

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РЕСПУБЛИКИ КАЗАХСТАН

Рудненский индустриальный институт

Смирнова С.В.

МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ

Учебное пособие

для студентов специальности  
«Информационные системы»



Рудный 2013

**УДК 004(075.8)**  
**ББК 32.965Я73**  
**С50**

**Рецензенты:**

Клименко И.С. – д.т.н., профессор КИнЭУ им. М.Дулатова  
Таджигитов А.А. – к.ф. – м.н., доцент, СКГУ им. М.Козыбаева  
Зарубин М.Ю. – к. т. н., доцент, Рудненский индустриальный институт

Рекомендовано к изданию Ученым Советом РГКП «РИИ», протокол № 10 от 31.05.2013 г.

Смирнова С.В.

**С 50** Методы обработки информации: учебное пособие. – Рудный, 2013. – 80 с.

**ISBN 9965-845-70-0**

В учебном пособии изложены теоретические вопросы некоторых разделов дисциплины «Методы обработки информации»: понятие статистики как науки, группировка, вариационные ряды и их статистические показатели. Для самопроверки знаний в каждой теме представлены контрольные вопросы. Приведены решения задач, тесты для контроля знаний.

Учебное пособие предназначено для студентов специальности «Информационные системы» и молодых преподавателей вузов. Пособие будет активным помощником студентам и позволит им рационально организовать самостоятельную работу по изучению данной дисциплины, активизировать участие в обсуждении вопросов, рассматриваемых на практических занятиях и СРСП.

УДК 004(075.8)  
ББК 32.965Я73

© Рудненский индустриальный институт

## СОДЕРЖАНИЕ

Введение	4
1 Статистика как наука	6
1.1 Теоретические основы статистики как науки	6
1.2 Основные понятия и категории статистической науки в целом	13
1.3 Метод статистики	20
2 Группировка	26
2.1 Задачи сводки и её основное содержание	26
2.2 Задачи группировок и их значение в статистическом исследовании	28
2.3 Виды статистических группировок	32
2.4 Ряды распределения и группировки	35
2.5 Многомерные группировки	37
2.6 Метод группировки и его место в системе статистических методов	39
2.7 Основные проблемы, возникающие при построении группировок	41
3 Вариационные ряды и их характеристики	43
3.1 Вариационные ряды и их графическое изображение	43
3.2 Средние величины	48
3.3 Показатели вариации	51
3.4 Упрощенный способ расчета средней арифметической и дисперсии	54
3.5 Начальные и центральные моменты вариационного ряда	55
Упражнения	63
Тестовые задания для самоконтроля	65
Глоссарий	73
Ответы к тестовым заданиям	78
Список литературы	79

## Введение

Практическая деятельность современных предприятий дает большой поток информации о производственных процессах. Эта информация должна быть своевременно систематизирована, обобщена и обработана с целью получения новых научных и практических выводов. Предварительная обработка информации наилучшим образом может быть осуществлена, в том числе и методами математической статистики.

Дисциплина «Методы обработки информации» дает основные представления о математических методах обработки экспериментальных данных в компьютерных системах, методах построения математических моделей, методах решений параметрической идентификации моделей.

Статистическими данными называют сведения о числе объектов, обладающих теми или иными признаками. Например, данные о прочностных признаках большого числа образцов одной и той же горной породы, данные о производительности труда, о себестоимости добычи одной тонны полезного ископаемого, о содержании металла в отобранных пробах и т.д.

Любое статистическое исследование выражается в следующих основных ступенях:

- массовое статистическое наблюдение. Сбор информации;
- сводка и группировка статистических данных. Формирование совокупности;
- построение статистических показателей и их анализ.

Целью учебного пособия по дисциплине «Методы обработки информации» является помощь студентам в изучении некоторых разделов основных методов обработки экспериментальных данных. Студент должен уметь находить дополнительные пути к повышению культуры производства, эффективности использования оборудования и росту производительности труда, снижению себестоимости и увеличению рентабельности.

Вся продуктивная деятельность человека, так или иначе, связана с обработкой информации. Процесс развития общества неотделим от становления все более полных и эффективных методов обработки информации. Каждая область науки и в большой степени различные отрасли деятельности (образование, экономика, экология, добывающие отрасли, транспорт, связь, медицинская диагностика, управление и т.д.) представляют собой совокупность идей и методов, предназначенных для целенаправленной и эффективной обработки той информации, за которую ответственна данная область.

Конкретизация понятий «целенаправленной и эффективной обработки» произошла под влиянием развития методов и технических устройств, предназначенных для обработки информации и способных выполнять ее быстро. Идеи, принципы и алгоритмы, которые в настоящее время составляют методологию обработки информации, уже сегодня позволили сделать существенный прорыв в технологии обработки информации (наглядный пример – Internet).

Методы обработки и принципы их реализации для каждой области имеют свои специфические особенности, которые, прежде всего, обуславливаются конкретным видом носителя информации, методами кодирования и способами представления результатов обработки. Вследствие этого устройства обработки информации для различных областей часто оказываются внешне непохожими друг на друга. Но за этой внешней непохожестью скрывается одинаковая методология и принципы построения систем обработки, что является для нас определяющим и составляет предмет изучения в данном курсе.

Основу методов обработки информации составляют вычислительная математика, теория информации и математическая статистика. Современные методы математической статистики и теории информации используют сложный математический аппарат, но базируются, тем не менее, на простых исходных положениях, вытекающих из практических задач. Одна из задач курса состоит в том, чтобы за сложностью математических методов не потерять тот смысл, который заложен в современных принципах обработки информации, и выделить те решения, которые используются на практике сегодня.

Под информацией понимают некоторую совокупность сведений, представляющих интерес для человека или полезных для функционирования интеллектуальных систем.

И практически всегда при получении информации (человеком или системой) возникает потребность ее обработки.

# 1 Статистика как наука

## 1.1 Теоретические основы статистики как науки

Первоначально под статистикой понимали совокупность знаний о государстве, затем статистикой стали называть сбор и обработку данных о массовых явлениях, и наконец, особую отрасль науки. Статистика как совокупность сведений, статистика как параметр совокупности, статистика как род деятельности, наконец, статистика как наука о массовых явлениях — вот вехи зарождения, становления и развития статистики.



Сам термин «статистика» происходит от латинского слова *status* (статус), которое означает «положение, состояние вещей». От корня этого слова образовались слова *stato* — государство и *statista* — знаток государства. От того же корня образовалось и существительное *statistica*.

Начало статистической практики относится примерно ко времени возникновения государства. Первой опубликованной статистической информацией можно считать глиняные таблички Шумерского царства (III — II тысячелетия до н. э.). Для учёта имущества в Шумере и Эламе использовали систему глиняных шариков. Изначально каждый шарик обозначал один объект (корову, барана и т. п.) Затем размер и форма комочка стали значащими. А вскоре на них стали появляться пометки (след пальца, засечка), изменяющие их «вес». Эти шарики складывались в глиняный контейнер, который запечатывался цилиндрической печатью, идентифицирующей владельца. Таким образом, например, если контейнер содержал информацию о количестве голов скота в стаде, его необходимо было разбить, чтобы провести подсчёт находящихся в нём шариков. К 3300 году до н. э. на поверхности контейнера вместе с печатью владельца стало появляться краткое описание его содержимого. При этом необходимость разбивать контейнер при каждой проверке отпала. Постепенно ставшие бесполезными шарики исчезли, контейнер из сферического стал плоским. Так появились первые глиняные таблички с первыми числами: кружочками и уголками, выдавленными в глине, форма и размер которых указывали на обозначаемый объект и его количество.



Вначале под статистикой понимали описание экономического и политического состояния государства или его части. Например, к 1792 году относится определение: «статистика описывает состояние государства в настоящее время или в некоторый известный момент в прошлом». И в настоящее время деятельность государственных статистических служб вполне укладывается в это определение.

Однако постепенно термин «статистика» стал использоваться более широко. По Наполеону Бонапарту, «статистика — это бюджет вещей, а без бюджета нет благополучия».



Тем самым статистические методы были признаны полезными не только для административного управления, но и для применения на уровне отдельного предприятия. Согласно формулировке 1833 года, «цель статистики заключается в представлении фактов в наиболее сжатой форме». Во 2-й половине XIX — начале XX веков сформировалась научная дисциплина — математическая статистика, являющаяся частью математики.

В XX веке статистику часто рассматривают, прежде всего, как самостоятельную научную дисциплину. Статистика есть совокупность методов и принципов, согласно которым проводится сбор, анализ, сравнение, представление и интерпретация числовых данных. В 1954 г. академик АН УССР Б. В. Гнеденко дал следующее определение: «Статистика состоит из трёх разделов:

- сбор статистических сведений, то есть сведений, характеризующих отдельные единицы каких-либо массовых совокупностей;
- статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе данных массового наблюдения;
- разработка приёмов статистического наблюдения и анализа статистических данных».

Как отрасль общественной науки и практической деятельности по получению, обработке, анализу и публикации информации она исследует количественное выражение закономерностей жизни общества в конкретных условиях места и времени. Сюда относятся, например, закономерности роста или снижения



экономических, демографических,

правовых и других данных, характеризующих жизнь общества за определенный период времени.

Эти закономерности статистика выражает с помощью статистических показателей, из чего следует, что статистика — вместе с тем и учение о системе показателей, т.е. количественных характеристик, определяющих состояние (уровень) - тех или иных явлений.

Статистикой называют также различного рода числовые, или цифровые, данные, характеризующие различные стороны жизни государства и деятельности предприятий: производство, экономику, население, т.е. с точки зрения сведений к ней относят только то, что получает количественное выражение.



Например, беспристрастная статистика свидетельствует о демографическом взрыве во многих частях света. Если в начале 19 века население достигало 1 миллиарда, в начале 20 века на Земле уже жили 1,65 миллиард человек, то сейчас — около 7 миллиардов. Статистика позволяет судить, насколько населенным будет наш мир через 30 лет, например, и насколько обострятся проблемы охраны окружающей среды и конкуренции за источники природных ресурсов.

Та же статистика свидетельствует, что мужчины в Казахстане живут в среднем 64,4 года, а женщины — 74,59 года, а общая продолжительность жизни - 69,83 года, в России мужчины в среднем живут 64,7 года, женщины - 76,1 года, средняя продолжительность жизни составила 70,4 года.



Кроме того, статистика — это отрасль (форма) практической деятельности. Цель данной отрасли — результат указанного процесса «ведения». О людях, собирающих и обрабатывающих статистические сведения, данные или статистическую информацию, говорят, что они занимаются статистикой, ведут статистику (например, медицинскую, уровня жизни, преступности и др.). И тот, для кого она род деятельности, — это статистик, а не статист.

Наконец (и это самое важное), статистика (в широком смысле) — наука, изучающая все массовые явления, к какой бы области они не относились, обладающие признаками совокупности.

В более специальном (узком) смысле статистика — наука, исследующая с количественной стороны массовые общественные явления, и в то же время — метод изучения каждой конкретной совокупности. Таковым она является для каждой общественной науки, поскольку в результате исследования

обнаруживает присущие их природе последовательности, повторяемости, тенденции, закономерности, направления развития и измеряет их действие.

Эта тема имеет большое значение не только для курса теории обработки информации, но и для всех статистических дисциплин вообще. В ней излагаются важнейшие вопросы статистической науки: о предмете статистической науки, ее методе, теоретических основах, задачах и др. В результате изучения темы студент должен получить ясное представление о том, что изучает статистика, ее место в системе наук, теоретические основы, важнейшие принципы, категории и понятия, основные задачи статистики на современном этапе.

Изучение темы должно вооружить студента пониманием основ теории статистики и статистической методологии.

При рассмотрении материала темы важно уяснить необходимость привлечения массовых данных для объективного познания действительности;



ведущую роль социально-экономических категорий в статистическом исследовании.

Необходимо хорошо усвоить такие важнейшие понятия статистической науки, как статистическая совокупность, единица совокупности, признаки и их классификация, вариация признаков, статистический показатель. Без них невозможно обойтись в дальнейшем изучении других статистических дисциплин, в которых применяются

понятия, термины, показатели, формулы теории статистики, но не разъясняются их сущность, смысл и значение, поскольку это составляет задачу теории статистики.

Статистика, вернее ее методы исследования, широко применяется в различных областях человеческих знаний. Однако, как любая наука, она требует определения предмета ее исследования. В связи с этим различают статистику, занимающуюся изучением социально-экономических явлений, которая относится к циклу общественных наук, и статистику, занимающуюся закономерностями явлений природы, которая относится к наукам естественным.

Авторы большинства современных отечественных вузовских учебников по теории статистики (общей теории статистики) под статистикой понимают предметную общественную науку, т.е. науку, имеющую свои особые предмет и метод познания. Поэтому в процессе изучения темы необходимо уяснить, что статистика – общественная наука, которая изучает количественную сторону качественно определенных массовых социально-экономических явлений и процессов, их структуру и распределение, размещение в пространстве,

движение во времени, выявляя действующие количественные зависимости, тенденции и закономерности, причем в конкретных условиях места и времени.

При изучении темы необходимо в первую очередь обратить особое внимание на определение предмета, метода и задач статистики, на уяснение сущности и содержания статистической науки, отличающих ее от других социально-экономических наук, а также от математики. Вопросы предмета и метода статистики являются исходными, с них начинается первое знакомство с основами статистической науки.

В процессе изучения данной темы важно уяснить, что статистика как наука исследует не отдельные факты, а массовые социально-экономические явления и процессы, выступающие как множество отдельных факторов, обладающих как индивидуальными, так и общими признаками.

Объект статистического исследования (в каждом конкретном случае) в статистике называют статистической совокупностью. Статистическая совокупность – это множество единиц, обладающих массовостью, однородностью, определенной целостностью, взаимозависимостью состояния отдельных единиц и наличием вариации. Например, в качестве особых объектов статистического исследования, т.е. статистических совокупностей, может выступать множество коммерческих банков, зарегистрированных на территории РК, множество акционерных обществ, множество граждан какой-либо страны и т.д. Важно помнить, что статистическая совокупность состоит из реально существующих материальных объектов.

Каждый отдельно взятый элемент данного множества называется единицей статистической совокупности. Единицы статистической совокупности характеризуются общими свойствами, именуемыми в статистике признаками, т.е. под качественной однородностью совокупности понимается сходство единиц (объектов, явлений, процессов) по каким-либо существенным признакам, но различающихся по каким-либо другим признакам. Например, из названных совокупностей множество коммерческих банков наряду с качественной определенностью (принадлежностью к разряду кредитных учреждений) обладает различиями по размеру объявленных уставных фондов, численности работающих, сумме активов и т.д.

Качественная определенность совокупности, хотя и имеет объективную основу, устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями и познавательными задачами.

Единицы совокупности наряду с общими для всех единиц признаками, обуславливающими качественную определенность совокупности, также обладают индивидуальными особенностями и различиями, отличающими их



друг от друга, т.е. существует вариация признаков. Она обусловлена различным сочетанием условий, которые определяют развитие элементов множества. Например, уровень производительности труда работников банка определяется его возрастом, квалификацией, отношением к труду и т.д. Именно наличие вариации предопределяет необходимость статистики. Необходимо помнить, что вариация признака может отражаться статистическим распределением единиц совокупности.

Важно также помнить, что статистика как наука изучает, прежде всего, количественную сторону общественных явлений и процессов в конкретных условиях места и времени, т.е. предметом статистики выступают размеры и количественные соотношения социально-экономических явлений, закономерности их связи и развития.

Количественную характеристику статистика выражает через определенные числа, которые называют статистическими показателями. Статистический показатель отражает результат измерения у единиц совокупности и совокупности в целом.

Однако тогда возникает вопрос, чем же статистика отличается от математики?

Основное отличие состоит в том, что статистика изучает количественную сторону качественно определенных массовых общественных явлений в данных условиях места и времени. При этом качественную определенность единичных явлений обычно определяют сопряженные науки.

При изучении данной темы важно выяснить теоретические основы статистики и проблему применения закона больших чисел.

Теоретическую основу любой науки, в том числе и статистики, составляют понятия и категории, в совокупности которых выражаются основные принципы данной науки. В статистике к важнейшим категориям и понятиям относятся: совокупность, вариация, признак, закономерность.

Важнейшим свойством статистической совокупности является ее неразложимость. Это означает, что дальнейшее дробление индивидуальных явлений не вызывает потери их качественной основы. Исчезновение или ликвидация одного или ряда явлений не разрушает качественной основы статистической совокупности в целом. Так, население страны или города останется населением, несмотря на постоянно происходящие процессы механического и естественного движения населения.

Количественные изменения значения признака при переходе от одной единицы совокупности к другой называются вариацией. Вариация возникает под воздействием случайных, прежде всего внешних причин.

Статистические совокупности имеют определенные свойства, носителями которых выступают единицы (отдельные элементы) совокупности (явления), обладающие определенными признаками. По форме внешнего выражения признаки делятся на:

- атрибутивные (описательные)
- количественные.

Атрибутивные (качественные) признаки не поддаются прямому количественному (числовому) выражению.

Количественные признаки делятся на дискретные (прерывные) и непрерывные.

Отличие количественных признаков от качественных состоит в том, что первые можно выразить итоговыми значениями, например общий объем перевозок грузов предприятиями транспорта и т.д., вторые – только числом единиц совокупности, например число театров по видам деятельности.

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

Статистическая же закономерность в статистике рассматривается как количественная закономерность изменения в пространстве и времени массовых явлений и процессов общественной жизни, состоящих из множества элементов (единиц совокупности). Она свойственна не отдельным единицам

совокупности, а всей их массе, или совокупности в целом.

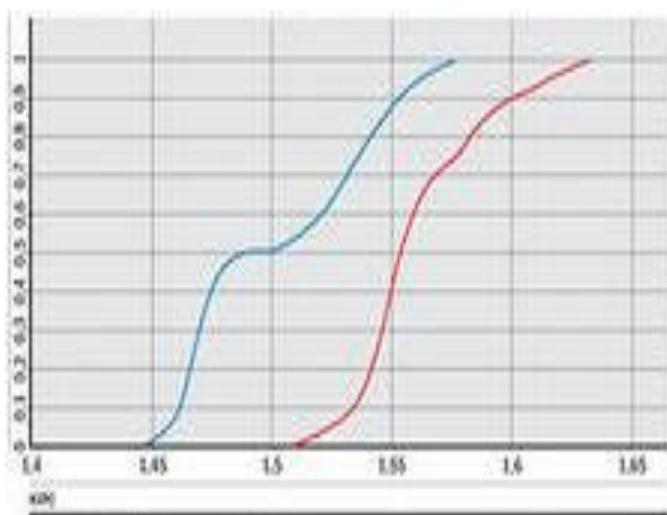
Статистическая закономерность – это форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности, если причины (условия), порождающие события, не изменяются или изменяются незначительно.

Статистическая закономерность устанавливается на основе анализа

массовых данных. Она возникает в результате действия объективных законов.

Так как статистическая закономерность обнаруживается в итоге массового статистического наблюдения, это обуславливает ее взаимосвязь с законом больших чисел.

Сущность закона больших чисел заключается в том, что в числах, суммирующих результат массовых наблюдений, выступают определенные правильности, которые не могут быть обнаружены на небольшом числе факторов. Закон больших чисел порожден свойствами массовых явлений. Важно помнить, что тенденции и закономерности, вскрытые с помощью закона больших чисел, имеют силу лишь как массовые тенденции, но не как законы для каждого отдельного, индивидуального случая.



## 1.2 Основные понятия и категории статистической науки в целом

К основным понятиям и категориям статистической науки относятся следующие: совокупность, признак, показатель, система показателей и др.

Статистическая совокупность - множество элементов одного и того же вида сходных между собой по одним признакам и различающимся по другим. Например: это совокупность отраслей экономики, совокупность вузов и т.п.

Отдельные элементы статистической совокупности называют ее единицами. В рассмотренных выше примерах единицами совокупности являются соответственно отрасли, вуз (один).

Единицы совокупности обладают, как правило, многими признаками.

Признак - свойство единиц совокупности, выражающее их сущность и имеющее способность варьировать, т.е. изменяться. Признаки, принимающие единичное значение у отдельных единиц совокупности называются варьирующими, а сами значения вариантами.

Варьирующие признаки подразделяются на атрибутивные или качественные. Признак называется атрибутивным или качественным, если его отдельные значения (варианты) выражаются в виде состояния или свойств, присущих явлению. Варианты атрибутивных признаков выражаются в словесной форме. Примерами таких признаков могут служить - хозяйственный.

Признак называется количественным, если его отдельные значения выражаются в виде чисел. Например: заработная плата, стипендия, возраст.

По характеру варьирования количественные признаки делятся на дискретные и непрерывные.

Дискретные - такие количественные признаки, которые могут принимать только вполне определенное, как правило, целое значение.

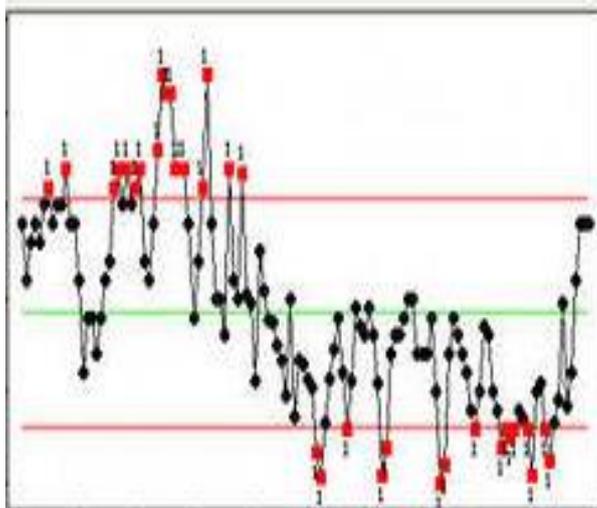
Непрерывными - являются такие признаки, которые в определенных пределах могут принимать значение, как целое, так и дробное.

Различаются также признаки основные и второстепенные.

Основные признаки характеризуют главное содержание и сущность изучаемого явления или процесса.

Второстепенные признаки дают дополнительную информацию и непосредственно связаны с внутренним содержанием явления.

В зависимости от целей конкретного исследования одни и те же признаки в одних и тех же случаях могут быть основными, а в других второстепенными.



Статистический показатель - это категория отображающая размеры и количественные соотношения признаков социально-экономических явлений и их качественной определенности в конкретных условиях места и времени.

Следует различать содержание статистического показателя и его конкретное числовое выражение. Содержание, т.е. качественная определенность, состоит в том, что показатели всегда характеризуют социально-экономические категории (население, экономика, финансовые институты и т.д.). Количественные размеры статистических показателей, т.е. их числовые значения зависят, прежде всего, от времени и места объекта, который подвергается статистическому исследованию.

Социально-экономические явления, как правило, не могут быть охарактеризованы каким-либо одним показателем, Например: уровнем жизни населения. Для комплексной всесторонней характеристики исследуемых явлений необходима научно обоснованная система статистических показателей. Такая система не является постоянной. Она постоянно совершенствуется исходя из потребностей общественного развития.

В науку термин статистика ввел немецкий ученый Готфрид Ахенваль в 1746 году, предложив заменить название курса «Государствоведение», преподававшегося в университетах Германии, на «Статистику», положив тем самым начало развитию статистики как науки и учебной дисциплины.



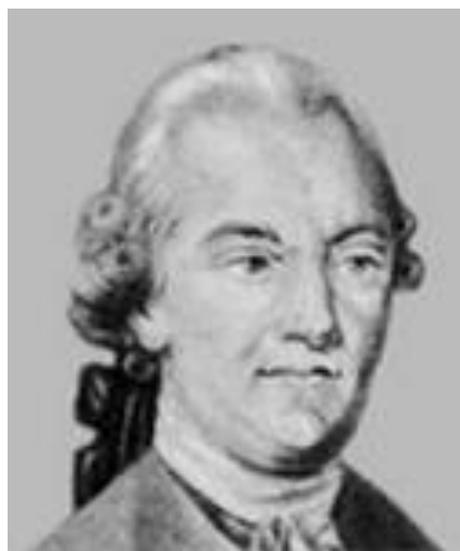
Готфрид Ахенваль (1719-1772)

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

Представители описательной школы (Герман Конринг, Готфрид Ахенваль, Август Людвиг Шлецер) своей задачей считали описание достопримечательностей государства: территории, населения, климата, политического устройства, вероисповедания, торговли и т.п. – без анализа закономерностей и связей между явлениями.



Герман Конринг (1606-1681)



Август Людвиг Шлецер (1735-1809)

Представители школы политических арифметиков (Уильям Петти, Джон Граунт, Эдмунд 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 государственоведения и политической арифметики. 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 Фрэнсис Гальтон, Фрэнсис Эджворт, Карл Пирсон, Одни Дж. Юл, Вильям Госсет, Рональд Фишер, Морис Дж.Кендэл, итальянец Коррадо Джини, русские – Пафнутий Львович Чебышёв, Андрей Андреевич Марков, Александр Михайлович Ляпунов, Александр Иванович и Александр Александрович Чупров и пр.

В настоящее время термин статистика употребляется в 4 значениях:

- наука, изучающая количественную сторону массовых явлений и процессов в неразрывной связи с их качественным содержанием – учебный предмет в высших и средних специальных учебных заведений;
- совокупность цифровых сведений, характеризующих состояние массовых явлений и процессов общественной жизни; статистические данные, представляемые в отчетности предприятий, организаций, отраслей экономики,

а также публикуемых в сборниках, справочниках, периодической печати и в сети Интернет, которые являются результатом статистической работы;

- отрасль практической деятельности («статистический учет») по сбору, обработке, анализу и публикации массовых цифровых данных о самых различных явлениях и процессах общественной жизни;

- 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.

Основные категории, используемые в статистике.

Статистическая совокупность – множество социально-экономических объектов или явлений общественной жизни, объединенных качественной основой, но отличающихся друг от друга отдельными признаками, т.е. однородных в одном отношении, но разнородных в другом. Таковы, например, совокупность домохозяйств, семей, предприятий, фирм и т.п.

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

Признак единицы совокупности – свойства единицы совокупности, которые различаются способами их измерения и другими особенностями.

Статистический показатель – понятие, отображающее количественные характеристики (размеры) или соотношения признаков общественных явлений. Статистические показатели можно подразделить на:

- первичные (объемные), которые характеризуют либо общее число единиц совокупности (объем совокупности), либо сумму значений какого-либо признака (объем признака) и выражаются абсолютными величинами;

- вторичные (расчетные), которые задаются на единицу первичного показателя и выражаются относительными и средними величинами. Статистические показатели могут быть плановыми, отчетными и прогнозными.

Система статистических показателей – совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями. Она охватывает все стороны общественной жизни как на макро-, так и на микроуровне. С изменением условий жизни общества меняются и системы статистических показателей, совершенствуется методология их расчета.

Совокупность приемов, пользуясь которыми статистика исследует свой предмет, составляет метод статистики.

Выделяют три группы статистических методов (три этапа статистического исследования):

- статистическое наблюдение - научно организованный сбор сведений, заключающийся в регистрации тех или иных фактов, признаков, относящихся к каждой единице изучаемой совокупности;

- сводка и группировка - обработка собранных первичных данных, включающая их группировку, обобщение и оформление в таблицах;
- статистический анализ - на основе итоговых данных сводки рассчитываются различные обобщающие показатели в виде средних и относительных величин, выявляются определенные закономерности в распределениях, динамике показателей и т.п.

Таким образом, любое законченное статистическое исследование проходит в три этапа, между которыми, могут быть перерывы во времени.

Для того чтобы изучить массовые явления и процессы общественной жизни, следует прежде всего собрать о них необходимые сведения - статистические данные.

Статистические данные - это составная часть глобальной информационной системы, которая формируется в соответствии с концепцией информатизации.

Государственная политика в сфере формирования информационных ресурсов и информатизации направлена на создание условий эффективного и качественного информационного обеспечения решения стратегических и оперативных задач социального и экономического развития страны.



Формирование информационной базы статистического исследования социальных явлений и процессов - сложный, многоступенчатый процесс.

В этом процессе выделяются следующие обязательные этапы, исследования: статистическое наблюдение, сводка и группировка собранного материала, обработка и анализ сводных статистических данных (информации). Последние два этапа составляют научную обработку статистических данных (информации).

Перечисленные этапы неразрывно связаны между собой, проводятся по предварительному плану и только в своей совокупности составляют законченный вид всякого статистического исследования.



Статистическое наблюдение - первый этап статистического исследования. Это планомерное, научно организованное и, как правило, систематическое получение данных (сбор сведений) о массовых явлениях и процессах социальной и экономической жизни путем регистрации существенных признаков каждой единицы их

совокупности.

Статистическое наблюдение это целенаправленный, научно организованный и методически контролируемый учет признаков и свойств массовых явлений, событий, фактов - основополагающий способ сбора данных во всех сферах общественной жизни, в том числе и в сфере реализации государственных мер социального контроля над правонарушениями.

Сформированные в процессе наблюдения данные - исходный статистический материал для решения последующих задач статистического исследования. От качества проведения статистического наблюдения зависят окончательные данные, которыми статистика характеризует изучаемые явления. Значимость этого этапа исследования определяется тем, что использование только объективной и достаточно полной информации, полученной в результате статистического наблюдения, на его последующих этапах в состоянии обеспечить научно обоснованные выводы о характере и закономерностях развития изучаемого объекта.

Полученные в результате наблюдения «статистические данные» представляют собой, образно говоря, лишь «сырье». Информацией они становятся только в процессе научной обработки, анализа.

Результаты исследования ценны лишь в том случае, если базируются на фактическом материале. Именно поэтому изучение общественных явлений должно опираться на прочный фундамент из точных и бесспорных фактов. А для того чтобы это был действительно фундамент, при исследовании общественных явлений необходимо брать не отдельные факты или фактики (как это иногда бывает в средствах массовой информации), а всю совокупность относящихся к рассматриваемому процессу фактов и свойств, их выражающих.

Общая характеристика статистического наблюдения может быть представлена следующими этапами.

Этап подготовки наблюдения включает разработку плана его проведения, и содержит ряд программно-методологических и организационных вопросов.

К программно-методологическим вопросам относятся: определение цели (задачи), объекта, единиц наблюдения, совокупности и измерения; состава признаков, подлежащих регистрации; разработка программы наблюдения и документов для сбора данных, а также выбор места и времени наблюдения.

К организационным - выбор вида и способа наблюдения, определение круга лиц и организаций, отвечающих за проведение наблюдения, подбор, обучение и инструктаж кадров, размножение и рассылка документов первичного учета и форм статистической отчетности, установление сроков представления материалов, а также другие практические вопросы, связанные с проведением статистического наблюдения.

Непосредственный сбор массовых данных - это заполнение статистических



формуляров - документов первичного учета (переписных листов, статистических карточек, анкет, бланков и т.д.) и статистической отчетности. Он начинается с их рассылки и заканчивается сдачей после заполнения в органы (подразделения), проводящие наблюдение.

Полученные данные на этапе их подготовки к автоматизированной обработке подвергаются проверке на качество путем арифметического и логического контроля, основанного на знании взаимосвязей между показателями и качественными признаками.

Точностью статистического наблюдения называют степень соответствия величины какого-либо показателя (значение какого-либо признака), определенной по материалам статистического наблюдения, действительной его величине.

Расхождение между расчетным и действительным значением изучаемых величин называют ошибкой наблюдения.

Точность данных - это основное требование, предъявляемое к статистическому наблюдению. Чтобы избежать ошибок наблюдения, предупредить, выявить и исправить их возникновение, необходимо:

- обеспечить качественное обучение персонала, который будет проводить исследование;
- организовать специальные, частичные или сплошные контрольные проверки правильности заполнения статистических формуляров;
- провести логический и арифметический контроль полученных данных после окончания сбора информации.

В зависимости от причин возникновения различают ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации - это отклонения между значением показателя, полученного в ходе статистического наблюдения, и фактическим, действительным его значением. Этот вид ошибок может быть и при сплошном, и при несплошном наблюдениях.

В отличие от ошибок регистрации ошибки репрезентативности характерны только для несплошного наблюдения. Они возникают потому, что отобранная и обследованная совокупность недостаточно точно воспроизводит (репрезентирует) всю исходную совокупность в целом.

Отклонение значения показателя обследованной совокупности от его величины по исходной совокупности называется ошибкой репрезентативности.

Ошибки репрезентативности также бывают случайные и систематические. Случайные ошибки возникают, если отобранная совокупность неполно воспроизводит всю совокупность в целом. Ее величина может быть оценена.

Систематические ошибки репрезентативности появляются вследствие нарушения принципов отбора единиц из исходной совокупности, которые должны быть подвергнуты наблюдению. Для выявления и устранения допущенных при регистрации ошибок может применяться счетный и

логический контроль собранного материала. Репрезентативности (так же, как и ошибки регистрации) могут быть случайными и систематическими.

Счетный контроль заключается в проверке точности арифметических расчетов, применявшихся при составлении отчетности или заполнении формуляров обследования.

Логический контроль заключается в проверке ответов на вопросы программы наблюдения путем их логического осмысления или путем сравнения полученных данных с другими источниками по этому же вопросу.

Примером логического сопоставления могут служить листы переписи населения. Так, например, в переписном листе двухлетний мальчик показан женатым, а девятилетний ребенок - грамотным. Ясно, что полученные ответы на вопросы неверны. Подобные записи требуют уточнения сведений и исправления допущенных ошибок. Примером сравнения могут быть сведения о заработной плате работников промышленного предприятия, которые имеются в отчете по труду и в отчете по себестоимости продукции. В торговле примером такого логического контроля может служить сопоставление сведений о фонде оплаты труда, содержащихся как в отчетности по труду, так и в отчете по издержкам обращения.

После получения статистических формуляров следует, прежде всего, провести проверку полноты собранных данных, т. е. определить, все ли отчетные единицы заполнили статистические формуляры, и значения всех ли показателей отражены в них. Следующим этапом контроля точности информации является арифметический контроль. Он основывается на использовании количественных связей между значениями различных показателей. Например, если среди собранных данных имеются сведения о численности промышленно-производственного персонала, выработке товарной продукции в среднем на одного работающего и стоимости товарной продукции, то произведение первых двух показателей должно дать значение третьего показателя. Если арифметический контроль покажет, что данная зависимость не выполняется, это будет свидетельствовать о недостоверности собранных данных. Поэтому в программу статистического наблюдения целесообразно включать показатели, которые дают возможность провести арифметический контроль.

Обычно для исправления ошибок, выявленных в ходе логического контроля, требуется повторно обратиться к источнику сведений.

### 1.3 Метод статистики

Типовые примеры раннего этапа применения статистических методов описаны в Библии, в Ветхом Завете. Там, в частности, приводится число воинов в различных племенах. С математической точки зрения дело сводилось к подсчёту числа попаданий значений наблюдаемых признаков в определённые градации.

Сразу после возникновения теории вероятностей (Паскаль, Ферма, XVII век) вероятностные модели стали использоваться при обработке статистических данных. Например, изучалась частота рождения мальчиков и девочек. Было установлено отличие вероятности рождения мальчика от 0,5, анализировались причины того, что в парижских приютах эта вероятность не та, что в самом Париже, и т. д.

В 1794 г. (по другим данным — в 1795 г.) немецкий математик Карл Гаусс формализовал один из методов современной математической статистики — метод наименьших квадратов. В XIX веке значительный вклад в развитие практической статистики внёс бельгиец Кетле, на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей, таких, как доля самоубийств среди всех смертей.

Первая треть XX века прошла под знаком параметрической статистики. Изучались методы, основанные на анализе данных из параметрических семейств распределений, описываемых кривыми семейства Пирсона. Наиболее популярным было нормальное распределение. Для проверки гипотез использовались критерии Пирсона, Стьюдента, Фишера. Были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

Разработанную в первой трети XX века теорию анализа данных называют параметрической статистикой, поскольку её основной объект изучения — это выборки из распределений, описываемых одним или небольшим числом параметров. Наиболее общим является семейство кривых Пирсона, задаваемых четырьмя параметрами. Как правило, нельзя указать каких-либо веских причин, по которым распределение результатов конкретных наблюдений должно входить в то или иное параметрическое семейство. Исключения хорошо известны: если вероятностная модель предусматривает суммирование независимых случайных величин, то сумму естественно описывать нормальным распределением; если же в модели рассматривается произведение таких величин, то итог, видимо, приближается логарифмически нормальным распределением и так далее. Главная особенность любой науки, дающая ей право на самостоятельное существование как особой отрасли знания, заключается в предмете познания, в принципах и методах его изучения которые в совокупности образуют её методологию.

Общая теория статистики разрабатывает наиболее общие понятия, категории статистической науки, которые имеют общестатистический смысл (например, «закономерность», «показатель», «признак», «статистическая совокупность» и т.д.) и методы изучения социально-экономических явлений.

Она выясняет природу этих категорий и разрабатывает научную методологию их построения (определения, вычисления), а также общие принципы, методы, стадии статистического исследования: наблюдения, сводки, группировки, анализа.

Категории, показатели и методы общей теории статистики являются методологической основой, ядром всех отраслей статистики. Именно поэтому

изучение любой отраслевой статистики начинают с уяснения общетеоретических положений статистики — принципов и методов статистического исследования.

Предмет исследования статистики как общественной науки – массовые явления социально – экономической и производственной жизни; она изучает количественную сторону этих явлений в неразрывной связи с их качественным содержанием в конкретных условиях места и времени.

Первая особенность статистики как науки — то, что она исследует для определенных условий времени и места не отдельные факты, а массовые явления и процессы как множества отдельных фактов, обладающих индивидуальными и общими признаками.

Статистическая совокупность — множество элементов, обладающих массовостью, некоторыми общими, но не обязательно системными свойствами, существенными характеристиками — однородностью, определенной целостностью, взаимозависимостью состояний отдельных элементов и наличием вариации признаков, их характеризующих.

Например, в качестве особых объектов статистического исследования, т.е. статистических совокупностей, могут быть: граждане какой-либо страны, региона; производственная деятельность предприятий и др.

При этом нельзя забывать, что статистическая совокупность — это реально существующие явления, факты, объекты.

Элементы, множество которых образует изучаемую статистикой совокупность, называют единицами статистической совокупности. Определить статистическую совокупность — значит определить входящие в нее элементы, т.е. единицы статистической совокупности. Единицы совокупности характеризуются разного рода общими свойствами, признаками, показателями, примет, знаков, по которым можно узнать, определить что-либо.

Признаки, варьирующиеся от единицы к единице, составляют специфическую черту совокупности, делающую ее предметом изучения статистики. Соответственно, их называют статистическими признаками. Метод статистики предполагает следующую последовательность действий:

- разработка статистической гипотезы,
- статистическое наблюдение,
- сводка и группировка статистических данных,
- анализ данных,
- интерпретация данных.

Прохождение каждой стадии связано с использованием специальных методов, объясняемых содержанием выполняемой работы.

Как в любой науке, теоретическую основу статистики составляют понятия и категории, в совокупности которых выражаются основные принципы статистики. Статистическая совокупность, статистическая закономерность, вариация, признак — вот важнейшие категории и понятия статистики.

Научное познание и практическая деятельность должны, прежде всего, установить качество, или сущность, изучаемых явлений, их природу,

внутреннюю определенность, благодаря которой они отличаются, отграничиваются друг от друга и которая делает их тем, что они есть.

Качество отражает устойчивое взаимоотношение составных элементов объекта, которое характеризует его специфику, дающую возможность отличать один объект от других. Именно благодаря качеству каждый объект существует и мыслится как нечто отграниченное от других объектов.

Вместе с тем следует отметить, что однородность статистической совокупности относительна и вовсе не означает полного соответствия всех единиц совокупности, а лишь подразумевает близость основного свойства, качества, типичности. Одна и та же совокупность единиц, к примеру, может быть однородна по одному признаку и неоднородна по- другому.

Статистическая закономерность — это форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности, если причины (условия), порождающие события, не изменяются или изменяются незначительно.

Статистические закономерности устанавливаются на основе обобщения и анализа данных о массовых явлениях в противоположность динамическим закономерностям, проявляющимся в отдельных явлениях, как это возможно в объектах естественных наук (биологии, механике, физике).

Методология (учение о методе) — это совокупность принципов и способов организации и построения теоретической и практической деятельности, а также учение об этой системе. Отсюда методология статистическая — совокупность взаимосвязанных специфических методов, способов и приемов исследования, применяемых в статистике.

С этих позиций статистику рассматривают обычно как науку о методах изучения массовых явлений и процессов. Однако отношение статистической науки к статистическому методу нельзя сводить к простому утверждению, что статистика — наука, которая пользуется для решения задач статистическим методом. Каково же их соотношение?

И наука, и метод исходят из рассмотрения массовых явлений. Эта внешняя связь приводит к расхождению между ними, поскольку в статистической науке массовое явление имеет одно значение, а в статистическом методе — другое. Для науки оно служит самоцелью, а для метода — лишь средством достижения цели.

Статистика непосредственно заинтересована в исследованных ею совокупностях, так как они представляют собой предмет, изучаемый статистикой. Статистический метод, напротив, пользуется совокупностями лишь для того, чтобы с помощью дальнейших логических операций, совершаемых на основе рассмотрения совокупностей, прийти к выводам о причинных связях. Тем не менее, хотя массовые явления ими рассматриваются с разных точек зрения между статистической наукой и статистическим методом тесные отношения, до известной степени оправдывающие сходство в их наименовании. Это сходство логически закономерно, так как, в конечном счете,

имеет общие задачи, решение которых — цель и статистической науки, и статистического метода.



Метод науки (исследования) в самом общем значении слова — способ, путь познания и преобразования действительности. Верная картина изучаемого объекта может быть получена лишь при правильном подходе к нему, лишь при правильном методе. Характеризуя роль правильного метода в научном познании, английский философ, основоположник эмпиризма Фрэнсис Бэкон сравнивал его со светильником, освещающим путнику дорогу в темноте. Он метко сказал: даже хромой, идущий по дороге, опережает того, кто бежит без дороги.

Фрэнсис Бэкон (1561—1626)

Исследовать явление методами статистики — значит исследовать его как явление массовое. То есть наблюдать множество его элементов или наблюдать само явление во множестве его повторений в пространстве или времени, характеризовать результаты наблюдений в их совокупности статистическими показателями анализировать их с учетом формы проявления закономерностей в массовых фактах, с учетом действующих в них общих законов. В статистической совокупности обнаруживаются сходство и различия, типичность явлений, она используется для вычисления обобщающих показателей.

Статистический метод направлен на познание массовых процессов объективной действительности. Он противопоставляется методу индукции (процесс рассуждения от частного к общему). Статистический метод можно успешно применять во всех областях научной работы, где речь идет о множественности причин и следствий.

Методы статистики принято делить на две основные группы: методы статистического наблюдения и методы обработки и анализа статистических данных (т.е. результатов наблюдения).

Процесс статистического исследования проходит три основные стадии (иногда их называют этапами).

Первая стадия — массовое статистическое наблюдение. Здесь в результате регистрации фактов по научной и тщательно разработанной программе получают объективные данные об изучаемых социально-экономических явлениях.

Вторая стадия — свodka и обработка, группировка (классификация) и систематизация материалов, собранных в результате массового статистического наблюдения.

Третья стадия — анализ показателей, полученных в результате сводки и обработки статистических материалов.

Все три стадии статистического исследования находятся между собой в тесной связи и требуют взаимосогласованности при их реализации. Отсутствие одной из них ведет к разрыву целостности статистического исследования.

Контрольные вопросы:

1. Назовите в качестве примера сферы общественной жизни, изучаемые статистикой.
2. Сформулируйте определение статистики как науки и дайте ему соответствующее обоснование.
3. Дайте характеристику основным чертам определения предмета статистики:
  - а) почему статистика является общественной наукой?
  - б) почему статистика изучает количественную сторону общественных явлений в связи с их качественным содержанием?
  - в) почему статистика изучает массовые явления?
  - г) почему каждое статистическое исследование должно опираться на изучение всех относящихся к данному вопросу фактов?
4. К каким видам (количественным или атрибутивным) относятся следующие признаки:
  - а) количество работников на фирме;
  - б) родственные связи членов семьи;
  - в) пол и возраст человека;
  - г) социальное положение вкладчика в Народном банке;
  - д) этажность жилых помещений;
  - е) количество детей в семье;
  - ж) розничный товароборот торговых объединений.
5. Укажите, какие совокупности можно выделить в высшем учебном заведении для статистического изучения?
6. Укажите, какие можно выделить статистические совокупности кредитных учреждений; сферы потребительского рынка; крестьянских хозяйств.
7. Какими количественными и атрибутивными признаками можно охарактеризовать совокупность студентов вуза?
8. Исследуется совокупность коммерческих банков Астаны. Какими количественными и качественными признаками можно ее охарактеризовать?
9. Назовите наиболее существенные варьирующие признаки, характеризующие студенческую группу.
10. Назовите основные факторные признаки, определяющие вариацию успеваемости студентов.
11. Какими показателями можно охарактеризовать совокупность жителей города?
12. Приведите перечень показателей, которыми можно было бы при статистическом обследовании полно охарактеризовать следующие явления:
  - а) население;

б) потребительский рынок;

в) промышленность;

г) транспорт и связь.

13. Какими признаками – прерывными или непрерывными – являются:

а) численность населения страны;

б) количество браков и разводов;

в) производство продукции легкой промышленности в стоимостном выражении;

г) капитальные вложения в стоимостном выражении;

д) процент выполнения плана реализованной продукции;

е) число посадочных мест в самолете;

ж) урожайность зерновых культур в центнерах с 1 га.

14. К каким видам (качественным или количественным) относятся следующие признаки:

а) тарифный разряд рабочего;

б) балл успеваемости;

в) форма собственности;

г) вид школы (начальная, неполная средняя и т.д.);

д) национальность;

е) состояние в браке.

15. Назовите, какие понятия, категории и методы излагаются в отрасли статистической науки – общей теории статистики.

## 2 ГРУППИРОВКА

### 2.1 Задачи сводки и её основное содержание

Собранный в процессе статистического наблюдения статистический материал нуждается в определённой обработке, сведений разрозненных данных воедино. Научно организованная обработка материалов наблюдения (по заранее разработанной программе), включая в себя кроме обязательного контроля собранных данных систематизацию, классификацию (группировку) материала, составление таблиц, получение итогов и производных показателей (средних, относительных величин), именуется в статистике сводкой и представляет собой второй этап статистического исследования.

В результате сводки достигается возможность по данным, относящимся к отдельным единицам наблюдения, характеризовать совокупность в целом. Так, например, на основе отчётов отдельных промышленных предприятий получают сведения о многих показателях отдельных отраслей и всей промышленности в целом: о выпуске продукции, числе занятых, производительности труда и т.п. Получение различного рода структурных характеристик (например, определение доли отдельных отраслей промышленности в общем, объеме промышленного производства) также возможно только на основе данных, полученных в результате сводки.

Таким образом, целью сводки является сведение воедино материалов статистического наблюдения и получение обобщающих статистических показателей, отражающих сущность социально-экономических и иных явлений и определённые статистические закономерности.

Различают следующие виды статистической сводки:

- простая статистическая сводка - это такая сводка, когда производится только подсчёт общих итогов статистического наблюдения;
- вторичная - это обработка и подсчёт данных, полученных в результате первичной сводки;
- децентрализованная - это способ организации сводки данных, состоящий в том, что обработка первичных данных, полученных в результате статистического наблюдения, производится на местах. Она проходит несколько этапов. Сначала данные сводятся в пределах административно-территориальных единиц (района, области, края, автономной республики) соответствующими статистическими органами, затем полученные сводные итоги представляются в республиканские органы для дальнейшего их обобщения.
- ручная - это способ выполнения статистической сводки, при котором все основные операции (главным образом подсчёт групповых и общих итогов) осуществляется вручную.
- механизированная - это способ выполнения статистической сводки, при котором все операции осуществляются на основе применения электронно-вычислительных машин (ЭВМ).
- первичная - это сводка, при которой обработка и подсчёт данных происходит непосредственно в процессе статистического наблюдения.
- централизованная - это способ организации сводки статистических данных, при котором все первичные материалы, полученные в результате статистического наблюдения, сосредотачиваются в центральном органе, где и подвергаются сводке.

Статистическая сводка проводится по определённой программе. Причём эта программа должна быть разработана ещё до сбора статистических данных, т.е. практически одновременно с составлением плана статистического наблюдения.

Разработать программу сводки - это, значит, определить, какие группы и подгруппы будут выделены в изучаемой совокупности, какие показатели в виде итогов, средних или относительных величин должны быть подсчитаны для групп и в целом по совокупности, в каких таблицах будет оформлен результат сводки. Выделение данных групп должно быть обоснованным, а не формальным. Сводка кроме получения итоговых и групповых показателей даёт основу для последующего анализа и выявления различного рода закономерностей. Она упорядочивает статистический материал, полученный при наблюдении, классифицирует и систематизирует его.

Упорядоченные в результате сводки статистические совокупности выражаются часто в виде статистических рядов.

В зависимости от того, по какому принципу группируются единицы статистической совокупности, статистические ряды могут носить разный характер. Если единицы совокупности систематизируются по какому-либо имманентному их признаку, то они образуют ряды распределения (например, распределение населения по полу, возрасту, национальности и т.п., распределение промышленных предприятий по отраслям, размерам производства и т.п.). К рядам распределения близко примыкают и так называемые территориальные, или географические, ряды, которые характеризуют распределение, какой либо совокупности по отдельным территориальным единицам.

Если единицы совокупности наблюдаются в течение длительного периода, то они могут быть систематизированы по времени. Получаемые в результате такой систематизации данные именуют рядами динамики.

Анализ различного рода рядов распределения и рядов динамики составляет основу статистического анализа, направленного на выявление статистических закономерностей.

## 2.2 Задачи группировок и их значение в статистическом исследовании

Сводка статистических данных производится таким образом, чтобы наиболее существенные различия между группами не затушёвывались, а наоборот, выделялись с целью их изучения. В то же время объединение в группы сходных между собой однотипных явлений помогает выявить такие их черты и особенности, которые при изучении отдельно каждого явления могли бы остаться незамеченными. Следовательно, научное исследование массовых общественных явлений невозможно без разграничения групп, существенно различающихся между собой, и объединения в группы явлений, сходных в существенном отношении. Поэтому статистическая группировка является основой научной сводки.

Статистическая группировка представляет собой разделение совокупностей общественных явлений на однородные, типичные группы по существенным для них признакам. Группировка является центральным моментом любой сводки. Именно благодаря группировкам материал наблюдения принимает систематизированный вид. Способствуя выявлению наиболее существенных черт и признаков изучаемых явлений, выделению и характеристике основных общественно-экономических явлений, типов предприятий, хозяйств и групп населения, метод группировок служит вместе с тем основой научного применения других методов статистики - средних величин, индексов и т.д. С помощью абсолютных, относительных, средних величин или индексов можно правильно характеризовать только однородные, однотипные группы явлений, выделенные в состав изучаемой совокупности методом группировок.

Важнейшей задачей статистических группировок является выделение существующих в действительности общественно-экономических типов

явлений. На основе теоретического анализа изучаемой совокупности явлений выделяются главные, наиболее характерные группы, типы явлений, из которых она состоит, определяются существенные различия между ними, а также признаки, являющиеся общими для каждой группы. К числу таких группировок относятся группировки предприятий и хозяйств по формам собственности, группировка населения по общественным группам, группировка предприятий и организаций по отраслям народного хозяйства и др.

Признаки, положенные в основу группировки, называются группировочными, а группировка единиц совокупности по тому или иному признаку приводит к рядам распределения.

Таким образом, метод группировок даёт возможность глубоко изучать общественные явления, отражать действительность такой, какой она есть, во всей её сложности и переплетении существенных черт и особенностей.

Группировка - это процесс образования групп единиц совокупности, однородных в каком либо существенном отношении, а также имеющих одинаковые или близкие значения группировочного признака.

Для осуществления статистической группировки устанавливают признак, по которому единицы совокупности распределяют по группам, число групп и их обозначение (границы). Каждая единица совокупности в зависимости от значения у неё группировочного признака

Группировочные признаки могут иметь количественное выражение (например, возраст, заработная плата, число детей в семье, урожайность отдельных сельскохозяйственных культур и т.д.). Эти признаки называют количественными, а ряды распределения, построенные по этим признакам, называют вариационными рядами.

Другие признаки не имеют количественного выражения. Они отражают определённые свойства, качества единиц совокупности. Эти признаки условно называют качественными (например, пол, национальность, семейное положение и т.п.). Группировки, построенные по таким качественным признакам, называют атрибутивными рядами распределения.

Группировки, при помощи которых выявляется взаимосвязь между явлениями, называют аналитическими. При построении таких группировок, прежде всего из двух взаимосвязанных один рассматривается как фактор (т.е. влияющий на другой), а второй - как результат влияния первого. Однако следует иметь в виду, что понятие факторного и результативного признаков рассматривается для каждого конкретного случая особо, ибо то, что служит факторным признаком в одном случае, в другом может выступать как результативный.

Чтобы при помощи группировки выявить зависимость между показателями, необходимо разгруппировать единицы совокупности по факторному признаку и для каждой выделенной группы рассматривать среднее значение результативного показателя, а затем проследить за изменением последнего от группы к группе. Примером может служить группировка фермерских хозяйств по площади земельных угодий. Статистическую

группировку можно строить как по одному, так и по нескольким признакам. Группировка по одному признаку называется простой. Группировка по нескольким признакам называется комбинационной. Комбинационные группировки приобретают особое значение в тех случаях, когда для выделения отдельных групп особенно социальных типов и форм явлений одного признака бывает недостаточно. Тогда приходится один признак брать в сочетании с другими. В статистической практике широко применяются вторичные группировки, к которым относятся группировки, которые формируются на уже обработанном ранее статистическом материале, т.е. в данном случае происходит перегруппировка уже ранее сгруппированного материала. К вторичной группировке прибегают тогда: когда из большого числа первоначально обработанных групп надо получить меньшее число более крупных, более характерных групп; когда в целях сравнения нужно привести в сопоставимый вид, по-разному сгруппированный материал.

В зависимости от сущности изучаемых явлений и задач исследования в статистике применяются различные приёмы группировок: по различным признакам, с различными интервалами, с большим или меньшим числом выделяемых групп; различны, могут быть и показатели, используемые для характеристики каждой группы. Конкретные приёмы группировок должны быть основаны на научном анализе изучаемых явлений, их характерных (типичных) черт и особенностей развития. Только при этом условии группировки из технического средства обобщения собранных сведений превращаются в научный метод статистики, в средство глубокого изучения действительности.

Каждая единица совокупности обладает рядом признаков, изучение которых является задачей науки, в том числе и статистики. Признак, который служит основанием для распределения явлений на группы, называется группировочным признаком, или основанием группировки. Например, стаж работы, вид выпускаемой продукции, уровень рентабельности, возрастная группа и т.д.

Группировочный признак может иметь количественное выражение (стаж работы, возраст и т.д.). При этом все единицы совокупности, обладающие таким признаком (например, рабочие определённого возраста), распределяются на группы в зависимости от размера величины данного признака. Группировочный признак может иметь также и качественное значение. В этом случае единицы совокупности, относимые к разным группам, различаются между собой не величиной признака, (не может быть количественно измерен), а его характером, содержанием, например профессия, образование, уровень качества продукции и т.д.

Правильный выбор группировочного признака имеет решающее значение в процессе группировки статистических данных, от этого зависит научная ценность группировки, и правильность тех выводов, которые могут быть сделаны на её основе. Поэтому при выборе группировочного признака нужно сначала выяснить сущность изучаемого явления или процесса и выявить, что

является главным, решающим для его развития. Затем необходимо установить из каких наиболее характерных, типичных групп состоит изучаемое явление, и какими признаками эти группы различаются между собой. Если признак группировки выбран правильно, то в результате группировки будут не надуманные группы, а реально существующие в действительности и различающиеся между собой наиболее существенными чертами.

Если группировка производится по признаку, имеющему количественное выражение и колеблющемуся в каждой группе в определённых пределах, то для выделения отдельных групп необходимо установить интервалы. Интервалом называется разность между наибольшими и наименьшими значениями признаков в каждой группе.

Определение интервала также важно, как и выбор группировочного признака. При этом необходимо исходить из сущности изучаемого явления и задач исследования. Главное при выборе интервалов состоит в том, чтобы в результате группировки были выделены группы, реально существующие в действительности, чтобы различия между ними при этом не стирались, и не происходило искусственного объединения в одной группе существенно различающихся между собой явлений.

С установлением интервалов связано и определение числа групп. Число групп должно быть по возможности невелико, так как при слишком большом числе групп различия между ними сглаживаются, и группировка перестаёт соответствовать своему назначению: выделять существенные различия между явлениями с целью их изучения. Кроме того, число групп всегда определяется конкретными условиями и зависит от колеблемости признака положенного в основу группировки: при большом колебании признака возрастает и число групп, выделяемых при группировке.

После установления числа групп необходимо определить границы интервалов и выбрать показатели для характеристики групп. Число показателей не должно быть велико, но группы должны быть охарактеризованы достаточно полно.

Если единицы совокупности распределяются по группам по одному признаку, то полученные группировки называются простыми.

Если группировка производится по двум или более признакам, то такие группировки называются сложными или комбинационными.

Необходимость применения сложных группировок объясняется тем, что выделенные группы по одному группировочному признаку не всегда бывают однородны и требуют дальнейшей более детальной группировки по другим отличительным признакам.

Результаты обработки статистических данных, как правило, оформляются в виде статистических таблиц. Табличная форма позволяет изложить материал наиболее удобно, компактно, наглядно и рационально. В каждой статистической таблице имеется ряд горизонтальных строк и вертикальных граф (столбцов), имеющих заголовки, объясняющие, что характеризуется

какими показателями. Строки и графы при пересечении образуют клетки, которые заполняются статистическими данными.

В статистических таблицах различают подлежащее и сказуемое.

Подлежащим в таблице является, те объекты или их части, а также отдельные периоды времени и территории, которые характеризуются рядом показателей.

Сказуемым в таблице являются все те показатели, которыми характеризуется подлежащее.

Статистические таблицы могут быть простыми и сложными.

К простым относятся перечневые таблицы, в которых подлежащим служит перечень отдельных объектов (заводов, агрохозяйства, министерства и т.д.), и динамические таблицы, где подлежащими являются отдельные годы, месяцы или иные периоды времени.

В сложных таблицах подлежащее представляет собой совокупность, расчленённую на группы по одному или нескольким признакам.

Таблицы, в подлежащем которых имеет место группировка по одному признаку, называются групповыми. При наличии в подлежащем группировки по двум или более признакам таблица именуется комбинационной. Таким образом, деление таблиц на простые и комбинационные основано на степени расчленённости подлежащего.

### 2.3 Виды статистических группировок

Исследование какого-либо объекта всегда необходимо разбить на ряд этапов.

Первым этапом является сбор и накопление необходимых знаний о данном объекте. Когда таких наблюдений не много, то можно произвести упорядочивание в порядке убывания или возрастания, то есть ранжировать (построить ранжированные ряды). Когда наблюдений много - их необходимо группировать. Под группировкой понимают расчленение единиц совокупности на группы, однородные по какому-либо признаку, при этом характеризуют их системой показателей.

Среди группировочных признаков выделяют:

1. Количественные (объём, доход, рентабельность, возраст)
2. Качественные (форма собственности предприятия, пол человека, вид экономической деятельности, семейное положение) группировочные признаки.

Варианты - значения группировочного признака.

Группировка, выполненная по одному признаку, называется простой. Среди простых группировок особо выделяют ряды распределения.

Ряд распределения - это группировка, в которой для характеристики групп, расположенных упорядочено по значению признака, применяется один показатель - численность группы.

Ряды распределения:

1. Атрибутивные - построенные по качественному признаку
2. Вариационные - построенные по количественному признаку

Вариационные ряды могут быть дискретными или интервальными. Дискретный ряд распределения - это ряд, в котором варианты выражены одним числом.

Интервальный ряд распределения - это ряд, в котором значения признака заданы в виде интервала. Интервал представляет собой промежуток между максимальным и минимальным значениями признака в группе. Если этот промежуток (величина интервала) не меняется, то такие интервалы называют равными.

Если величина интервала постепенно увеличивается или уменьшается в арифметической или геометрической прогрессии, то интервалы получаются неравными. Также они могут быть открытыми, когда имеется только верхняя или нижняя граница, либо закрытыми когда имеются обе границы.

Если группировка проводится по нескольким признакам, она называется сложной.

Особым видом группировки является классификация, представляющая собой устойчивую номенклатуру классов и групп, образованных на основе сходства и различия у единиц изучаемого объекта.

По своему содержанию группировки делятся на 3 вида:

1. Типологические группировки: выделение социально-экономических типов (например, группировки хозяйствующих субъектов по формам собственности)
2. Структурные группировки: характеризующие структуру изучаемых объектов (например, группировка населения по половозрастному, национальному составу)
3. Аналитические группировки: характеризующие взаимосвязи между отдельными признаками изучаемого явления.

Типологическая группировка решает задачу выявления и характеристики социально-экономических типов (частных подсовокупностей) путем разделения качественно однородной совокупности на классы, социально-экономические типы, однородные группы единиц в соответствии с правилами научной группировки.

Примерами типологической группировки могут служить группировки секторов экономики, хозяйствующих субъектов по формам собственности.

Признаки, по которым производится распределение единиц изучаемой совокупности на группы, называются группировочными признаками, или основанием группировки. Выделить типичное можно не по любому признаку, а только по определенному, который должен изменяться в зависимости от условий места и времени. Для правильного выбора группировочных признаков необходимо предварительно выявить возможные типы, четко формулировать познавательную задачу.

Если группировочными признаками выступают признаки атрибутивные (форма собственности, отрасль производства и т.д.), то образовать группы сравнительно просто.

Выделение типов на основе количественного признака состоит в определении групп с учетом границ перехода количественного признака в новое качество, в новый тип явления.

Однако во всех случаях типологических группировок выбор группировочных признаков всегда должен быть основан на анализе качественной природы исследуемого явления. Экономический анализ сущности и закономерности развития явления должен быть направлен на то, чтобы в соответствии с целью и задачами исследования положить в основание группировки существенные признаки. При этом следует иметь в виду, что один и тот же материал при различных приемах группировки может привести к диаметрально противоположным выводам. Раскрыть закономерности экономического развития помогут те группировки, которые исходят из реально существующих закономерностей.

Структурной группировкой называется группировка, в которой происходит разделение выделенных с помощью типологической группировки типов явлений, однородных совокупностей на группы, характеризующие их структуру по какому-либо варьирующему признаку.

К структурным относится группировка населения по размеру среднедушевого дохода, группировка хозяйств по объему продукции, структура депозитов по сроку их привлечения.

Анализ структурных группировок, взятых за ряд периодов или моментов времени, показывает изменение структуры изучаемых явлений, т.е. структурные сдвиги. В изменении структуры общественных явлений отражаются важнейшие закономерности их развития.

Аналитические (факторные) группировки, в частности, исследуют связи и зависимости между изучаемыми явлениями и их признаками. В основе аналитической группировки лежит факторный признак, и каждая выделенная группа характеризуется средними значениями результативного признака. Так, группируя достаточно большое число рабочих по факторному признаку  $X$  - квалификации (разряду) с указанием их заработной платы, можно заметить прямую зависимость результативного признака  $Y$  - средней месячной заработной платы рабочих от квалификации: чем выше квалификация, тем выше и средняя месячная плата (хотя у отдельных рабочих с более высоким разрядом она может быть ниже).

Если группы, образованные по одному признаку, делятся на подгруппы по второму, а последние - на подгруппы по третьему и т.д., т.е. в основании группировки лежит несколько признаков, взятых в комбинации, то такая группировка называется комбинационной (например, дополнив простую группировку населения по возрастным группам группировкой по полу, получим комбинационную группировку). Комбинационная группировка позволяет выявить и сравнить различия и связи между исследуемыми признаками, которые нельзя обнаружить на основе изолированных группировок по ряду группировочных признаков. Однако при изучении влияния большого числа признаков применение комбинационных группировок

становится невозможным, поскольку чрезмерное дробление информации затушевывает проявление закономерностей. Даже при наличии большого массива первичной информации приходится ограничиваться двумя-четырьмя признаками.

В практике проведения статистического исследования, для определения количества групп, наибольшее распространение получила формула Стерджесса

$$k = 1 + 3,322 \cdot \lg n, \quad (1)$$

где  $k$  - число групп;

$n$  - число единиц совокупности.

После определения числа групп необходимо определить интервалы группировки. Интервал - это значение варьирующего признака, лежащего в определенных границах. Нижней границей интервала называется наименьшее значение признака в интервале и наоборот.

В случае если вариация признака проявляется в сравнительно узких границах и распределение носит более или менее равномерный характер, то целесообразно формировать группировку с равными интервалами.

Величина равного интервала определяется по следующей формуле:

$$h = \frac{x_{\max} - x_{\min}}{n}, \quad (2)$$

где  $h$  - размах вариации.

Группировки с неравными интервалами применяются, когда значение признака варьируют неравномерно и в значительных размерах, что характерно для большинства социально - экономических явлений, особенно при анализе макроэкономических показателей.

## 2.4 Ряды распределения и группировки

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

Статистический ряд распределения - это такое распределение единиц статистической совокупности по значению, какого либо признака, при котором каждому значению или группе значений этого признака соответствует некоторое число единиц совокупности. Статистический ряд может быть формализован (представлен) как в табличном, так и графическом виде.

В зависимости от признака, положенного в основу образования ряда распределения, различают атрибутивные и вариационные ряды распределения.

Атрибутивными называют ряды распределения, построенные по качественным признакам.

Вариационными называют ряды распределения, построенные по количественному признаку.

Любой вариационный ряд состоит из двух элементов: вариантов и частот либо частостей.

Вариантами называют отдельные значения признака, которые он принимает в вариационном ряду.

Частоты - это численности отдельных вариантов или каждой группы вариационного ряда, т. е. это числа, показывающие, как часто встречаются те или иные варианты в ряду распределения.

Частостями называют частоты, выраженные в долях единицы или в процентах к итогу. Соответственно сумма частостей равна 1 или 100%.

В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды.

В дискретном вариационном ряду величина количественного признака принимает только целые значения.

В случае непрерывной вариации (интервального вариационного ряда) величина признака у единиц совокупности может принимать в определенных пределах любые значения, отличающиеся друг от друга на сколь угодно малую величину.

Средняя величина - обобщающая характеристика количественного признака на определенный момент времени в расчете на единицу совокупности.

Статистическая средняя является объективной и типичной, если она рассчитывается для качественно однородной совокупности массовых явлений.

При помощи средней происходит сглаживание различий в величине признака, которые возникают по тем или иным причинам у отдельных единиц наблюдения.

Средняя величина является отражением значений изучаемого признака, следовательно, имеет ту же единицу измерения.

Каждая средняя величина характеризует изучаемую совокупность по какому-либо одному признаку. Чтобы получить полное и всестороннее представление об изучаемой совокупности по ряду существенных признаков, в целом необходимо располагать системой средних величин, которые могут описать явление с разных сторон

Различие индивидуальных значений признака внутри изучаемой совокупности в статистике называется вариацией признака. Она возникает в результате того, что его индивидуальные значения складываются под совокупным влиянием разнообразных факторов, которые по-разному сочетаются в каждом отдельном случае.

Средняя величина даёт обобщающую характеристику признака в изучаемой совокупности, но не отражает строения совокупности.

Отдельные значения изучаемого признака могут располагаться около средней величины различным образом. Типичность средней величины зависит от того, насколько сильно отклоняются индивидуальные значения от среднего. Чем меньше эти отклонения, тем лучше средняя величина представляет изучаемую совокупность.

Для оценки вариации используются абсолютные и относительные показатели.

К абсолютным показателям вариации относятся размах вариации, отражающий максимальное различие значений признака, среднее линейное отклонение, дисперсия и среднее квадратическое отклонение, которые характеризуют отклонение отдельных значений признака от средней величины.

Чем меньше эти показатели, тем лучше средняя величина описывает изучаемую совокупность в целом.

К относительным показателям вариации относится коэффициент вариации. Он характеризует силу вариации, также используется для сравнения силы вариации нескольких признаков в одной совокупности или одного признака в нескольких совокупностях. Кроме того, коэффициент вариации позволяет оценить однородность совокупности. Если коэффициент вариации превышает 33%, можно считать, что совокупность не является однородной.

Ряды распределения удобно анализировать при помощи их графического изображения, позволяющего судить и о форме распределения. Наглядное представление о характере изменения частот вариационного ряда отражают полигон и гистограмма.

Полигон используется при изображении дискретных вариационных рядов.

Для его построения в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжированные (упорядоченные) значения варьирующего признака, а по оси ординат наносится шкала для выражения частот.

Гистограмма применяется для изображения интервального ряда. При построении гистограммы на оси абсцисс откладываются величины интервалов, а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. Высота столбиков в случае равных интервалов должна быть пропорциональна частотам.

## 2.5 Многомерные группировки

Использование в статистических исследованиях ЭВМ и статистической теории распознавания образов позволило разработать метод группировки совокупности единиц одновременно по множеству характеризующих признаков. Такие группировки получили название многомерных.

Многомерная группировка или многомерная классификация основана на измерении сходства или различия между объектами (единицами): единицы, отнесенные к одной группе (классу), различаются между собой меньше, чем единицы, отнесенные к различным группам (классам). Мерой близости (сходства) между объектами могут служить различные критерии. Самой распространенной мерой близости является евклидово расстояние между объектами, представленными точками в  $n$ - мерном пространстве. Чем меньше это расстояние, тем больше близость.

Задача многомерной группировки сводится к выделению сгущений точек (объектов) в  $n$  - мерном пространстве. Группы (кластеры) формируются на основании близости объектов одновременно по всему комплексу признаков, описывающих объект. Нахождение этих групп осуществляется методами кластерного анализа на ЭВМ

Многомерные группировки позволяют решать целый ряд таких важных задач экономико-статистического исследования, как формирование однородных совокупностей, выбор существенных признаков, выделение типичных групп объектов и др. в зависимости от вида группировочных признаков различают группировки по атрибутивным и количественным признакам. Если атрибутивный признак имеет мало разновидностей, то количество групп определяется числом этих разновидностей. Таковы, например, группировки населения по полу, семейному положению, образованию; распределение числа групп при группировке по варьирующему количественному признаку (например, распределение населения по уровню доходов, потреблению отдельных продуктов питания и др.) требует специальных расчетов.

Цель многомерных группировок - классификация данных на основе множества признаков, то есть выделение групп статистических единиц, однородных по нескольким признакам одновременно. В процессе такой группировки решаются, например, задачи типизации - выделяются самостоятельные экономические или социальные типы явлений. Так, приемами многомерной классификации можно всю совокупность промышленных предприятий разбить на «мелкие», «средние» и «крупные», используя следующие признаки: численность промышленно- производственного персонала, объем продукции, стоимость ОПФ, потребление материальных ресурсов и т.д.

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

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

При выполнении многомерных группировок могут быть использованы два основных подхода:

- первый заключается в том, что рассчитывается обобщающий показатель по совокупности группировочных признаков и проводится простая группировка по этому обобщающему показателю;
- второй подход состоит в использовании методом кластерного анализа.

Мерой близости объектов, то есть мерой сходства единиц совокупности, могут служить различные критерии.

Выделяют три типа мер сходства:

- коэффициенты подобия;
- коэффициенты связи;

- показатели расстояния.

## 2.6 Метод группировки и его место в системе статистических методов

Отдельные единицы статистической совокупности объединяются в группы при помощи метода группировки. Это позволяет «сжать» информацию, полученную в ходе наблюдения, и на этой основе выявить закономерности, присущие изучаемому явлению.

Группировка является одним из самых сложных в методологическом плане этапов статистического исследования.

Причины, обуславливающие необходимость проведения группировки и определяющие ее место в системе статистических методов, кроются в своеобразии объекта статистического исследования. Он представляет собой комплекс частных совокупностей, которые могут быть качественно и глубоко различны, обладать различными свойствами, степенью сложности, характером развития.

Невозможность статистической характеристики объекта исследования без выделения групп легко показать на примере совокупности промышленных предприятий. Каждое промышленное предприятие имеет индивидуальные особенности: год образования, место положения, состав установленного оборудования и т.д. Без преодоления этих индивидуальных черт исследовать закономерности развития промышленности, которые теряются в многочисленных характеристиках, отличающих одно предприятие от другого, нельзя. Поэтому предприятия следует объединить в группы по отрасли промышленности, назначению выпускаемой продукции, численности занятых и форме собственности и т.д. Таким образом, в показателях, исчисленных по достаточно большим группам, произойдет погашение случайного и выявление общего, существенного для развития исследуемого явления.

Итак, группировки являются важнейшим статистическим методом обобщения данных, основой для правильного исчисления статистических показателей.

С помощью метода группировок решаются следующие задачи:

- выделение социально-экономических типов явлений;
- изучение структуры явления и структурных сдвигов, происходящих в нем;
- выявление связи и зависимости между явлениями.

Статистические группировки можно классифицировать по следующим признакам: целям и задачам, числу группировочных признаков, упорядоченности исходных статистических данных.

В зависимости от цели и задач исследования статистические группировки бывают типологическими, структурными и аналитическими.

Типологическая группировка - это разделение исследуемой качественно разнородной совокупности на классы, социально-экономические типы, однородные группы единиц в соответствии с правилами научной группировки. При этом под однородностью понимается подчинение всех единиц

совокупности одному закону развития в отношении рассматриваемого свойства.

Типологические группировки позволяют проследить зарождение, развитие и отмирание различных типов явлений (развитие различных форм собственности, формирование новых слоев населения). При построении типологической группировки особое внимание уделяется идентификации типов и выбору группировочного признака. Вопрос об основании группировки должен решаться исходя из сущности изучаемого явления.

Однако формирование типов явлений связано с конкретными условиями места и времени. Сущность того или иного типа явления или процесса может проявляться и раскрываться как в одном, так и во множестве признаков. Одни и те же значения группировочных признаков в различных сочетаниях с другими признаками могут определять принадлежность единиц наблюдения то к одному, то к другому типу. Поэтому для выделения социально-экономических типов целесообразнее всего рассматривать не отдельные, изолированные признаки, а их совокупность, характеризующую изучаемое явление с различных сторон.

При построении типологической группировки в качестве группировочного признака могут выступать как количественные, так и атрибутивные (качественные) признаки.

Примером типологической группировки по атрибутивному признаку является группировка предприятий и организаций по формам собственности.

Информационная ценность типологической группировки повышается, если она носит условный характер, так как группировка может быть универсальной, т.е. одновременно выделяя типы, показывать структуру совокупности для оценки типов явлений используется не один, а несколько показателей, характеризующих каждую выделенную группу.

Структурная группировка разделяет однородную в качественном отношении совокупность единиц по определенным, существенным признакам на группы, характеризующие ее состав и структуру.

Структурные группировки применяются в изучении практически всех социально-экономических явлений и процессов. С их помощью исследуется состав населения по полу, возрасту, месту проживания; состав коммерческих банков по уставному фонду, капиталу, численности работающих и т.д. В качестве группировочных признаков, так же как и при построении типологической группировки, могут рассматриваться количественные и атрибутивные признаки.

При группировке по атрибутивному признаку группы отличаются друг от друга по характеру признака. Число групп, на которые делится изучаемая совокупность, как правило, определяется числом градаций атрибутивного признака.

Практическое применение структурных группировок позволяет на локальном уровне раскрыть структуру совокупности, проанализировать изучаемые явления и процессы, изменение их во времени и закономерности

изменения состава совокупности во времени, если совокупности прослеживаются за ряд последовательных периодов времени.

В статистике признаки делятся на факторные и результативные. Факторными называются признаки, под воздействием которых изменяются другие, результативные, признаки.

Особенностью аналитической группировки является то, что в основание группировки кладется факторный признак, затем подсчитывается количество единиц совокупности и общее суммарное значение результативного признака по каждой выделенной группе и даже производится расчет среднего значения результативного признака по выделенным группам.

Деление рассмотренных группировок, в зависимости от цели и решаемых задач, на три и отражать закономерности изменения значений признака в зависимости от другого.

По числу группировочных признаков различаются простые группировки (один признак) и сложные (два и более признаков).

## 2.7 Основные проблемы, возникающие при построении группировок

Важнейшей проблемой при построении группировки, является выбор группировочного признака или основание группировки.

Группировочный признак - варьирующий признак по которому производится объединение единиц совокупности в группы. По характеру варьирования, признаки разделяются, как известно, на: атрибутивные и количественные. Это деление определяет особенности решения второй проблемы группировок, а именно - определение числа выделяемых групп.

При выборе в качестве группировочных некоторых атрибутивных признаков, может быть выделено только строго определенное количество групп. При группировке предприятий по прибыли может быть выделено 3 группы. Для многих атрибутивных признаков разрабатываются устойчивые группировки, называемые классификацией. Например: классификация отраслей экономики, классификация занятий населения и др.

При группировке по количественному признаку, вопрос о количестве границы групп следует решать исходя из сущности изучаемого социально-экономического явления. При этом следует принимать во внимание такой показатель, как размах вариаций. Чем больше размах варьирования, тем больше образуется групп и наоборот. Необходимо также принимать во внимание численность единиц совокупности по которой строится группировка. При небольшом объеме совокупности, нецелесообразно образовывать большое число групп, т.к. в этом случае в группах не будет достаточного числа единиц для выявления статистических закономерностей.

Существенным вопросом при группировке по количественному признаку является определение интервалов. Показатели числа групп и

величины интервалов находятся в обратной зависимости. Чем больше величина интервалов - тем меньше требуется групп и наоборот.

Вопрос о количестве групп, а следовательно, и об интервалах группировки решается по-разному при типологических группировках и при выделении групп внутри типов. Изучая количественную сторону массовых общественных явлений, статистика, опираясь на конкретные положения экономической теории, должна в процессе группировки наметить точки перехода количества в новое качество; на основе анализа количественных изменений группировочных признаков наметить точки перехода одного качества в другое. При типологической группировке интервалы должны намечаться таким образом, чтобы они отграничивали социально-экономические типы, установленные на основе экономической теории.

Теоретико-экономический анализ изучаемого явления должен быть предпосылкой научной статистической группировки, но вместе с тем использование аппарата современных статистических методов позволяет оценить степень однородности выделенных групп, производить отбор существенных группировочных признаков, совершенствовать методику определения величины интервалов группировки. Количество выделяемых групп может зависеть и от характера вариации изучаемого показателя.

Если в качестве группировочного используется дискретный признак, то число выделяемых групп соответствует количеству вариантов значения признака, если их число не очень велико. Например, распределение рабочих предприятия по тарифным разрядам, группировка семей по размеру и т.д.

При анализе разнородных данных, например при анализе материала, собранного в различные периоды времени, относящегося к различным отраслям народного хозяйства, возникает необходимость применения вторичной группировки.

Кроме того, методом вторичной группировки пользуются также для того, чтобы показать интенсивность процессов и явлений в разнообразных условиях, например, когда нужно показать степень укрупнения колхозов в различных районах, причем исходные данные представлены различными группировками.

Контрольные вопросы:

1. Что такое статистическая сводка?
2. Что такое статистическая группировка?
3. Какие виды группировок используются в статистике?
4. В чем особенность аналитической группировки?
5. Чем статистическая классификация отличается от группировки?
6. Что такое ряд распределения?
7. Какие ряды распределения Вам известны?
8. Как определяется число групп в ряду распределения?

9. Как определить величину равного интервала в интервальном ряду распределения?

### 3 Вариационные ряды и их характеристики

#### 3.1 Вариационные ряды и их графическое изображение

Установление статистических закономерностей, присущих массовым случайным явлениям, основано на изучении статистических данных — сведений о том, какие значения принял в результате наблюдений интересующий нас признак (случайная величина  $X$ ).

Пример 1. Необходимо изучить изменение выработки на одного рабочего механического цеха в отчетном году по сравнению с предыдущим. Получены следующие данные о распределении 100 рабочих цеха по выработке в отчетном году (в процентах к предыдущему году):

97,8; 97,0; 101,7; 132,5;...; 142,3; 104,2; 141,0; 122,1.



100 значений

Различные значения признака (случайной величины  $X$ ) называют вариантами (обозначаем их через  $x$ ).

Рассмотрение и осмысление этих данных (особенно при большом числе наблюдений  $n$ ) затруднительно, и по ним практически нельзя представить характер распределения признака (случайной величины  $X$ ).

Первый шаг к осмыслению имеющегося статистического материала — это его упорядочение, расположение вариантов в порядке возрастания (убывания), т.е. ранжирование вариантов ряда:

$x_{\min} = 94,0; 94,2; \dots; 142,3; 141,0 = x_{\max}.$



$n = 100$  значений.

В таком виде изучать выработку рабочих тоже не очень удобно из-за обилия числовых данных. Поэтому разобьем варианты на отдельные интервалы, т.е. проведем их группировку.

Число интервалов  $m$  следует брать не очень большим, чтобы после группировки ряд не был громоздким, и не очень малым, чтобы не потерять особенности распределения признака. Согласно формуле Стерджеса (1) определяем рекомендуемое число интервалов:

$$k = (141,0 - 97,0) / (1 + 3,322 \lg 100) = 5,76 (\%).$$

Примем  $k = 6,0(\%)$ . За начало первого интервала рекомендуется брать величину  $x_{нач} = x_{min} - \frac{k}{2}$ . В данном случае  $x_{нач} = 97,0 - \frac{6,0}{2} = 94,0(\%)$ .

Сгруппированный ряд представим в виде таблицы 1.

Таблица 1- Вариационный ряд

№	Выработка в отчетном году (в процентах)	Частота (количество работающих) $n_i$	Частость (доля рабочих)	Накопленная частота $n_i^{нак}$	Накопленная частость $w_i^{нак} = \frac{n_i^{нак}}{n}$
1	94,0-100,0	3	0,03	3	0,03
2	100,0-106,0	7	0,07	10	0,1
3	106,0-112,0	11	0,11	21	0,21
4	112,0-118,0	20	0,2	41	0,41
5	118,0-124,0	28	0,28	69	0,69
6	124,0-130,0	19	0,19	88	0,88
7	130,0-136,0	10	0,1	98	0,98
8	136,0-142,0	2	0,02	100	1
$\Sigma$		100	1		

Числа, показывающие, сколько раз встречаются варианты из данного интервала, называют частотами (обозначаем  $n_i$ ), а отношение их к общему числу наблюдений — частостями или относительными частотами, т.е.

$$w_i = \frac{n_i}{n}. \quad (3)$$

Частоты и частости называются весами.

Вариационным рядом называют ранжированный в порядке возрастания (или убывания) ряд вариантов с соответствующими им весами(частотами или частостями).

Если просмотр первичных, несгруппированных данных делал затруднительным представление об изменчивости значений признака, то полученный теперь вариационный ряд позволяет выявить закономерности распределения рабочих по интервалам выработки. Мы видим, например, что выработка колеблется от 94,0 до 142,0%, наибольшее число рабочих (28, или 0,28 от общего числа) увеличили выработку до 118,0—124,0%, уменьшили выработку (в пределах от 94,0 до 100%) 3 рабочих и т.п.

При изучении вариационных рядов наряду с понятием частоты используется понятие накопленной частоты (обозначаем  $n_i^{нак}$ ). Накопленная частота показывает, сколько наблюдалось вариантов со значением признака,

меньшим  $x$ . Отношение накопленной частоты  $n_i^{нак}$  к общему числу наблюдений  $n$  называют накопленной частотой. Обозначают  $w_i^{нак}$ .

Накопленные частоты (частоты) для каждого интервала находятся последовательным суммированием частот (частостей) всех предшествующих интервалов, включая данный (см. таблицу 1). Например, для  $x = 124$  накопленная частота  $n_i^{нак} = 3 + 7 + 11 + 20 + 28 = 69$ , т.е. 69 рабочих имели выработку, меньшую 124%.

Для задания вариационного ряда достаточно указать варианты и соответствующие им частоты (частости) или накопленные частоты (частости) (в таблице 1 приведены и те, и другие).

Напомним, что вариационный ряд называют дискретным, если любые его варианты отличаются на постоянную величину, и — непрерывным (интервальным), если варианты могут отличаться один от другого на сколь угодно малую величину. При сравнительно небольшом объеме выборки, как правило, составляется дискретный вариационный ряд.

Так, вариационный ряд, представленный в таблице 1, — интервальный (проценты выработки условно округлены до десятых долей). Примером дискретного ряда является распределение 50 рабочих механического цеха по тарифному разряду (таблица 2).

Таблица 2 – Вариационный ряд

Тарифный разряд	1	2	3	4	5	6
Частота (количество рабочих)	2	3	6	8	22	9

Для графического изображения вариационных рядов наиболее часто используются полигон, гистограмма, кумулятивная кривая.

Полигон, как правило, служит для изображения дискретного вариационного ряда и представляет собой ломаную, в которой концы отрезков прямой имеют координаты  $(x_i, n_i), i = 1, 2, \dots, m$ .

Гистограмма служит только для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений признака  $k_i = x_{i+1} - x_i, i = 1, 2, \dots, m$ , и высотами, равными частотам (частостям)  $n_i (w_i)$  интервалов. Если соединить середины верхних оснований прямоугольников отрезками прямой, то можно получить полигон того же распределения.

Кумулятивная кривая (кумулята) — кривая накопленных частот (частостей). Для дискретного ряда кумулята представляет ломаную, соединяющую точки  $(x_i, n_i^{нак})$  или  $(x_i, w_i^{нак})$  ( $x_j, w_j^{нак}$ ). Для интервального вариационного ряда ломаная начинается с точки, абсцисса которой равна началу первого интервала, а ордината — накопленной частоте (частости), равной нулю. Другие точки этой ломаной соответствуют концам интервалов. Весьма важным является понятие эмпирической функции распределения.

Эмпирической функцией распределения  $F_n(x)$  называется относительная частота (частость) того, что признак (случайная величина  $X$ ) примет значение, меньшее заданного  $x$ , т.е.

$$F_n(x) = w(X < x) = w_x^{нак} \quad (4)$$

Другими словами, для данного  $x$  эмпирическая функция распределения представляет накопленную частость

$$w_i^{нак} = \frac{n_i^{нак}}{n}. \quad (5)$$

Пример 2. Построить полигон (гистограмму), кумуляту и эмпирическую функцию распределения рабочих:

- а) по тарифному разряду по данным таблицы 2;
- б) по выработке по данным таблицы 1.

Решение. На рисунке 1 и 2 изображены полигон (гистограмма), кумулята и эмпирическая функция распределения соответственно для дискретного (таблица 2) и интервального (таблица 1) вариационных рядов.

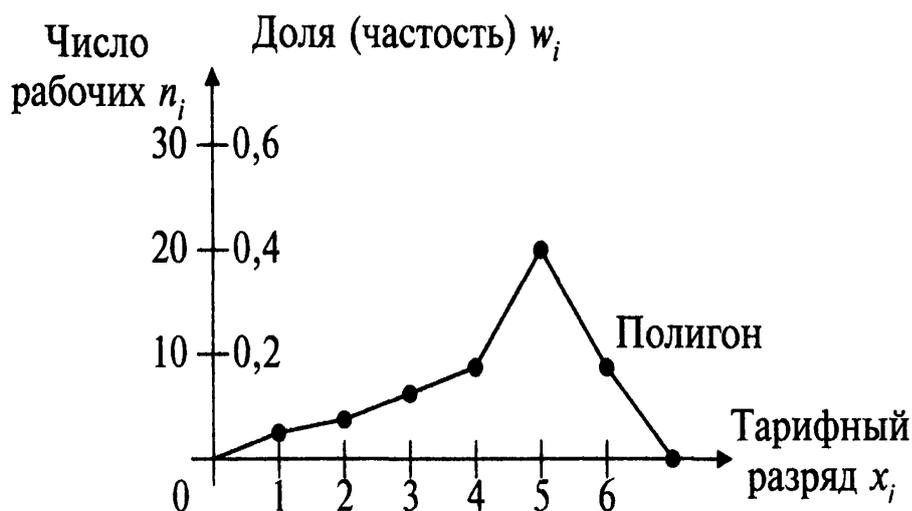


Рисунок 1 – Полигон

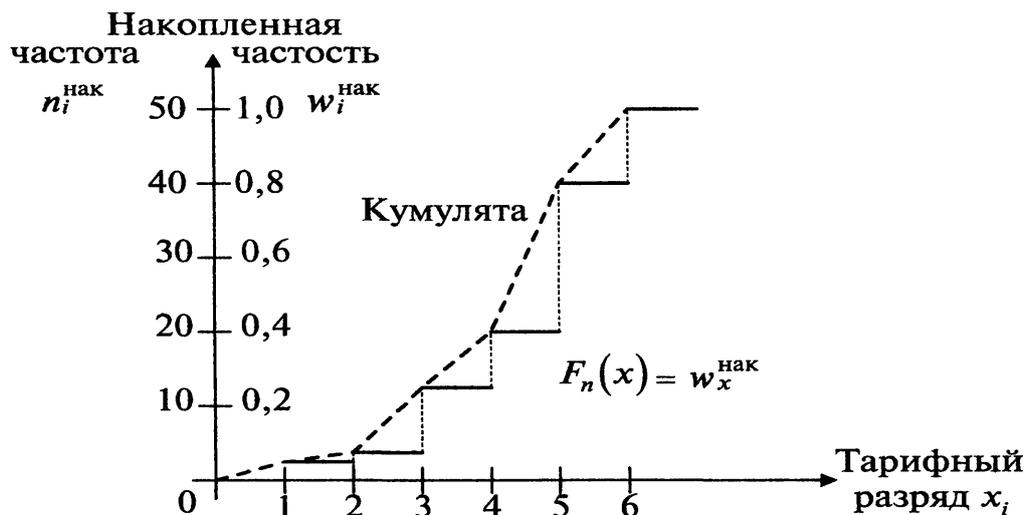


Рисунок 2 – Кумулята

Для интервального вариационного ряда (таблица 1) имеем лишь значения функции распределения  $F_n(x)$  на концах интервала (см. последнюю графу таблицы 1). Поэтому для графического изображения этой функции целесообразно ее доопределить, соединив точки графика, соответствующие концам интервалов, отрезками прямой.

В результате полученная ломаная совпадет с кумулятой (см. рисунок 4).

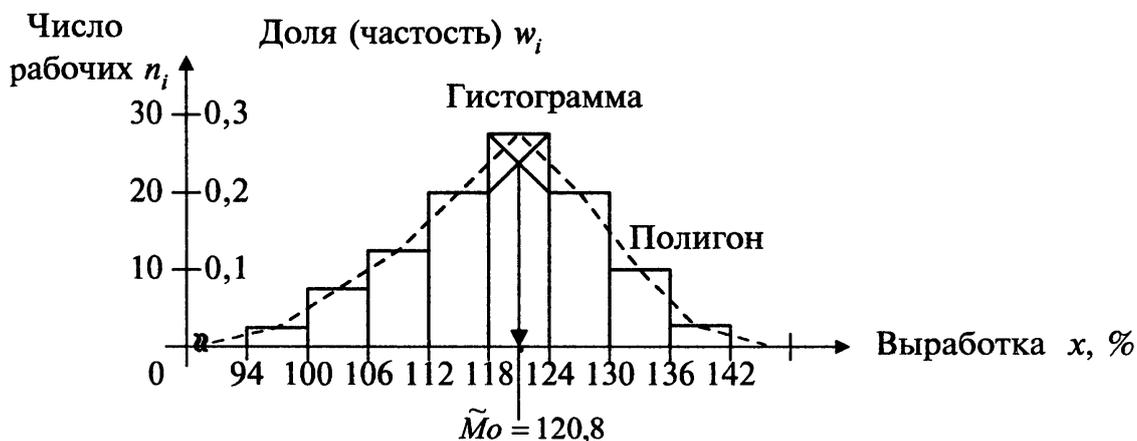


Рисунок 3 – Гистограмма

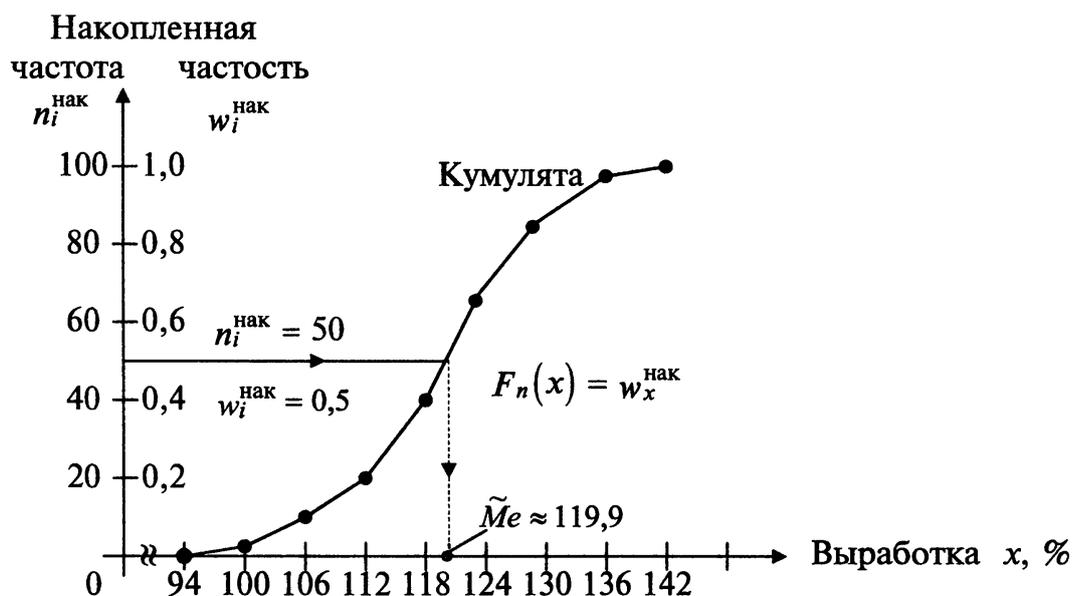


Рисунок 4 - Кумулята

Вариационный ряд является статистическим аналогом (реализацией) распределения признака (случайной величины  $X$ ). В этом смысле полигон (гистограмма) аналогичен кривой распределения, а эмпирическая функция распределения — функции распределения случайной величины  $X$ .

Вариационный ряд содержит достаточно полную информацию об изменчивости (вариации) признака. Однако обилие числовых данных, с помощью которых он задается, усложняет их использование.

В то же время на практике часто оказывается достаточным знание лишь сводных характеристик вариационных рядов: средних или характеристик центральной тенденции; характеристик изменчивости (вариации) и др. Расчет статистических характеристик представляет собой второй после группировки этап обработки данных наблюдений.

### 3.2 Средние величины

Средние величины характеризуют значения признака, вокруг которого концентрируются наблюдения или, как говорят, центральную тенденцию распределения. Наиболее распространенной из средних величин является средняя арифметическая.

Средней арифметической вариационного ряда называется сумма произведений всех вариантов на соответствующие частоты, деленная на объем выборки:

$$\tilde{x} = \frac{\sum_{i=1}^m x_i n_i}{n}, \quad (6)$$

где  $x_i$  - варианты дискретного ряда или середины интервалов интервального вариационного ряда;  $n_i$  - соответствующие им частоты;  $n$  — число неповторяющихся вариантов или число интервалов:

$$n = \sum_{i=1}^m n_i . \quad (7)$$

Очевидно, что

$$X = \sum_{i=1}^m x_i w_i . \quad (8)$$

Пример 3. Найти среднюю выработку рабочих по данным таблицы 1.  
Решение. По формуле (6) для интервального вариационного ряда

$$\tilde{x} = \frac{97 \cdot 3 + 103 \cdot 7 + \dots + 133 \cdot 10 + 139 \cdot 2}{100} = 119,2 (\%),$$

где числа 97,103,..., 133,139 — середины соответствующих интервалов.

Отметим основные свойства средней арифметической, аналогичные свойствам математического ожидания случайной величины:

1. Средняя арифметическая постоянной равна самой постоянной.

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

$$\frac{\sum_{i=1}^m (kx_i)n_i}{n} = k \frac{\sum_{i=1}^m x_i n_i}{n} .$$

3. Если все варианты увеличить (уменьшить) на одно и то же число, то средняя арифметическая увеличится (уменьшится) на то же число:

$$\frac{\sum_{i=1}^m (x_i + c)n_i}{n} = \frac{\sum_{i=1}^m x_i n_i}{n} + c .$$

4. Средняя арифметическая отклонений вариантов от средней арифметической равна нулю:

$$\sum_{i=1}^m (x_i - \tilde{x})n_i = 0 .$$

5. Средняя арифметическая алгебраической суммы нескольких признаков равна такой же сумме средних арифметических этих признаков.

Кроме рассмотренных средних величин, называемых аналитическими, в статистическом анализе применяют структурные, или порядковые, средние. Из них наиболее широко применяются медиана и мода.

Медианой  $\tilde{M}_e$  вариационного ряда называется значение признака, приходящееся на середину ранжированного ряда наблюдений.

Для дискретного вариационного ряда с нечетным числом членов медиана равна срединному варианту, а для ряда с четным числом членов — полусумме двух срединных вариантов.

Пример 4. Найти медиану распределения рабочих по тарифному разряду по данным таблицы 2.

Решение.  $n = 50$  — четное, следовательно, срединных вариантов два:

$$x_{25} = 5 \text{ и } x_{26} = 5. \text{ Поэтому } \tilde{M}_e = \frac{x_{25} + x_{26}}{2} = \frac{5 + 5}{2} = 5 (\%).$$

Для интервального вариационного ряда находится медианный интервал, на который приходится середина ряда, а значение медианы на этом интервале находят с помощью линейного интерполирования. Не приводя соответствующей формулы, отметим, что медиана может быть приближенно

найдена с помощью кумуляты как значение признака, для которого  $n_x^{\text{нак}} = \frac{n}{2}$

или  $w_x^{\text{нак}} = \frac{1}{2}$ .

Достоинство медианы как меры центральной тенденции заключается в том, что на нее не влияет изменение крайних членов вариационного ряда, если любой из них, меньший медианы, остается меньше ее, а любой, больший медианы, продолжает быть больше ее. Медиана предпочтительнее средней арифметической для ряда, у которого крайние варианты по сравнению с остальными оказались чрезмерно большими или малыми.

Модой  $\tilde{M}_o$  вариационного ряда называется вариант, которому соответствует наибольшая частота.

Например, для вариационного ряда таблицы 2 мода  $\tilde{M}_o = 5$ , так как этому варианту соответствует наибольшая частота  $n_i = 22$ . Для интервального ряда находится модальный интервал, имеющий наибольшую частоту, а значение моды на этом интервале определяют с помощью линейного интерполирования. Однако проще моду можно найти графическим путем с помощью гистограммы.

Особенность моды как меры центральной тенденции заключается в том, что она не изменяется при изменении крайних членов ряда, т.е. обладает определенной устойчивостью к вариации признака.

Пример 5. Найти медиану и моду распределения рабочих по выработке по данным таблицы 1.

Решение. На рисунке 4 проведем горизонтальную прямую  $y=0,5$  (или  $y=50$ ), соответствующую накопленной частоте  $w_x^{нак} = F_n(x) = 0,5$  (или накопленной частоте  $n_x^{нак} = 50$ ), до пересечения с графиком эмпирической функции распределения (или кумулятой). Абсцисса точки пересечения и будет медианой вариационного ряда:  $\tilde{M}_e = 119,9(\%)$ .

На гистограмме распределения (рисунок 3) находим прямоугольник с наибольшей частотой (частостью). Соединяя отрезками прямых вершины этого прямоугольника с соответствующими вершинами двух соседних прямоугольников (рисунок 3), получим точку пересечения этих отрезков (диагоналей), абсцисса которой и будет модой вариационного ряда:  $\tilde{M}_o = 120,8(\%)$ .

### 3.3 Показатели вариации

Средние величины, рассмотренные выше, не отражают изменчивости (вариации) значений признака.

Простейшим (и весьма приближенным) показателем вариации является вариационный размах  $R$ , равный разности между наибольшим и наименьшим вариантами ряда:

$$R = x_{\max} - x_{\min}. \quad (9)$$

Наибольший интерес представляют меры вариации (рассеяния) наблюдений вокруг средних величин, в частности, вокруг средней арифметической.

Средним линейным отклонением вариационного ряда называется средняя арифметическая абсолютных величин отклонений вариантов от их средней арифметической:

$$d = \frac{\sum_{i=1}^m |x_i - \bar{x}| n_i}{n}. \quad (10)$$

(Заметим, что «простая» сумма отклонений  $\sum_{i=1}^m |x_i - \bar{x}| n_i$  не может характеризовать вариацию признака, ибо согласно свойству 4 средней арифметической эта сумма равна нулю для любого вариационного ряда).

Дисперсией  $s^2$  вариационного ряда называется средняя арифметическая квадратов отклонений вариантов от их средней арифметической:

$$s^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 n_i}{n}. \quad (11)$$

Дисперсию  $s^2$  часто называют эмпирической или выборочной, подчеркивая, что она (в отличие от дисперсии случайной величины  $\sigma^2$ ) находится по опытным или статическим данным.

Желательно в качестве меры вариации (рассеяния) иметь характеристику, выраженную в тех же единицах, что и значения признака. Такой характеристикой является среднее квадратическое отклонение  $s$  — арифметическое значение корня квадратного из дисперсии:

$$s = \sqrt{\frac{\sum_{i=1}^m (x_i - \tilde{x})^2 n_i}{n}}. \quad (12)$$

Рассматривается также безразмерная характеристика — коэффициент вариации, равный процентному отношению среднего квадратического отклонения к средней арифметической:

$$v = \frac{s}{\bar{x}} \cdot 100\% \quad (\bar{x} \neq 0). \quad (13)$$

Если коэффициент вариации признака, принимающего только положительные значения, высок (например, более 100%), то, как правило, это свидетельствует о неоднородности значений признака.

Отметим основные свойства дисперсии, аналогичные свойствам дисперсии случайной величины:

1. Дисперсия постоянной равна нулю.
2. Если все варианты увеличить (уменьшить) в одно и то же число  $k$  раз, то дисперсия увеличится (уменьшится) в  $k^2$  раз:

$$\frac{\sum_{i=1}^m (x_i k - \tilde{x} k)^2 n_i}{n} = k^2 \frac{\sum_{i=1}^m (x_i - \tilde{x})^2 n_i}{n}.$$

3. Если все варианты увеличить (уменьшить) на одно и то же число, то дисперсия не изменится:

$$\frac{\sum_{i=1}^m [(x_i + c) - (\tilde{x} + c)]^2 n_i}{n} = \frac{\sum_{i=1}^m (x_i - \tilde{x})^2 n_i}{n}.$$

4. Дисперсия равна разности между средней арифметической квадратов вариантов и квадратом средней арифметической:

$$s^2 = \tilde{x}^2 - (\tilde{x})^2. \quad (14)$$

5. Если ряд состоит из нескольких групп наблюдений, то общая дисперсия равна средней арифметической групповых дисперсий и межгрупповой дисперсии:

$$s^2 = \bar{s}_i^2 + \delta^2, \quad (15)$$

где  $s^2$  - общая дисперсия(дисперсия всего ряда);

$$\bar{s}_i^2 = \frac{\sum_{i=1}^l s_i^2 n_i}{n}$$

- средняя арифметическая групповых дисперсий;

$$s_i^2 = \frac{\sum_{j=1}^m (x_j - \tilde{x}_i)^2 n_j}{n_i};$$

$$\delta^2 = \frac{\sum_{i=1}^l (\tilde{x}_i - \tilde{x})^2 n_i}{n}$$

- межгрупповая дисперсия.

Формула (15) носит название «правило сложения дисперсий».

Пример 6. Вычислить дисперсию, среднее квадратическое отклонение и коэффициент вариации распределения рабочих по выработке по данным таблицы 1.

Решение. В примере 3 было получено  $\tilde{x} = 119,2$  (%). По определению дисперсия

$$s^2 = \frac{(97 - 119,2)^2 \cdot 3 + (103 - 119,2)^2 \cdot 7 + \dots + (139 - 119,2)^2 \cdot 2}{100} = 87,48.$$

Среднее квадратическое отклонение

$$s = \sqrt{87,48} = 9,35 \text{ (\%)},$$

коэффициент вариации

$$v = \frac{9,35}{119,2} \cdot 100\% = 7,8 (\%).$$

Среди признаков, изучаемых статистикой, есть и такие, которым свойственны лишь два взаимно исключаящих значения. Это альтернативные признаки. Им придается соответственно два количественных значения: варианты 1 и 0. Частотой варианты 1, которая обозначается  $p$ , является доля единиц, обладающих данным признаком. Разность  $1 - p = q$  является частотой варианты 0. Таким образом,

$x_i$	$w_i$
1	$p$
0	$q$

Средняя арифметическая альтернативного признака

$$\tilde{x} = \frac{1 \cdot p + 0 \cdot q}{p + q} = p, \text{ так как } p + q = 1$$

Дисперсия альтернативного признака

$$s^2 = \frac{(1 - p)^2 p + (0 - p)^2 q}{p + q} = pq.$$

Таким образом, дисперсия альтернативного признака равна произведению доли единиц, обладающих данным признаком, и доли единиц, не обладающих этим признаком. Если значения 1 и 0 встречаются одинаково часто, т. е.  $p=q$ , дисперсия достигает своего максимума  $pq=0,25$ . Дисперсия альтернативного признака используется в выборочных обследованиях, например, качества продукции.

### 3.4 Упрощенный способ расчета средней арифметической и дисперсии

Вычисление средней арифметической  $\tilde{x}$  и дисперсии  $s^2$  вариационного ряда можно упростить, если использовать не первоначальные варианты  $x_i$  ( $i = 1, 2, \dots, m$ ), а новые варианты

$$u_i = \frac{x_i - c}{k}, \tag{16}$$

где  $c$  и  $k$  — специально подобранные постоянные.

Согласно свойствам 2 и 3 средней арифметической и дисперсии

$$\bar{u} = \frac{\tilde{x} - c}{k}, \quad (17)$$

откуда

$$\tilde{x} = \bar{u}k + c. \quad (18)$$

Или

$$\tilde{x} = \frac{\sum_{i=1}^m u_i n_i}{n} \cdot k + c. \quad (19)$$

Формулы (17) – (19) дадут заметное упрощение расчетов, если в качестве постоянной  $k$  взять величину (ширину) интервала по  $x$ , а в качестве  $c$  — середину серединного интервала. Если серединных интервалов два (при четном числе интервалов), то в качестве  $c$  рекомендуется взять середину одного из этих интервалов, например, имеющего большую частоту.

### 3.5 Начальные и центральные моменты вариационного ряда

Средняя арифметическая и дисперсия вариационного ряда являются частными случаями более общего понятия — моментов вариационного ряда.

Начальный момент  $\tilde{v}_k$   $k$ -го порядка вариационного ряда определяется по формуле:

$$\tilde{v}_k = \frac{\sum_{i=1}^m x_i^k n_i}{n}. \quad (20)$$

Очевидно, что  $\tilde{v}_1 = \tilde{x}$ , т.е. средняя арифметическая является начальным моментом первого порядка вариационного ряда.

Центральный момент  $\tilde{\mu}_k$   $k$ -го порядка вариационного ряда определяется по формуле:

$$\tilde{\mu}_k = \frac{\sum_{i=1}^m (x_i - \tilde{x})^k n_i}{n}. \quad (21)$$

С помощью моментов распределения можно описать не только среднюю тенденцию, рассеяние, но и другие особенности вариации признака.

Очевидно, в силу свойства 4 средней арифметической, что  $\tilde{\mu}_1 = 0$ , а  $\tilde{\mu}_2 = s^2$ , т.е. центральный момент первого порядка для любого распределения равен нулю, а второго порядка является дисперсией вариационного ряда.

Коэффициентом асимметрии вариационного ряда называется число

$$\tilde{A} = \frac{\tilde{\mu}_3}{s^3} = \frac{\sum_{i=1}^m (x_i - \tilde{x})^3 n_i}{ns^3}. \quad (22)$$

Если  $\tilde{A} = 0$ , то распределение имеет симметричную форму, т.е. варианты, равноудаленные от  $\tilde{x}$ , имеют одинаковую частоту. При  $\tilde{A} > 0$  ( $\tilde{A} < 0$ ) говорят о положительной (правосторонней) или отрицательной (левосторонней) асимметрии.

Экссесом (или коэффициентом эксцесса) вариационного ряда называется число

$$\tilde{E} = \frac{\tilde{\mu}_4}{s^4} - 3 = \frac{\sum_{i=1}^m (x_i - \tilde{x})^4 n_i}{ns^4} - 3. \quad (23)$$

Экссес является показателем «крутости» вариационного ряда по сравнению с нормальным распределением. Экссес нормально распределенной случайной величины равен нулю.

Если  $\tilde{E} > 0$  ( $\tilde{E} < 0$ ), то полигон вариационного ряда имеет более крутую (пологую) вершину по сравнению с нормальной кривой.

Пример 9. Вычислить коэффициент асимметрии и эксцесс распределения рабочих по выработке по данным таблице 1.

Решение. Коэффициент асимметрии и эксцесс вариационного ряда, приведенного в таблице 1, найдем по формулам (22) и (23):

$$\tilde{A} = \frac{(97 - 119,2)^3 \cdot 3 + (103 - 119,2)^3 \cdot 7 + \dots + (139 - 119,2)^3 \cdot 2}{100 \cdot 9,35^3} = -0,302.$$

$$\tilde{E} = \frac{(97 - 119,2)^4 \cdot 3 + (103 - 119,2)^4 \cdot 7 + \dots + (139 - 119,2)^4 \cdot 2}{100 \cdot 9,35^4} - 3 = 0,286.$$

В силу того, что коэффициент асимметрии  $\tilde{A}$  отрицателен и близок нулю, распределение рабочих по выработке обладает незначительной левосторонней асимметрией, а поскольку эксцесс  $\tilde{E}$  близок нулю, рассматриваемое распределение по крутости приближается к нормальной кривой.

Средняя арифметическая  $\tilde{x}$ , дисперсия  $s^2$  и другие характеристики вариационного ряда являются статистическими аналогами математического ожидания  $M(X)$ , дисперсии  $\sigma^2$  и соответствующих характеристик случайной величины  $X$ .

Начальные и центральные моменты вариационного ряда можно также вычислить упрощенным способом через формулы (17) - (19).

Пример 10. Имеются следующие данные о среднесуточной добыче угля 100 действующих забоев (т/д) (таблица 3):

Таблица 3 – Исходные данные к задаче

434	451	471	450	424	432	440	443	415	423
442	444	452	403	458	402	412	427	417	422
438	442	482	432	416	477	731	432	412	462
496	468	424	438	452	446	418	474	432	452
466	488	452	489	451	422	442	492	473	402
481	468	404	498	467	401	440	449	417	425
444	498	466	442	483	462	492	492	435	449
472	448	436	424	412	443	464	455	463	472
421	442	451	455	405	475	438	447	447	436
438	448	444	445	457	414	425	474	443	408

Требуется:

1. Составить интервальный и соответствующий ему дискретный вариационные ряды.
2. Построить гистограмму и полигон частот распределения.
3. Найти моду и медиану (исходя из дискретного ряда)
4. Найти эмпирические функции распределения непрерывного и дискретного вариационных рядов: построить их графики.
5. Вычислить выборочные среднюю, дисперсию, средне квадратичное отклонение, асимметрию и эксцесс.

Решение.

- 1) Составим интервальный ряд (таблица 4) и соответствующий ему дискретный вариационные ряды (таблица 5):

Таблица 4 - Интервальный ряд

Среднесуточная добыча (интервалы)	400-410	410-420	420-430	430-440	440-450	450-460	460-470	470-480	480-490	490-500
Частота (кол-во забоев) $n_i$	7	9	10	13	21	12	9	8	5	6
$n_i^{\text{нак}}$	7	16	26	39	60	72	81	89	94	100
Частость $W_i$	0,07	0,09	0,1	0,13	0,21	0,12	0,09	0,08	0,05	0,06

Продолжение таблицы 4

$W_{i\text{нак}}$	0,07	0,16	0,26	0,39	0,6	0,72	0,81	0,89	0,94	1
-------------------	------	------	------	------	-----	------	------	------	------	---

Таблица 5 - Дискретный ряд

Среднесуточная добыча (т/д), $x_i$	405	415	425	435	445	455	465	475	485	495
$n_i$	7	9	10	13	21	12	9	8	5	6
$W_i$	0,07	0,09	0,1	0,13	0,21	0,12	0,09	0,08	0,05	0,06
$W_{i\text{нак}}$	0,07	0,76	0,26	0,39	0,6	0,72	0,81	0,89	0,94	1

2) Построим гистограмму и полигон относительных частот распределения (рисунок 5).

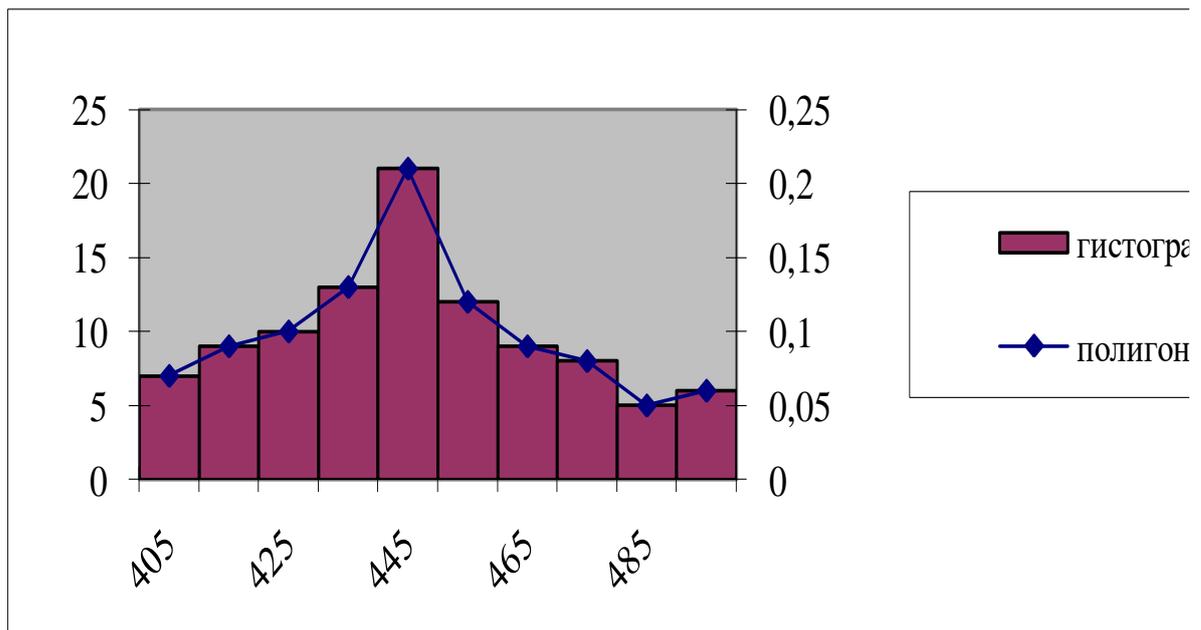


Рисунок 5 - Гистограмма и полигон относительных частот распределения

3) Найдем моду, медиану (исходя из дискретного ряда):

$$\tilde{M}_o = 445, \quad \tilde{M}_e = \frac{445 + 455}{2} = 450.$$

4) Используя формулу (4) найдем эмпирические функции распределения непрерывного и дискретного вариационных рядов и построим их графики.

Для непрерывного ряда:

$$F_n(x) = \begin{cases} 0, & x < 405 \\ 0,07, & 405 \leq x < 415 \\ 0,16, & 415 \leq x < 425 \\ 0,26, & 425 \leq x < 435 \\ 0,39, & 435 \leq x < 445 \\ 0,6, & 445 \leq x < 455 \\ 0,72, & 455 \leq x < 465 \\ 0,81, & 465 \leq x < 475 \\ 0,89, & 475 \leq x < 485 \\ 0,94, & 485 \leq x < 495 \\ 1, & x \geq 495 \end{cases}$$

Построим ему соответствующий график (рисунок 6):

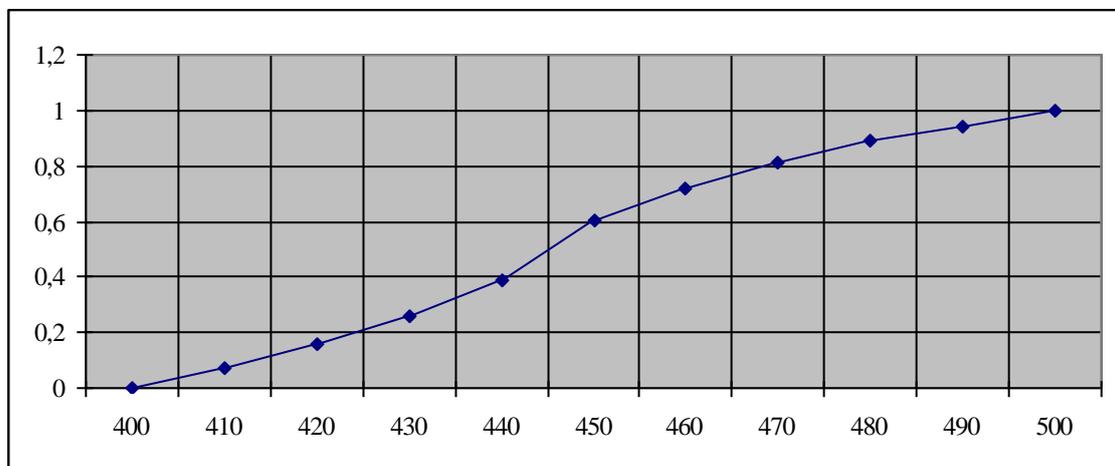


Рисунок 6 - График эмпирической функции распределения непрерывного вариационного ряда

Для дискретного ряда эмпирическая функция распределения имеет вид:

$$F_n(x) = \begin{cases} 0, & x < 405 \\ 0,07, & 405 \leq x < 415 \\ 0,16, & 415 \leq x < 425 \\ 0,26, & 425 \leq x < 435 \\ 0,39, & 435 \leq x < 445 \\ 0,6, & 445 \leq x < 455 \\ 0,72, & 455 \leq x < 465 \\ 0,81, & 461 \leq x < 475 \\ 0,89, & 475 \leq x < 485 \\ 0,94, & 485 \leq x < 495 \\ 1, & x \geq 495. \end{cases}$$

График данной функции может быть построен в MathCad (рисунок 6). Следует помнить, что MathCad не совсем корректно строит графики ступенчатых функций, соединяя отрезками прямых значения функции в точке скачка. Более точный график функции распределения представляет собой отрезки, параллельные оси абсцисс, с «выколотым» правым концом.

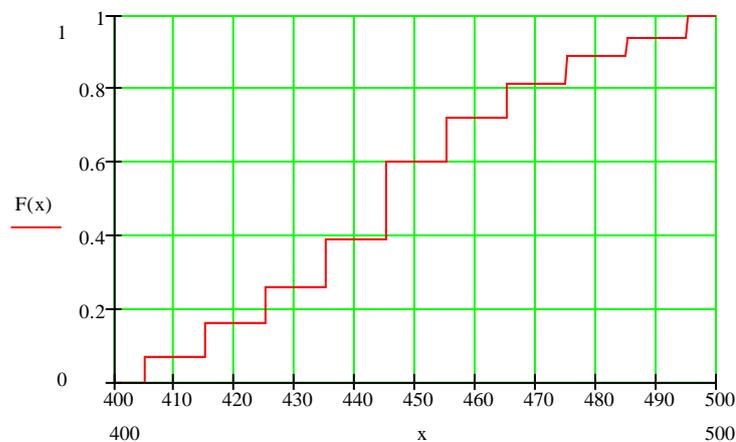


Рисунок 6 - График эмпирической функции для дискретного вариационного ряда

5) Найдем числовые характеристики: выборочную среднюю, дисперсию, среднее арифметическое отклонение, асимметрию, эксцесс. Составим расчетную таблицу (таблица 6):

Таблица 6 - Расчетная таблица

$x_i$	$n_i$	$u_i$	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i (u_i + 1)^4$
405	7	-4	-28	112	-448	1792	567
415	9	-3	-27	81	-243	729	144
425	10	-2	-20	40	-80	160	10
435	13	-1	-13	13	-13	13	0
445	21	0	0	0	0	0	21
455	12	1	12	12	12	12	192
465	9	2	18	36	72	144	729
475	8	3	24	72	216	648	2048
485	5	4	20	80	320	1280	3125
495	6	5	30	150	750	3750	7776
$\Sigma$	100	5	16	596	586	8528	14612

Контроль произведенных расчетов осуществляется по формуле:

$$\sum_{i=1}^{10} n_i (U_i + 1)^4 = \sum_{i=1}^{10} n_i \cdot U_i^4 + 4 \sum_{i=1}^{10} n_i \cdot U_i^3 + 6 \sum_{i=1}^{10} n_i \cdot U_i^2 + 4 \sum_{i=1}^{10} n_i \cdot U_i + n \quad (23)$$

$$14612 = 8528 + 4 \cdot 586 + 6 \cdot 596 + 4 \cdot 16 + 100.$$

$$14612 = 14612.$$

Найдем начальные центральные моменты по упрощенным формулам:

$$\tilde{v}_1 = \frac{\sum_{i=1}^{10} n_i \cdot u_i}{n} = \frac{16}{100} = 0,16, \quad \tilde{v}_2 = \frac{\sum_{i=1}^{10} n_i \cdot u_i^2}{n} = \frac{596}{100} = 5,96$$

$$\tilde{v}_3 = \frac{\sum_{i=1}^{10} n_i \cdot u_i^3}{n} = \frac{586}{100} = 5,86, \quad \tilde{v}_4 = \frac{\sum_{i=1}^{10} n_i \cdot u_i^4}{n} = \frac{8528}{100} = 85,28$$

$$\tilde{\mu}_3 = (\tilde{v}_3 - 3 \cdot \tilde{v}_1 \cdot \tilde{v}_2 + 2\tilde{v}_1^3) \cdot h^3 = (5,86 - 3 \cdot 0,16 \cdot 5,96 + 2 \cdot 0,16^3) \cdot 10^3 = 3007,392$$

$$\tilde{\mu}_4 = (\tilde{v}_4 - 4 \cdot \tilde{v}_1 \cdot \tilde{v}_3 + 6\tilde{v}_1^2 \cdot \tilde{v}_2 - 3(\tilde{v}_1)^4) h^4 =$$

$$= (85,28 - 4 \cdot 0,16 \cdot 5,86 + 6 \cdot 0,16^2 \cdot 5,96 - 3 \cdot 0,16^4) \cdot 10^4 = 899438,8992.$$

$$\tilde{x} = \tilde{v}_1 \cdot h + c = 0,16 \cdot 10 + 445 = 446,6.$$

$$s^2 = (\tilde{v}_2 - (\tilde{v}_1)^2) \cdot h^2 = (5,96 + 0,16^2) \cdot 100 = 593,44.$$

$$s = \sqrt{593,44} = 24,36.$$

Найдем асимметрию:

$$\tilde{A} = \frac{\tilde{\mu}_3}{s^3} = \frac{3007,392}{24,36^2} = 0,208 \quad (\text{асимметрия мало отличается от кривой Гаусса,}$$

асимметрия правосторонняя).

Найдем эксцесс:

$$\tilde{E} = \frac{\tilde{\mu}_4}{s^4} - 3 = \frac{899438,8992}{24,36^4} - 3 = -0,446 \quad (\text{т.к. величина мало } \Rightarrow \text{ кривая близка к}$$

кривой Гаусса, знак «-» означает, что кривая более пологая).

На основании полученных данных можно сделать предположение, что признак  $X$  распределен по нормальному закону.

Контрольные вопросы:

1. Что называется случайной выборкой и просто выборкой из генеральной совокупности и в чём состоит различие между ними?
2. Что называют вариационным рядом?
3. Дайте определение статистического ряда и интервального статистического ряда.
4. Что такое эмпирическая функция распределения?
5. Что называют гистограммой относительных частот и полигоном частот?
6. Как гистограмма связана с законом распределения непрерывной генеральной совокупности?
7. По каким формулам вычисляются выборочные начальные и центральные моменты?
8. Запишите формулы для подсчёта выборочных моментов по группированным данным.

## Упражнения

Задача 1. Распределение сотрудников супермаркета по стажу работы не менее 4 лет характеризуется данными, приведенными в таблице 7.

Рассчитайте размах вариации, среднее линейное отклонение, дисперсию, среднее квадратическое отклонение и коэффициент вариации. Сделайте выводы.

Таблица 7 – Исходные данные

Стаж, лет	4	5	6	7	8	9
Число сотрудников	8	12	17	11	9	4

Задача 2. Имеются следующие данные о товарообороте продовольственных магазинов города (таблица 8).

Определите размах вариации, среднее линейное отклонение, дисперсию, среднее квадратическое отклонение и коэффициент вариации. Проанализируйте полученные результаты.

Таблица 8 – Исходные данные

Группы магазинов по товарообороту млн. руб.	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80– 90	90 – 100	100 – 110
Число магазинов	3	4	8	13	15	21	17	9	5	2

Задача 3. Совместные предприятия одной из отраслей промышленности сгруппированы по стоимости реализованной продукции и услуг за год. Результаты представлены в таблице 9.

Определите: среднюю из групповых дисперсий, межгрупповую дисперсию, общую дисперсию (по правилу сложения дисперсий); эмпирическое корреляционное отношение. Дайте экономическую интерпретацию полученных результатов.

Таблица 9 – Исходные данные

Стоимость реализованной продукции и услуг	Число предприятий	Средняя стоимость реализованной продукции и услуг по группе	Групповая дисперсия
3,5 – 6,5	9	5,59	6,13
6,5 – 9,5	10	7,06	6,51
9,5 и выше	11	12,20	72,16

Задача 4. Результаты статистического обследования фирм, участвовавших в междугородной выставке, представлены в таблице 10.

Оцените степень соответствия данного распределения нормальному распределению на основе показателей асимметрии и эксцесса.

Таблица 10 – Исходные даны

Расходы на рекламу в % к общим расходам фирмы	0,5-1,0	1,0-1,5	1,5-2,0	2,0-2,5	2,5-3,0	3,0-3,5	3,5-4,0	4,0-4,5
Количество фирм	46	123	525	228	35	28	12	3

Задача 5. Имеются следующие данные о ежемесячном объеме выработки 100 рудодобывающих бригад, т.

252	250	244	224	232	240	224	244	226	253
248	216	230	254	258	202	225	224	252	234
242	212	231	251	204	246	232	282	242	252
296	242	254	218	226	252	238	224	248	260
276	254	282	242	270	254	260	232	268	242
244	276	224	240	272	268	281	234	268	232
271	212	234	262	204	261	254	266	278	248
252	262	256	264	272	242	244	246	253	234
236	264	252	248	247	268	229	239	262	212
278	242	254	263	261	266	254	264	248	251

Требуется:

1. Составить интервальный и соответствующий ему дискретный вариационные ряды.
2. Построить гистограмму и полигон частотей (относительных частот) распределения.
3. Найти моду и медиану (исходя из дискретного ряда).
4. Найти эмпирические функции распределения непрерывного и дискретного вариационных рядов; построить их графики.
5. Вычислить выборочную среднюю, дисперсию, среднее квадратическое отклонение, асимметрию и эксцесс.

## Тесты для самоконтроля

1. Дана выборка

10   12   10   12   12   14   15   16   16   15   10   12  
           16   18   20   15   20   16   15

Статистическое распределение имеет вид:

а)

X <sub>i</sub>	10	12	14	15	16	18	20
N <sub>i</sub>	4	3	2	3	4	1	2

б)

X <sub>i</sub>	10	12	14	15	16	20
N <sub>i</sub>	4	3	2	4	4	2

в)

X <sub>i</sub>	10	12	14	15	16	18	20
N <sub>i</sub>	3	4	1	4	4	1	2

г)

X <sub>i</sub>	10	12	14	15	16	18	20
N <sub>i</sub>	3	4	2	3	4	1	2

д)

X <sub>i</sub>	10	12	14	15	16	18	20
N <sub>i</sub>	3	2	2	3	2	1	2

2.

2. Дана выборка

10   12   10   12   12   14   15   14   16   16   15  
           10   12   16   18   20   15   15   20   16

Объём её равен

а)20;

б)19;

в)1;

г)0;

д)18.

3. Дана выборка

10   12   10   12   12   14   15   14   16   16   15   10   12

Объём её равен

- а)12;
- б)13;
- в)10;
- г)16;
- д)20.

4. Дана выборка

10    12    10    12    12    14    15    14    16    16    15    10    12  
      16    18    20    15    15    20    16

Выборочная средняя равна

- а)14,0;
- б)14,4;
- в)15,6;
- г)12,8;
- д)16,0.

5. Дана выборка

10    12    10    12    12    14    15    14    16    16    15    10  
      12    16    18    20    15    15    20    16

Выборочная дисперсия равна

- а)6,792;
- б)8,54;
- в)8,44;
- г)7,2;
- д)1.

6. Полигон частот – это

- а) ломаная;
- б) ступенчатая фигура;
- в) кривая;
- г) прямая линия;
- д)парабола.

7. Начальный момент  $\tilde{V}_1$  для выборки

10    12    10    12    12    14    15    14    16    16    15    10    12  
      16    18    20    15    15    20    16

равен:

- а) 2,8;
- б) 1,4;
- в) 2;
- г) 14,4;
- д) 28.

8. Дана выборка

5 5 1 4 5 1 1 1 4 1

Выборочная средняя равна

- а) 16,1;
- б) 0,71;
- в) 2,8;
- г) 28;
- д) 1,4.

9. Эмпирическая функция  $F_n(x)$  определяет

- а) вероятность событий;
- б) среднее значение отклонения;
- в) среднее арифметическое значение признака;
- г) квадратный корень из выборочной функции;
- д) относительную частоту события.

10. Дана выборка

1 2 4 5 1 1 2 2 4

Объём выборки равен

- а) 22;
- б) 8;
- в) 11;
- г) 10;
- д) 9.

11. Эксцесс нормального распределения равен

- а) 3;
- б) 0;
- в) 1;
- г) -1;
- д) -3.

12. Гистограмма – это

- а) ломаная;
- б) ступенчатая функция;
- в) кривая;
- г) прямая линия;
- д) парабола.

13. Дана выборка

2    3    2    4    5    5    2    3    2    2

Статистическое распределение имеет вид

а)  $\frac{x_i | 1 | 2 | 3 | 4}{n_i | 2 | 3 | 5 | 4};$

б)  $\frac{x_i | 2 | 3 | 4 | 5}{n_i | 5 | 2 | 1 | 2};$

в)  $\frac{x_i | 2 | 3 | 4 | 5}{n_i | 0.5 | 0.2 | 0.1 | 0.2};$

г)  $\frac{x_i | 2 | 3 | 4 | 5}{n_i | 4 | 2 | 1 | 3};$

д)  $\frac{x_i | 2 | 3 | 4 | 5}{n_i | 5 | 2 | 2 | 1}.$

14. Для выборки 1    1    1    1    1    1    1    1    1    1

Выборочная средняя равна

- а) 10;
- б) 1;
- в) 0;
- г) 0,1;
- д) 1,1 .

15. Для выборки 1    1    1    1    1    1    1    1    1    1

Начальный момент  $\tilde{V}_1$  равен

- а) 1,1;
- б) 11;
- в) 1;
- г) 10;
- д) 0.

16. Для выборки 1 1 1 1 1 2 2 2 2 2  
 Распределение относительных частот

а)  $\frac{x_i | 1 | 2}{n_i | 0.5 | 0.5}$  ;

б)  $\frac{x_i | 1 | 2}{n_i | 0.1 | 0.2}$  ;

в)  $\frac{x_i | 1 | 2}{n_i | 5 | 5}$  ;

г)  $\frac{x_i | 1 | 2}{n_i | 0.55 | 0.55}$  ;

д)  $\frac{x_i | 1 | 2}{n_i | 0.55 | 0.45}$ .

17. Пусть из 2000 изделий отобрано для обследования 100 изделий. Тогда объем генеральной совокупности равен:

а) 100;

б) 2000;

в) 20;

г) 2100;

д) 1900.

18. Для исследования производительности труда из 1000 предприятий горной промышленности отобрано 50. Тогда объем выборки равен

а) 20;

б) 50;

в) 1000;

г) 950;

д) 1050.

19. Относительная частота  $w_i$  определяется по формуле

а)  $w_i = n_i x_i$  ;

б)  $w_i = \frac{n_i}{n}$  ;

в)  $w_i = n n_i$  ;

г)  $w_i = \frac{x_i}{n_i}$  ;

д)  $w_i = \frac{n_i}{x_i}$ .

20. Дано распределение

$x_i$	4	7	8	12	17
$n_i$	2	4	5	6	3

Тогда мода вариационного ряда равна:

- а) 5;
- б) 12;
- в) 6;
- г) 17;
- д) 7.

21. Выборка задана таблицей распределения

$x_i$	1	2	3	5
$n_i$	15	20	10	5

Тогда выборочная дисперсия равна

- а) 1,36;
- б) 0,8;
- в) 1,166;
- г) 2,2;
- д) 1.

22. Выборка задана в виде распределения частот

$x_i$	4	7	8	12	17
$n_i$	2	4	5	6	3

Тогда медиана равна

- а) 8;
- б) 5;
- в) 7;
- г) 12 ;
- д) 1.

23. Выборка задана таблицей распределения

$x_i$	1	2	3	5
$n_i$	15	20	10	5

Тогда среднее квадратическое отклонение равно:

- а) 1,36;
- б) 2,2 ;

- в) 50;
- г) 1,166 ;
- д) 1.

24. Пусть из 5000 результатов эксперимента отобрано для анализа 50 . Тогда объем генеральной совокупности равен:

- а) 50;
- б) 100;
- в) 5000;
- г) 5050;
- д) 4950.

25.  $F_n(x)$  обладает тем свойством, что:

- а) значения  $F_n(x)$  принадлежат отрезку  $[0,1]$ ;
- б)  $F_n(x)$  является убывающей функцией;
- в)  $F_n(x)$  является невозрастающей функцией;
- г) значения  $F_n(x)$  принадлежат промежутку  $(-\infty,+\infty)$ ;
- д) значения  $F_n(x)$  не принадлежат отрезку  $[0,1]$ ;

26. Из 100 шахт горной промышленности для проведения эксперимента отобрали 21 шахту. Тогда объем выборки равен:

- а) 100;
- б) 1;
- в) 21;
- г) 79 ;
- д) 121.

27. Выборка задана в виде распределения частот

$x_i$	1	7	8	12	100
$n_i$	2	4	15	6	5

Тогда мода этого вариационного ряда равна:

- а) 8;
- б) 15;
- в) 100;
- г) 1;
- д) 12.

28. Выборка задана в виде распределения частот

$x_i$	2	4	7	8	10
$n_i$	3	2	6	11	5

Тогда медиана этого вариационного ряда равна:

- а) 7;
- б) 11;

- в) 6;
- г) 10 ;
- д) 7,2.

29. Чему равна сумма всех относительных частот

- а) 1;
- б) 0;
- в) объему выборки;
- г) 10;
- д) значению, полученному в результате расчетов.

30. Чему равна сумма всех частот

- а) 1;
- б) 0;
- в) объему выборки;
- г) 10;
- д) значению, полученному в результате расчетов.

## Глоссарий

Новые понятия	Содержание
Статистика	Наука, исследующая с количественной стороны массовые общественные явления
Статистическая совокупность	Множество единиц, обладающих массовостью, однородностью, определенной целостностью, взаимозависимостью состояния отдельных единиц и наличием вариации
Объект статистического исследования	Статистическая совокупность
Методология статистическая	Совокупность взаимосвязанных специфических методов, способов и приемов исследования, применяемых в статистике
Закономерность	Повторяемость, последовательность и порядок изменений в явлениях
Статистическая закономерность	Форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности, если причины (условия), порождающие события, не изменяются или изменяются незначительно
Статистический показатель	Категория отображающая размеры и количественные соотношения признаков социально-экономических явлений и их качественной определенности в конкретных условиях места и времени
Точность статистического наблюдения	Степень соответствия величины какого-либо показателя, определенной по материалам статистического наблюдения
Ошибка наблюдения	Расхождение между расчетным и действительным значением изучаемых величин
Ошибка регистрации	Отклонение между значением показателя, полученного в ходе статистического наблюдения, и фактическим, действительным его значением
Ошибка репрезентативности	Отклонение значения показателя обследованной совокупности от его величины по исходной совокупности
Признак	Свойство единиц совокупности, выражающее их сущность и имеющее способность варьировать, т.е. изменяться

Новые понятия	Содержание
Метод статистики	Совокупность приемов, пользуясь которыми статистика исследует свой предмет
Сводка	Научно организованная обработка материалов наблюдения, включая в себя контроль собранных данных, систематизацию, классификацию материала, составление таблиц, получение итогов и производных показателей
Формула Стерджесса	$k = 1 + 3,322 \cdot \lg n,$ <p>где <math>k</math> - число групп;  <math>n</math> - число единиц совокупности</p>
Размах вариации, $h$	$h = \frac{x_{\max} - x_{\min}}{n}$
Статистический ряд распределения	Распределение единиц статистической совокупности по значению, какого либо признака, при котором каждому значению или группе значений этого признака соответствует некоторое число единиц совокупности
Атрибутивный ряд распределения	Ряд распределения, построенный по качественным признакам
Вариационный ряд распределения	Ряд распределения, построенный по количественному признаку
Объем выборки	Сумма всех частот
Вариант	Наблюдавшееся значение признака
Частоты	Числа, показывающие, как часто встречаются те или иные варианты в ряду распределения
Частоты (относительные частоты)	Частоты, выраженные в долях единицы или в процентах к итогу
Структура вариационного ряда	Состоит из двух элементов: вариантов и частот либо частостей.
Статистическая группировка	Разделение совокупностей общественных явлений на однородные, типичные группы по существенным для них признакам
Группировка	Процесс образования групп единиц совокупности, однородных в каком либо существенном отношении, а также имеющих одинаковые или близкие значения группировочного признака

Новые понятия	Содержание
Аналитические группировки	Группировки, при помощи которых выявляется взаимосвязь между явлениями
Интервал	Значение варьирующего признака, лежащего в определенных границах
Формула, для вычисления относительной частоты	$w_i = \frac{n_i}{n}$
Накопленная частота	Частота, которая показывает, сколько наблюдалось вариантов со значением признака, меньшим $x$ . Обозначается $n_i^{нак}$
Накопленная частость	Отношение накопленной частоты $n_i^{нак}$ к общему числу наблюдений $n$ . Обозначают $w_i^{нак}$ .
Полигон	Служит для изображения дискретного вариационного ряда и представляет собой ломаную, в которой концы отрезков прямой имеют координаты $(x_i, n_i), i = 1, 2, \dots, m$ .
Гистограмма	Служит для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений признака $k_i = x_{i+1} - x_i, i = 1, 2, \dots, m$ , и высотами, равными частотам (частостям) $n_i (w_i)$ интервалов
Кумулята	Кривая накопленных частот (частостей)
Эмпирическая функция распределения	Относительная частота (частость) того, что признак (случайная величина $X$ ) примет значение, меньшее заданного $x$ , т.е. $F_n(x) = w(X < x) = w_x^{нак}$
Средняя арифметическая	Сумма произведений всех вариантов на соответствующие частоты, деленная на объем выборки $\tilde{x} = \frac{\sum_{i=1}^m x_i n_i}{n}$

Новые понятия	Содержание
Медиана	Значение признака, приходящееся на середину ранжированного ряда наблюдений Обозначается $\tilde{M}_e$
Мода	Вариант, которому соответствует наибольшая частота Обозначается $\tilde{M}_o$
Дисперсия	Средняя арифметическая квадратов отклонений вариантов от их средней арифметической $s^2 = \frac{\sum_{i=1}^m (x_i - \tilde{x})^2 n_i}{n}$
Среднее квадратическое отклонение	Арифметическое значение корня квадратного из дисперсии $s = \sqrt{\frac{\sum_{i=1}^m (x_i - \tilde{x})^2 n_i}{n}}$
Коэффициент вариации	Процентное отношение среднего квадратического отклонения к средней арифметической $v = \frac{s}{\tilde{x}} \cdot 100\% \quad (x \neq 0)$
Начальный момент $k$ -го порядка	Величина, определяемая по формуле $\tilde{\nu}_k = \frac{\sum_{i=1}^m x_i^k n_i}{n}$
Центральный момент $k$ -го порядка	Величина, определяемая по формуле $\tilde{\mu}_k = \frac{\sum_{i=1}^m (x_i - \tilde{x})^k n_i}{n}$
Коэффициент асимметрии	Число, определяемое по формуле $\tilde{A} = \frac{\tilde{\mu}_3}{s^3} = \frac{\sum_{i=1}^m (x_i - \tilde{x})^3 n_i}{ns^3}$

Новые понятия	Содержание
Эксцесс	Число, определяемое по формуле $\tilde{E} = \frac{\tilde{\mu}_4}{s^4} - 3 = \frac{\sum_{i=1}^m (x_i - \tilde{x})^4 n_i}{ns^4} - 3$
Средняя арифметическая альтернативного признака	Величина, которую находят по формуле $\tilde{x} = \frac{1 \cdot p + 0 \cdot q}{p + q} = p$
Дисперсия альтернативного признака	Величина, которую находят по формуле $s^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = pq$

### Ответы к тестовым заданиям

1	в	16	а
2	а	17	б
3	б	18	б
4	б	19	б
5	б	20	б
6	а	21	в
7	г	22	а
8	в	23	г
9	д	24	в
10	д	25	а
11	а	26	в
12	б	27	а
13	б	28	а
14	б	29	а
15	в	30	в

## Список литературы

1. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие / В.Е. Гмурман. – М.: Высшая школа, 2001. – 400 с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика: учебник / В.Е. Гмурман. - М.: Высшая школа, 1977. – 356 с.
3. Вентцель, Е.С. Теория вероятностей: учебник / Е.С. Вентцель, Л.А. Овчаров. - М.: Наука, 2000. – 448 с.
4. Королёв, В. Ю. Теория вероятностей и математическая статистика: учебник для студ. вузов / В.Ю. Королёв. - М.: Проспект, 2008. - 160 с.
5. Соколов, Г. А. Теория вероятностей: учебник для студ. вузов / Г.А. Соколов, Н.А. Чистякова. - М.: Экзамен, 2005. - 416 с.
6. Ефимова, М.Р. Практикум по общей теории статистики: учеб. пособие / М.Р. Ефимова, О.И. Танченко, Е.В. Петрова - М.: Финансы и статистика, 2001. - 256 с.
7. Шмойлова, Р.А. Практикум по теории статистики: учеб. пособие / Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова. - М.: Финансы и статистика, 2006. - 416 с.
8. Лысенко, С.Н. Общая теория статистики: учебник / С.Н. Лысенко, И.А. Дмитриева. - М.: Форум, 2008. – 208 с.
9. Дубина, И. Н. Математико-статистические методы в эмпирических социально-экономических исследованиях: учеб. пособие / И. Н. Дубина. - М.: Финансы и статистика, 2010. – 416 с.
10. Елисеева, И.И. Статистика. Теория и практика : учеб. пособие / И.И. Елисеева. – СПб.: Питер, 2010. – 368 с.
11. Орлов, А.И. Статистика: учебник / А.И. Орлов. – М.: Экзамен, 2004. – 483 с.
12. Тюрин, Ю.Н. Математическая статистика: учебник для вузов / Ю.Н. Тюрин, Э.К. Лецкий. - М.: МГТУ им. Н.Э. Баумана, 2001. – 424 с.
13. Айвазян, С.А. Прикладная статистика и основы эконометрики: учебник / С.А. Айвазян, В.С.Мхитарян. – М.: ЮНИТИ, 1998. – 1000 с.
14. Айвазян, С.А. Прикладная статистика: учебник / С.А. Айвазян, В.М. Бухштабер, Е.С.Енюков. - М.: ЮНИТИ, 1989. – 608 с.
15. Айвазян, С.А. Прикладная статистика исследование зависимостей: учебник / С.А. Айвазян. - М.: Финансы и статистика, 1985. – 488 с.
16. Кибзун, А.И. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами: учеб. пособие / А.И. Кибзун, Е.Р. Горяинова, А.В. Наумов, А.Н. Сиротин. - М.: ФИЗМАТЛИТ, 2002. – 224 с.
17. Лялин, В.С. Общая теория статистики: учебник для вузов / В.С. Лялин.- СПб: ИВЭСЭП, 2004. – 196 с.
18. Чернова, Т.В. Экономическая статистика: учеб. пособие / Т.В. Чернова. - Таганрог: Изд-во ТРТУ, 1999. – 140 с.
19. Гусаров, В. М. Статистика: учеб. пособие для вузов / В. М. Гусаров. - М.: ЮНИТИ-ДАНА, 2001. – 463 с.

20. Андронов, А.М. Теория вероятностей и математическая статистика: учебник для вузов / А.М. Андронов, Л.Я. Гринглаз, Е.А. Копытов. - М.: ЮНИТИ, 2005. – 464 с.
21. Толстик, Н.В. Статистика: учебник / Н.В. Толстик. – М.: Феникс, 2010. – 246 с.
22. Ефимова, М.Р. Социально-экономическая статистика: учебник / М.Р. Ефимова. - М.: Юрайт, 2013. – 278 с.
23. Рогатных, Е.Б. Элементарная статистика: теоретические основы и практические задания: учебник / Е.Б. Рогатных. – М.: Экзамен, 2009. - 424 с.
24. Елисеева, И.И. Статистика: учебник для вузов / И.И. Елисеева. - М.: Юрайт, 2010. – 278 с.
25. Дианов, Д.В. Прикладная статистика: учебник для вузов / Д.В. Дианов. – М.: Элит, 2006. – 284 с.
26. Елисеева, И.И. Общая теория статистики: учебник для вузов / И.И. Елисеева, М.М. Юзбашев. - М.: Финансы и статистика, 2009. - 656 с.
27. Мелкумов, Я.С. Социально-экономическая статистика: учебно-методическое пособие / Я.С. Мелкумов. - М.: ИМПЭ-ПАБЛИШ, 2007. 200 с.
28. Башина, О.Э. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: учебник для вузов / О.Э. Башина, А.А. Спирина. - М.: Финансы и статистика, 2008. - 440 с.
29. Боровков, А.А. Математическая статистика: учебник для вузов / А. А. Боровков. – М.: Лань, 2010. – 704 с.
30. Павлушков, И.В. Основы высшей математики и математической статистики: учебник / И. В. Павлушков, Л. В. Розовский. - М. : "ГЭОТАР-Медиа", 2007. - 424 с.