграммы динамики и диаграммы распределения и концентрации величин. Статистические карты по графическому образу делят на картограммы и картодиаграммы [10].

В соответствии с графическими знаками различают точечные, линейные, плоскостные, объемные и фигурные графики.

Основные типы диаграмм сравнения:

1. *На столбиковых* диаграммах статистические данные изображаются в виде вытянутых по вертикали прямоугольников (рис. 1.3). Построение столбиковой диаграммы требует применения вертикальной масштабной шкалы.

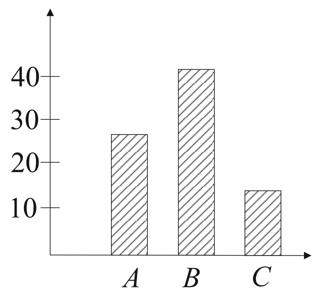


Рис. 1.3 – Пример столбиковой диаграммы

2. *Ленточные диаграммы* состоят из прямоугольников, расположенных горизонтально. В этом случае шкала — горизонтальная ось (рис. 1.4).

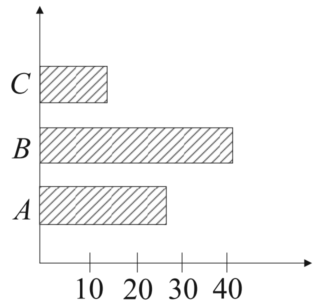


Рис. 1.4 – Пример ленточной диаграммы

3. *Плоские диаграммы* отображают величину изучаемого показателя площадью фигуры.

4. *Знаки В. Е. Варзара* — прямоугольные плоские диаграммы для графического изображения трех показателей, один из которых является произведением двух других.

5. *Объемные диаграммы* — отображаемая величина показывается объемом фигуры.

6. *Радиальные диаграммы* строятся на базе полярных координат. Началом отсчета в них служит центр окружности, а носителями масштабных шкал являются радиусы круга. Обычно в основе радиальных диаграмм лежат повторяющиеся годовые циклы с помесячными или поквартальными данными. Могут быть замкнутыми — отображается один цикл или спиральными — отображаются непрерывно несколько циклов.

7. *Фигурные диаграммы*. При их построении статистические данные изображаются рисунками — символами, которые в наибольшей степени соответствуют существу отображаемых явлений.

8. *Структурные диаграммы* служат для отображения структур статистических совокупностей по соотношению удельных весов. Структурные диаграммы могут быть представлены в виде круговых (секторных) диаграмм или структурного квадрата. В этих диаграммах площадь круга (квадрата) принимается за величину всей изучаемой совокупности, а площади отдельных частей отображают удельный вес ее составных частей.

При построении круговой структурной диаграммы исходят из соотношения 1% = 3,6°, квадратной — 1% = одной клетке.

9. *Диаграммы* распределения и концентрации — это графическое отображение вариационных рядов распределения. К диаграммам распределения относятся полигон распределения и гистограмма.

Диаграммы концентрации представлены кумулятой (кривой сумм), огивой, кривой Лоренца.

При построении *кумуляты* по оси абсцисс откладываются значения изучаемых признаков, а по оси ординат — соответствующие признакам суммы накопленных частот.

Огива строится по тем же данным, что и кумулята, только на оси абсцисс откладываются суммы накопленных частот, а по оси ординат — соответствующие значения признака.

Координатами *кривой Лоренца* являются: по оси абсцисс — суммы накопленных частот $\sum f_i$ в процентах к итогу $\left((\sum f_i / \sum f) \cdot 100\right)$; по оси ординат — суммы накопленных объемов признака $(\sum x_i f_i)$, в процентах к итоговому суммарному объему $\left((\sum x_i \cdot f_i)/(\sum x \cdot f)\right)$. Построенная в системе координат по оси абсцисс от 0 до 100% и по оси ординат от 0 до 100% кривая Лоренца сравнивается с диагональю квадрата, как это показано на рисунке 1.5. Степень концентрации показывается степенью отклонения от диагонали.

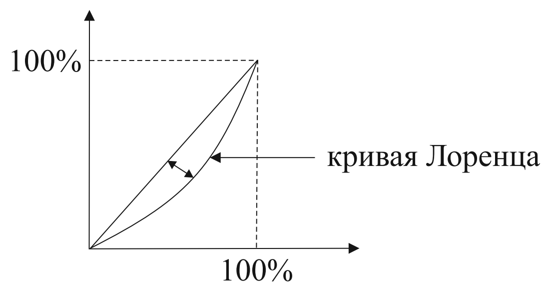


Рис. 1.5 – Кривая Лоренца

1.4 Организация статистики в Российской Федерации

Главным статистическим органом РФ является государственный комитет Российской Федерации по статистике (Госкомстат России). В положении о Госкомстате России (1994 г.) сказано, что он является федеральным органом исполнительной власти, осуществляющим руководство российской статистикой. Госкомстат России, его органы в республиках, краях, областях, районах и городах, подведомственные предприятия, организации, учебные заведения составляют единую систему государственной статистики.

Основными задачами Госкомстата являются: предоставление официальной статистической информации Президенту, правительству, федеральным органам законодательной и исполнительной власти, общественности, международным организациям; разработка научнообоснованной статистической методологии, соответствующей международным стандартам; координация статистической деятельности федеральных и территориальных органов исполнительной власти; разработка и анализ экономико-статистической информации; составление национальных счетов балансовых расчетов [11].

В соответствии с возложенными на него задачами Госкомстат России организует сбор необходимой информации, ее обработку и хранение; согласовывает программы проведения отраслевых (ведомственных) статистических наблюдений; обеспечивает функционирование Единого государственного регистра предприятий и организаций, общероссийских классификаторов; взаимодействует с региональными и отраслевыми информационно-вычислительными системами; внедряет новейшие технологии обработки информации; выпускает справочные информационно-аналитические издания.

Госкомстат и его местные органы утверждают минимум показателей и форм статистической отчетности, проводят работу по совершенствованию организации и методологии учета и статистики.

Государственные статистические органы РФ всех уровней независимы от административного давления.

Госкомстат России вправе получать статистическую отчетность и иные необходимые материалы от всех юридических и других хозяйствующих субъектов; издавать в установленном порядке постановления и инструкции по вопросам статистики и осуществлять контроль за их исполнением.

Правовая ответственность предприятий, учреждений и организаций за нарушением дисциплины в области статистической отчетности определена Законом Российской Федерации N 2761-1 от 13 мая 1992 г. «Об ответственности за нарушение порядка предоставления государственной статистической отчетности».

Наряду с общегосударственной статистикой имеет место ведомственная статистика, основанная на первичном учете, который ведется на отдельных предприятиях. Данные ведомственной статистики необходимы прежде всего для планирования деятельности подведомственных предприятий. Ведомственная статистика проводится под единым методологическим началом и руководством Госкомстата России [12].

Контрольные вопросы по главе 1

1. Что такое статистическое наблюдение?
2. Какие формы, виды и способы статистического наблюдения Вы знаете?
3. Что понимается под статистической сводкой и группировкой данных?
4. Какие виды группировок Вы знаете?
5. Какие признаки могут лежать в основе группировок?
6. Что представляет собой статистическая таблица?
7. Какие существуют виды статистических графиков? Что они отображают?
8. Каковы основные задачи Госкомстата?

Глава 2

АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ

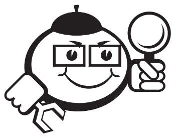
2.1 Абсолютные величины

Результаты статистических наблюдений представляют собой ***абсолютные величины****, отражающие уровень развития какого-либо явления или процесса.*

Абсолютные величины всегда имеют свою единицу измерения (размерность), присущую изучаемому явлению [13].

Широко распространены следующие *виды единиц измерения*:

1) *натуральные*, представленные мерой, счетом, весом и т. д. (например, штуки, тонны, метры);
2) *условно-натуральные* (например, общая масса энергоносителей — дрова, торф, каменный уголь, нефтепродукты, природный газ — измеряется в *т.у.т.* — тонны условного топлива, поскольку каждый его вид имеет разную теплотворную способность, а за стандарт принято 29,3 МДж/кГ);
3) *трудовые*, предназначенные для измерения количества отработанного времени (например, человеко-часы, человеко-дни, станко-часы и т. д.);
4) *стоимостные*, позволяющие соизмерить в денежной форме товары, которые нельзя соизмерить в натуральной форме (доллары США, рубли и т. д.).



Пример

Перевести в тонны условного топлива 23,8 млн т нефти с теплотворной способностью 45 Мдж/кГ.

Решение:

Учитывая стандартную теплотворную способность 29,3 МДж/кГ, определяем:

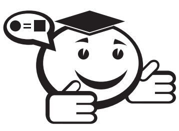
$$\frac{23{,}8 \cdot 45}{29{,}3} = 36{,}55 \text{ млн т. у. т.}$$

Количество единиц с одинаковым значением признака обозначается f и называется *частотой*[1]. Очевидно, что, суммируя число всех единиц с одинаковыми значениями признака, получаем N, то есть:

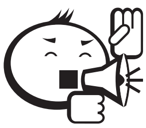
$$\sum f = N.$$

Анализируя абсолютные величины, например статистические данные о торговле, необходимо сопоставлять эти данные во времени и пространстве, исследовать закономерности их изменения и развития, изучать структуру совокупностей. С помощью абсолютных величин эти задачи невыполнимы, в этом случае необходимо использовать относительные величины.

2.2 Относительные величины



Относительная величина *— это результат деления (сравнения) двух абсолютных величин. В числителе дроби стоит величина, которую сравнивают, а в знаменателе — величина, с которой сравнивают (база сравнения).*



Относительные величины могут быть выражены в виде:

- *коэффициентов*, которые показывают, во сколько раз сравниваемая абсолютная величина больше базисной. В данном случае база сравнения (знаменатель относительного показателя) принимается за единицу;
- *процентов* (%) — в случае, если основание принимается за 100;
- *промилле* (‰) — в случае, если основание принимается за 1000.

[1] f — это начальная буква англ. слова *frequency* — частота.

Выбор той или иной формы относительной величины зависит от ее абсолютного значения:

- если сравниваемая величина больше базы сравнения в 2 раза и более, то выбирают форму коэффициента;
- если относительная величина близка к единице, как правило, ее выражают в процентах;
- если относительная величина значительно меньше единицы (близка к нулю), ее выражают в промилле.

Различают следующие *виды относительных величин*:

1. ***Индекс динамики*** характеризует изменение какого-либо явления во времени. Он представляет собой отношение значений одной и той же абсолютной величины в разные периоды времени. Данный индекс определяется по формуле:

$$i_Д = \frac{X_1}{X_0},$$

где цифры означают: 1 — отчетный или анализируемый период, 0 — прошлый или базисный период.

Критериальным значением индекса динамики служит единица (или 100%), то есть если $i_Д > 1$, то имеет место рост (увеличение) явления во времени; если $i_Д = 1$ — стабильность; если $i_Д < 1$ — наблюдается спад (уменьшение) явления. Еще одно название индекса динамики — *индекс изменения*, вычитая из которого единицу (100%), получают *темп изменения (динамики)*[1] с критериальным значением 0, который определяется по формуле:

$$T = i_Д - 1.$$

Если $T > 0$, то имеет место рост явления; $T = 0$ — стабильность; $T < 0$ — спад.

Разновидностями индекса динамики являются индексы планового задания и выполнения плана, рассчитываемые для планирования различных величин и контроля их выполнения [14].

Пример

Рассчитать индекс и темп изменения, если в марте произведено продукции 138 тонн, а в феврале — 108 тонн.

Решение:

1. Индекс изменения (динамики) по формуле: $i_Д = 138/108 = 1{,}278$ или 127,8% — рост, т. к. $i_Д > 1$.
2. Темп изменения по формуле: $T = 1{,}278 - 1 = 0{,}278$ или 27,8% — рост, т. к. $t_Д > 0$.

[1] Часто встречается и другое название темпа изменения — *темп прироста*.

2. ***Индекс планового задания*** — это отношение планового значения признака к базисному. Он определяется по формуле:

$$i_{\text{ПЗ}} = \frac{X_1'}{X_0},$$

где X_1' — планируемое значение; X_0 — базисное значение признака.

3. ***Индекс выполнения плана***. Для определения процента выполнения плана необходимо рассчитать отношение наблюдаемого значения признака к плановому (оптимальному, максимально возможному) значению по формуле:

$$i_{\text{ВП}} = \frac{X_1}{X_1'}.$$

Пример

Рассчитать индексы планового задания, выполнения плана и динамики, если выпуск продукции в отчетном году составил 20 млн рублей. На следующий год планировалось 28 млн рублей, а фактически получено 26 млн рублей.

Решение:

1. Индекс планового задания по формуле: $i_{\text{ПЗ}} = 28/20 = 1{,}4$.
2. Индекс выполнения плана по формуле: $i_{\text{ВП}} = 26/28 = 0{,}928$.
3. Индекс динамики по формуле $i_{\text{Д}} = 26/20 = 1{,}3$ или 130% — рост, т. к. $i_{\text{Д}} > 1$.

4. ***Индекс структуры (доля)*** — это отношение какой-либо части объекта (совокупности) ко всему объекту. Он определяется по формуле:

$$i_{\text{СТ}} = d = \frac{f}{\sum f}.$$

5. ***Индекс координации*** — это отношение какой-либо части объекта к другой его части, принятой за основу (базу сравнения). Он определяется по формуле:

$$i_{\text{К}} = \frac{f}{f_{\text{б}}}.$$

Пример

В составе ВВП региона 136,5 млрд рублей, произведено товаров на 75,4 млрд рублей, оказано услуг на 51,6 млрд рублей и собрано налогов 9,5 млрд рублей. Рассчитать относительные величины структуры и координации, приняв за основу производство товаров.

Решение:

1. Индексы структуры (доли) по формуле:
   - товары $i_{\text{СТ}} = 75{,}4/136{,}5 = 0{,}552$ или 55,2%;
   - услуги $i_{\text{СТ}} = 51{,}6/136{,}5 = 0{,}378$ или 37,8%;
   - налоги $i_{\text{СТ}} = 9{,}5/136{,}5 = 0{,}07$ или 7%. Контроль: $0{,}552 + 0{,}378 + 0{,}07 = 1$.
2. Индексы координации по формуле:
   - услуги $i_{\text{К}} = 51{,}6/75{,}4 = 0{,}684$;
   - налоги $i_{\text{К}} = 9{,}5/75{,}4 = 0{,}126$.

6. ***Индекс сравнения*** — это сравнение (соотношение) разных объектов по одинаковым признакам. Он определяется по формуле:

$$i_{\text{С}} = \frac{X_{\text{А}}}{X_{\text{Б}}},$$

где А, Б — сравниваемые объекты.

Пример

Запасы воды в озере Байкал составляют 23 000 км^3, а в Ладожском озере 911 км^3. Рассчитать относительные величины сравнения этих озер.

Решение:

1. Индекс сравнения озер Байкал с Ладожским по формуле: $i_{\text{С}} = 23\,000/911 = 25{,}25$.
2. Индекс сравнения Ладожского озера с Байкалом по той же формуле: $i_{\text{С}} = 911/23\,000 = 0{,}0396$ или $1/25{,}25 = 0{,}0396$.

7. ***Индекс интенсивности*** — это соотношение разных признаков одного объекта между собой. Он определяется по формуле:

$$i_{\text{ИН}} = \frac{X}{Y},$$

где X — один признак объекта; Y — другой признак этого же объекта.

Например, показатели выработки продукции в единицу рабочего времени, затрат на единицу продукции, цены единицы продукции и т. д.

Пример

Рассчитать относительную величину интенсивности ВВП в сумме 276 611 млн $ на душу населения в 147 млн человек.

Решение:

Показатель интенсивности по формуле $i_{\text{ИН}} = 276\ 611/147 = 1881{,}7$ $/чел.

Контрольные вопросы по главе 2

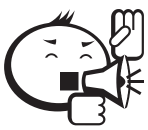
1. Что такое абсолютные величины? В каких единицах они измеряются?
2. Что представляют собой относительные величины? В каких единицах они измеряются?
3. Какие виды относительных величин существуют? Как они рассчитываются?

Глава 3

СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАЦИИ

3.1 Понятие средней величины

Статистическая совокупность содержит некоторое количество статистических величин, имеющих, как правило, разные значения и признаки, что делает невозможным сравнение нескольких совокупностей в целом. Для этой цели применяется ***средняя величина*** *как обобщающий показатель совокупности*, характеризующий уровень изучаемого явления или процесса.



Средняя величина всегда обобщает количественное выражение признака и погашает индивидуальные различия статистических величин совокупности, вызванные случайными обстоятельствами. Но по значению средней величины нельзя делать принципиальные выводы.

Главное значение средних величин состоит в их обобщающей функции, то есть замене множества различных индивидуальных значений признака средней величиной, характеризующей всю совокупность явлений [15].

3.2 Виды средних величин

Средняя арифметическая простая

Виды средних величин различаются прежде всего тем, какое свойство, какой параметр исходной варьирующей массы индивидуальных значений признака должен быть сохранен неизменным.

Средней арифметической величиной *называется такое среднее значение признака, при вычислении которого общий объем признака в совокупности сохраняется неизменным. Исходя из определения, формула средней арифметической величины имеет вид:*

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{\sum X_i}{N}.$$

Иначе можно сказать, что средняя арифметическая величина — среднее слагаемое. При ее вычислении общий объем признака мысленно распределяется поровну между всеми единицами совокупности.

По этой формуле вычисляются средние величины первичных признаков, если известны индивидуальные значения признака.

Средняя арифметическая взвешенная

Пример

Если изучаемая совокупность велика, исходная информация чаще представляет собой ряд распределения или группировку, как, например, в таблице 3.1.

Таблица 3.1 – Распределение студентов группы дневного отделения по возрасту

Возраст студентов, X	17	18	19	20	21
Количество студентов, f	3	5	7	4	2

Средний возраст должен представлять собой результат равномерного распределения общего (суммарного) возраста всех студентов. Общий (суммарный) возраст всех студентов, согласно исходной информации (табл. 3.1), можно получить как сумму произведений значений признака в каждой группе X_i, на число студентов с таким возрастом f_i (частоты).

Получим формулу:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i f_i}{\sum_{i=1}^{N} f_i},$$

где i *— число групп.*

Такую форму средней арифметической величины называют ***взвешенной арифметической средней***[1]. *В качестве весов здесь выступают количество единиц совокупности в разных группах. Название «вес» выражает тот факт, что разные значения признака имеют неодинаковую «важность» при расчете средней величины.*

Пример

«Важнее», весомее возраст студентов 18, 19, 20 лет, а такие значения возраста как 17, 20 или 21 при расчете средней не играют большой роли — их «вес» мал.

По формуле среднеарифметической взвешенной по данным таблицы 3.1 имеем:

$$\overline{X} = \frac{17 \cdot 3 + 18 \cdot 5 + 19 \cdot 7 + 20 \cdot 4 + 21 \cdot 2}{21} = \frac{396}{21} = 18{,}857 \text{ лет.}$$

Как видим, средняя арифметическая величина может быть дробным числом, если даже индивидуальные значения признака могут принимать только целые значения. Ничего необычного для метода средних в этом не заключено, так как из сущности средней не следует, что она обязана быть реальным значением признака, которое могло бы встретиться у какой-либо единицы совокупности.

Если при группировке значения усредняемого признака заданы интервалами, то при расчете средней арифметической величины в качестве значения признака в группах принимают середины этих интервалов, то есть исходят из предположения о равномерном распределении единиц совокупности по интервалу значений признака.

Для открытых интервалов в первой и последней группах, если таковые есть, значения признака надо определить экспертным путем исходя из сущности, свойств признака и совокупности. При отсутствии возможности экспертной оценки значения признака в открытых интервалах, для нахождения недостающей границы открытого интервала применяют размах (разность между значениями конца и начала интервала) соседнего интервала (*принцип «соседа»*) [11].

[1]Обычно (в т. ч. и в дальнейшем в данном пособии) в статистических формулах пределы суммирования не ставятся, а подразумеваются, т. е. подразумеваются именно такие пределы, как в формуле — с 1-ой группы по N-ю (последнюю).

Пример

Например, по данным таблицы 3.2 минимальную и максимальную величину веса студентов определить затруднительно, поэтому воспользуемся принципом «соседа» — применим размах соседнего интервала, который у второго и предпоследнего составляет 10 кг, значит, первый интервал будет от 50 до 60 кг, а последний — от 80 до 90 кг. Середины интервалов определяем как полусумму нижней и верхней границ интервалов.

Таблица 3.2 – Распределение группы студентов по весу

Группы студентов по весу, кг	**Количество студентов, чел.**	**Середина интервала X_i'**	$X_i' f_i$
До 60	6	55	330
60–70	8	65	520
70–80	5	75	375
Более 80	2	85	170
Итого	21	66,429	1395

Средний вес студентов, рассчитанный с заменой точных значений признака в группах серединами интервалов, составил:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i' f_i}{\sum_{i=1}^{N} f_i} = \frac{1395}{21} = 66{,}429 \text{ кг},$$

что и записано в итоговую строку в 3-м столбце таблицы 3.2. Следует обратить внимание, что объемный показатель (количество студентов) — это сумма, а итог по столбцам относительных показателей или средних групповых величин — средняя.

Свойства средних величин

Средняя арифметическая величина обладает *свойствами*, знание которых полезно как при ее использовании, так и при ее расчете.

1. Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие частоты:

$$\bar{x} \cdot \sum f_i = \sum x_i \cdot f_i.$$

2. Сумма отклонений индивидуальных значений признака от средней арифметической равна нулю:

$$\sum (x_i - \bar{x}) \cdot f_i = 0.$$

3. Если из всех вариантов признака вычесть одно и то же число, то средняя уменьшится на это число.

$$x_i - A = x', \quad \overline{x} = \overline{x}' + A,$$

где A — произвольное число.

4. Если все варианты признака разделить на одно и то же число, то средняя уменьшится в это число раз.

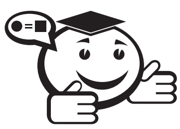
$$\frac{x_i}{k} = x', \quad \overline{x} = \overline{x}' \cdot k,$$

где k — произвольное число.

5. Средняя арифметическая может быть рассчитана методом условного нуля (методом моментов), при котором данные преобразуются до образования нуля последовательными операциями вычитания и деления вариантов признака на определенные числа (A и K):

$$x' = \frac{x - A}{k}, \quad \overline{x} = \overline{x}' \cdot k + A.$$

Средняя квадратическая

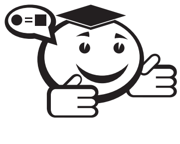


Если при замене индивидуальных величин признака на среднюю величину необходимо сохранить неизменную сумму квадратов исходных величин, то средняя будет являться ***квадратической средней величиной****. Ее формула следующая:*

$$\overline{X}_{\text{кв}} = \sqrt{\frac{\sum\limits_{i=1}^{N} X_i^2}{N}}.$$

Главной сферой применения квадратической средней является измерение вариации признака в совокупности.

Средняя геометрическая

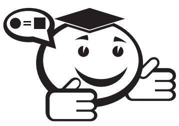


Если при замене индивидуальных величин признака на среднюю величину необходимо сохранить неизменным произведение индивидуальных величин, то следует применить ***геометрическую среднюю величину****, имеющую следующий вид:*

$$\overline{X}_{\text{геом}} = \sqrt[N]{X_1 \cdot X_2 \cdot \ldots \cdot X_N} = \sqrt[N]{\prod_{i=1}^{N} X_i}.$$

Основное применение средняя геометрическая находит при определении средних относительных изменений, о чем сказано в теме «Ряды динамики». Геометрическая средняя величина дает наиболее точный результат осреднения, если задача также состоит в нахождении такого значения признака, который качественно был бы равноудален как от максимального, так и от минимального значений признака.

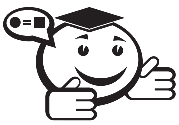
Средняя гармоническая



Когда статистическая информация не содержит частот f по отдельным вариантам X_i совокупности, а представлена как их произведение Xf, тогда применяется формула ***средней гармонической взвешенной****, для получения которой обозначим $Xf = w$, откуда $f = w/X$, и, подставив эти обозначения в формулу, получим формулу:*

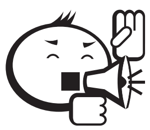
$$\overline{X}_{\text{гарм}} = \frac{\sum w_i}{\sum \frac{w_i}{X_i}} = \frac{w_1 + w_2 + \ldots + w_N}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots + \frac{w_N}{x_N}}.$$

Таким образом, средняя гармоническая взвешенная применяется тогда, когда неизвестны действительные веса f, а известно $w = Xf$.



В тех случаях, когда вес каждого варианта $w = 1$, то есть индивидуальные значения X встречаются по 1 разу, применяется формула ***средней гармонической простой****:*

$$\overline{X}_{\text{гарм}} = \frac{1 + 1 + \ldots + 1}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_N}} = \frac{N}{\sum \frac{1}{X_i}}.$$



Все рассмотренные выше виды средних величин принадлежат к общему типу *степенных средних*, имеющему следующий вид:

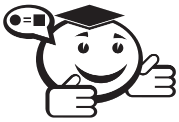
$$\overline{X} = \sqrt[m]{\frac{\sum X_i^m}{N}}.$$

При $m = 1$ получаем среднюю арифметическую; при $m = 2$ — среднюю квадратическую; при $m = 3$ — среднюю кубическую; при $m = 0$ — среднюю геометрическую; при $m = -1$ — среднюю гармоническую.

Чем выше показатель степени m, тем больше значение средней величины (если индивидуальные значения признака варьируют). В итоге, можно построить следующее соотношение, которое называется *правилом мажорантности средних*:

$$\overline{X}_{\text{ГМ}} \leqslant \overline{X}_{\text{геом}} \leqslant \overline{X}_{\text{арифм}} \leqslant \overline{X}_{\text{КВ}} \leqslant \overline{X}_{\text{куб}}.$$

Средняя хронологическая



В случае если необходимо посчитать среднее значение ряда данных, распределенных во времени, используют формулу:

- ***средней хронологической простой**, если моменты времени равноотстоят друг от друга:*

$$\overline{y} = \frac{\dfrac{y_1}{2} + y_2 + \ldots + y_{n-1} + \dfrac{y_n}{2}}{n-1};$$

- ***средней хронологической взвешенной**, если расстояния между уровнями, ряда динамики различны:*

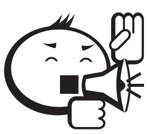
$$\overline{y} = \frac{(y_1 + y_2) \cdot t_1 + (y_2 + y_3) \cdot t_2 + \ldots + (y_{n-1} + y_n) \cdot t_{n-1}}{2(t_1 + t_2 + \ldots + t_{n-1})}.$$

3.3 Статистическое изучение вариации

Признаки, изучаемые статистикой, варьируются (отличаются друг от друга) у различных единиц совокупности в один и тот же период или момент времени.

Причиной *вариации* являются разные условия существования разных единиц совокупности. Например, огромное число причин влияет на масштабы внешней торговли различных стран мира.

Для управления и изучения вариации статистикой разработаны специальные методы исследования вариации, система показателей, с помощью которой вариация измеряется, характеризуются ее свойства.



Первым этапом статистического изучения вариации является построение *ряда распределения* (или *вариационного ряда*) — упорядоченного распределения единиц совокупности по возрастающим (чаще) или по убывающим (реже) значениям признака и подсчет числа единиц с тем или иным значением признака.

Существует три *вида* ряда распределения [16]:

1) *ранжированный ряд* — это перечень отдельных единиц совокупности в порядке возрастания изучаемого признака; если численность единиц совокупности достаточно велика, ранжированный ряд становится громоздким, и в таких случаях ряд распределения строится с помощью группировки единиц совокупности по значениям изучаемого признака (если признак принимает небольшое число значений, то строится дискретный ряд, а в противном случае — интервальный ряд);

2) *дискретный ряд* — это таблица, состоящая из двух столбцов (строк) — конкретных значений варьирующего признака X_i и числа единиц совокупности с данным значением признака f_i — частот; число групп в дискретном ряду определяется числом реально существующих значений варьирующего признака;
3) *интервальный ряд* — это таблица, состоящая из двух столбцов (строк) — интервалов варьирующего признака X_i и числа единиц совокупности, попадающих в данный интервал (частот), или долей этого числа в общей численности совокупностей (частостей).

Построим ряд распределения выручки по торговым точкам, для чего необходимо провести статистическое наблюдение, то есть собрать первичный статистический материал, который представляет собой величину выручки.

Результаты наблюдения за отчетный период представим в виде ранжированного по возрастанию величины выручки ряда распределения (табл. 3.3).

Таблица 3.3 – Выручка по 35 торговым точкам, тыс. д. е.

№ торговой точки	**Выручка (тыс. д. е.)**	**№ торговой точки**	**Выручка (тыс. д. е.)**	**№ торговой точки**	**Выручка (тыс. д. е.)**
1	24,16	13	54,12	25	65,31
2	27,06	14	54,91	26	69,24
3	29,12	15	55,74	27	71,39
4	31,17	16	55,91	28	77,12
5	37,08	17	56,07	29	79,12
6	39,11	18	56,80	30	84,34
7	41,58	19	56,93	31	86,89
8	44,84	20	57,07	32	91,74
9	46,80	21	58,39	33	96,01
10	48,37	22	59,61	34	106,84
11	51,44	23	59,95	35	111,16
12	52,56	24	62,05	Итого	2100,00

Построим интервальный ряд распределения выручки по торговым точкам, для чего необходимо выбрать оптимальное число групп (интервалов признака) и установить длину (размах) интервала. Поскольку при анализе ряда распределения сравнивают частоты в разных интервалах, необходимо, чтобы длина интервалов была постоянной[1]. Оптимальное число групп выбирается так, чтобы в достаточной мере отразилось разнообразие значений признака в совокупности и в то же время закономерность распределения, а его форма не искажалась случайными колебаниями частот. Если групп будет слишком мало, не проявится закономерность вариации; если групп будет чрезмерно много, случайные скачки частот исказят форму распределения.

[1] Если приходится иметь дело с интервальным рядом распределения с неравными интервалами, то для сопоставимости нужно частоты или частости привести к единице интервала. Полученное значение называется *плотностью* ρ, то есть $\rho = f/h$.

Чаще всего число групп в ряду распределения определяют по ***формуле Стерждесса****:*

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

где k *— число групп (округляемое до ближайшего целого числа);* N *— численность совокупности.*

Из формулы Стерджесса видно, что число групп — функция объема данных (N).

Зная число групп, рассчитывают длину (размах) интервала[1] по формуле:

$$h = \frac{X_{\max} - X_{\min}}{k},$$

где $X_{\max}$ и $X_{\min}$ — максимальное и минимальное значения в совокупности.

В нашем примере про внешнеторговый оборот (ВО) по формуле Стерждесса определим число групп:

$$k = 1 + 3{,}322 \lg 35 = 1 + 3{,}322 \cdot 1{,}544 = 6{,}129 \approx 6.$$

Рассчитаем длину (размах) интервала по формуле:

$$h = \frac{111{,}16 - 24{,}16}{6} = \frac{87}{6} = 14{,}5 \text{ тыс. д. е.}$$

Теперь построим интервальный ряд с 6 группами с интервалом 14,5 тыс. д. е. (см. первые 3 столбца табл. 3.4).

Существенную помощь в анализе ряда распределения и его свойств оказывает графическое изображение. Интервальный ряд изображается столбиковой диаграммой, в которой основания столбиков, расположенные по оси абсцисс, — это интервалы значений варьирующего признака, а высоты столбиков — частоты, соответствующие масштабу по оси ординат. Графическое изображение распределения таможенных постов в выборке по величине ВО приведено на рисунке 3.1. Диаграмма такого типа называется *гистограммой*[2].

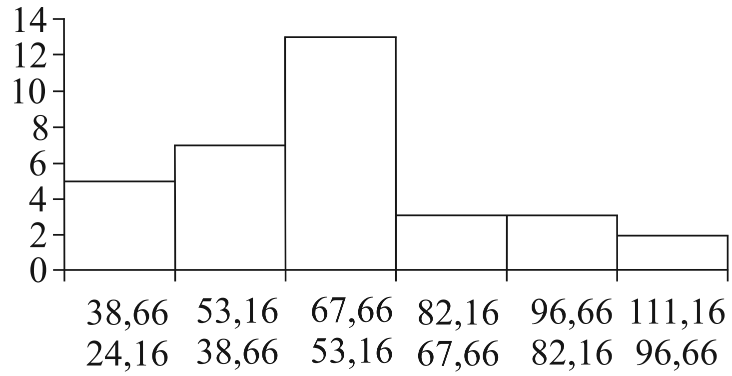


Рис. 3.1 – Гистограмма распределения

[1]Единицы совокупности, имеющие значение признака, равное границе интервала, включаются в тот интервал, где это точное значение впервые указывается.

[2]От греч. «*гистос*» — ткань, строение.

Таблица 3.4 – Интервальный ряд распределения выручки по торговым точкам (тыс. д. е.)

i	**Группы постов по величине ВО** X_i	**Число постов** f_i	**Середина интервала** X_i'	$X_i' f_i$	**Накопл. частота** f_i'	$\lvert X_i' - \overline{X}\rvert f_i$	$(X_i' - \overline{X})^2 f_i$	$(X_i' - \overline{X})^3 f_i$	$(X_i' - \overline{X})^4 f_i$
1	24,16–38,66	5	31,41	157,05	5	147,071	4326,001	–127 246,23	3 742 856,97
2	38,66–53,16	7	45,91	321,37	12	104,400	1557,051	–23 222,31	346 344,16
3	53,16–67,66	13	60,41	785,33	25	5,386	2,231	–0,92	0,38
4	67,66–82,16	4	74,91	299,64	29	56,343	793,629	11 178,84	157 461,90
5	82,16–96,66	4	89,41	357,64	33	114,343	3268,572	93 434,47	2 670 891,13
6	96,66–111,16	2	103,91	207,82	35	86,171	3712,758	159 966,81	6 892 284,32
	Итого	35		2128,85		513,714	13 660,243	114 110,66	13 809 838,86

Данные таблицы 3.4 и рисунка 3.1 показывают характерную для многих признаков форму распределения: чаще встречаются значения средних интервалов признака, реже — крайние (малые и большие) значения признака. Форма этого распределения близка к нормальному закону распределения, которое образуется, если на варьирующую переменную влияет большое число факторов, ни один из которых не имеет преобладающего значения [17].

Если имеется дискретный ряд распределения или используются середины интервалов (как в нашем примере про ВО — в таблице 3.4 в 4-м столбце рассчитаны середины интервалов как полусумма значений начала и конца интервала), то графическое изображение такого ряда называется *полигоном* (см. рис. 3.2)[1], которое получается соединением прямыми точек с координатами X_i и f_i.

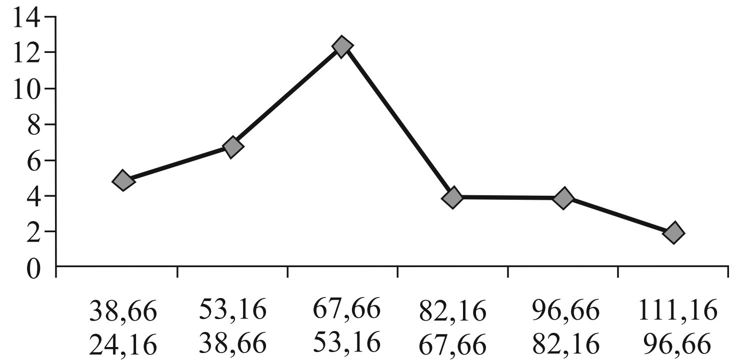


Рис. 3.2 – Полигон распределения

Вторым этапом статистического изучения вариации является расчет характеристик ряда распределения, которые описывают количественно его структуру, строение.

Медиана — *величина варьирующего признака, делящая совокупность на две равные части — со значением признака меньше медианы и со значением признака больше медианы*[2].

Пример

В нашем примере медиана — это 18-я торговая точка из 35-ти с величиной выручки 56,8 тыс. д. е. Из этого примера видно принципиальное различие между медианой и средней величиной: медиана не зависит от значений на краях ранжи-

[1] От греч. слов «*поли*» и «*гонос*» — многоугольник.

[2] При четном числе единиц совокупности за медиану принимают полусумму из двух центральных вариант.

рованного ряда. Даже если бы выручка 35-й торговой точки была в 10 раз больше, величина медианы не изменилась бы.

Поэтому медиану часто используют как более надежный показатель типичного значения признака, нежели средняя арифметическая, если ряд значений неоднороден, включает резкие отклонения от средней.

В интервальном ряду распределения для нахождения ***медианы*** *применяется формула:*

$$\mathrm{Me} = X_0 + h\frac{0{,}5\sum f - f'_{\mathrm{Me}-1}}{f_{\mathrm{Me}}},$$

где Me *— медиана;* X_0 *— нижняя граница интервала, в котором находится медиана;* h *— величина (размах) интервала;* $f'_{\mathrm{Me}-1}$ *— накопленная частота в интервале, предшествующем медианному;* f_{Me} *— частота в медианном интервале.*

Пример

В таблице 3.4 медианным является среднее из 35-ти значений, т. е. 18-е от начала значение выручки. Как видно из столбца накопленных частот (6-й столбец), оно находится в третьем интервале. Тогда:

$$\mathrm{Me} = 53{,}16 + 14{,}5 \cdot \frac{0{,}5 \cdot 35 - 12}{13} = 59{,}30 \text{ тыс. д. е.}$$

Аналогично медиане вычисляются значения признака, делящие совокупность на четыре равные по численности части — *квартили*, которые обозначаются заглавной латинской буквой Q с подписным значком номера квартиля. Ясно, что Q_2 совпадает с Me. Для первого и третьего квартилей приводим формулы и расчет по данным таблицы 3.4.

$$Q_1 = X_0 + h\frac{0{,}25\sum f - f'_{Q_1-1}}{f_{Q_1}} = 38{,}66 + 14{,}5 \cdot \frac{0{,}25 \cdot 35 - 5}{7} = 43{,}43 \text{ тыс. д. е.};$$

$$Q_3 = X_0 + h\frac{0{,}75\sum f - f'_{Q_3-1}}{f_{Q_3}} = 67{,}66 + 14{,}5 \cdot \frac{0{,}75 \cdot 35 - 25}{4} = 72{,}19 \text{ тыс. д. е.}$$

Так как Q_2 = Me = 59,30 тыс. д. е., видно, что различие между первым квартилем и медианой (−15,87) больше, чем между медианой и третьим квартилем (12,89). Этот факт свидетельствует о наличии некоторой несимметричности в средней области распределения, что заметно и на рисунке 3.2.

*Значения признака, делящие ряд на 5 равных частей, называются **квинтилями**, на 10 частей — **децилями**, на 100 частей — **перцентилями**. Эти характеристики применяются при необходимости подробного изучения структуры ряда распределения*[1].

Безусловно, важное значение имеет такая величина признака, которая встречается в изучаемом ряду распределения чаще всего. Такую величину принято называть *модой*.

*В дискретном ряду **мода** определяется без вычисления как значение признака с наибольшей частотой.*

Обычно встречаются ряды с одним модальным значением признака. Если в ряду распределения встречаются два или несколько равных (и даже несколько различных, но больших, чем соседние) значений признака, то он считается соответственно бимодальным или мультимодальным. Это свидетельствует о неоднородности совокупности, возможно, представляющей собой агрегат нескольких совокупностей с разными модами. В интервальном ряду распределения интервал с наибольшей частотой является модальным. Внутри этого интервала находят условное значение признака, вблизи которого *плотность распределения* (число единиц совокупности, приходящихся на единицу измерения варьирующего признака) достигает максимума. Это условное значение и считается *точечной модой*. Логично предположить, что такая точечная мода располагается ближе к той из границ интервала, за которой частота в соседнем интервале больше частоты в интервале за другой границей модального интервала.

Отсюда получаем обычно применяемую формулу:

$$\mathrm{Mo} = X_0 + h\frac{f_{\mathrm{Mo}} - f_{\mathrm{Mo}-1}}{(f_{\mathrm{Mo}} - f_{\mathrm{Mo}-1}) + (f_{\mathrm{Mo}} - f_{\mathrm{Mo}+1})},$$

где Mo — *мода;* X_0 — *нижнее значение модального интервала;* f_{Mo} — *частота в модальном интервале;* $f_{\mathrm{Mo}-1}$ — *частота в предыдущем интервале;* $f_{\mathrm{Mo}+1}$ — *частота в следующем интервале за модальным;* h — *величина интервала.*

[1]Получите формулы и произведите их расчет (по аналогии с формулами для расчета квартилей) самостоятельно.

Пример

По данным таблицы 3.4 рассчитаем точечную моду:

$$\text{Mo} = 53{,}16 + 14{,}5 \cdot \frac{13-7}{(13-7)+(13-4)} = 58{,}96 \text{ тыс. д. е.}$$

К изучению структуры ряда распределения средняя арифметическая величина также имеет отношение, хотя основное значение этого обобщающего показателя другое. В интервальном ряду распределения выручки по торговым точкам средняя арифметическая рассчитывается как взвешенная по частоте середина интервалов X (расчет числителя — в 5-м столбце таблицы 3.4) по формуле:

$$\overline{X} = \frac{\sum X_i' f_i}{\sum f_i} = \frac{2128{,}85}{35} = 60{,}82 \text{ тыс. д. е.}$$

Различие между средней арифметической величиной (60,82), медианой (59,30) и модой (58,96) в нашем примере невелико. Чем ближе распределение по форме к нормальному закону, тем ближе значения медианы, моды и средней величины между собой.

Третьим этапом статистического изучения вариации является расчет показателей размера и интенсивности вариации.

Простейшим показателем является размах вариации.

Размах вариации — *абсолютная разность между максимальным и минимальным значениями признака из имеющихся в изучаемой совокупности значений:*

$$H = X_{\max} - X_{\min}.$$

Поскольку величина размаха характеризует лишь максимальное различие значений признака, она не может измерять закономерную силу его вариации во всей совокупности.

Показателем силы вариации выступает ***средний модуль отклонений****, или* ***среднее линейное отклонение****:*

$$l = \frac{\sum \left| X_i - \overline{X} \right|}{N}.$$

Пример

В нашем примере (по данным табл. 3.4) среднее линейное отклонение вычисляется как взвешенное по частоте отклонение по модулю середин интервалов от средней арифметической величины (расчет числителя произведен в 7-м столбце таблицы 3.4), т. е.:

$$l = \frac{\sum \left| X_i' - \overline{X} \right| f_i}{\sum f_i} = \frac{513{,}714}{35} = 14{,}678 \text{ тыс. д. е.}$$

Это означает, что в среднем величина выручки в изучаемой совокупности торговых точек отклонялась от средней выручки на 14,678 тыс. д. е.

Простота расчета и интерпретации составляют положительные стороны показателя l, однако математические свойства модулей «плохие»: их нельзя поставить в соответствие с каким-либо вероятностным законом, в том числе и с нормальным распределением, параметром которого является не средний модуль отклонений, а *среднее квадратическое отклонение*, обозначаемое малой греческой буквой сигма (σ).

***Среднее квадратическое отклонение** вычисляется для ранжированного ряда по формуле:*

$$\sigma = \sqrt{\frac{\sum \left(X_i - \overline{X} \right)^2}{N}}$$

и для интервального ряда по формуле:

$$\sigma = \frac{\sum \left(X_i' - \overline{X} \right)^2 f_i}{\sum f_i}.$$

Пример

В нашем примере по данным таблицы 3.4 среднее квадратическое отклонение выручки составило (расчет числителя произведен в 8-м столбце таблицы 3.4):

$$\sigma = \sqrt{\frac{13\ 660{,}243}{35}} = \sqrt{390{,}293} = 19{,}756 \text{ тыс. д. е.}$$

Среднее квадратическое отклонение по величине в реальных совокупностях всегда больше среднего модуля отклонений. Разница между ними тем больше, чем больше в изучаемой совокупности резких, выделяющихся отклонений, что служит индикатором «засоренности» совокупности неоднородными с основной массой элементами.

Квадрат среднего квадратического отклонения представляет собой *дисперсию* отклонений. На использовании дисперсии основаны практически все методы математической статистики [17].

Формула ***дисперсии*** *для несгруппированных данных (простая дисперсия) имеет вид:*

$$\sigma^2 = \frac{\sum\left(X_i - \overline{X}\right)^2}{N} = \overline{X^2} - \overline{X}^2;$$

и для сгруппированных (взвешенная дисперсия):

$$\sigma^2 = \frac{\sum\left(X_i' - \overline{X}\right)^2 f}{\sum f_i} = \overline{X^2} - \overline{X}^2.$$

Еще одним показателем силы вариации, характеризующим ее не по всей совокупности, а лишь в ее центральной части, служит *среднее квартильное расстояние (отклонение)*, т. е. средняя величина разности между квартилями.

Среднее квартильное расстояние (отклонение) *определяется по формуле:*

$$q = \frac{(Q_3 - Q_2) + (Q_2 - Q_1)}{2} = \frac{Q_3 - Q_1}{2}.$$

Пример

В нашем примере:

$$q = \frac{72{,}19 - 43{,}43}{2} = 14{,}38 \text{ тыс. д. е.}$$

Сила вариации в центральной части совокупности, как правило, меньше, чем в целом по всей совокупности. Соотношение между средним линейным отклонением и средним квартильным расстоянием служит для изучения структуры вариации: большое значение такого соотношения свидетельствует о наличии слабоварьирующего «ядра» и сильно рассеянного вокруг него окружения в изучаемой совокупности.

Пример

Для нашего примера соотношение

$$\frac{l}{q} = 1{,}021,$$

что говорит о совсем незначительном различии силы вариации в центральной части совокупности и на ее периферии.

Для оценки интенсивности вариации и для сравнения ее в разных совокупностях и тем более для разных признаков необходимы *относительные показатели вариации*, которые вычисляются как отношение абсолютных показателей силы вариации, рассмотренных ранее, к средней арифметической величине признака, то есть показатели:

- *относительный размах вариации*: $\rho = H/\overline{X}$;
- *линейный коэффициент вариации*: $\lambda = l/\overline{X}$;
- *квадратический коэффициент вариации*: $\nu = \sigma/\overline{X}$;
- *относительное квартильное расстояние*: $d = q/\overline{X}$.

Пример

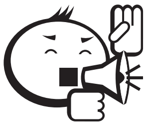
В нашем примере эти показатели составляют:

- $\rho = 87/60{,}82 = 1{,}43$, или 143%;
- $\lambda = 14{,}678/60{,}82 = 0{,}241$, или 24,1%;
- $\nu = 19{,}756/60{,}82 = 0{,}32$, или 32%;
- $d = 14{,}38/60{,}82 = 0{,}236$, или 23,6%.

Оценка степени интенсивности вариации возможна только для каждого отдельного признака и совокупности определенного состава, она состоит в сравнении наблюдаемой вариации с некоторой обычной ее интенсивностью, принимаемой за норматив[1]. Так, для совокупности торговых точек вариация выручки может быть определена как слабая, если $\nu < 25\%$, умеренная при $25\% < \nu < 50\%$ и сильная при $\nu > 50\%$.

[1] Максимально возможные значения показателей вариации: $l_{\max} = 2\overline{X} - 2\overline{X}/N$; $\sigma_{\max} = \bar{x}\sqrt{N-1}$; $\lambda_{\max} = 2 - 2/N$; $\nu_{\max} = \sqrt{N-1}$.

Различная сила, интенсивность вариации обусловлены объективными причинами, поэтому нельзя говорить о каком-либо универсальном критерии вариации (например, 33%), так как для разных явлений и признаков этот критерий различен.



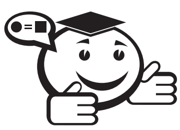
Четвертым этапом статистического изучения вариации является расчет моментов распределения и показателей его формы.

Для дальнейшего изучения характера вариации используются средние значения разных степеней отклонений отдельных величин признака от его средней арифметической величины. Эти показатели называются *центральными моментами распределения* порядка, соответствующего степени, в которую возводятся отклонения (табл. 3.5), или просто моментами [18].

Таблица 3.5 – Центральные моменты

Порядок момента	**Формула**	
	по несгруппированным данным	**по сгруппированным данным**
Первый μ_1	$\frac{\sum\left(X_i-\overline{X}\right)}{N}=0$	$\frac{\sum\left(X_i'-\overline{X}\right)f_i}{\sum f_i}=0$
Второй μ_2	$\frac{\sum\left(X_i-\overline{X}\right)^2}{N}=\overline{X^2}-\overline{X}^2=\sigma^2$	$\frac{\sum\left(X_i'-\overline{X}\right)^2 f_i}{\sum f_i}=\overline{X^2}-\overline{X}^2=\sigma^2$
Третий μ_3	$\frac{\sum\left(X_i-\overline{X}\right)^3}{N}$	$\frac{\sum\left(X_i'-\overline{X}\right)^3 f_i}{\sum f_i}$
Четвертый μ_4	$\frac{\sum\left(X_i-\overline{X}\right)^4}{N}$	$\frac{\sum\left(X_i'-\overline{X}\right)^4 f_i}{\sum f_i}$

Величина третьего момента μ^3 зависит, как и его знак, от преобладания положительных кубов отклонений над отрицательными кубами либо наоборот. При нормальном и любом другом строго симметричном распределении сумма положительных кубов строго равна сумме отрицательных кубов, поэтому на основе третьего момента строится показатель, характеризующий степень асимметричности распределения.



Коэффициент асимметрии:

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

Пример

В нашем примере показатель асимметрии составил (расчет числителя произведен в 9-м столбце таблицы 3.4):

$$As = \frac{114\,110{,}66}{19{,}756^3 \cdot 35} = 0{,}423 > 0, \text{ т. е. асимметрия значительна.}$$

Английский статистик К. Пирсон на основе разности между средней арифметической величиной и модой предложил другой показатель асимметрии:

$$As_{\Pi} = \frac{\overline{X} - \text{Mo}}{\sigma}.$$

Пример

В нашем примере по данным таблицы 3.4 показатель асимметрии составил:

$$As = \frac{60{,}82 - 58{,}96}{19{,}756} = 0{,}09.$$

Показатель асимметрии Пирсона зависит от степени асимметричности в средней части ряда распределения, а коэффициент асимметрии — от крайних значений признака. Таким образом, в нашем примере в средней части распределения наблюдается меньшая асимметрия, чем по краям. Распределения с сильной правосторонней и левосторонней асимметрией показаны на рисунке 3.3.

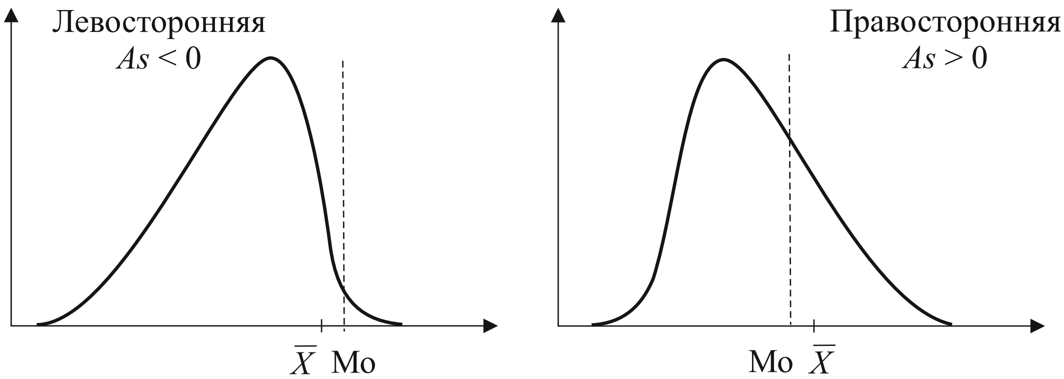


Рис. 3.3 – Асимметрия распределения

С помощью момента четвертого порядка характеризуется еще более сложное свойство рядов распределения — *эксцесс* (от англ. «излишество»).

Показатель эксцесса рассчитывается по формуле:

$$Ex = \frac{\mu_4}{\sigma_4} - 3.$$

Чаще всего эксцесс интерпретируется как «крутизна» распределения. Чтобы показать, в чем состоит эксцесс распределения, и правильно его интерпретировать, нужно сравнить ряды с одинаковой силой вариации (одной и той же величиной σ) и разными показателями эксцесса. Чтобы не смешать эксцесс с асимметрией, все сравниваемые ряды должны быть симметричными. Такое сравнение изображено на рисунке 3.4.

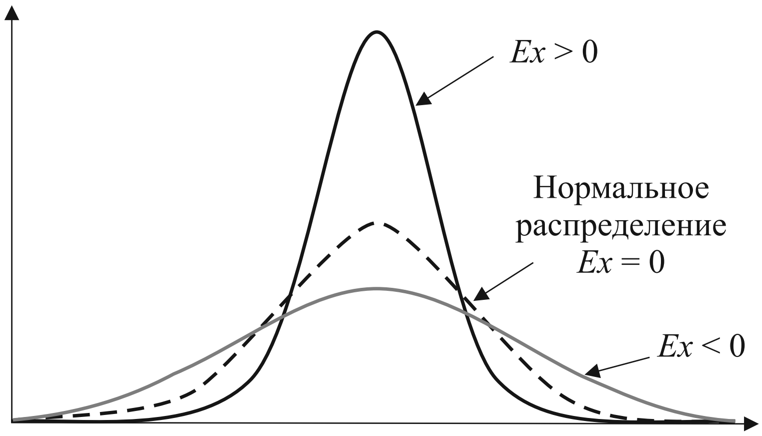


Рис. 3.4 – Эксцесс распределения

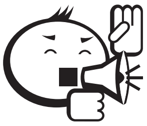
Наличие положительного эксцесса означает наличие слабоварьирующего «ядра» и сильно рассеянного вокруг него окружения в изучаемой совокупности. Отрицательный эксцесс означает отсутствие такого «ядра».

Пример

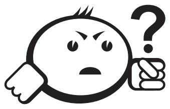
В нашем примере эксцесс составил (расчет числителя произведен в 10-м столбце таблицы 3.4):

$$Ex = \frac{13\ 809\ 838{,}86}{19{,}756^4 \cdot 35} - 3 = -0{,}41,$$

т. е. величина ВО варьирует сильнее, чем при нормальном распределении.



Пятым этапом статистического изучения вариации является проверка соответствия ряда распределения теоретическому (нормальному, логнормальному, биномиальному, распределению Пуассона и др.) с помощью критериев согласия, среди которых чаще всего применяют критерии Пирсона χ^2, Колмогорова и Романовского.



Контрольные вопросы по главе 3

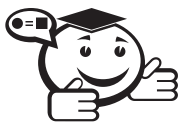
1. Что такое средняя величина? Что она характеризует?
2. Какие виды средних Вы знаете? Как они рассчитываются?
3. Какие свойства средних величин Вам известны?
4. Что такое вариация? Как она оценивается?
5. Какие существуют этапы изучения вариации?
6. Какие существуют виды распределения?

Глава 4

РЯДЫ ДИНАМИКИ

4.1 Понятие о рядах динамики

Одной из важнейших задач статистики является изучение изменений анализируемых показателей во времени, то есть их динамика. Эта задача решается при помощи анализа рядов динамики (временных рядов).



Ряд динамики *— это числовые значения определенного статистического показателя в последовательные моменты или периоды времени (т. е. расположенные в хронологическом порядке).*

Числовые значения того или иного статистического показателя, составляющего ряд динамики, называют *уровнями* ряда (будем обозначать их y). Первый член ряда y_1 называют начальным (базисным) уровнем, а последний y_n — конечным. Моменты или периоды времени, к которым относятся уровни, обозначают через t.

Ряды динамики, как правило, представляют в виде таблицы (см. табл. 4.1) или графически (см. рис. 4.1), причем по оси абсцисс строится шкала времени t, а по оси ординат — шкала уровней ряда y.

Таблица 4.1 – Внешнеторговый оборот (ВО) России за период 2000–2006 гг.

Год	2000	2001	2002	2003	2004	2005	2006
Млрд долл. США	149,9	155,6	168,3	212,0	280,6	368,9	468,4

Данные таблицы 4.1 и рисунка 4.1 наглядно иллюстрируют ежегодный рост внешнеторгового оборота (ВО) в России за период 2000–2006 гг.

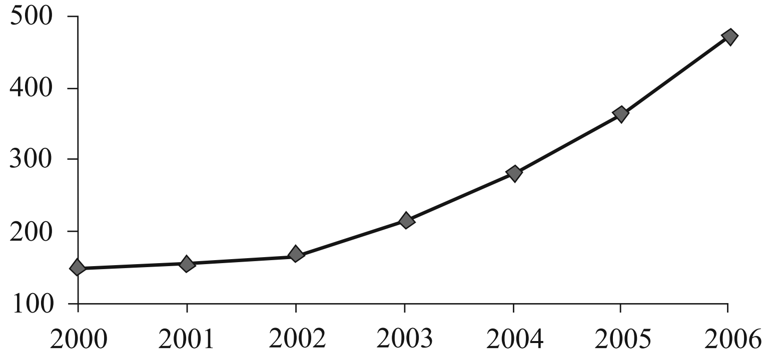


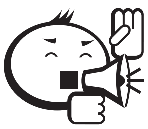
Рис. 4.1 – Внешнеторговый оборот (ВО) России за период 2000–2006 гг.

4.2 Показатели изменения уровней ряда динамики

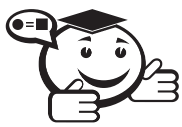
Цепные и базисные показатели динамики

Анализ рядов динамики начинается с определения того, как именно изменяются уровни ряда (увеличиваются, уменьшаются или остаются неизменными) в абсолютном и относительном выражении. Чтобы проследить за направлением и размером изменений уровней во времени, для рядов динамики рассчитывают *показатели изменения уровней ряда динамики* [10]:

- абсолютное изменение (абсолютный прирост);
- относительное изменение (темп роста или индекс динамики);
- темп изменения (темп прироста).



Все эти показатели могут определяться *базисным* способом, когда уровень данного периода сравнивается с первым (базисным) периодом, либо *цепным* способом — когда сравниваются два уровня соседних периодов.



Абсолютное изменение (абсолютный прирост) *уровней рассчитывается как разность между двумя уровнями ряда:*

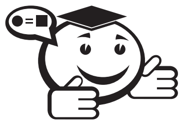
- *для базисного способа сравнения по формуле:*

$$\Delta y_i^{\text{Б}} = y_i - y_1$$

- *или для цепного по формуле:*

$$\Delta y_i^{\text{Ц}} = y_i - y_{i-1}.$$

Оно показывает, на сколько (в единицах показателей ряда) уровень одного (i-ого) периода отличается от уровня предшествующего или первоначального периода соответственно.



Темп роста (индекс динамики) *— относительное изменение уровней ряда, которое рассчитывается как отношение (деление) двух уровней:*

- *для базисного способа сравнения по формуле:*

$$T_{pi}^{\text{Б}} = \frac{y_i}{y_1}$$

- *или для цепного по формуле:*

$$T_{pi}^{\text{Ц}} = \frac{y_i}{y_{i-1}}.$$

Относительное изменение показывает, во сколько раз уровень данного периода больше уровня какого-либо предшествующего периода (при $i_i > 1$) или какую его часть составляет (при $i_i < 1$). Относительное изменение может выражаться в виде *коэффициентов*, то есть простого кратного отношения (если база сравнения принимается за единицу), и в *процентах* (если база сравнения принимается за 100 единиц) путем домножения относительного изменения на 100%.

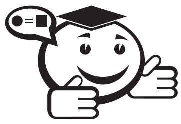
Темп прироста уровней — относительный показатель, показывающий, на сколько процентов данный уровень больше (или меньше) другого, принимаемого за базу сравнения. Он рассчитывается путем вычитания из относительного изменения 100%, соответственно для цепных и базисных показателей, то есть по формулам:

- темп прироста базисный:

$$T_{\text{пр}_i}^{\text{Б}} = T_{p_i}^{\text{Б}} - 100\%;$$

- темп прироста цепной:

$$T_{\text{пр}_i}^{\text{Ц}} = T_{p_i}^{\text{Ц}} - 100\%.$$



Абсолютное значение одного процента прироста *представляет собой отношение абсолютного прироста к темпу прироста и определяется по формуле:*

$$A = 0{,}01 y_{i-1}.$$

Абсолютное значение одного процента прироста определяется только для цепных приростов. Среднее абсолютное значение одного процента прироста не рассчитывается.

Пример

В таблице 4.2 рассчитаны цепные и базисные показатели динамики.

Таблица 4.2 – Анализ динамики ВО России

Год	y	$\Delta y_i^{\text{Б}}$	$\Delta y_i^{\text{Ц}}$	$T_{p_i}^{\text{Б}}$	$T_{p_i}^{\text{Ц}}$	$T_{\text{пр}_i}^{\text{Ц}}$, %	$T_{\text{пр}_i}^{\text{Б}}$, %	A
2000	149,9							
2001	155,6	5,7	5,7	1,038	1,038	3,8	3,8	1,499
2002	168,3	18,4	12,7	1,123	1,082	12,3	8,2	1,556
2003	212,0	62,1	43,7	1,414	1,260	41,4	26,0	1,683
2004	280,6	130,7	68,6	1,872	1,324	87,2	32,4	2,120
2005	368,9	219,0	88,3	2,461	1,315	146,1	31,5	2,806
2006	468,4	318,5	99,5	3,125	1,270	212,5	27,0	3,689
Итого	1803,7		318,5		3,125			

Между базисными и цепными абсолютными изменениями существует взаимосвязь: сумма цепных абсолютных изменений равна последнему базисному изменению, то есть:

$$\sum_{i=1}^{n} \Delta y_i^{\text{Ц}} = \Delta y_n^{\text{Б}}.$$

Пример

В нашем примере про ВО подтверждается правильность расчета абсолютных изменений по формуле: $\sum_{i=1}^{n} \Delta y_i^{\text{Ц}} = 318{,}5$, которая рассчитана в итоговой строке 4-го столбца, а $\Delta y_n^{\text{Б}} = 318{,}5$ — в предпоследней строке 3-го столбца.

В столбце 5 рассчитаны базисные относительные изменения, а в столбце 6 — цепные относительные изменения.

Между базисными и цепными относительными изменениями существует взаимосвязь: произведение цепных относительных изменений равно последнему базисному изменению, то есть

$$\prod_{i=1}^{n} i_i^{\text{Ц}} = i_n^{\text{Б}}.$$

Пример

В нашем примере про ВО подтверждается правильность расчета относительных изменений:

$$\prod_{i=1}^{n} i_i^{\text{Ц}} = 1{,}038 \cdot 1{,}082 \cdot 1{,}260 \cdot 1{,}324 \cdot 1{,}315 \cdot 1{,}270 = 3{,}125$$

рассчитано по данным 6-го столбца, а $i_n^{\text{Б}} = 3{,}125$ — в предпоследней строке 5-го столбца.

Средние показатели ряда динамики

Каждый ряд динамики можно рассматривать как некую совокупность n меняющихся во времени показателей, которые можно обобщить в виде средних величин. Такие обобщенные (средние) показатели особенно необходимы при сравнении динамики изменений того или иного показателя ВЭД в разные периоды, в разных странах и т. д. [14].

Обобщенной характеристикой ряда динамики служит прежде всего средний уровень ряда $\bar{y}$. Для разных видов рядов динамики он рассчитывается неодинаково.

Ряды динамики бывают равномерные (с равными интервалами времени между уровнями), для которых средний уровень определяется по простой формуле средней величины, и неравномерные (с неравными интервалами), для которых используются формулы средних взвешенных (по интервалам времени) величин.

В интервальном ряду динамики (в котором время задано в виде промежутков времени, к которым относятся уровни) $\bar{y}$ определяется по формуле средней арифметической, а в моментном ряду (в котором время задано в виде конкретных моментов времени или дат, к которым относятся уровни) — по формуле средней хронологической. В таблице 4.3 приводятся виды рядов динамики и соответствующие формулы для расчета их среднего уровня $\bar{y}$.

Пример

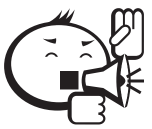
В нашем примере про ВО России за период 2000–2006 гг. имеем равномерный интервальный ряд динамики, поэтому его средний уровень определяем:

$$\bar{y} = \frac{1803{,}7}{7} = 257{,}671,$$

то есть ВО России в период 2000–2006 гг. составлял ежегодно в среднем 257,671 млрд долл. США.

Таблица 4.3 – Виды средних величин, применяемых при расчете среднего уровня

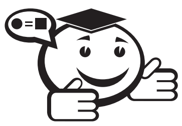
Вид ряда динамики	**Название средней величины**	**Формула средней величины**
Равномерный интервальный	Арифметическая простая	$\bar{y} = \dfrac{\sum\limits_{i=1}^{n} y_i}{n}$
Равномерный моментный	Хронологическая простая	$\bar{y} = \dfrac{\dfrac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \dfrac{y_n}{2}}{n-1} = \dfrac{\dfrac{y_1 + y_n}{2} + \sum\limits_{i=2}^{n-1} y_i}{n-1}$
Неравномерный интервальный	Арифметическая взвешенная	$\bar{y} = \dfrac{\sum\limits_{i=1}^{n} y_i t_i}{\sum\limits_{i=1}^{n} t_i}$
Неравномерный моментный	Хронологическая взвешенная	$\bar{y} = \dfrac{\sum\limits_{i=1}^{n-1} \left(y_i + y_{i+1}\right) t_i}{2\sum\limits_{i=1}^{n-1} t_i}$



Кроме среднего уровня ряда рассчитываются и другие средние показатели:

- средний абсолютный прирост;
- средний темп роста;
- средний темп прироста.

Каждый из этих показателей может рассчитываться базисным и цепным способом.



Средний абсолютный прирост *— это частное от деления последнего базисного абсолютного изменения на количество изменений уровней или частное от деления суммы всех цепных абсолютных изменений на количество изменений:*

$$\Delta\bar{y} = \frac{\Delta y_n^{\text{Б}}}{n-1} = \frac{\sum\limits_{i=1}^{n} \Delta y_i^{\text{Ц}}}{n-1}.$$

По знаку средних абсолютных изменений также судят о характере изменения явления в среднем: рост, спад или стабильность. Очевидно, что числители формулы равны между собой, значит, среднее абсолютное изменение не зависит от способа расчета (базисный или цепной), так как результат получится одинаковый.

Пример

В нашей задаче:

$$\Delta\overline{y} = \frac{318{,}5}{6} = 53{,}083,$$

то есть ежегодно в среднем ВО растет на 53,083 млрд долл.

Наряду со средним абсолютным изменением рассчитывается и среднее относительное.

Средний темп роста *показывает, во сколько раз в среднем изменялось значение ряда динамики в течение рассматриваемого периода и определяется по формуле:*

$$\overline{T_p} = \sqrt[n-1]{T^{\text{Б}}_{p_i}} = \sqrt[n-1]{\frac{y_n}{y_1}}.$$

Пример

В нашем примере про ВО:

$$\overline{T_p} = \sqrt[6]{3{,}125} = 1{,}209,$$

то есть ежегодно в среднем в период 2000–2006 гг. ВО России растет в 1,209 раза.

Вычитанием 100% из среднего относительного изменения образуется соответствующий ***средний темп прироста****:*

$$\overline{T_{\text{пр}}} = \overline{T_p} - 100\%.$$

По его знаку также можно судить о характере изменения изучаемого явления, отраженного данным рядом динамики.

Пример

В нашем примере:

$$\overline{T}_{\text{пр}} = 1{,}209 - 1 = 0{,}209,$$

то есть ежегодно в среднем в период 2000–2006 гг. ВО России растет на 20,9%.

4.3 Методы выявления основной тенденции (тренда) в рядах динамики

Одна из основных задач изучения рядов динамики — это выявление тренда.

***Тренд** — это основная тенденция (закономерность) в изменении уровней ряда.*

Закономерность в изменении уровней ряда в одних случаях проявляется наглядно, в других — может маскироваться колебаниями случайного или неслучайного характера. Поэтому, чтобы сделать правильные выводы о закономерностях развития того или иного показателя, надо суметь отделить тренд от колебаний, вызванных случайными кратковременными причинами. На основании выделенного тренда можно экстраполировать (прогнозировать) развитие явления в будущем. С этой целью (устранить колебания, вызванные случайными причинами) ряды динамики подвергают *обработке*.

Существует несколько методов обработки рядов динамики, помогающих выявить основную тенденцию изменения уровней ряда, а именно: метод укрупнения интервалов, метод скользящей средней и аналитическое выравнивание. Во всех методах вместо фактических уровней при обработке ряда рассчитываются иные (расчетные) уровни, в которых тем или иным способом взаимопогашается действие случайных факторов и тем самым уменьшается колеблемость уровней. Последние в результате становятся как бы «выравненными», «сглаженными» по отношению к исходным фактическим данным. Такие методы обработки рядов динамики называются *сглаживанием*, или *выравниванием*, рядов динамики.

Простейший метод сглаживания уровней ряда — *укрупнение интервалов* для определения итогового значения или средней величины исследуемого показателя.

Этот метод особенно эффективен, если первоначальные уровни ряда относятся к коротким промежуткам времени.

Пример

Например, если имеются данные о ежесуточном производстве мороженого на предприятии за месяц, то, естественно, в таком ряду возможны значительные колебания уровней, так как чем меньше период, за который приводятся данные, тем больше влияние случайных факторов. Чтобы устранить это влияние, рекомендуется укрупнить интервалы времени, например до 5 или 10 дней, и для этих укрупненных интервалов рассчитать общий или среднесуточный объем производства (соответственно по пятидневкам или декадам).

В ряду с укрупненными интервалами времени закономерность изменения уровней будет более наглядной.

По своей сути метод *скользящей средней* похож на метод укрупнения интервалов, но в данном случае фактические уровни заменяются средними уровнями, рассчитанными для последовательно подвижных (скользящих) укрупненных интервалов, охватывающих m уровней ряда.

Каждое звено скользящей средней — это уровень за соответствующий период, который относится к середине выбранного периода, если число уровней скольжения нечетное. Например, для ряда динамики, представленного уровнями y_1, y_2, y_3, y_4, y_5, скольжение 3-уровневой средней будет иметь вид:

$$y'_2 = \frac{y_1 + y_2 + y_3}{3}; \quad y'_3 = \frac{y_2 + y_3 + y_4}{3}; \quad y'_4 = \frac{y_3 + y_4 + y_5}{3}.$$

Нахождение скользящей средней по четному числу уровней несколько сложнее, так как средняя может быть отнесена к середине между двумя периодами, находящимися в середине интервала сглаживания.

Чтобы ликвидировать такой сдвиг, применяют способ *центрирования*. Центрирование заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенному периоду.

$$y'_{2-3} = \frac{y_1 + y_2 + y_3 + y_4}{4}; \quad y'_{3-4} = \frac{y_2 + y_3 + y_4 + y_5}{4}.$$

Центрированием получаем $y'_3 = \dfrac{y_{2-3} + y_{3-4}}{2}$.

Наиболее совершенным методом обработки рядов динамики в целях устранения случайных колебаний и выявления тренда является *выравнивание уровней ряда по аналитическим формулам* (или *аналитическое выравнивание*).

Суть аналитического выравнивания заключается в замене эмпирических (фактических, исходных) уровней y_t теоретическими $\widehat{y}_t$, которые рассчитаны по определенному уравнению, принятому за математическую модель тренда, где теоретические уровни рассматриваются как функция времени: $\widehat{y}_t = f(t)$.

При этом каждый фактический уровень y_i рассматривается обычно как сумма двух составляющих:

$$y_t = f(t) + \varepsilon_t,$$

где $f(t) = \widehat{y}_t$ — систематическая составляющая, отражающая тренд и выраженная определенным уравнением; ε_t — случайная величина, вызывающая колебания уровней вокруг тренда.

Задача аналитического выравнивания сводится к следующему:

1) определение на основе фактических данных формы (вида) гипотетической функции $\widehat{y}_t = f(t)$, способной наиболее адекватно отразить тенденцию развития исследуемого показателя;
2) нахождение по эмпирическим данным параметров указанной функции (уравнения);
3) расчет по найденному уравнению теоретических (выравненных) уровней.

В аналитическом выравнивании наиболее часто используются простейшие функции, представленные в таблице 4.4, где обозначено $\widehat{y}_t$ — теоретические (выравненные) уровни (читается как «игрек, выравненный по t»); t — условное обозначение времени (1, 2, 3, ...); a_0, a_1, a_2,... — параметры аналитической функции; k — число гармоник (при выравнивании по ряду Фурье).

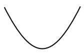
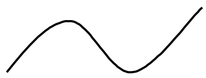
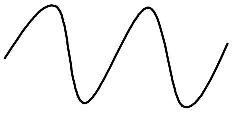
Выбор той или иной функции для выравнивания ряда динамики осуществляется на основании графического изображения эмпирических данных. Если по тем или иным причинам уровни эмпирического ряда трудно описать одной функцией, следует разбить анализируемый период на отдельные части и затем выровнять каждую часть по соответствующей кривой [7].

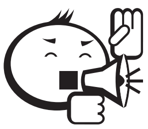
Нередко один и тот же ряд можно выровнять по разным аналитическим функциям и получить довольно близкие результаты. В нашем примере про ВО России можно произвести выравнивание и по прямой линии, и по параболе. Чтобы решить вопрос о том, использование какой кривой дает лучший результат, обычно сопоставляют суммы квадратов отклонений эмпирических уровней от теоретических (*остатки*), рассчитанных по разным функциям, то есть:

$$\sum(\widehat{y}_t - y_t)^2.$$

Та функция, при которой эта сумма минимальна, считается наиболее адекватной, приемлемой. Однако сравнивать непосредственно суммы квадратов отклонений можно в том случае, если сравниваемые уравнения имеют одинаковое число параметров. Если же число параметров k разное, то каждую сумму квадратов делят на разность $(n - k)$, выступающую в роли числа степеней свободы, и сравнивают уже квадраты отклонений уровней, рассчитанные на одну степень свободы (т. е. остаточные дисперсии на одну степень свободы).

Таблица 4.4 – Виды математических функций, используемые при выравнивании

Название функции	Вид функции	Формула
Прямая линия		$\widehat{y}_t = a_0 + a_1 t$
Парабола 2-го порядка	 или 	$\widehat{y}_t = a_0 + a_1 t + a_2 t^2$
Парабола 3-го порядка		$\widehat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$
Гипербола		$\widehat{y}_t = a_0 + \dfrac{a_1}{t}$
Показательная		$\widehat{y}_t = a_0 a_1^t$
Степенная		$\widehat{y}_t = a_0 t^{a_1}$
Ряд Фурье		$\widehat{y}_t = a_0 + \sum\limits_{k=1}^{m} (a_k \cos kt + b_k \sin kt)$



Параметры искомых уравнений (a_0, a_1, a_2,...) при аналитическом выравнивании могут быть определены по-разному, но наиболее распространенным методом является *метод наименьших квадратов* (*МНК*). При этом методе учитываются все эмпирические уровни и должна обеспечиваться минимальная сумма квадратов отклонений эмпирических значений уровней y от теоретических уровней $\widehat{y}_t$:

$$\sum(\widehat{y}_t - y_t)^2 \to \min.$$

В частности, при выравнивании по прямой параметры a_0 и a_1 отыскиваются по МНК следующим образом. В вместо $\widehat{y}_t$ записываем его конкретное выражение $a_0 + a_1 t$. Тогда

$$S = \sum(a_0 + a_1 t - y_t)^2 \to \min.$$

Дальнейшее решение сводится к задаче на экстремум, т. е. к определению того, при каком значении a_0 и a_1 функция двух переменных S может достигнуть минимума. Как известно, для этого надо найти частные производные S по a_0 и a_1, приравнять их к нулю и после элементарных преобразований решить систему двух уравнений с двумя неизвестными.

В соответствии с вышеизложенным найдем частные производные:

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 2\sum(a_0 + a_1 t - y_t) = 0; \\ \dfrac{\partial S}{\partial a_1} = 2\sum(a_0 + a_1 t - y_t)t = 0. \end{cases}$$

Сократив каждое уравнение на 2, раскрыв скобки и перенеся члены с y в правую сторону, а остальные — оставив в левой, получим систему нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum t = \sum y_t; \\ a_0 \sum t + a_1 \sum t^2 = \sum y_t t, \end{cases}$$

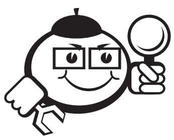
где n — количество уровней ряда; t — порядковый номер в условном обозначении периода или момента времени; y — уровни эмпирического ряда.

Эта система и, соответственно, расчет параметров a_0 и a_1 упрощаются, если отсчет времени ведется от середины ряда[1]. Например, при *нечетном* числе уровней (как в нашем примере про ВО России — 7 уровней) серединная точка времени (год, месяц) принимается за нуль, тогда предшествующие периоды обозначаются соответственно −1, −2, −3 и т. д., а следующие за средним (центральным) — соответственно 1, 2, 3 и т. д. При *четном* числе уровней два серединных момента (периода) времени обозначают −1 и +1, а все последующие и предыдущие соответственно через два интервала: ±3, ±5, ±7 и т. д.

При таком порядке отсчета времени (от середины ряда) $\sum t = 0$, поэтому система уравнений для нахождения коэффициентов a_0 и a_1 упрощается до следующих двух уравнений, каждое из которых решается самостоятельно:

$$\begin{cases} na_0 = \sum y_t \Rightarrow a_0 = \dfrac{\sum y_t}{n}; \\ a_1 \sum t^2 = \sum y_t t \Rightarrow a_1 = \dfrac{\sum y_t t}{\sum t^2}. \end{cases}$$

Как видим, при такой нумерации периодов параметр a_0 представляет собой средний уровень равномерного интервального ряда.



Пример

Определим параметры уравнения прямой для нашего примера про ВО России, для чего исходные данные и все расчеты необходимых сумм представим в таблице 4.5.

Из таблицы 4.5 получаем, что:

$$a_0 = \frac{1803{,}7}{7} = 257{,}671,$$

$$a_1 = \frac{1494{,}4}{28} = 53{,}371.$$

Отсюда искомое уравнение тренда:

$$\hat{y}_t = 257{,}671 + 53{,}371t.$$

[1] При расчете параметров уравнения тренда на ЭВМ необходимость вести отсчет от середины ряда динамики отпадает. Например, для получения уравнения тренда в *Microsoft Office Excel* необходимо построить его график с помощью «Мастера диаграмм», после чего вызвать контекстное меню, нажав на правую кнопку мыши на построенном графике, и выбрать пункт «Добавить линию тренда», в появившемся окне выбрать подходящую математическую функцию и установить галочку «показывать уравнение на диаграмме».

Таблица 4.5 – Вспомогательные расчеты для линейного тренда

Год	y	t	t^2	yt	$\widehat{y}_t$	$(\widehat{y}_t - y)^2$	$(\widehat{y}_t - \bar{y})^2$	$(y - \bar{y})^2$
2000	149,9	–3	9	–449,7	97,557	2739,775	25 636,584	11 614,681
2001	155,6	–2	4	–311,2	150,929	21,822	11 394,038	10 418,577
2002	168,3	–1	1	–168,3	204,300	1296,000	2848,509	7987,252
2003	212	0	0	0	257,671	2085,879	0,000	2085,879
2004	280,6	1	1	280,6	311,043	926,768	2848,509	525,719
2005	368,9	2	4	737,8	364,414	20,122	11 394,038	12 371,795
2006	468,4	3	9	1405,2	417,786	2561,806	25 636,584	44 406,531
Итого	1803,7	0	28	1494,4	1803,700	9652,171	79 758,263	89 410,434

В 6-м столбце таблицы 4.5 приведены теоретические (трендовые) уровни, рассчитанные по этому уравнению, а в итоге 7-го столбца — остатки. Для иллюстрации построим график эмпирических и трендовых уровней (рис. 4.2).

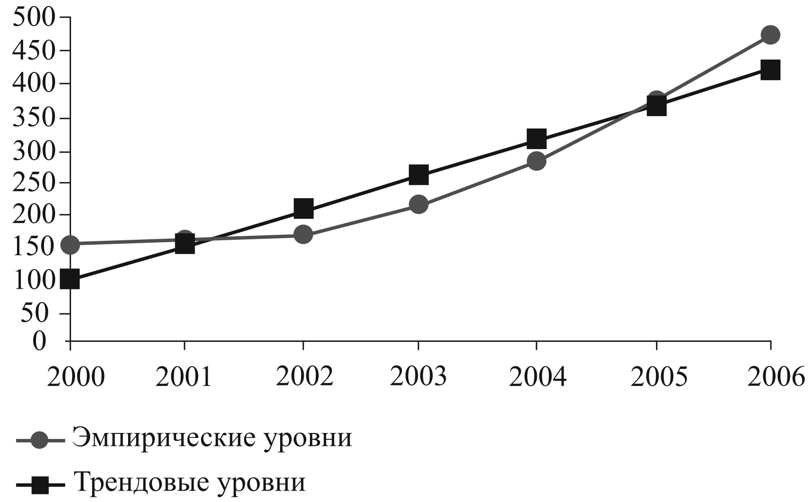
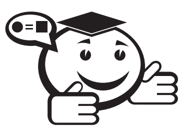


Рис. 4.2 – Эмпирические и трендовые уровни ряда динамики ВО России

4.4 Оценка адекватности тренда и прогнозирование



*Для найденного уравнения тренда необходимо провести оценку его **надежности (адекватности)**, что осуществляется обычно с помощью критерия Фишера, сравнивая его расчетное значение F_p с теоретическим (табличным) значением F_T (Приложение В). При этом расчетный критерий Фишера определяется по формуле:*

$$F_p = \frac{(n-k)\sum(\widehat{y}_t - \bar{y})^2}{(k-1)\sum(\widehat{y}_t - y_t)^2},$$

где k — число параметров (членов) выбранного уравнения тренда.

Для проверки правильности расчета сумм для F_p можно использовать следующее равенство:

$$\sum(y_t - \bar{y})^2 = \sum(\widehat{y}_t - y_t)^2 + \sum(\widehat{y}_t - \bar{y})^2.$$

Сравнение расчетного и теоретического значений критерия Фишера ведется при заданном уровне значимости (вероятности сделать неверный прогноз) с учетом степеней свободы: $\nu_1 = k - 1$ и $\nu_2 = n - k$. При условии $F_p > F_T$ считается, что выбранная математическая модель ряда динамики адекватно отражает обнаруженный в нем тренд [10].

Пример

В нашем примере про ВО это равенство соблюдается (необходимые суммы рассчитаны в трех последних столбцах таблицы 4.5):

$$89\ 410{,}434 = 9652{,}171 + 79\ 758{,}263.$$

Проверим тренд на адекватность:

$$F_p = \frac{79\ 758{,}263 \cdot 5}{9652{,}171 \cdot 1} = 41{,}32 > F_T,$$

значит, модель адекватна и ее можно использовать для прогнозирования ($F_T = 6{,}61$ находим по Приложению В в 1-ом столбце $[\nu_1 = k - 1 = 2 - 1 = 1]$ и 5-й строке $[\nu_2 = n - k = 5]$).

Как уже было отмечено ранее, в нашем примере можно произвести выравнивание не только по прямой линии, но и по параболе, чего делать не будем, так как уже найденный линейный тренд адекватно описывает тенденцию.

При составлении прогнозов уровней социально-экономических явлений обычно оперируют не точечной, а интервальной оценкой, рассчитывая так называемые *доверительные интервалы прогноза*. Границы интервалов определяются по формуле (4.1):

$$\widehat{y}_t \pm t_\alpha \sigma_{\widehat{y}}, \qquad (4.1)$$

где $\widehat{y}_t$ — точечный прогноз, рассчитанный по модели тренда; t_α — *коэффициент доверия по распределению Стьюдента* при уровне значимости α и числе степеней свободы $\nu = n - 1$ (Приложение Б)[1]; $\sigma_{\widehat{y}}$ — *ошибка аппроксимации*, определяемая по формуле:

$$\sigma_{\widehat{y}} = \sqrt{\frac{\sum(\widehat{y}_t - y_t)^2}{n - k}}.$$

[1] Используется при малом количестве уровней ($n < 30$), в противном случае ($n > 30$) вместо t_α используют коэффициент доверия t нормального закона распределения (Приложение А).

Пример

Спрогнозируем ВО России на 2007 и 2008 годы с вероятностью 0,95 (значимостью 0,05), для чего найдем ошибку аппроксимации по формуле:

$$\sigma_{\tilde{y}} = \sqrt{\frac{9652{,}171}{7-2}} = 43{,}937$$

и найдем коэффициент доверия по распределению Стьюдента по Приложению Б:

$$t_\alpha = 2{,}4469 \text{ при } \nu = 7 - 1 = 6.$$

Прогноз на 2007 и 2008 годы с вероятностью 0,95:

$$Y_{2007} = \left(257{,}671 + 53{,}371 \cdot 4\right) \pm 2{,}4469 \cdot 43{,}937$$

или

$$363{,}6 < Y_{2007} < 578{,}7 \text{ млрд долл.};$$

$$Y_{2008} = \left(257{,}671 + 53{,}371 \cdot 5\right) \pm 2{,}4469 \cdot 43{,}937$$

или

$$417{,}0 < Y_{2008} < 632{,}0 \text{ млрд долл.}$$

Как видно из полученных прогнозов, доверительный интервал достаточно широк (из-за достаточно большой величины ошибки аппроксимации). Более точный прогноз можно получить при выравнивании по параболе 2-го порядка.

Контрольные вопросы по главе 4

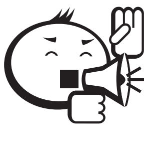
1. Что такое ряд динамики? Что он характеризует?
2. Какие виды рядов динамики Вы знаете?
3. Назовите цепные и базисные показателя рядов динамики. Как они рассчитываются и что показывают?
4. Назовите средние показатели рядов динамики. Как они рассчитываются и что показывают?
5. Какие Вы знаете методы выравнивания рядов динамики?
6. Что такое тренд и как определить его параметры?
7. Как оценить адекватность тренда?

Глава 5

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗЕЙ

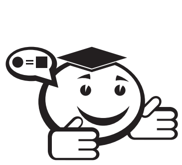
5.1 Понятие корреляционной зависимости

Один из наиболее общих законов объективного мира — закон всеобщей связи и зависимости между явлениями. Естественно, что, исследуя явления в самых различных областях, статистика неизбежно сталкивается с зависимостями как между количественными, так и между качественными показателями, признаками. Ее задача — обнаружить (выявить) такие зависимости и дать им количественную характеристику.



Среди взаимосвязанных признаков (показателей) одни могут рассматриваться как определенные факторы, влияющие на изменение других (*факторные*), а вторые (*результативные*) — как следствие, результат влияния первых.

Существует два вида связи между отдельными признаками: функциональная и стохастическая (статистическая), частным случаем которой является корреляционная.



Связь между двумя переменными x и y называется ***функциональной****, если определенному значению переменной x строго соответствует одно или несколько значений другой переменной y, и с изменением значения x значение y меняется строго определенно.*

Такие связи обычно встречаются в точных науках.

Пример

Например, известно, что площадь квадрата равна квадрату его стороны ($S = a^2$). Это соотношение характерно для каждого единичного случая (квадрата), это так называемая *жестко детерминированная* связь. Такие связи можно встретить и в области экономических явлений.

Например, при простой сдельной оплате труда связь между оплатой труда y и количеством изготовленных изделий x при фиксированной расценке за одну деталь, например 5 руб., легко выразить формулой $y = 5x$.

Для изучения функциональных связей применяется *индексный метод*, который рассматривается в главе 6.

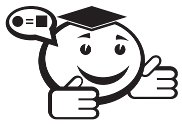
Существуют и иного рода связи, где взаимно действуют многие факторы, комбинация которых приводит к вариации значений результативного признака (показателя) при одинаковом значении факторного признака. Например, при изучении зависимости величины таможенных платежей, поступающих в федеральный бюджет, от количества товаров, перемещаемых через таможенную границу государства (или от стоимостного товарооборота), последние будут рассматриваться как факторный признак, а величина таможенных платежей — как результативный. Между ними нет жестко детерминированной связи, т. е. при одном и том же количестве перемещенных через таможенную границу товаров (или стоимости товарооборота) величина таможенных платежей, перечисленных разными таможнями, будет различной, так как кроме количества товаров, перемещаемых через таможенную границу государства (или стоимость товарооборота), на величину таможенных платежей влияет много других факторов (различная номенклатура товаров, для которых применяются различные таможенные пошлины, сборы и льготы; различные таможенные режимы перемещения товаров через таможенную границу и др.), комбинация которых вызывает вариацию величины таможенных платежей.

Там, где взаимодействует множество факторов, в том числе и случайных, выявить зависимости, рассматривая единичный случай, невозможно. Такие связи можно обнаружить только при массовом наблюдении как статистические закономерности[1]. *Выявленная таким образом связь именуется* ***стохастической***[2].

[1] Проявление стохастических связей подвержено *действию закона больших чисел*: лишь в достаточно большом числе единиц индивидуальные особенности сгладятся, случайности взаимопогасятся и зависимость, если она имеет существенную силу, проявится достаточно отчетливо.

[2] Термин «стохастический» происходит от греч. «*stochos*» — мишень. Стреляя в мишень, даже хороший стрелок редко попадает в ее центр, выстрелы ложатся в некоторой близости от него. Другими словами, стохастическая связь означает приблизительный характер значений признака.

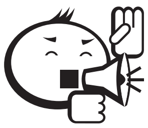
Корреляционная связь[1] — понятие более узкое, чем стохастическая связь, это ее частный случай. Именно корреляционные связи являются предметом изучения статистики.



Корреляционная связь *— это связь, проявляющаяся при большом числе наблюдений в виде определенной зависимости между средним значением результативного признака и признаками-факторами.*

Другими словами, корреляционную связь условно можно рассматривать как своего рода функциональную связь средней величины одного признака (результативного) со значением другого (или других) [3]. При этом, если рассматривается связь средней величины результативного показателя y с одним признаком-фактором x, корреляция называется *парной*, а если факторных признаков 2 и более $(x_1, x_2, \ldots, x_m)$ — *множественной*[2].

По характеру изменений x и y в парной корреляции различаются *прямая* и *обратная* связь. При прямой связи значения обоих признаков изменяются в одном направлении, т. е. с увеличением (уменьшением) значений x увеличиваются (уменьшаются) и значения y. При обратной связи значения факторного и результативного признаков изменяются в разных направлениях.



Изучение корреляционных связей сводится в основном к решению следующих задач[3]:

1) выявление наличия (отсутствия) корреляционной связи между изучаемыми признаками;
2) измерение тесноты связи между двумя (и более) признаками с помощью специальных коэффициентов (эта часть исследования именуется корреляционным анализом);
3) определение уравнения регрессии — математической модели, в которой среднее значение результативного признака y рассматривается как функция одной или нескольких переменных — факторных признаков (эта часть исследования именуется регрессионным анализом).

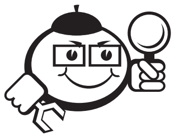
[1]Термин «корреляция» ввел в статистику английский биолог и статистик Ф. Гальтон в конце XIX в., под которым понималась «как бы связь», т. е. связь в форме, отличающейся от функциональной. Еще ранее этот термин применил француз Ж. Кювье в палеонтологии, где под законом корреляции частей животных он понимал возможность восстановить по найденным в раскопках частям облик всего животного.

[2]Множественная корреляция изучается в курсе эконометрики на основе применения компьютерных программ (напр., специальная надстройка к *Excel*, *SPSS* и др.), в курсе статистики изучается только парная корреляция.

[3]Теория статистики: учебник для вузов под. ред. Р. А. Шмойловой. — изд. 4-е, доп., перераб. — М.: Финансы и статистика, 2007.

Общий термин «*корреляционно-регрессионный анализ*» подразумевает всестороннее исследование корреляционных связей (т. е. решение всех трех задач).

Корреляционно-регрессионный анализ находит широкое применение в статистике. Рассмотрим его практическое применение на примере данных таможенной статистики внешней торговли России в 2006 году[1] — таблица 5.1.



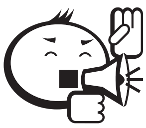
Пример

Таблица 5.1 – Величина внешнеторгового оборота и таможенных платежей

Месяц	Оборот, млрд долл.	Платеж, млрд руб.
Январь	27,068	172,17
Февраль	29,889	200,90
Март	34,444	231,83
Апрель	33,158	232,10
Май	37,755	233,40
Июнь	37,554	236,99
Июль	37,299	246,53
Август	40,370	253,62
Сентябрь	37,909	256,43
Октябрь	38,348	261,89
Ноябрь	39,137	259,36
Декабрь	46,298	278,87

В качестве факторного признака x примем стоимостной внешнеторговый товарооборот в млрд долл. США, а в качестве результативного признака y — величину таможенных платежей в федеральный бюджет в млрд руб.

5.2 Методы выявления и оценки корреляционной связи



Для выявления наличия и характера корреляционной связи между *двумя* признаками в статистике используется ряд методов.

1. ***Рассмотрение параллельных данных*** (значений x и y в каждой из n единиц). Единицы наблюдения необходимо расположить по возрастанию значений фактор-

[1]Чалиев А. А., Овчаров А. О. Таможенная статистика: учебно-методическое пособие. — Нижний Новгород: ННГУ, 2008.

ного признака x (как в табл. 5.1) и затем сравнить с ним (визуально) поведение результативного признака y.

В нашей задаче в 6 случаях по мере увеличения значений x увеличиваются и значения y, а в 5 случаях этого не происходит, поэтому затруднительно говорить о прямой связи между x и y.

2. ***Графический метод*** — это графическое изображение корреляционной зависимости. Для этого, имея n взаимосвязанных пар значений x и y и пользуясь прямоугольной системой координат, каждую такую пару изображают в виде точки на плоскости с координатами x и y. Совокупность полученных точек представляет собой *корреляционное поле* (рис. 5.1).

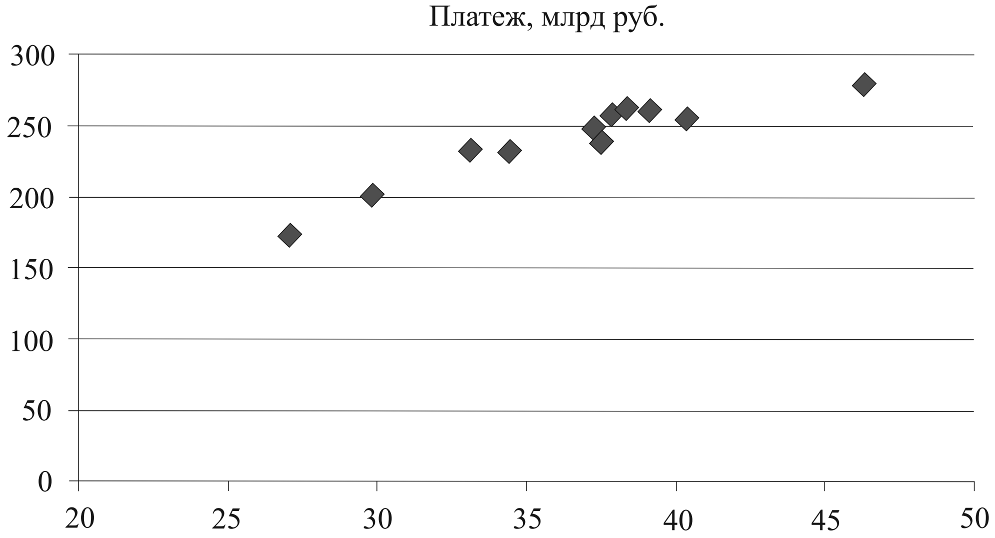


Рис. 5.1 – Корреляционное поле

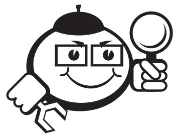
Визуально анализируя график, можно предположить характер зависимости между признаками x и y. В нашей задаче можно выдвинуть гипотезу о наличии прямой зависимости между величиной стоимостного внешнеторгового товарооборота и величиной таможенных платежей в федеральный бюджет.

3. ***Коэффициент корреляции знаков (Фехнера)*** — простейший показатель тесноты связи, основанный на сравнении поведения отклонений индивидуальных значений каждого признака (x и y) от своей средней величины. При этом во внимание принимаются не величины отклонений ($x_i - \bar{x}$) и ($y_i - \bar{y}$), а их знаки («+» или «–»). Определив знаки отклонений от средней величины в каждом ряду, рассматривают все пары знаков и подсчитывают число их совпадений (C) и несовпадений (H). Тогда коэффициент Фехнера рассчитывается как отношение разности чисел пар совпадений и несовпадений знаков к их сумме, т. е. к общему числу наблюдаемых единиц:

$$K_{\Phi} = \frac{\sum C_i - \sum H_i}{\sum C_i + \sum H_i}.$$

Очевидно, что если знаки всех отклонений по каждому признаку совпадут, то $K_{\Phi} = 1$, что характеризует наличие прямой связи. Если все знаки не совпадут, то $K_{\Phi} = -1$ (обратная связь). Если же $\sum C_i = \sum H_i$, то $K_{\Phi} = 0$. Итак, как и любой показатель тесноты связи, коэффициент Фехнера может принимать значения от 0 до ±1.

Однако если $K_{\Phi} = 1$, то это ни в коей мере нельзя воспринимать как свидетельство функциональной зависимости между x и y.



Пример

Средние значения факторного и результативного признаков определяем по формуле средней арифметической простой:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{439{,}229}{12} = 36{,}602;$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2864{,}09}{12} = 238{,}674.$$

В двух последних столбцах таблицы 5.2 приведены знаки отклонений каждого x и y от своей средней величины. Число совпадений знаков — 10, а несовпадений — 2. Определяем коэффициент корреляции знаков (Фехнера):

$$K_{\Phi} = \frac{10-2}{10+2} = \frac{8}{12} = 0{,}667.$$

Таблица 5.2 – Вспомогательная таблица для расчета коэффициента Фехнера

№ п/п	x	y	$x-\bar{x}$	$y-\bar{y}$
1	27,068	172,17	–	–
2	29,889	200,90	–	–
3	33,158	232,10	–	–
4	34,444	231,83	–	–
5	37,299	246,53	+	+
6	37,554	236,99	+	–
7	37,755	233,40	+	–
8	37,909	256,43	+	+
9	38,348	261,89	+	+
10	39,137	259,36	+	+
11	40,370	253,62	+	+
12	46,298	278,87	+	+
Итого	439,229	2864,09		

Обычно такое значение показателя тесноты связи характеризует заметную прямую зависимость между x и y, однако следует иметь в виду, что поскольку K_{Φ} зависит только от знаков и не учитывает величину самих отклонений x и y от их средних величин, то он практически характеризует не столько тесноту связи, сколько ее наличие и направление.

4. ***Линейный коэффициент корреляции*** — самый распространенный измеритель линейной зависимости между двумя количественными признаками x и y. Он

основан на предположении, что при *полной независимости*[1] *признаков* x и y отклонения значений факторного признака от средней $(x-\bar{x})$ носят случайный характер и должны случайно сочетаться с различными отклонениями $(y-\bar{y})$. При наличии значительного перевеса совпадений или несовпадений таких отклонений делается предположение о наличии связи между x и y [15].

В отличие от $K_Ф$ в линейном коэффициенте корреляции учитываются не только знаки отклонений от средних величин, но и значения самих отклонений, выраженные для сопоставимости в единицах среднего квадратического отклонения t:

$$t_x = \frac{x-\bar{x}}{\sigma_x} \quad \text{и} \quad t_y = \frac{y-\bar{y}}{\sigma_y}.$$

__Линейный коэффициент корреляции r__ представляет собой среднюю величину из произведений нормированных отклонений для x и y:

$$r = \frac{\sum\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right)}{n} = \frac{\sum t_x t_y}{n},$$

или

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

Числитель формулы, деленный, представляющий собой среднее произведение отклонений значений двух признаков от их средних значений, называется *коэффициентом ковариации* — это мера совместной вариации факторного x и результативного y признаков:

$$\text{cov}(x, y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n} = \overline{(x_i-\bar{y})(y_i-\bar{y})}.$$

Недостатком коэффициента ковариации является то, что он не нормирован, в отличие от линейного коэффициента корреляции. Очевидно, что линейный коэффициент корреляции представляет собой частное от деления ковариации между x и y на произведение их средних квадратических отклонений:

$$r = \frac{\text{cov}(x, y)}{\sigma_x\sigma_y}.$$

Путем несложных математических преобразований можно получить и другие модификации формулы линейного коэффициента корреляции, например:

$$r = \frac{\overline{xy}-\bar{x}\,\bar{y}}{\sigma_x\sigma_y},$$

[1] Данное условие означает отсутствие автокорреляции в коррелируемых рядах динамики, проверка на данное условие изучается на дневной форме обучения (при необходимости см. конспект лекций для дневного отделения).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}},$$

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - (\sum x_i)^2\right]\left[n\sum y_i^2 - (\sum y_i)^2\right]}},$$

$$r = \frac{\sum x_i y_i - \sum x_i \frac{\sum y_i}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]\left[\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right]}}.$$

Линейный коэффициент корреляции может принимать значения от −1 до +1, причем знак определяется в ходе решения.

- Если $r > 0$, то зависимость между x и y прямая.
- Если $r < 0$, то зависимость между x и y прямая обратная.
- Если $r = 0$, то линейная зависимость между x и y отсутствует.
- При $r = 1$ существует функциональная зависимость между x и y.

Следовательно, всякое промежуточное значение r от 0 до 1 характеризует степень приближения корреляционной связи между x и y к функциональной. Существует эмпирическое правило (шкала Чэддока) для оценки тесноты связи, представленное в таблице 5.3.

Таблица 5.3 – Шкала Чэддока

$\lvert r \rvert$	**Теснота связи**
менее 0,1	отсутствует линейная связь
0,1 ÷ 0,3	слабая
0,3 ÷ 0,5	умеренная
0,5 ÷ 0,7	заметная
более 0,7	сильная (тесная)

Таким образом, коэффициент корреляции при линейной зависимости служит как мерой тесноты связи, так и показателем, характеризующим степень приближения корреляционной зависимости между x и y к линейной. Поэтому близость значения r к 0 в одних случаях может означать отсутствие связи между x и y, а в других свидетельствовать о том, что зависимость не линейная.

В нашей задаче для расчета r построим вспомогательную таблицу 5.4.

Таблица 5.4 – Вспомогательные расчеты линейного коэффициента корреляции

№ п/п	x	y	$(x-\bar{x})^2$	$(y-\bar{y})^2$	t_x	t_y	$t_x t_y$	$(x-\bar{x})(y-\bar{y})$	xy
1	27,068	172,17	90,905	4422,804	–1,993	–2,408	4,799	634,078	4660,298
2	29,889	200,90	45,070	1426,888	–1,403	–1,368	1,919	253,594	6004,700
3	33,158	232,10	11,864	43,220	–0,720	–0,238	0,171	22,644	7695,972
4	34,444	231,83	4,659	46,843	–0,451	–0,248	0,112	14,773	7985,153
5	37,299	246,53	0,485	61,714	0,146	0,284	0,041	5,472	9195,322
6	37,554	236,99	0,906	2,836	0,199	–0,061	–0,012	–1,603	8899,922
7	37,755	233,40	1,328	27,817	0,241	–0,191	–0,046	–6,079	8812,017
8	37,909	256,43	1,707	315,270	0,273	0,643	0,176	23,199	9721,005
9	38,348	261,89	3,047	538,975	0,365	0,841	0,307	40,525	10 042,958
10	39,137	259,36	6,424	427,904	0,530	0,749	0,397	52,430	10 150,572
11	40,37	253,62	14,195	223,378	0,788	0,541	0,426	56,310	10 238,639
12	46,298	278,87	94,004	1615,705	2,027	1,455	2,950	389,722	12 911,123
Итого	439,229	2864,09	274,594	9153,353			11,241	1485,066	10 6317,681

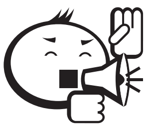
$$\sigma_x = \sqrt{\frac{274{,}594}{12}} = 4{,}784;$$

$$\sigma_y = \sqrt{\frac{9153{,}353}{12}} = 27{,}618.$$

Тогда линейный коэффициент корреляции:

- $r = 11{,}241/12 = 0{,}937$;
- $r = 1485{,}066/(12 \cdot 4{,}784 \cdot 27{,}618) = 0{,}937$;
- $r = (106\,317{,}681/12 - 36{,}602 \cdot 238{,}674)/(4{,}784 \cdot 27{,}618) = 0{,}937$.

Найденное значение свидетельствует о том, что связь между величиной стоимостного внешнеторгового товарооборота и величиной таможенных платежей в федеральный бюджет очень близка к функциональной (сильная по шкале Чэддока).



Проверка коэффициента корреляции на значимость (существенность).

Интерпретируя значение коэффициента корреляции, следует иметь в виду, что он рассчитан для ограниченного числа наблюдений и подвержен случайным колебаниям, как и сами значения x и y, на основе которых он рассчитан. Другими словами, как любой выборочный показатель, он содержит случайную ошибку и не всегда однозначно отражает действительно реальную связь между изучаемыми показателями. Для того чтобы оценить существенность (значимость) самого r и соответственно реальность измеряемой связи между x и y, необходимо рассчитать среднюю квадратическую ошибку коэффициента корреляции σ_r. Оценка существенности (значимости) r основана на сопоставлении значения r с его средней квадратической ошибкой: $|r|/\sigma_r$.

Существуют некоторые особенности расчета σ_r в зависимости от числа наблюдений (объема выборки) — n.

1. Если число наблюдений достаточно велико ($n > 30$), то σ_r рассчитывается по формуле:

$$\sigma_r = \frac{1 - r^2}{\sqrt{n}}.$$

Обычно, если $|r|/\sigma_r > 3$, то r считается значимым (существенным), а связь — реальной. Задавшись определенной вероятностью, можно определить *доверительные пределы (границы)* $r = (r \pm t\sigma_r)$, где t — коэффициент доверия, рассчитываемый по интегралу Лапласа (см. Приложение А).

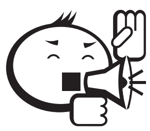
2. Если число наблюдений небольшое ($n < 30$), то σ_r рассчитывается по формуле:

$$\sigma_r = \frac{\sqrt{1 - r^2}}{\sqrt{n - 2}},$$

а значимость r проверяется на основе t-критерия Стьюдента, для чего определяется расчетное значение критерия по формуле и сопоставляется с $t_{\text{ТАБЛ}}$.

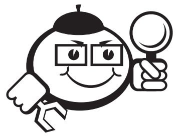
$$t_{\text{РАСЧ}} = \frac{|r|}{\sigma_r} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Табличное значение $t_{\text{ТАБЛ}}$ находится по таблице распределения t-критерия Стьюдента (см. Приложение Б) при уровне значимости $\alpha = 1 - \beta$ и числе степеней свободы $\nu = n - 2$.



Если $t_{\text{РАСЧ}} > t_{\text{ТАБЛ}}$, то r считается значимым, а связь между x и y — реальной.

В противном случае ($t_{\text{РАСЧ}} < t_{\text{ТАБЛ}}$) считается, что связь между x и y отсутствует и значение r, отличное от нуля, получено случайно.



Пример

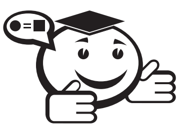
В нашей задаче число наблюдений небольшое, значит, оценивать существенность (значимость) линейного коэффициента корреляции будем по следующей формуле:

$$\sigma_r = \frac{\sqrt{1-0{,}937^2}}{\sqrt{12-2}} = \frac{0{,}349}{3{,}162} = 0{,}110;$$

$$t_{\text{РАСЧ}} = \frac{|r|}{\sigma_r} = \frac{0{,}937}{0{,}110} = 8{,}482.$$

Из Приложения Б видно, что при числе степеней свободы $\nu = 12 - 2 = 10$ (в 10-й строке) и вероятности $\beta = 95\%$ (уровень значимости $\alpha = 1 - \beta = 0{,}05$) $t_{\text{ТАБЛ}} = 2{,}2281$, а при вероятности 99% ($\alpha = 0{,}01$) $t_{\text{ТАБЛ}} = 3{,}169$, значит, $t_{\text{РАСЧ}} > t_{\text{ТАБЛ}}$, что дает возможность считать линейный коэффициент корреляции $r = 0{,}937$ значимым.

5. ***Подбор уравнения регрессии***[1].



Уравнение регрессии *представляет собой математическое описание изменения взаимно коррелируемых величин по эмпирическим (фактическим) данным.*

Уравнение регрессии должно определить, каким будет среднее значение результативного признака y при том или ином значении факторного признака x, если

[1] Термин «регрессия» ввел в статистику Ф. Гальтон, который, изучив большое число семей, установил, что в группе семей с высокорослыми отцами сыновья в среднем ниже ростом, чем их отцы, а в группе семей с низкорослыми отцами сыновья в среднем выше отцов, т. е. отклонение роста от среднего в следующем поколении уменьшается — регрессирует.

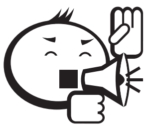
остальные факторы, влияющие на y и не связанные с x, не учитывать, т. е. абстрагироваться от них. Другими словами, уравнение регрессии можно рассматривать как вероятностную гипотетическую функциональную связь величины результативного признака y со значениями факторного признака x.

Уравнение регрессии можно также назвать *теоретической линией регрессии*. Рассчитанные по уравнению регрессии значения результативного признака называются *теоретическими*. Они обычно обозначаются $\widehat{y}_x$ или $\bar{y}_x$ (читается: «игрек, выравненный по икс») и рассматриваются как функция от x, т. е. $\widehat{y}_x = f(x)$.

Найти в каждом конкретном случае тип функции, с помощью которой можно наиболее адекватно отразить ту или иную зависимость между признаками x и y, — одна из основных задач регрессионного анализа. Выбор теоретической линии регрессии часто обусловлен формой эмпирической линии регрессии; теоретическая линия как бы сглаживает изломы эмпирической линии регрессии. Кроме того, необходимо учитывать природу изучаемых показателей и специфику их взаимосвязей [18].

Для аналитической связи между x и y могут использоваться виды уравнений, приведенные в таблице 4.4 (при условии замены t на x). Обычно зависимость, выражаемую уравнением прямой, называют *линейной* (или *прямолинейной)*, а все остальные — *криволинейными зависимостями*.

Выбрав тип функции, по эмпирическим данным определяют параметры уравнения. При этом отыскиваемые параметры должны быть такими, при которых рассчитанные по уравнению теоретические значения результативного признака $\widehat{y}_x$ были бы максимально близки к эмпирическим данным.



Существует несколько методов нахождения параметров уравнения регрессии. Наиболее часто используется *метод наименьших квадратов* (*МНК*). Его суть заключается в следующем требовании: искомые теоретические значения результативного признака $\widehat{y}_x$ должны быть такими, при которых бы обеспечивалась минимальная сумма квадратов их отклонений от эмпирических значений, т. е.

$$S = \sum (y_i - \widehat{y}_x)^2 \to \min .$$

Поставив данное условие, легко определить, при каких значениях a_0, a_1 и т. д. для каждой аналитической кривой эта сумма квадратов отклонений будет минимальной. Данный метод уже использовался нами в пп. 4.3. Методы выявления основной тенденции (тренда) в рядах динамики, поэтому для нахождения параметров теоретической линии регрессии, заменив параметр t на x, получим:

$$\begin{cases} na_0 + a_1 \sum x_i = \sum y_i; \\ a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i. \end{cases}$$

Выразив из первого уравнения системы a_0, получим:

$$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n} = \bar{y} - a_1 \bar{x}.$$

Подставив во второе уравнение системы, затем разделив обе его части на n, получим:

$$(\bar{y} - a_1\bar{x})\,\frac{\sum x_i}{n} + a_1\frac{\sum x_i^2}{n} = \frac{\sum x_i y_i}{n}.$$

Применяя 3 раза формулу средней арифметической, получим:

$$(\bar{y} - a_1\bar{x})\,\bar{x} + a_1\overline{x^2} = \overline{xy}.$$

Раскрыв скобки и перенеся члены без a_1 в правую часть уравнения, выразим a_1:

$$a_1 = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x^2}.$$

Параметр a_1 в уравнении линейной регрессии называется *коэффициентом регрессии*, который показывает, на сколько изменяется значение результативного признака y при изменении факторного признака x на единицу.

Пример

Исходные данные и расчеты для нашего примера представим в таблице 5.5.

$$a_1 = \frac{106\,317{,}681/12 \;- 36{,}602 \cdot 238{,}674}{4{,}784^2} = 5{,}407.$$

$$a_0 = 238{,}674 - 5{,}407 \cdot 36{,}602 = 40{,}767.$$

Отсюда получаем уравнение регрессии:

$$\widehat{y}_x = 40{,}72 + 5{,}4082x,$$

подставляя в которое вместо x эмпирические значения факторного признака (2-й столбец таблицы 5.5), получаем выравненные по прямой линии теоретические значения результативного признака $\widehat{y}_x$ (6-й столбец таблицы 5.5). Для иллюстрации различий между эмпирическими и теоретическими линиями регрессии построим график (рис. 5.2).

Таблица 5.5 – Вспомогательные расчеты для нахождения уравнения регрессии

№п/п	x	y	x^2	xy	$\widehat{y}_x$	$(y-\widehat{y}_x)^2$	$(\widehat{y}_x-\overline{y})^2$
1	27,068	172,17	732,677	4660,298	187,124	223,612	2657,453
2	29,889	200,90	893,352	6004,700	202,377	2,181	1317,497
3	33,158	232,10	1099,453	7695,972	220,052	145,147	346,774
4	34,444	231,83	1186,389	7985,153	227,006	23,274	136,153
5	37,299	246,53	1391,215	9195,322	242,443	16,706	14,202
6	37,554	236,99	1410,303	8899,922	243,821	46,669	26,495
7	37,755	233,40	1425,440	8812,017	244,908	132,441	38,864
8	37,909	256,43	1437,092	9721,005	245,741	114,256	49,940
9	38,348	261,89	1470,569	10 042,958	248,115	189,761	89,122
10	39,137	259,36	1531,705	10 150,572	252,381	48,710	187,871
11	40,370	253,62	1629,737	10 238,639	259,048	29,459	415,076
12	46,298	278,87	2143,505	12 911,123	291,100	149,580	2748,498
Итого	439,229	2864,09	16 351,437	10 6317,681	2864,115	1121,795	8027,945

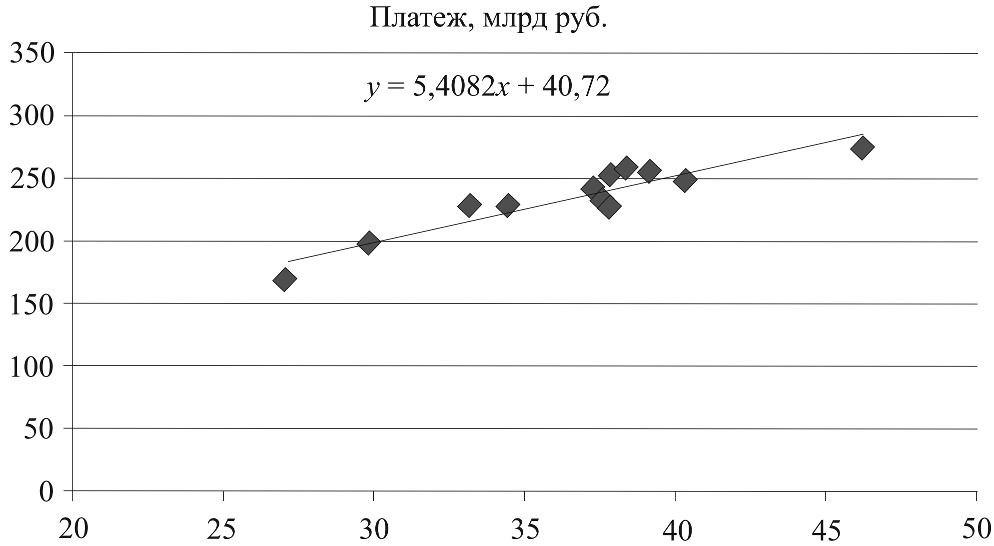


Рис. 5.2 – График эмпирических данных и теоретической линии регрессии

Из рисунка 5.2 видно, что небольшие различия между эмпирическими данными и теоретической линией регрессии существуют, поэтому необходимо *оценить существенность* коэффициента регрессии и уравнения связи, для чего определяют среднюю ошибку параметров уравнения регрессии и сравнивают их с этой ошибкой.

. . . .

Расчет ошибок параметров уравнения регрессии основан на использовании остаточной дисперсии, характеризующей расхождение (отклонение) между эмпирическими и теоретическими значениями результативного признака. Для линейного уравнения регрессии ($\widehat{y}_x = a_0 + a_1 x$) средние ошибки параметров a_1 и a_2 определяются соответственно:

$$\mu_{a_0} = \frac{\sigma_{\text{ОСТ}}}{\sqrt{n-2}},$$

$$\mu_{a_0} = \frac{\sigma_{\text{ОСТ}}}{\sigma_x \sqrt{n-2}},$$

$$\sigma_{\text{ОСТ}} = \sqrt{\frac{\sum (y_i - \widehat{y}_x)^2}{n}}.$$

Значимость параметров проверяется путем сопоставления его значения со средней ошибкой. Обозначим это соотношение как t:

$$t_{a_i} = \frac{a_i}{\mu_{a_i}},$$

При большом числе наблюдений ($n > 30$) параметр a_i считается значимым, если $t_{a_i} > 3$.

Если выборка малая ($n < 30$), то значимость параметра a_i проверяется путем сравнения с табличным значением t-критерия Стьюдента при числе степеней свободы $\nu = n-2$ и заданном уровне значимости α (Приложение Б). Если рассчитанное значение t больше табличного, то параметр считается значимым.

Пример

В нашем примере:

$$\sigma_{\text{ОСТ}} = \sqrt{\frac{1121{,}795}{12}} = 9{,}669.$$

Находим среднюю ошибку параметра a_0:

$$\mu_{a_0} = \frac{9{,}669}{\sqrt{12-2}} = 3{,}06.$$

Теперь находим среднюю ошибку параметра a_1:

$$\mu_{a_0} = \frac{9{,}669}{4{,}784\sqrt{12-2}} = 0{,}639.$$

Теперь для параметра a_0:

$$t_{a_0} = \frac{40{,}767}{3{,}06} = 13{,}3.$$

И по той же формуле для параметра a_1:

$$t_{a_1} = \frac{5{,}407}{0{,}639} = 8{,}46.$$

Так как выборка малая, то, задавшись стандартной значимостью $\alpha = 0{,}05$, находим в 10-й строке Приложения Б табличное значение $t_\alpha = 2{,}23$, которое значительно меньше полученных значений 13,3 и 8,46, что свидетельствует о значимости обоих параметров уравнения регрессии.

Наряду с проверкой значимости отдельных параметров осуществляется *проверка значимости уравнения регрессии* в целом, или, что то же самое, проверка адекватности модели с помощью критерия Фишера по Приложению В.

Пример

Данный метод уже использовался нами для проверки адекватности уравнения тренда в предыдущей теме, поэтому в нашем примере получим:

$$F_P = \frac{(12-2)\cdot 8027{,}945}{(2-1)\cdot 1121{,}795} = 71{,}56.$$

Сравнивая расчетное значение критерия Фишера $F_P = 71{,}56$ с табличным $F_T = 4{,}96$, определяемым по Приложению В при числе степеней свободы $\nu_1 = k - 1 =$

$= 2 - 1 = 1$ и $\nu_2 = n - k = 12 - 2 = 10$ (т. е. 1-й столбец и 10-я строка) и стандартном уровне значимости $\alpha = 0{,}05$, можно сделать вывод, что уравнение регрессии значимо.

6. ***Коэффициент эластичности***.

Коэффициент эластичности *показывает, на сколько процентов изменяется в среднем результативный признак y при изменении факторного признака x на 1%. Он рассчитывается на основе уравнения регрессии:*

$$Э = \frac{\partial \widehat{y}_x}{\partial x} \frac{x}{\widehat{y}_x},$$

где $\partial \widehat{y}_x / \partial x$ — *первая производная уравнения регрессии y по x.*

Коэффициент эластичности — величина переменная, т. е. изменяется с изменением значений фактора x. Так, для линейной зависимости $\widehat{y}_x = a_0 + a_1 x$:

$$Э = a_1 \frac{x}{a_0 + a_1 x}.$$

Пример

Применительно к рассмотренному уравнению регрессии, выражающему зависимость величины таможенных платежей в федеральный бюджет от величины стоимостного внешнеторгового оборота ($\widehat{y}_x = 40{,}767 + 5{,}407x$), коэффициент эластичности:

$$Э = \frac{5{,}407x}{40{,}767 + 5{,}407x}.$$

Подставляя в данное выражение разные значения x, получаем и разные значения Э. Так, например, при $x = 40$ коэффициент эластичности

$$Э = \frac{5{,}407 \cdot 40}{40{,}767 + 5{,}407 \cdot 40} = 0{,}84,$$

а при $x = 50$ соответственно

$$Э = \frac{5{,}407 \cdot 50}{40{,}767 + 5{,}407 \cdot 50} = 0{,}87 \text{ и т. д.}$$

Это значит, что при увеличении внешнеторгового товарооборота x с 40 до 40,4 млрд долл. (т. е. на 1%) величина таможенных платежей возрастет в среднем на 0,84% прежнего уровня; при увеличении x с 50 до 50,5 млрд долл. (т. е. на 1%) y возрастет на 0,87% и т. д.

Контрольные вопросы по главе 5

1. Что такое функциональная зависимость?
2. Что такое стохастическая зависимость?
3. Что такое корреляционная связь?
4. Как определяется линейный коэффициент корреляции?
5. Что такое уравнение регрессии?
6. В чем суть метода наименьших квадратов?
7. Как оценить адекватность параметров уравнения регрессии?

Глава 6

ИНДЕКСЫ

__Индекс__ — относительная величина, показывающая, во сколько раз уровень изучаемого явления в данных условиях отличается от уровня того же явления в других условиях.

В статистическом анализе индексы используются не только для сопоставления уровней явлений, но и для установления значимости причин, вызывающих их изменение.

6.1 Индивидуальные индексы

Если анализируются простые явления или не имеет значения структура сложных явлений, то применяются индивидуальные индексы.

Общий вид индивидуального индекса:

$$i = \frac{a_1}{a_0},$$

где 0 — базисный период; 1 — отчетный период.

Например, такие простые явления, как:

- q — количество проданного товара,
- p — его цена за единицу,

своим произведением образуют такое сложное явление, как выручка от продаж:

$$Q = q \cdot p.$$

Сравнение их значений по отдельности для конкретного товара в отчетном периоде времени относительно какого-либо базисного периода и дает индивидуальные индексы:

- количества товара $i_q = q_1/q_0$;
- его цены $i_p = p_1/p_0$;
- выручки от продаж $i_Q = Q_1/Q_0$.

Очевидно, что индивидуальный индекс сложного явления формируется из таких индексов простых его составляющих по типологической формуле его определения. То есть

$$i_Q = i_q \cdot i_p.$$

Подставив сюда индивидуальный индекс выручки, записываем:

$$\frac{Q_1}{Q_0} = i_q \cdot i_p,$$

откуда получаем, что

$$Q_1 = i_q \cdot i_p \cdot Q_0.$$

Общее абсолютное изменение выручки равняется:

$$\Delta Q = Q_1 - Q_0,$$

а ее изменение от каждого фактора определяется следующим образом. От изменения количества товара при постоянной цене ($i_p = 1$) оно равно

$$\Delta Q_q = i_q Q_0 - Q_0 = (i_q - 1)Q_0 = (q_1 - q_0)p_0,$$

а при изменении цены оно определяется по формулам:

$$\Delta Q_p = Q_1 - Q_0 - \Delta Q_q = i_q(i_p - 1)Q_0 = q_1(p_1 - p_0).$$

Пример

Например, если выручка от продаж возросла с $Q_0 = 8$ млн руб. в предыдущем периоде до $Q_1 = 12{,}18$ млн руб. в последующем при увеличении количества проданного товара на 5% ($i_q = 1{,}05$) и повышении цены на 45% ($i_p = 1{,}45$), то можно по формуле записать, что

$$Q_1 = 1{,}05 \cdot 1{,}45 \cdot 8 = 12{,}18 \text{ млн руб.}$$

При этом весь прирост выручки в сумме

$$\Delta Q = 12{,}18 - 8 = 4{,}18 \text{ млн руб.}$$

вызван увеличением обоих факторов.

За счет изменения количества проданного товара он равен:

$$\Delta Q_q = (1{,}05 - 1) \cdot 8 = 0{,}4 \text{ млн руб.},$$

а за счет изменения цены:

$$\Delta Q_p = 1{,}05 \cdot (1{,}45 - 1) \cdot 8 = 3{,}78 \text{ млн руб.}$$

Для контроля отмечаем, что сумма факторных изменений выручки равна общему:

$$0{,}4 + 3{,}78 = 4{,}18 \text{ млн руб.}$$

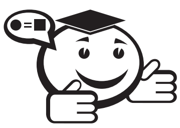
Данные расчеты осуществлены исходя из того, что в основной формуле выручки количество товара — первый фактор, а цена — второй.

6.2 Агрегатные (сводные) индексы

Значительно сложнее, если необходимо соизмерить не отдельный элемент (цену, объем выпущенных одноименных и т. п.), а всю совокупность в целом.

В этом случае необходимо использовать общие индексы (I), которые могут быть агрегатными или средними.

В *агрегатных индексах* для разнородной совокупности находится такой общий показатель (агрегат), в котором можно объединить все ее элементы [11].



Агрегатный показатель *представляет собой произведение качественного показателя (показывает экономическую сущность явления) и количественного (представляет объем изучаемой совокупности).*

Объективность общим индексам придает их запись в агрегатном виде, предложенная *Ласпейресом* и *Пааше*.

Агрегатный общий индекс Ласпейреса для количества товаров как первого фактора выручки определяется по формуле:

$$I_q^{\text{Л}} = \frac{\sum q_1 p_0}{\sum q_0 p_0}.$$

Аналогично можно записать агрегатный общий индекс Ласпейреса для цен как первого фактора выручки, то есть

$$I_p^{\text{Л}} = \frac{\sum p_1 q_0}{\sum p_0 q_0}.$$

В формулах Ласпейреса знаменатели по существу одинаковые, представляя собой выручку базисного периода, а числители разные. В индексе количества это отчетная выручка в базисных ценах, в индексе цен, наоборот — базисная выручка в отчетных ценах.

Агрегатные общие индексы Пааше применяются ко вторым факторам мультипликативных моделей. Поэтому такой индекс для цен как второго фактора выручки определяется по формуле:

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

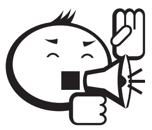
Аналогично можно записать агрегатный общий индекс Пааше для количества товаров как второго фактора выручки, то есть

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

В формулах Пааше числители по существу одинаковые, представляя собой выручку отчетного периода, а знаменатели аналогичны числителям формул Ласпейреса.

Произведения количественного индекса Ласпейреса и ценового индекса Пааше, а также ценового индекса Ласпейреса и количественного индекса Пааше дают общий индекс выручки

$$I_Q = I_q^{\text{Л}} I_p^{\text{П}} = I_p^{\text{Л}} I_q^{\text{П}}.$$

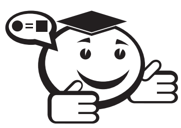


Однако вид этих формул показывает, что однофакторные индексы Ласпейреса и Пааше не равны между собой. То есть не равными являются количественные индексы Ласпейреса и Пааше и ценовые. Американский экономист Гершенкрон обширными расчетами установил, что по одному и тому же фактору индекс Ласпейреса обычно больше индекса Пааше, и это открытие названо *эффектом Гершенкрона*.

Но в статистике должно быть одно значение индекса, поэтому американский экономист Фишер предложил применять среднюю геометрическую величину из индексов Ласпейреса и Пааше, определяя ее по формулам:

для количества товаров $I_q^{\text{Ф}} = \sqrt{I_q^{\text{Л}} I_q^{\text{П}}}$;

для цен $I_p^{\text{Ф}} = \sqrt{I_p^{\text{Л}} I_p^{\text{П}}}$.



На современном этапе экономических расчетов ***при построении агрегатных индексов существует правило: если индексируется (соизмеряется) качественный показатель, то весами к нему берется количественный в отчетном уровне; если индексируется количественный показатель, то весами берется качественный показатель в базисном уровне.***

Согласно этому правилу ***агрегатный индекс объема продукции*** *(индекс физического объема) рассчитывается по формуле Ласпейреса:*

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

Агрегатный индекс объема продукции показывает, во сколько раз изменилась стоимость продукции за счет изменения объема выпуска продукции.

А ***агрегатный индекс цен*** *будет определяться по формуле Пааше:*

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}.$$

Агрегатный индекс цен показывает, во сколько раз изменилась стоимость продукции за счет изменения цен.

Агрегатный индекс товарооборота (I_{pq}) *показывает, во сколько раз изменилась стоимость выпущенной продукции. Его можно определить двумя способами:*

$$I_{pq} = I_p \cdot I_q \text{ или } I_{pq} = \frac{\sum q_1 p_1}{\sum q_0 p_0}.$$

Для того, чтобы определить *абсолютное влияние факторов (в ден. единицах) на изменение товарооборота*, нужно из числителя соответствующего индекса вычесть его знаменатель.

Так, абсолютное изменение товарооборота определяется по формуле:

$$\Delta Q = Q_1 - Q_0 = \sum q_1 p_1 - \sum q_0 p_0.$$

В том числе за счет изменения объемов выпуска продукции:

$$\Delta Q(q) = \sum q_1 p_0 - \sum q_0 p_0 = \sum (q_1 - q_0) p_0$$

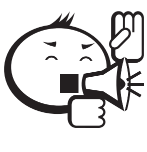
и под влиянием изменения цен:

$$\Delta Q(p) = \sum q_1 p_1 - \sum q_1 p_0 = \sum q_1 (p_1 - p_0).$$

Общее изменение товарооборота равно сумме изменений за счет факторов:

$$\Delta Q = \Delta Q(q) + \Delta Q(p).$$

6.3 Общие индексы как средние из индивидуальных



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

Используя их формулы, можем записывать, что $q_1 = q_0 i_q$ и $p_1 = p_0 i_p$, а также, что $q_0 = q_1/i_q$ и $p_0 = p_1/i_p$. Подставив отчетные значения количества товара и цены в формулу общего индекса выручки, получим

$$I_Q = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_0 i_q p_0 i_p}{\sum q_0 p_0} = \frac{\sum q_0 p_0 i_Q}{\sum q_0 p_0}.$$

Значит, общий индекс выручки можно определять только через ее базисные значения с умножением в числителе на индивидуальный индекс выручки по конкретному товару.

Теперь подставим базисные значения количества товара и цены в формулу общего индекса выручки. Тогда получим

$$I_Q = \frac{\sum q_1 p_1}{\sum q_1 p_1 / i_Q}.$$

Значит, общий индекс выручки можно определять только через ее отчетные значения с делением в знаменателе на индивидуальный индекс выручки по конкретному товару.

6.4 Индексы переменного состава, постоянного состава и индексы структурных сдвигов

Вышеизложенные общие индексы применимы к изучению явлений, образованных как разными, так и однородными процессами. В последнем случае динамику итога можно показать через общие индексы отдельных факторов.

Общая динамика среднего уровня называется ***индексом переменного состава*** $(I_{\text{п. с}})$

$$I_{\text{п. с}} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0}.$$

Если значения $q_i / \sum q_i$ заменить на показатель структуры d_i, то индекс переменного состава примет вид:

$$I_{\text{п. с}} = \frac{\sum p_1 d_1}{\sum p_0 d_0}.$$

Индексы среднего уровня, рассчитанные исходя из постоянной структуры, называются ***индексами постоянного состава*** $I_{\text{ф. с}}$*, которые показывают изменение среднего уровня под влиянием изменений индивидуальных значений признака*

$$I_{\text{ф. с}} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_1}{\sum q_1} = \frac{\sum p_1 d_1}{\sum p_0 d_1}.$$

Влияние структурных сдвигов на динамику среднего уровня определяется ***индексом структурных сдвигов*** $(I_{\text{с. с}})$

$$I_{\text{с. с}} = \frac{\sum p_0 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} = \frac{\sum p_0 d_1}{\sum p_0 d_0}.$$

. . . .

Рассчитанные индексы находятся во взаимосвязи

$$I_{\text{п. с}} = I_{\text{ф. с}} \cdot I_{\text{с. с}}.$$

6.5 Базисные и цепные индексы

При исчислении индексов за ряд периодов можно рассчитывать цепные и базисные индексы. Причем если для всех индексов берутся веса на уровне какого-то одного периода, то такой индексный ряд называется индексным рядом с постоянными весами. Если веса берутся на уровне разных периодов, — то с переменными весами. Правило для индексного ряда с постоянными весами — произведение цепных индексов равно соответствующему базисному индексу.

В каждом отдельном индексе веса в его числителе и знаменателе обязательно фиксируются на одном и том же уровне. Если же строится ряд индексов, то веса в нем могут быть либо постоянными для всех индексов ряда, либо переменными.

Ряд базисных индексов объема продукции:

$$I_q^{1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_q^{2/0} = \frac{\sum q_2 p_0}{\sum q_0 p_0}; \quad I_q^{3/0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}; \ \dots$$

имеет постоянные веса (p_0).

Постоянные веса (p_0) имеет и ряд цепных индексов:

$$I_q^{1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad I_q^{2/1} = \frac{\sum q_2 p_0}{\sum q_1 p_0}; \quad I_q^{3/2} = \frac{\sum q_3 p_0}{\sum q_2 p_0}; \ \dots$$

Ряд базисных индексов цен:

$$I_p^{1/0} = \frac{\sum q_1 p_1}{\sum q_1 p_0}; \quad I_p^{2/0} = \frac{\sum q_2 p_2}{\sum q_2 p_0}; \quad I_p^{3/0} = \frac{\sum q_3 p_3}{\sum q_3 p_0}; \ \dots$$

Ряд цепных индексов цен:

$$I_p^{1/0} = \frac{\sum q_1 p_1}{\sum q_1 p_0}; \quad I_p^{2/1} = \frac{\sum q_2 p_2}{\sum q_2 p_1}; \quad I_p^{3/2} = \frac{\sum q_3 p_3}{\sum q_3 p_2}; \ \dots$$

Для индексов динамики с постоянными весами имеет силу взаимосвязь между цепными и базисными темпами роста (индексами).

Таким образом, использование постоянных весов в течение ряда лет позволяет переходить от цепных индексов к базисным, и наоборот. Поэтому ряды индексов объема продукции и объема проданных товаров строятся в статистической практике с постоянными весами. Так, в индексах объема продукции в качестве постоянных весов используются цены, зафиксированные на уровне, который был установлен на 1 января какого-либо базисного года. Такие цены, используемые в течение ряда лет, называются сопоставимыми (фиксированными).

Использование в индексах объема продукции (товаров) сопоставимых цен позволяет путем простого суммирования получать итоги за несколько лет. Сопоставимые цены не должны сильно отличаться от действующих (текущих) цен, поэтому их периодически пересматривают, переходя к новым сопоставимым ценам. Чтобы иметь возможность исчислять индексы объема продукции за длительные периоды, в течение которых применялись различные сопоставимые цены, продукцию одного года оценивают как в прежних, так и в новых фиксированных ценах. Индекс за длительный период исчисляют цепным методом, т. е. путем перемножения индексов за отдельные отрезки этого периода [9].

Ряды индексов качественных показателей, которые экономически правильно взвешивать по весам текущего периода, строятся с переменными весами.

Пример

Процесс определения всевозможных индексов и факторного анализа сложного явления рассмотрим на примере фирмы, выпускающей два вида продукции. Исходные данные приведены в таблице 6.1.

Таблица 6.1 – Результаты работы двух фирм по выпуску однородного продукта

Изделие	**Базисный период**		**Отчетный период**	
	Количество продукта q_0, тыс. ед.	**Цена p_0, руб./ед.**	**Количество продукта q_1, тыс. ед.**	**Цена p_1, руб./ед.**
1	100	20	140	15
2	150	22	160	25
Итого	250	21,20	300	20,23

В таблице 6.1 итоговая цена представляет собой среднюю арифметическую взвешенную величину.

Так, для базисного периода она равна

$$\overline{p_0} = \frac{\sum q_0 p_0}{\sum q_0} = \frac{100 \cdot 20 + 150 \cdot 22}{100 + 150} = \frac{5300}{250} = 21{,}20 \text{ руб./ед.}$$

Для отчетного периода средняя цена равняется

$$\overline{p_1} = \frac{\sum q_1 p_1}{\sum q_1} = \frac{14 \cdot 15 + 160 \cdot 25}{140 + 160} = \frac{6100}{300} = 20{,}23 \text{ руб./ед.}$$

Найдем индивидуальные индексы:

1) количества товара

- по изделию A:

$$i_q = \frac{q_1}{q_0} = \frac{140}{100} = 1{,}4;$$

- по изделию *B*:

$$i_q = \frac{q_1}{q_0} = \frac{160}{150} = 1{,}067;$$

2) его цены
- по изделию *A*:

$$i_p = \frac{p_1}{p_0} = \frac{15}{20} = 0{,}667;$$

- по изделию *B*:

$$i_p = \frac{p_1}{p_0} = \frac{25}{22} = 1{,}136;$$

3) выручки от продаж
- по изделию *A*:

$$i_Q = \frac{Q_1}{Q_0} = \frac{140 \cdot 15}{100 \cdot 20} = \frac{2100}{2000} = 1{,}05;$$

- по изделию *B*:

$$i_Q = \frac{Q_1}{Q_0} = \frac{160 \cdot 25}{150 \cdot 22} = \frac{4000}{3300} = 1{,}212.$$

Далее вычислим агрегатные индексы количества, цен и товарооборота, а также абсолютное изменение товарооборота под влиянием факторов цены и количества:

1) индекс товарооборота:

$$I_{pq} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{140 \cdot 15 + 160 \cdot 25}{100 \cdot 20 + 150 \cdot 22} = \frac{6100}{5300} = 1{,}151;$$

- абсолютное изменение товарооборота:

$$\Delta Q = Q_1 - Q_0 = \sum q_1 p_1 - \sum q_0 p_0 = 6100 - 5300 = 800 \text{ тыс. руб.};$$

2) индекс количества:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{140 \cdot 20 + 160 \cdot 22}{100 \cdot 20 + 150 \cdot 22} = \frac{6320}{5300} = 1{,}192;$$

- абсолютное изменение товарооборота за счет изменения объемов выпуска продукции:

$$\Delta Q(q) = \sum q_1 p_0 - \sum q_0 p_0 = 6320 - 5300 = 1020 \text{ тыс. руб.};$$

3) индекс цен:

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{140 \cdot 15 + 160 \cdot 25}{140 \cdot 20 + 160 \cdot 22} = \frac{6100}{6320} = 0{,}965;$$

- абсолютное изменение товарооборота под влиянием изменения цен:

$$\Delta Q(p) = \sum q_1 p_1 - \sum q_1 p_0 = 6100 - 6320 = -220;$$

4) контроль по формулам:

$$I_{pq} = I_p \cdot I_q = 1{,}192 \cdot 0{,}965 = 1{,}151;$$

$$\Delta Q = \Delta Q(q) + \Delta Q(p) = 1020 - 220 = 800.$$

Вывод. Товарооборот увеличился на 15,1% или 800 тыс. руб., в том числе за счет изменения количества выпуска на 19,2% (1020 тыс. руб.), и снизился на 3,5% (220 тыс. руб.) под влиянием цен.

Сумма долей количества продукта по периодам должна равняться единице:

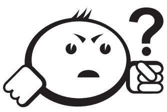
$$\sum d_0 = 0{,}4 + 0{,}6 = 1; \quad \sum d_1 = 0{,}467 + 0{,}533 = 1.$$

Индекс переменного состава:

$$I_{\Pi} = \frac{\sum d_1 p_1}{\sum d_0 p_0} = \frac{0{,}467 \cdot 15 + 0{,}533 \cdot 25}{21{,}2} = \frac{20{,}33}{21{,}2} = 0{,}959.$$

Индекс структурных сдвигов:

$$I_d = \frac{\sum d_1 p_0}{\sum d_0 p_0} = \frac{0{,}467 \cdot 20 + 0{,}533 \cdot 22}{0{,}4 \cdot 20 + 0{,}6 \cdot 22} = \frac{21{,}07}{21{,}2} = 0{,}994.$$



Контрольные вопросы по главе 6

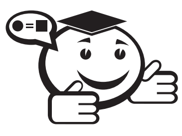
1. Что такое индекс? Что он показывает?
2. Какие виды индексов Вам известны?
3. В чем отличие индивидуальных индексов от агрегатных?
4. В чем отличие индексов Пааше от индексов Ласпейреса?
5. Влияние каких факторов, качественных или количественных, определяется в первую очередь?
6. Как при помощи индексов определить абсолютное изменение результативного показателя под влиянием факторов?

Глава 7

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

7.1 Понятие выборочного наблюдения

Выборочный метод используется, когда применение сплошного наблюдения физически невозможно из-за огромного массива данных или экономически нецелесообразно. Физическая невозможность имеет место, например, при изучении пассажиропотоков, рыночных цен, семейных бюджетов. Экономическая нецелесообразность имеет место при оценке качества товаров, связанной с их уничтожением. Например, дегустация, испытание кирпичей на прочность и т. п. Выборочное наблюдение используется также для проверки результатов сплошного.



Статистические единицы, отобранные для наблюдения, составляют ***выборочную*** *совокупность, или* ***выборку****, а весь их массив —* ***генеральную*** *совокупность (ГС).*

При этом число единиц в выборке обозначают n, во всей ГС — N. Отношение n/N называется относительным размером, или *долей выборки*.

Качество результатов выборочного наблюдения зависит от *репрезентативности* выборки, т. е. от того, насколько она представительна в ГС. Для обеспечения репрезентативности выборки необходимо соблюдать принцип случайности отбора единиц, который предполагает, что на включение единицы ГС в выборку не может повлиять какой-либо иной фактор, кроме случая [6].

7.2 Способы формирования выборки

1. *Собственно случайный* отбор: все единицы ГС нумеруются, а выпавшие в результате жеребьевки номера соответствуют единицам, попавшим в вы-

борку, причем число номеров равно запланированному объему выборки. На практике вместо жеребьевки используют генераторы случайных чисел. Данный способ отбора может быть *повторным* (когда каждая единица, отобранная в выборку, после проведения наблюдения возвращается в ГС и может быть вновь подвергнута обследованию) и *бесповторным* (когда обследованные единицы в ГС не возвращаются и не могут быть обследованы повторно). При повторном отборе вероятность попадания в выборку для каждой единицы ГС остается неизменной, а при бесповторном отборе она меняется (увеличивается), но для оставшихся в ГС после отбора из нее нескольких единиц вероятность попадания в выборку одинакова.

2. *Механический* отбор: отбираются единицы генеральной совокупности с постоянным шагом N/n. Так, если одна генеральная совокупность содержит 100 тыс. ед., а требуется выбрать 1 тыс. ед., то в выборку попадет каждая сотая единица.
3. *Стратифицированный* (расслоенный) отбор осуществляется из неоднородной генеральной совокупности, когда ее предварительно разбивают на однородные группы, после чего производят отбор единиц из каждой группы в выборочную совокупность случайным или механическим способом пропорционально их численности в генеральной совокупности.
4. *Серийный* (гнездовой) отбор: случайным или механическим способом выбирают не отдельные единицы, а определенные серии (гнезда), внутри которых производится сплошное наблюдение.

7.3 Средняя ошибка выборки

После завершения отбора необходимого числа единиц в выборку и регистрации предусмотренных программой наблюдения изучаемых признаков этих единиц, переходят к расчету обобщающих показателей. К ним относят среднюю величину изучаемого признака и долю единиц, обладающих каким-либо значением этого признака. Однако если ГС произвести несколько выборок, определив при этом их обобщающие характеристики, то можно установить, что их значения будут различными, кроме того, они будут отличаться и от реального их значения в ГС, если такое определить с помощью сплошного наблюдения. Другими словами, обобщающие характеристики, рассчитанные по данным выборки, будут отличаться от их реальных значений в ГС, поэтому введем следующие условные обозначения (табл. 7.1).

Разность между значением обобщающих характеристик выборочной и генеральной совокупностей называется *ошибкой выборки*, которая подразделяется на ошибку *регистрации* и ошибку *репрезентативности*. Первая возникает из-за неправильных или неточных сведений по причинам непонимания существа вопроса, невнимательности регистратора при заполнении анкет, формуляров и т. п. Она достаточно легко обнаруживается и устраняется. Вторая возникает из-за несоблюдения принципа случайности отбора единиц в выборку. Ее сложнее обнаружить и устранить, она гораздо больше первой, и потому ее измерение является основной задачей выборочного наблюдения.

Таблица 7.1 – Условные обозначения

Показатель	Совокупность	
	генеральная	**выборочная**
Число единиц совокупности	N	n
Среднее значение	$\overline{X}$	$\widetilde{X}$
Доля единиц, обладающих каким-либо значением признака	d	$\widetilde{d}$
Доля единиц, не обладающих каким-либо значением признака	$1 - d$	$1 - \widetilde{d}$
Дисперсия	σ^2	$\widetilde{\sigma}^2$

Для измерения ошибки выборки определяется ее средняя ошибка по формуле для повторного отбора:

$$\mu = \sqrt{\frac{\widetilde{\sigma}^2}{n}}$$

и по формуле — для бесповторного:

$$\mu = \sqrt{\frac{\widetilde{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)}.$$

Из формул видно, что средняя ошибка меньше у бесповторной выборки, что и обусловливает ее более широкое применение.

7.4 Предельная ошибка выборки

Учитывая, что на основе выборочного обследования нельзя точно оценить изучаемый параметр ГС, необходимо найти пределы, в которых он находится. В конкретной выборке разность $\left|\widetilde{X}_i - \overline{X}\right|$ может быть больше, меньше или равна μ. Каждое из отклонений $\left|\widetilde{X}_i - \overline{X}\right|$ от μ имеет определенную вероятность. При выборочном обследовании реальное значение $\overline{X}$ в ГС неизвестно. Зная среднюю ошибку выборки, с определенной вероятностью можно оценить отклонение выборочной средней от генеральной и установить пределы, в которых находится изучаемый параметр (в данном случае среднее значение) в генеральной совокупности.

Отклонение выборочной характеристики от генеральной называется *предельной ошибкой выборки* Δ. Она определяется в долях средней ошибки с заданной вероятностью, т. е.

$$\Delta = t\mu,$$

где t — *коэффициент доверия*, зависящий от вероятности, с которой определяется предельная ошибка выборки.

Вероятность появления определенной ошибки выборки находят с помощью теорем теории вероятностей. Согласно теореме Чебышёва *при достаточно большом объеме выборки и ограниченной дисперсии генеральной совокупности (ГС) вероятность того, что разность между выборочной средней и генеральной средней будет сколь угодно мала, близка к единице*:

$$P\left(\left|\widetilde{X} - \overline{X}\right| \leqslant \xi\right) \to 1 \text{ при } n \to \infty.$$

А. М. Ляпунов доказал, что *независимо от характера распределения генеральной совокупности при увеличении объема выборки распределение вероятностей появления того или иного значения выборочной средней приближается к нормальному распределению* (центральная предельная теорема). Следовательно, вероятность отклонения выборочной средней от генеральной средней, т. е. вероятность появления заданной предельной ошибки, также подчиняется указанному закону и может быть найдена как функция от t с помощью интеграла вероятностей Лапласа:

$$P\left(\left|\widetilde{X} - \overline{X}\right| \leqslant t\mu\right) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt,$$

где $t = (\widetilde{X} - \overline{X})/\mu$ — нормированное отклонение выборочной средней от генеральной средней.

Значения P (интеграла Лапласа) для разных t рассчитаны и имеются в специальной таблице, которая приведена в Приложении А.

Вероятность, которая принимается при расчете выборочной характеристики, называется *доверительной*. Чаще всего принимают вероятность $P = 0{,}950$, которая означает, что только в 5 случаях из 100 ошибка может выйти за установленные границы. Задавшись конкретным уровнем вероятности, выбирают величину нормированного отклонения t по Приложению А и рассчитывают предельную ошибку выборки.

После расчета предельной ошибки находят *доверительный интервал* обобщающей характеристики ГС совокупности по формуле для среднего значения:

$$\overline{X} = \widetilde{X} \pm \Delta \text{ или } (\widetilde{X} - \Delta) \leqslant \overline{X} \leqslant (\widetilde{X} + \Delta)$$

и для доли единиц, обладающих каким-либо значением признака по формуле:

$$d = \widetilde{d} \pm \Delta \text{ или } (\widetilde{d} - \Delta) \leqslant d \leqslant (\widetilde{d} + \Delta).$$

Следовательно, при выборочном наблюдении определяется не одно, точное значение обобщающей характеристики ГС, а лишь ее доверительный интервал с заданным уровнем вероятности. И это серьезный недостаток выборочного метода статистики.

7.5 Необходимая численность выборки

Разрабатывая программу выборочного наблюдения, задаются конкретным значением предельной ошибки и уровнем вероятности. Неизвестной остается минимальная численность выборки, обеспечивающая заданную точность. Ее можно получить из формул средней и предельной ошибок в зависимости от типа выборки:

- для повторной выборки

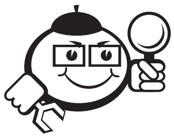
$$n = \frac{\widetilde{\sigma}^2 t^2}{\Delta^2};$$

- для бесповторной выборки

$$n = \frac{\tilde{\sigma}^2 t^2}{\Delta^2 + \tilde{\sigma}^2 t^2 / N}.$$

Вариация ($\tilde{\sigma}^2$) значений признака к началу выборочного наблюдения как правило неизвестна, поэтому ее берут приближенно одним из способов:

1) из предыдущих выборочных наблюдений;
2) по правилу «трех сигм», согласно которому в размахе вариации укладывается примерно шесть стандартных отклонений σ ($H/\sigma = 6$, отсюда $\tilde{\sigma}^2 = H^2/36$);
3) если приблизительно известна средняя величина изучаемого признака, то $\tilde{\sigma}^2 = \overline{X}^2/9$;
4) если неизвестна дисперсия доли единиц, обладающих каким-либо значением признака, то используется ее максимально возможная величина $\tilde{\sigma}^2 = 0{,}25$.



Пример

На предприятии в порядке случайной бесповторной выборки было опрошено 100 рабочих из 1000 и получены следующие данные об их доходе за месяц (табл. 7.2).

Таблица 7.2 – Результаты бесповторного выборочного наблюдения на предприятии

Доход, у. е.	до 300	300–500	500–700	700–1000	более 1000
Число рабочих	8	28	44	17	3

С вероятностью 0,950 определить:

1) *среднемесячный размер дохода работников данного предприятия;*
2) *долю рабочих предприятия, имеющих месячный доход более 700 у. е.;*
3) *необходимую численность выборки при определении среднемесячного дохода работников предприятия, чтобы не ошибиться более чем на 50 у. е.;*
4) *необходимую численность выборки при определении доли рабочих с размером месячного дохода более 700 у. е., чтобы при этом не ошибиться более чем на 5%.*

Решение:

Для расчета обобщающих характеристик выборки построим вспомогательную таблицу 7.3.

Рассчитаем средний доход в выборке: $\tilde{X} = 57\,100/100 = 571$ у. е. и дисперсию среднего выборочного дохода: $\tilde{\sigma}^2 = 4\,285\,900/100 = 42\,859$.

Таблица 7.3 – Вспомогательные расчеты для решения задачи

X	f	X'	$X'f$	$(X'-\widetilde{X})^2$	$(X'-\widetilde{X})^2 f$
до 300	8	200	1600	137 641	1 101 128
300–500	28	400	11 200	29 241	818 748
500–700	44	600	26 400	841	37 004
700–1000	17	850	14 450	77 841	1 323 297
более 1000	3	1150	3450	335 241	1 005 723
Итого	100		57 100		4 285 900

Теперь можно определить среднюю ошибку выборки:

$$\mu = \sqrt{\frac{42\,859}{100}\left(1 - \frac{100}{1000}\right)} = 19{,}640 \text{ у. е.}$$

В нашей задаче $\beta = 0{,}950$, значит, $t = 1{,}96$. Тогда предельная ошибка выборки:

$$\Delta = 1{,}96 \cdot 19{,}64 = 38{,}494 \text{ у. е.}$$

Для определения средней ошибки выборки при определении доли рабочих с доходами более 700 у. е. в ГС необходимо определить их долю: $w = 20/100 = 0{,}2$ или 20%, а затем ее дисперсию по формуле $\widetilde{\sigma}^2 = w(1 - w) = 0{,}2 \cdot (1 - 0{,}2) = 0{,}16$. Тогда можно рассчитать среднюю ошибку выборки:

$$\mu = \sqrt{\frac{0{,}16}{100}\left(1 - \frac{100}{1000}\right)} = 0{,}038 \text{ или } 3{,}8\%.$$

А затем и предельную ошибку выборки:

$$\Delta = 1{,}96 \cdot 0{,}038 = 0{,}075 \text{ или } 7{,}5\%.$$

Доверительный интервал среднего дохода:

$$571 - 38{,}494 \leqslant \overline{X} \leqslant 571 + 38{,}494$$

или

$$532{,}506 \text{ у. е.} \leqslant \overline{X} \leqslant 609{,}494 \text{ у. е.},$$

то есть средний доход всех рабочих предприятия с вероятностью 95% будет лежать в пределах от 532,5 до 609,5 у. е.

Аналогично определяем доверительный интервал для доли:

$$0{,}2 - 0{,}075 \leqslant p \leqslant 0{,}2 + 0{,}075$$

или

$$0{,}125 \leqslant p \leqslant 0{,}275,$$

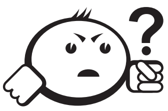
то есть доля рабочих с доходами более 700 у. е. на всем предприятии с вероятностью 95% будет лежать в пределах от 12,5% до 27,5%.

В нашей задаче выборка бесповторная, значит:

$$n_{\text{б/повт}} = \frac{42\ 859 \cdot 1{,}96^2}{50^2 + 42\ 859 \cdot 1{,}96^2/1000} = 62 \text{ чел.},$$

$$n_{\text{б/повт}} = \frac{0{,}16 \cdot 1{,}96^2}{0{,}05^2 + 0{,}16 \cdot 1{,}96^2/1000} = 197 \text{ чел.}$$

Таким образом, необходимо включить в выборку не менее 62 рабочих при определении среднего месячного дохода работников предприятия, чтобы не ошибиться более чем на 50 у. е., и не менее 197 рабочих при определении доли рабочих с размером месячного дохода более 700 у. е., чтобы при этом не ошибиться более чем на 5%.



Контрольные вопросы по главе 7

1. Что такое выборочное наблюдение?
2. Какие способы формирования выборки Вам известны?
3. Что такое средняя ошибка выборки?
4. Что такое предельная ошибка выборки?
5. Как определить необходимую численность выборки?

ЗАКЛЮЧЕНИЕ

Мы рассмотрели приемы сбора, обработки и анализа статистических данных, которые являются методологическим базисом любой статистической работы. В то же время необходимо отметить, что статистическое наблюдение не является обязательным этапом статистического исследования. Во многих случаях экономист-аналитик имеет дело с материалом, полученным из баз данных, бюллетеней информационных агентств, статистических сборников и других источников. Тогда работа должна начинаться с проверки полноты и качества данных, их группировки, а при отсутствии необходимости в этих этапах — с расчета индивидуальных и обобщающих показателей.

Рассмотренные приемы и методы с успехом могут использоваться не только в практике статистического анализа. Статистическая методология исследования в настоящее время заняла прочные позиции во многих областях знания. Статистические формулы находят применение в макро- и микроэкономике, оценке бизнеса и недвижимости, финансовом анализе, техническом анализе товарных и финансовых рынков.

Более того, подвергающийся статистической обработке материал не обязательно должен относиться к экономической области. В большинстве случаев, описанные приемы и показатели будут работоспособны и эффективны при обобщении и анализе технической, биологической, медицинской, демографической и социологической информации.

Рассматриваемые в пособии методы в большинстве случаев иллюстрированы практическими примерами. Подобные вычисления при небольших объемах совокупности или коротких динамических рядах не очень трудоемки. При работе же с большими массивами статистической информации необходимо использовать прикладное программное обеспечение, существенно ускоряющее и упрощающее все расчеты. Среди наиболее распространенных современных программных продуктов следует отметить пакеты *MS Excel* и *Statistica*.

ЛИТЕРАТУРА

[1] Социальная статистика : учебник / под ред. чл.-кор. РАН И. И. Елисеевой. — 3-е изд., перераб. и доп. — М. : Финансы и статистика, 2008. — 480 с.: ил.

[2] Экономическая статистика : учебник. — 3-е изд., перераб. и доп. / под ред. проф. Иванова Ю. Н. — М. : ИНФРА-М, 2007. — 736 с.

[3] Голуб Л. А. Социально-экономическая статистика : учеб. пособие для студ. высш. учеб. заведений / Л. А. Голуб. — М. : ВЛАДОС, 2009. — 272 с.

[4] Рудакова Р. П. Статистика / Р. П. Рудакова, Л. Л. Букин, В. И. Гаврилов. — 2-е изд. — СПб. : Питер, 2007 — 288 с.: ил.

[5] Салин В. Н. Система национальных счетов : учеб. пособие / В. Н. Салин, С. И. Кудряшова. — М. : Финансы и статистика, 2006. — 272 с.

[6] Береславская В. А. Теория статистики : учеб. пособие / В. А. Береславская, Н. М. Стрельникова, Л. А. Хинканина. — Йошкар-Ола : МарГТУ, 2008. — 136 с.

[7] Салин В. Н. Курс теории статистики для подготовки специалистов финансово-экономического профиля : учебник / В. Н. Салин, Э. Ю. Чурилова. — М. : Финансы и статистика, 2007. — 480 с.: ил.

[8] Ефимова М. Р. Общая теория статистики : учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. — 2-е изд., испр. и доп. — М. : ИНФРА-М, 2007. — 416 с.

[9] Социально-экономическая статистика. Практикум : учеб. пособие / под ред. В. Н. Салина, Е. П. Шпаковской. — М. : Финансы и статистика, 2006. — 192 с.

[10] Статистика : учеб. пособие / А. В. Багат [и др.] ; под ред. В. М. Симчеры. — М. : Финансы и статистика, 2009. — 368 с.: ил.

[11] Статистика : учеб. пособие / Л. П. Харченко [и др.] ; под ред. канд. экон. наук, проф. В. Г. Ионина. — 3-е изд., перераб. и доп. — М. : ИНФРА-М, 2008. — 445 с.

[12] Статистика : учебник / под ред. И. И. Елисеевой. — М. : Высшее образование, 2007. — 566 с.

[13] Теория статистики : учебник / Р. А. Шмойлова [и др.] ; под ред. Р. А. Шмойловой. — 5-е изд. — М. : Финансы и статистика, 2007. — 656 с.: ил.

[14] Экономика и статистика фирм : учебник / В. Е. Адамов [и др.] ; под ред. д-ра экон. наук, проф. С. Д. Ильенковой. — 3-е изд., перераб. и доп. — М. : Финансы и статистика, 2007. — 288 с.: ил.

[15] Экономическая статистика : учебник / под ред. Ю. Н. Иванова. — 2-е изд., доп. — М. : ИНФРА-М, 2009. — 480 с.

[16] Гусаров В. М. Статистика : учеб. пособие для вузов / В. М. Гусаров. — М. : ЮНИТИ-Дана, 2012. — 463 с.

[17] Орлов А. И. Прикладная статистика : учебник / А. И. Орлов. — М. : Экзамен, 2011. — 671 с.

[18] Практикум по статистике : учеб. пособие для вузов / под ред. В. М. Симчеры. — М. : Финстатинформ, 2013. — 259 с.

Приложение А

ЗНАЧЕНИЯ ИНТЕГРАЛА ЛАПЛАСА

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

Таблица А.1 – Значения интеграла Лапласа

t	Сотые доли									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,00	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,10	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,20	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,30	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2961	0,3035
0,40	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,50	0,3829	0,3899	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,60	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,70	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,80	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,90	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6680	0,6729	0,6778
1,00	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,10	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,20	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7995	0,8029
1,30	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,40	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,50	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,60	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,70	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,80	0,9281	0,9297	0,9312	0,9328	0,9342	0,9357	0,9371	0,9385	0,9399	0,9412
1,90	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
2,00	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9615	0,9625	0,9634

продолжение на следующей странице

Таблица А.1 — Продолжение

t	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
2,10	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,20	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,30	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,40	0,9836	0,9840	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,50	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,60	0,9907	0,9909	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9929
2,70	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,80	0,9949	0,9950	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,90	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,00	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,10	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,20	0,9986	0,9987	0,9987	0,9988	0,9988	0,9988	0,9989	0,9989	0,9990	0,9990
3,30	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,40	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,50	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997

Приложение Б

ЗНАЧЕНИЯ *t*-КРИТЕРИЯ СТЬЮДЕНТА

Таблица Б.1 – Значения *t*-критерия Стьюдента при уровне значимости α: 0,10, 0,05, 0,01

Число степеней свободы ν	α			Число степеней свободы ν	α		
	0,1	0,05	0,01		0,1	0,05	0,01
1	6,314	12,706	63,66	18	1,734	2,101	2,878
2	2,92	4,3027	9,925	**19**	1,729	2,093	2,861
3	2,353	3,1825	5,841	**20**	1,725	2,086	2,845
4	2,132	2,7764	4,604	**21**	1,721	2,08	2,831
5	2,015	2,5706	4,032	**22**	1,717	2,074	2,819
6	1,943	2,4469	3,707	**23**	1,714	2,069	2,807
7	1,895	2,3646	3,5	**24**	1,711	2,064	2,797
8	1,86	2,306	3,355	**25**	1,708	2,06	2,787
9	1,833	2,2622	3,25	**26**	1,706	2,056	2,779
10	1,813	2,2281	3,169	**27**	1,703	2,052	2,771
11	1,796	2,201	3,106	**28**	1,701	2,048	2,763
12	1,782	2,1788	3,055	**29**	1,699	2,045	2,756
13	1,771	2,1604	3,012	**30**	1,697	2,042	2,75
14	1,761	2,1448	2,977	**40**	1,684	2,021	2,705
15	1,753	2,1315	2,947	**60**	1,671	2	2,66
16	1,746	2,1199,	2,921	**120**	1,658	1,98	2,617
17	1,74	2,1098	2,898	∞	1,645	1,96	2,576

Приложение В

ЗНАЧЕНИЯ F-КРИТЕРИЯ ФИШЕРА

Таблица В.1 – Значения F-критерия Фишера при уровне значимости α = 0,05

ν_1	ν_2									
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**∞**
1	161,5	200	215,7	224,6	230,2	234	238,9	243,9	249	254,3
2	18,5	19	19,16	19,25	19,3	19,33	19,37	19,41	19,45	19,5
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,9	2,71
10	4,96	4,1	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,2	3,09	2,95	2,79	2,61	2,4
12	4,75	3,88	3,49	3,26	3,11	3	2,85	2,69	2,5	2,3
13	4,67	3,8	3,41	3,18	3,02	2,92	2,77	2,6	2,42	2,21
14	4,6	3,74	3,34	3,11	2,96	2,85	2,7	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,9	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,2	2,96	2,81	2,7	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,9	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,1	2,87	2,71	2,6	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,3	3,44	3,05	2,82	2,66	2,55	2,4	2,23	2,03	1,78

продолжение на следующей странице

Таблица В.1 — Продолжение

ν_1	ν_2									
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	∞
23	4,28	3,42	3,03	2,8	2,64	2,53	2,38	2,2	2	1,76
24	4,26	3,4	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,6	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,3	2,13	1,93	1,67
28	4,2	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,7	2,54	2,43	2,28	2,1	1,9	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2	1,79	1,52
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,4	2,29	2,13	1,95	1,72	1,44
60	4	3,15	2,76	2,52	2,37	2,25	2,1	1,92	1,7	1,39
70	3,98	3,13	2,74	2,5	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,1	2,71	2,47	2,32	2,2	2,04	1,86	1,64	1,28
100	3,94	3,09	2,7	2,46	2,3	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,6	1,21
150	3,9	3,06	2,66	2,43	2,27	2,16	2	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,8	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79.	1,55	1,1
400	3,86	3,02	2,63	2,4	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3	2,61	2,38	2,22	2,1	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,6	2,37	2,21	2,09	1,94	1,75	1,52	

ГЛОССАРИЙ

Абсолютная величина — величина, отражающая уровень развития какого-либо явления или процесса. Абсолютные величины всегда имеют свою единицу измерения (размерность), присущую изучаемому явлению.

Абсолютное значение одного процента прироста представляет собой отношение абсолютного прироста к темпу прироста.

Абсолютное изменение (абсолютный прирост) уровней рассчитывается как разность между двумя уровнями ряда.

Аналитическая группировка выявляет взаимосвязи между изучаемыми явлениями и их признаками. Особенность аналитической группировки состоит в том, что ее единицы сгруппированы по факторному признаку и каждая группа охарактеризована средними величинами результативного признака.

Вариация — отличие друг от друга значений признаков, у различных единиц совокупности в один и тот же период или момент времени.

Графический образ — геометрические знаки, с помощью которых изображаются статистические показатели.

Группировка данных — это разделение элементов изучаемой совокупности на однородные группы по определенным признакам.

Единица наблюдения — это часть объекта наблюдения, которая является носителем признаков, подлежащих изучению.

Единица совокупности — первичный элемент статистической совокупности, являющийся носителем признаков и основой ведущегося при обследовании счета.

Закрытый интервал — это интервал, у которого имеются нижняя и верхняя границы.

Индекс — относительная величина, показывающая, во сколько раз уровень изучаемого явления в данных условиях отличается от уровня того же явления в других условиях.

Индекс выполнения плана. Для определения процента выполнения плана необходимо рассчитать отношение наблюдаемого значения признака к плановому (оптимальному, максимально возможному).

Индекс динамики характеризует изменение какого-либо явления во времени. Он представляет собой отношение значений одной и той же абсолютной величины в разные периоды времени.

Индекс интенсивности — это соотношение разных признаков одного объекта между собой.

Индекс координации — это отношение какой-либо части объекта к другой его части, принятой за основу (базу сравнения).

Индекс планового задания — это отношение планового значения признака к базисному.

Индекс сравнения — это сравнение (соотношение) разных объектов по одинаковым признакам.

Индекс структуры (доля) — это отношение какой-либо части объекта (совокупности) ко всему объекту.

Комбинированная группировка — группировка, в которой расчленение совокупности на группы производится по нескольким признакам.

Корреляционная связь — это связь, проявляющаяся при большом числе наблюдений в виде определенной зависимости между средним значением результативного признака и признаками-факторами.

Коэффициент эластичности показывает, на сколько процентов изменяется в среднем результативный признак y при изменении факторного признака x на 1%.

Макет таблицы — заполненный заголовками скелет таблицы.

Медиана — величина варьирующего признака, делящая совокупность на две равные части — со значением признака меньше медианы и со значением признака больше медианы.

Мода определяется как значение признака с наибольшей частотой.

Объектом наблюдения называется совокупность, о которой должны быть собраны сведения.

Открытый интервал — это тот интервалы, у которого указана только одна граница: верхняя — для начального; нижняя — для последнего.

Относительная величина — это результат деления (сравнения) двух абсолютных величин. В числителе дроби стоит величина, которую сравнивают, а в знаменателе — величина, с которой сравнивают (база сравнения).

Отчетность — это форма статистического наблюдения, при которой предприятия или иные хозяйствующие субъекты в установленном виде и в определенные сроки предоставляют в статистические органы необходимую информацию.

Подлежащее таблицы — объект, характеристика которого представлена в таблице. Обычно, подлежащее таблицы дается в левой части, в наименовании строк.

Поле графика — часть плоскости, где расположены графические образы.

Признак единицы совокупности — свойства единицы совокупности, которые различаются способами их измерения и другими особенностями, что дает основание для их классификации.

Простая группировка — группировка, в которой группы выделены по одному признаку.

Размах вариации — абсолютная разность между максимальным и минимальным значениями признака из имеющихся в изучаемой совокупности значений.

Ряд динамики — это числовые значения определенного статистического показателя в последовательные моменты или периоды времени (т. е. расположенные в хронологическом порядке).

Система статистических показателей — совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями.

Сказуемое таблицы — система показателей, которыми характеризуется объект изучения, т. е. подлежащее таблицы, сказуемое, в большинстве случаев, формирует верхние заголовки и составляет содержание граф.

Средний темп роста показывает, во сколько раз в среднем изменялось значение ряда динамики в течение рассматриваемого периода.

Средняя величина — такое среднее значение признака, при вычислении которого общий объем признака в совокупности сохраняется неизменным.

Статистическая сводка — это научное обобщение первичного статистического материала, полученного в ходе наблюдения с помощью подсчетов, выполняемых по определенной схеме.

Статистическая совокупность — множество социально-экономических объектов или явлений общественной жизни, объединенных качественной основой, но отличающихся друг от друга отдельными признаками. Таковы, например, совокупность домохозяйств, семей, предприятий и т. п.

Статистическая таблица — таблица, содержащая сводную характеристику исследуемой совокупности по одному или нескольким существенным признакам, взаимосвязанным логикой статистического анализа.

Статистический график — это чертеж, на котором статистические совокупности, характеризуемые определенными показателями, описываются с помощью условных геометрических образов или знаков.

Статистический показатель — это понятие, отображающее количественные характеристики (размеры) или соотношения признаков общественных явлений.

Статистическое наблюдение — это научно-организованный сбор данных об изучаемой совокупности путем регистрации заранее намеченных признаков с целью получения обобщающих характеристик.

Структурная группировка — предназначена для изучения состава однородной совокупности по определенному варьирующему показателю.

Темп прироста — относительный показатель, показывающий, на сколько процентов данный уровень больше (или меньше) другого, принимаемого за базу сравнения.

Темп роста (индекс динамики) — относительное изменение уровней ряда, которое рассчитывается как отношение (деление) двух уровней.

Типологическая группировка — это расчленение совокупности на однородные группы по типам экономических явлений.

Тренд — основная тенденция (закономерность) в изменении уровней ряда.

Уравнение регрессии представляет собой математическое описание изменения взаимно коррелируемых величин по эмпирическим (фактическим) данным.

Факторный признак — это признак-причина, под воздействием которого изменяется *результативный* признак — признак-следствие.

Ценз — это определенная количественная характеристика, которая служит для ограничения объекта наблюдения.

Шапка таблицы — совокупность подлежащего и сказуемого таблицы.

Учебное издание

Подопригора Игнат Валерьевич

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебное пособие

Корректор Осипова Е. А.
Компьютерная верстка Мурзагулова Н. Е.

Издано в Томском государственном университете
систем управления и радиоэлектроники.
634050, г. Томск, пр. Ленина, 40
Тел. (3822) 533018.