**КОНСПЕКТ ЛЕКЦИЙ ПО ДИСЦИПЛИНЕ**

**«СТАТИСТИКА»**

**для студентов заочного факультета специальности**

**«Экономика и организация производства**

**(железнодорожный транспорт)»**

#### 1. Общие понятия статистической науки

**1.1 Предмет и метод статистики**

 **Предмет статистики.** Статистика изучает массовые общественные явления и их динамику при помощи статистических показателей, в постоянной связи с их содержанием, а также количественное выражение качественно определённых закономерностей общественного развития в конкретных условиях места и времени. Для изучения предмета статистики разработаны и применяются специфические приёмы, совокупность которых образует *методологию статистики.*

 Статистика оперирует определёнными *категориями*, т.е. понятиями, отражающими существенные, всеобщие свойства и основные отношения явлений действительности. К таким понятиям относятся: статистическая совокупность, единица совокупности, признак, статистический показатель, система статистических показателей, статистическая закономерность.

 *Статистическая совокупность* — это совокупность объектов или явлений общественной жизни, объединенных общей связью.

 Совокупности могут быть разнородными и однородными. *Единица совокупности* — это первичный элемент статистической совокупности, являющимся носителем признаков, подлежащих регистрации, и основой ведущегося при обследовании счёта.

 Единицы статистической совокупности характеризуются общими свойствами, именуемыми в статистике признаками. *Признак* — показатель, характеризующий свойства, характерные черты или особенности объектов (явлений), которые могут быть охарактеризованы рядом статистических величин. *Показатель* выражает единство количественной стороны явления (его меру), *признак* — отличительные особенности или сходство объектов статистической совокупности.

 Признаки, систематически принимающие различные значения у отдельных единиц совокупности, называются *варьирующими признаками.* Варьирующие признаки могут быть *количественными*, если их варианты выражаются числовыми значениями (возраст, стаж работы, оплата труда) и *атрибутивными*, представляющими собой смысловые понятия (профессия, социальная принадлежность).

 Количественные признаки могут быть *дискретными* (принимающими целые значения) и *непрерывными* (интервальными). В случае, когда варианты признака могут принимать одно из двух противоположных значений, говорят об *альтернативном признаке* (да, нет).

 Признаки могут быть *основные*, определяющие, например, экономическое содержание процессов, и *второстепенные*, внешние по отношению к сущности изучаемых явлений, т.е. непосредственно не связанные с внутренней структурой процессов. Признаки бывают первичные и вторичные. *Первичные признаки* лежат в основе программы сбора первичных статистических материалов. *Вторичные признаки* (производные показатели) — это признаки, получаемые в процессе обработки и анализа данных.

 Признаки, характеризующие статистическую совокупность, взаимосвязаны между собой, поэтому следует различать факторные и результативные признаки. *Факторные признаки* – это независимые признаки, оказывающие влияние на другие, связанные с ними признаки. *Результативные признаки* – это зависимые признаки, которые изменяются под влиянием факторных признаков. Статистическая совокупность состоит из массы отдельных единиц, разрозненных фактов. Задача статистики – Установление общих свойства единиц совокупности достигается с помощью расчёта обобщающих статистических показателей и их анализа.

 *Статистический показатель* — обобщенная количественная характеристика явлений и процессов в единстве с их качественной определенностью. Статистические показатели называются *натуральными,* когда они выражены в единицах счета или различных физических единицах измерения, и *денежными*, или стоимостными, когда они представляют денежную оценку экономических объектов или их элементов.

 Статистические показатели также условно делятся на *объемные* и *качественные*. Статистические показатели образуют систему показателей. *Система статистических показателей* — это совокупность взаимосвязанных между собой статистических показателей, всесторонне отображающих процессы общественной жизни в определенных условиях места и времени. Показатели в системе могут быть связаны как жестко *детерминированной связью* (например, связь основных средств, числа работников и объёма продукции предприятия), так и не жёсткой, свободной, т.е. *стохастической связью* (например, зависимость себестоимости перевозок от качественных показателей работы подвижного состава).

 *Статистическая закономерность* – количественная закономерность изменения в пространстве и во времени массовых явлений и процессов общественной жизни, состоящих из множества элементов (единиц совокупности). Она проявляется не в индивидуальном явлении, а в массе однородных явлений, при обобщении данных статистической совокупности, т.е. в среднем. Следовательно, это средняя закономерность массовых явлений и процессов.

 *Статистические методы* — это совокупность приемов, применяемых в процессе статистического исследования.

 *Статистическое исследование* — процесс изучения явлений на основе статистических методов. Статистические исследования делятся на *стадии*: статистическое наблюдение, сводка и обработка материалов, анализ данных.

#### 1.2 Статистическое наблюдение

Для проведения статистического исследования необходимо обладать статистической информацией. *Статистическая информация* (статистические данные) – первичный статистический материал о социально-экономических явлениях, формирующийся в процессе статистического наблюдения, который затем подвергается систематизации, сводке, анализу и обобщению.

 *Статистическое наблюдение* — это научно организованный сбор количественных данных о явлениях и процессах, происходящих в различных областях деятельности, с помощью учета первичных данных о каждом отдельном случае или факте, относящемся к изучаемому явлению.

 На рисунке 1 показаны формы, виды и способы статистического наблюдения.



Рис.1. Формы, виды и способы статистического наблюдения

 Статистическое наблюдение осуществляется в двух формах: путём предоставления отчётности и проведения специально организованных статистических наблюдений.

 *Отчётностью* называют такую организованную форму, при которой сведения поступают в виде обязательных отчётов в определённые сроки и по утверждённым формам. Отчетность подразделяется на срочную, текущую и годовую.

 Второй формой статистического наблюдения является - специально организованные статистические наблюдения. *Специально организованные статистические наблюдения* проводятся в тех случаях, когда необходимо получить сведения по показателям, не предусмотренным статистической отчетностью.

 Статистическое наблюдение подразделяется на виды – по *времен*и регистрации данных и по *степени охвата* единиц наблюдения.

 *По времени регистрации фактов* различают непрерывное, или текущее наблюдение, периодическое (регистрация по мере надобности) и единовременное.

 *По степени охвата единиц* совокупности различают сплошное и не сплошное наблюдения. Не сплошное наблюдение подразделяется на способ основного массива, монографическое и выборочное.

 Для получения первичных данных могут быть использованы непосредственные наблюдения, документальный учёт, опрос.

 К программно-методологическим вопросам статистического наблюдения относятся:

 установление и цели и задач наблюдения;

 определение объекта и единицы наблюдения;

 разработка программы наблюдения;

 выбор вида и способа наблюдения.

 *Единица наблюдения* — единица, о которой записываются данные, составляющие программы статистического изучения. От единицы наблюдения следует отличать *единицу совокупности*, т.е. первичный элемент объекта статистического наблюдения, признаки которого подлежат регистрации и который является основой ведущегося счета. Исходя из содержания объекта, цели и задач статистического наблюдения разрабатывается программа наблюдения.

*Программа наблюдения* — перечень вопросов (показателей), по которым регистрируются единицы наблюдения и на которые должны быть получены правильные, исчерпывающие ответы.

 *Формуляр* статистического наблюдения *-* это специальный документ, в котором регистрируются ответы на вопросы программы наблюдения. *Индивидуальный формуляр -* это формуляр, предназначенный для регистрации в нем ответов на вопросы программы наблюдения только об одной единице наблюдения*. Списочный формуляр* - это формуляр, предназначенный для регистрации в нем ответов на вопросы программы наблюдения о нескольких единицах наблюдения.

 К статистическим формулярам составляется *инструкция*, в которой подробно разъясняется, как следует заполнять статистический формуляр.

 Для успешного проведения статистического наблюдения разрабатывается *организационный* план. В нём указываются: органы наблюдения, время наблюдения, сроки наблюдения, необходимые подготовительные работы, порядок проведения наблюдения, приёма и сдачи материалов, получения и предоставления предварительных и окончательных итогов.

 Для правильной характеристики изучаемого объекта важное значение имеет установление времени наблюдения. В статистике различают объективное и субъективное время наблюдения. *Объективным временем* называется время, к которому относится данное наблюдение. *Субъективное время наблюдения* — это время производства наблюдения, т.е. период, в течение которого производится регистрация единиц совокупности.

 *Срок (период) наблюдения -* это время, в течение которого производится заполнение статистических формуляров, т.е. осуществляется регистрация единиц наблюдения по установленной программе. Срок наблюдения определяется рядом факторов. В первую очередь он зависит от специфики и особенностей объекта наблюдения. Так, чем крупнее объект наблюдения, тем, при прочих равных условиях, требуется больше времени для проведения статистического наблюдения над ним. Срок диктуется также программой наблюдения, ее объемом и сложностью признаков, подлежащих регистрации. Срок наблюдения, как правило, предполагает указание даты начала и завершения статистического наблюдения.

 *Критический момент* статистического наблюдения *-* это момент времени (конкретный год, день и час), по состоянию на который производится регистрация собираемых сведений в процессе статистического наблюдения.

 Расхождение между установленными статистическим наблюдением и действительными значениями изучаемых величин называется *ошибками наблюдения.*

 Различают ошибки регистрации и ошибки репрезентативности. *Ошибки регистрации* могут возникнуть как вследствие неправильного установления факта, так и вследствие неправильной записи. В результате проверки статистических данных могут быть обнаружены ошибки случайные и систематические. *Случайные ошибки регистрации -* это ошибки, которые возникают вследствие различных случайных причин. *Систематические ошибки регистрации —* это неточности, возникающие в силу определенных и постоянно действующих на протяжении всего статистического наблюдения в одном направлении факторов

 Систематические ошибки могут быть преднамеренные и непреднамеренные. *Преднамеренные систематические ошибки регистрации -* это ошибки, являющиеся результатом того, что опрашиваемый умышленно, сознательно сообщает регистратору неправильные данные. *Непреднамеренные ошибки* чаще всего возникают в результате небрежности или недостаточной квалификации счетного аппарата, плохой постановки первичного учета. *Ошибки регистрации* – это расхождение между зафиксированным при наблюдении статистическом значении признака и действительным его значением в результате неправильной, ошибочной регистрации ответа на вопрос статистического формуляра. Этот вид ошибок может быть и при сплошном и при не сплошном наблюдении.

 При не сплошном наблюдении зачастую возникают ошибки репрезентативности. *Ошибки репрезентативности -* это расхождение между значениями изучаемого признака или показателя в отобранной и обследованной части совокупности (выборочной) и его значениями во всей исходной (генеральной) совокупности. Ошибки репрезентативности могут быть случайными и систематическими. *Случайные* ошибки репрезентативности *-* это ошибки, возникающие в силу не сплошного характера статистического наблюдения. *Систематические* ошибки репрезентативности *–* это ошибки, возникающие вследствие нарушения принципов беспристрастного, непреднамеренного отбора единиц изучаемой (генеральной) совокупности, которые должны быть подвергнуты наблюдению.

 Для выявления и устранения допущенных при регистрации ошибок может применяться счётный и логический контроль собранного материала.

 *Счётный контроль* заключается в проверке точности арифметических расчётов. *Логический контроль* *–* во взаимном сопоставлении ответов на вопросы программы наблюдения путём их логического осмысления или путём сравнения полученных данных с другими источниками по тому же вопросу.

**1.3 Сводка и группировка статистических данных**

 *Статистическая сводка* – это научно организованная обработка материалов наблюдения, включающая в себя систематизацию, группировку данных, составление таблиц, подсчёт групповых и общих итогов, расчёт производных показателей (средних, относительных величин).

 Если производится подсчёт только общих итогов по изучаемой совокупности единиц наблюдения, то сводка называется *простой*.

 По технике или способу выполнения сводка может быть *ручной* либо *механизированной.*

 Статистическая сводка должна проводиться по программе и плану.

*Программа статистической сводки* предусматривает следующие этапы:

 - выбор группировочных признаков;

 - определение порядка формирования групп;

 - разработка системы статистических показателей для характеристики групп и

 - объекта в целом;

 - разработка макетов статистических таблиц для представления результатов

 сводки.

 *План статистической сводки* содержит указания о последовательности и сроках выполнения отдельных частей сводки, её исполнителях и порядке изложения и представления результатов.

 В сводке статистического материала отдельные единицы статистической совокупности объединяются в группы при помощи метода группировок.

 *Статистическая группировка* – это расчленение изучаемой совокупности на группы и подгруппы по определенным характерным достаточным признакам для глубокого и всестороннего изучения явлений. Особым видом группировок является *классификация*, представляющая собой устойчивую номенклатуру классов и групп, образованных на основе сходства и различия единиц изучаемого объекта.

 Метод группировок применяется для решения задач, возникающих в ходе научного статистического исследования:

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

 *Типологическая группировка* – это расчленение разнородной совокупности на отдельные качественно однородные группы, социально-экономические классы и типы в соответствии с правилами научной группировки и выявление на этой основе экономических типов явлений. *Структурной* 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.

 *Аналитическая группировка* – это группировка, выявляющая взаимосвязи и взаимозависимости между изучаемыми социально-экономическими явлениями и признаками, их характеризующими. В основе аналитической группировки лежит факторный признак, каждая выделенная группа характеризуется средними значениями результативного признака.

*Группировочный признак.* Выбор группировочного признака зависит от характера изучаемых явлений и целей группировки. Признаки, по которым производится распределение единиц изучаемой совокупности на группы, называются *группировочными признаками* или *основанием группировки*. Важнейший вопрос группировки — оптимальный выбор группировочного признака. Признак должен быть существенным, а не второстепенным или малозначительным.

 Группировочные признаки условно различаются на качественные и количественные. Группровочный признак называют *количественным*, если он выражается числом. Группировочный признак может иметь качественное выражение: группировка работников по полу, образованию. Признак, который характеризует свойство, качество данного явления и не имеет количественного выражения, называется а*трибутивным.*

Группировки по одному признаку называются *простыми*. Если для формирования групп берут два и более признаков, т.е. группы, образованные по одному признаку, подразделяются на подгруппы по другому, а полученные в результате этого подгруппы подразделяются (каждая в отдельности) еще на подгруппы и т.д., то такие группировки называются *комбинационными.*

 Определение числа групп и количественных границ признаков зависит от цели группировки и от того, с какими признаками приходится иметь дело. При группировке по атрибутивным (качественным) признакам статистическая совокупность распределяется на столько групп, сколько разновидностей имеет признак (по полу — на две группы, по национальному составу — на столько групп, сколько имеется национальностей, и т.д.).

 Если атрибутивный признак имеет большое количество разновидностей (профессии, наименование выпускаемой продукции, оборудования, товаров и т.д.), то для обоснованного объединения их в группы разрабатываются номенклатуры и классификации. *Номенклатура* — это твердо установленный полный подробный перечень отдельных видов изучаемой совокупности.

 *Количественным* называется признак, характеризующий размеры, величину совокупности и дающий возможность расчленить ее на группы по величине индивидуальных значений группировочного признака.

 *Признак пространства* – это адресный признак (адрес предприятий, фирмы и т.д.) При изучении изменений явлений во времени группируют по признаку времени. *Признак времени* позволяет установить хронологию событий (даты, годы, и т.д.)

 Признаки также бывают первичные и вторичные. *Первичные* признаки характеризуют абсолютные размеры изучаемых явлений (численность сотрудников компании, издержки производства, издержки обращения, стоимость основных фондов и т.д.), *вторичные* являются производными от первичных и показывают структуру группируемых явлений (фондовооруженность, производительность труда, себестоимость единицы продукции и т.д.).

 Если признак изменяется в широких пределах и имеет много различных значений, определяется интервала группировки. *Интервал* – это разность между наибольшим и наименьшим значением признака. Интервалы могут быть равными и неравными. Расчет равной величины интервала производится по формуле

i =,

 где Xmax, Xmin – наибольшее и наименьшее значение признака.

 **Пример.** Провести группировку рабочих вагонного депо по величине начисленной им за месяц заработной платы на основании следующих данных. Начислено заработной платы: по 70 у. е. – трём рабочим; по 80 у. е. – четырём; по 90 у. е. – трём; по 100 у. е. – шести; по 110 у. е. – пяти; по 120 у. е. – четырём; по 130 у. е. – одному; 150 у. е. - восьми. Требуется распределить рабочих по размеру заработной платы, установив четыре группы с равными интервалами.

 Решение. Находим разность между наибольшим и наименьшим значениями признака: 150 –70 = 80 у.е.; определяем величину интервала: 80 / (число групп), т.е. (80 / 4) = 20 у.е. В результате получаем следующие группы (табл.1).

## Таблица 1 - **Группировка рабочих по размеру месячной заработной платы**

|  |  |
| --- | --- |
| Группы рабочих по размерумесячной заработной платы, у.е. | Численность рабочих,человек |
| от 70 до 90 (до 90)от 90 до 110от 110 до 130от 130 до 150 (Свыше 130) | 7999 |
| Итого | 34 |

 Если мы несколько изменим запись групп в таблице, например, первый интервал запишем до 90, последний – свыше 130. В этом случае первый и последний интервалы называются открытыми. Оклад 90 у.е. служит верхней границей для первого интервала. Оклад 130 у.е. – нижней границей последнего интервала; второй интервал (от 90 до 110 у.е.) будет закрытым. Таким образом, **открытые** интервалы имеют одну какую-нибудь обозначенную границу, верхнюю или нижнюю, и неопределенные границы, **закрытые** - и верхнюю, и нижнюю. При расчётах надо помнить, что длина открытых интервалов принимается равной длине рядом стоящих закрытых интервалов.

 **Виды рядов распределения**. *Ряд распределения* – это упорядочение по определенному варьирующему признаку однородные группы единиц совокупности. В зависимости от признака, положенного в основание построения ряда распределения, различают атрибутивные и вариационные ряды распределения.

 *Атрибутивный ряд распределения* – это ряд распределения, построенный по качественным признакам, не имеющим числового выражения и характеризующим свойство, качество изучаемого социально-экономического явления.

 *Вариационный ряд распределения* строится по количественному признаку.

 В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды распределения. *Дискретный вариационный ряд* распределения – это ряд, в котором группы составлены по признаку, изменяющемуся дискретно и принимающему целые значения. График дискретного вариационного ряда называется полигоном. *Интервальный вариационный ряд распределения* – это ряд, в котором группировочный признак, составляющий основание группировки, может принимать в определенном интервале любые значения. График интервального вариационного ряда распределения называется гистограммой.

 Вариационные ряды распределения состоят из двух элементов: *вариант* и *частот*. Числовые значения количественного признака в вариационном ряду распределения называются *вариантами*. *Частоты* – это численности отдельных вариант или каждой группы вариационного ряда, т.е. это числа, показывающие, как часто встречаются те или иные варианты в ряду распределения. Сумма всех частот называется объёмом совокупности и определяет число элементов всей совокупности. Процесс накопления частот можно представить в виде графика, который называется кумулятой. Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то получим *огиву*.

 *Частости* – это частоты, выраженные в виде относительных величин (долях единиц или процентах). Сумма частостей равна единице или 100%.

### 1.4 Статистические таблицы, их элементы, виды и правила построения.

 Результаты статистического наблюдения сводки и группировки обычно представляются в форме таблиц. Основанием любой таблицы является сетка-скелет, в которой вертикальные столбцы называются графами, а горизонтальные — строками. Внешне таблицы представляют собой сетку из вертикальных и горизонтальных линий, в которой записываются числовые данные. Графы (сказуемое) и строки (подлежащее) образуют макет таблицы

 *Статистическое подлежащее таблицы* — это то, о чем говорится и что характеризуется в таблице. *Статистическое сказуемое таблицы* показывает, какими признаками характеризуется подлежащее.

 По построению подлежащего различают следующие виды таблиц: *простые, групповые, комбинационные*. В простой таблице подлежащее не делится на группы. Простые таблицы бывают перечневые (содержится перечень единиц, составляющих объект изучения), динамические (приводятся периоды времени), территориальные (даётся перечень территорий, стран, областей, городов). Часто они строятся в различном сочетании (перечневые и динамические, территориальные и динамические).

 *Групповыми называются* таблицы, в которых подлежащее разделено на группы по какому-либо одному признаку (табл. 2).

Таблица 2 - **Структура работников дистанции гражданских сооружений**

|  |  |  |
| --- | --- | --- |
| Категории работников | Количество человек | В % к общей численности |
| РабочиеСпециалистыСлужащиеРуководители | 38751809 | 73,439,6815,181,71 |
| Всего | 447 | 100,00 |

 *Комбинационными* называются такие таблицы, в которых подлежащее делится на группы не по одному, а по нескольким признакам, причем каждая группа, образованная по одному признаку, делится на подгруппы по другому признаку ( табл. 3.)

 При построении таблиц следует соблюдать следующие *правила*:

- таблица должна иметь небольшие размеры, чтобы ее удобно было читать и анализировать;

- название таблицы, заголовки подлежащего и сказуемого должны быть точными, краткими и ясными;

- в таблице должны быть точно обозначены единицы измерения, а также территория и период, к которым относятся приводимые данные;

- приводимые в подлежащем и сказуемом признаки должны быть расположены в логическом порядке с учётом необходимости их совместного рассмотрения.

Таблица 3 - **Группировка магазинов организации рабочего снабжения (ОРС) по размеру товарооборота и по площади торгового зала**

|  |  |  |  |
| --- | --- | --- | --- |
| Группы магазинов по размеру квартального товарооборота, млн. руб. | Площадь торгового зала, кв. м | Количество розничных предприятий, единиц | Розничный товарооборот, млн. руб. |
| До 10 | До ЗО | 1 | 1,2 |
|  | 30—50 | 4 | 14,2 |
|  | 50—100 | 2 | 9,3 |
|  | Свыше 100 | 3 | 28,4 |
| От 11 до 20 | До ЗО | — |  |
|  | 30—50 | 1 | 12,8 |
|  | 50—100 | 6 | 90,1 |
|  | Свыше 100 | 8 | 132,6 |
|  |  |  |  |

Обычный принцип размещения – от частного к общему, т.е. сначала показывают слагаемые, а в конце подводят итоги. Когда приводятся не все слагаемые, а лишь наиболее важные из них, применяется противоположный принцип: сначала показываются общие итоги, а затем выделяют наиболее важные части (« в том числе», «из них»). Следует различать «Итого» и «Всего». «Итого» является итогом для определённой части совокупности, а «Всего» - итог для совокупности в целом.

 - если число показателей подлежащего и сказуемого таблицы значительно, то строки и графы таблиц следует пронумеровать (например, графа 2 + графа 3);

 - таблица может сопровождаться примечаниями, в которых указываются источники данных, более подробно раскрывается содержание показателей, даются другие пояснения, а также оговорки в случае, если таблица содержит данные, полученные в результате вычислений.

 - при оформлении таблиц обычно применяются такие условные обозначения: при отсутствии данных следует ставить знак тире (-), а при отсутствии сведений — многоточие (…) или «нет сведений», х – если явление не имеет осмысленного содержания, Если сведения имеются, но числовое значение меньше принятой в таблице точности, оно выражается дробным числом (0,0);

 - в таблице должны быть подсчитаны итоги;

- цифровой материал должен даваться с одинаковой степенью точности.

#### 1.5 Статистические графики и правила их построения

*Графический метод* есть метод условных изображения статистических данных при помощи геометрических фигур, линий, точек и разнообразных символических образов.

В статистическом графике различают следующие основные элементы: поле графика, графический образ, пространственные и масштабные ориентиры, экспликация графика.

*Полем графика* называют место, на котором он выполнятся. Это листы бумаги, географические карты, план местности и т.п. Поле графика характеризуется его форматом (размерами и пропорциями сторон). Размер поля графика зависит от его назначения. Стороны поля статистического графика обычно находятся в определенной пропорции. Принято считать, что наиболее близким к оптимальному для зрительного восприятия является график, выполненный на поле прямоугольной формы с соотношением сторон от I : 1,3 до 1 : 1,5. Этот вариант именуется правилом «золотого сечения». Иногда используется и поле графика с равными сторонами - т.е. имеющее форму квадрата.

*Графический образ -* это символические знаки, с помощью которых изображаются статистические данные: линии, точки, плоские геометрические фигуры (прямоугольники квадраты, круги и т. д.).

Размещение графических образов на поле графика определяют. Они задаются координатной сеткой или контурными линиями и делят поле графика на части, соответствующие значениям изучаемых показателей.

Рис.1 Числовые интервалы в полярной системе координат .

В полярной системе координат (рис. 1) один из лучей, обычно правый горизонтальный, принимается за ось координат, относительно которой определяется угол луча. Второй координатой считается ее расстояние от центра сетки, называемое *радиусом.* В радиальных графиках лучи обозначают моменты времени, а окружности - величины изучаемого явления.

 На статистических картах пространственные ориентиры задаются контурной сеткой (контуры рек, береговая линия морей и океанов, границы государств) и определяют те территории, к которым относятся статистические величины.

 *Масштабные ориентиры* статистического графика определяются масштабом и системой масштабных шкал. Масштаб статистического графика— это мера перевода числовой величины в графическую. *Масштабной шкалой* называется линия, отдельные точки которой могут быть прочитаны как определенные числа. Шкала имеет большое значение в графике и включает три элемента: линию (или носитель шкалы), определенное число помеченных черточками точек, которые расположены на носителе шкалы в определённом порядке, цифровое обозначение чисел, соответствующих отдельным помеченным точкам. Как правило, цифровым обозначением снабжаются не все помеченные точки, а лишь некоторые из них , расположенные в определенном интервале. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними .

 Носитель шкалы может представлять собой как прямую, так и кривую линии. Поэтому различают шкалы *прямолинейные* (например, миллиметровая линейка) и криволинейные - дуговые и круговые (циферблат часов) . Графические и числовые интервалы бывают равными и неравными. Если на всем протяжении шкалы равным графическим интервалам соответствует равные числовые, то такая шкала называется *равномерной*. Когда же равным числовым интервалам соответствует неравные графические интервалы и наоборот, то перед нами *неравномерная* шкала. *Масштабом равномерной шкалы* называется длина отрезка принятого за единицу и измеренного в каких-либо мерах.

 Последний элемент графика – *экспликация*. Каждый график должен иметь словесное описание. Оно включает его содержание, подписи вдоль масштабных шкал и пояснения к отдельным частям графика.

 *Диаграммы* – наиболее распространенный способ графических изображений. Применяются диаграммы для наглядного сопоставления в различных аспектах независимых друг от друга совокупностей. *Статистические карты* – графики количественного распределения по конкретной территории.

 Статистические графики можно классифицировать по разным признакам: назначению (содержанию), способу построения и характеру графического образа. По *назначению (содержанию*) можно выделить графики сравнения, графики относительных величин (структуры, динамики). Особым видом графика являются диаграммы распределения величин, представленных вариационным рядом. Это гистограмма, полигон, огива, кумулята. По *способу построения* графики можно разделить на диаграммы, картодиаграммы и картограммы. По *характеру графического образа* различают графики точечные, линейные, полосовые, круговые, секторные, фигурные и объёмные.

#### 2 Статистические величины

**2.1 Абсолютные величины**

 Результаты статистического наблюдения регистрируются в виде **абсолютных** *величин*, которые отражают либо объём, либо уровень развития явления. Они имеют размерность (единицы измерения) и могут быть положительными и отрицательными. Абсолютные величины выражаются в различных единицах измерения — натуральных, стоимостных (денежных), условных, трудовых.

 *Натуральные* единицы измерения характеризуют величину и размер изучаемых явлений. Они выражаются в определенных единицах измерения: килограммах, штуках, тоннах, метрах, километрах и т.д., то есть либо в объемных измерителях, либо в весовых, либо в метрах длины и площади. Все эти единицы измерения принято называть натуральными, а учет в них – учетом в натуральном выражении. Натуральные единицы можно суммировать только по однородным продуктам.

 *Стоимостные* единицы находят широкое применение. Они применяются для оценки в стоимостном выражении многих статистических показателей: объема выпущенной продукции промышленности, размера розничного товарооборота и т.д. Показатели, выраженные в стоимостных измерителях, можно суммировать, получать по ним итоговые данные. При использовании стоимостных измерителей принимают во внимание изменение цен с течением времени. Этот недостаток стоимостных измерителей преодолевают применением «неизменных» или «сопоставимых» цен одного и того же периода.

 *Условные единицы измерения*. Необходимость в применении условных единиц вызывается тем, что в ряде случаев не все виды даже однородной продукции можно суммировать. Так, нельзя суммировать мыло, топливо, консервы и т.п., так как мыло, например имеет различный процент жирности, топливо — различную калорийность и т.п. Следовательно, условные единицы измерения применяются для учета однородной продукции различных разновидностей.

 *Трудовые единицы измерения* — это человеко-часы, человеко-дни, человеко-месяцы и т.п. Трудовые измерители характеризуют использование трудовых ресурсов или затраты труда в производственной деятельности, сфере обслуживания и т.п.

 По способу выражения абсолютные величины подразделяются на индивидуальные и суммарные.

 *Индивидуальные абсолютные величины* — это показатели, выражающие размеры количественных признаков отдельных единиц исследуемых объектов. *Суммарные абсолютные величины* характеризуют итоговое значение признака всех единиц изучаемой совокупности или отдельных ее групп и получаются в результате суммирования индивидуальных абсолютных величин.

 Абсолютные статистические показатели подразделяются на показатели объема и показатели уровня*. Показатели объема* позволяют характеризовать величину всей совокупности или ее частей. *Показатели уровня* характеризуют величину нагрузки единицы одной совокупности элементами другой совокупности.

**2.2 Относительные статистические величины**

 **Относительная**величина – это обобщающий показатель, который получаются в результате деления одной величины на другую и даёт числовую меру соотношения между ними.

 Величина, с которой производится сравнение (знаменатель дроби), обычно называется *базой сравнения* или *основанием*. Величина, находящаяся в числителе, называется *текущей*.

 Для удобства чтения и восприятия относительной величины выделяется её целая часть. В качестве базы сравнения может быть выбрана 1, тогда в результате сравнения получится коэффициент. Если за базу сравнения принять 100 – получиться процент %, 1000 – промилле %0 , 10000 – продецимилле %оо .

 Сопоставляемые величины могут быть как одноимёнными, так и разноимёнными (руб./чел.; ткм / км; руб./ м2).

 Относительные величины подразделяются на несколько видов: относительные величины динамики, планового задания, выполнения планового задания, относительные величины структуры, относительные величины координации, относительные величины сравнения и относительные величины интенсивности.

 *Относительная величина динамики(* iд) характеризует изменение (увеличение или снижение) показателей текущего периода по сравнению с прошлым периодом, называются *темпами роста*.

##### $$i\_{д}=\frac{Фактические данные текущего периода}{Фактические данные прошлого периода}∙100$$

 Существуют две схемы расчёта относительных величин динамики: цепная и базисная. *Относительные величины динамики цепные*, характеризуют изменения каждого последующего уровня ряда динамики по сравнению с уровнем ему предшествующим.

 *Относительные величины динамики базисные* показывают темпы развития за каждый данный отрезок времени по сравнению с уровнем, принятым за базу сравнения, чаще всего это бывает начальный уровень ряда динамики. Они широко используются при характеристике развития явлений за определённый период времени.

 *Относительная величина планового задания* (iпл.з) рассчитывается как отношение уровня, запланированного на предстоящий период, к уровню, фактически сложившемуся в этом периоде.

 *Относительная величина выполнения планового задания* ( iвып.пл ) представляет собой отношение фактически достигнутого в данном периоде уровня к запланированному.

 Например, по плану в 2013г. намечено произвести продукции на 3264,7 млн. у.е., фактически произведено на 3330,0 млн. у.е. Плановое задание выполнено на 3330,0/3264,7 100 = 102,0%.

 В плановом задании может устанавливаться величина прироста или снижения показателя. Например, планировалось снижение себестоимости единицы продукции на 24,2 у.е., а фактически снижение составило 27,5 у.е., плановое задание по снижению себестоимости выполнено на

27,5 : 24,2 100 = 113,6%.

 Относительные величины динамики, планового задания и выполнения планового задания связаны соотношением:

$$i\_{д} = i\_{пл.з}∙i\_{вып.пл}$$

 *Относительная величина структуры* характеризует состав изучаемой совокупности и определяется отношением отдельных частей к целому.

Доля, % =$\frac{Часть совокупности}{Вся совокупность}∙$ 100

 *Относительная величина координации(* iк) характеризует соотношение между частями (элементами) одной совокупности.

##### $$i\_{к}=\frac{Одна часть совокупности}{Другая часть той же совокупности}∙100$$

Относительная величина координации выражаются в виде коэффициентов. При исчислении относительных величин координации велико значение не только выбора базы сравнения, но и вообще выбора явлений, которые, могут быть сравнимы между собой.

 *Относительная величина сравнения* показывает соотношение одноименных величин, относящихся к разной территории или к разным объектам, за один и тот же период времени и применяется для сопоставления экономических показателей. Они должны быть сопоставимы между собой

 *Относительная величина интенсивности(* iи) показывает степень распространенности данного явления в изучаемой среде и образуется в результате сравнения разноименных, но определенным образом связанных между собой абсолютных величин.

##### $$i\_{и}=\frac{Одна совокупность, характеризующая явление}{Другая совокупность, характеризующая среду }∙100$$

Относительные величины интенсивности, в отличии от других видов относительных величин, всегда выражаются именованными числами, например, плотность населения 8,6 чел./км2. Разновидностью относительных величин интенсивности являются относительные показатели уровня экономического развития, характеризующие уровни ВВП, ВНП, национального дохода и др. показателей на душу населения и играющие важную роль в оценке развития экономики страны.

1.3 Средние величины и показатели вариации

 **Средняя величина** – это обобщающая характеристика однородной совокупности явлений по определённому признаку.

 В статистике используются различного вида средние величины: средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая, средняя хронологическая и т.д. Чтобы определить *среднюю арифметическую простую*, нужно сумму всех значений данного признака Σx разделить на число единиц n, обладающим этим признакам. Она применяется в тех случаях, когда имеются не сгруппированные индивидуальные значения признака.

$$\overbar{X}\_{ар} = \frac{Σx}{n}$$

*Средняя арифметическая взвешенная* есть частное от деления суммы произведений вариантов и соответствующих им частот Σxf на сумму всех частот Σ f. Частоты (f), фигурирующие в формуле средней, принято называть *весами,* вследствие чего средняя арифметическая, вычисленная с учетом весов, и получила название взвешенной.

$$\overbar{X}\_{ар} = \frac{Σxf}{Σ f}$$

 В отдельных случаях веса могут быть представлены не абсолютными, а относительными величинами (в процентах или долях единицы). Тогда формула средней будет иметь вид

$$\overbar{X}\_{ар} = \frac{Σxd}{Σ d},$$

где d – частость, т.е. доля каждой частоты в общей сумме всех частот.

Если частоты подсчитывают в долях (коэффициентах), то Σ d = 1, формула средней арифметической взвешенной имеет вид:

$\overbar{X}\_{ар} = Σxd$.

Часто приходится исчислять среднюю по групповым средним или по средним отдельных частей совокупности (частным средним), т.е. *среднюю из средних*. Вычисление *средней арифметической взвешенной из групповых средних*xгр осуществляется по формуле

$$\overbar{X}\_{ар} = \frac{Σx\_{гр}f}{Σ f}. $$

Если значения осредняемого признака заданы в виде интервалов («от – до»), т.е. интервальных рядов распределения, то при расчёте средней арифметической величины в качестве значения признаков в группах принимают середины этих интервалов, они рассчитываются как простая средняя между верхней и нижней границами каждого интервала. Если в ряду распределения есть открытые интервалы (первый и последний), их величина условно приравниваются к интервалам, примыкающим к ним. При таком исчислении средней допускается *некоторая неточность*, поскольку делается предположение о равномерности распределения признака внутри группы. Однако ошибка будет тем меньше, чем уже интервал и чем больше единиц в интервале.

Пример. Рассчитаем среднемесячную заработную плату рабочих по формуле средней арифметической взвешенной. После того, как найдены середины интервалов, варианты умножают на частоты (веса) и сумму произведений (тыс. д. е.) делят на сумму частот (весов). (табл. 4).

$$\overbar{X}\_{ар} = \frac{Σxf}{Σ f} = \frac{7290 }{100} = 72,9 тыс. д.е.$$

Т а б л и ц а 4 - **Распределение рабочих по уровню ежемесячной оплаты труда.**

|  |  |  |  |
| --- | --- | --- | --- |
| Группы рабочих по оплате труда, тыс. д.е. | Число рабочих, чел.f | Середина интервала, тыс. д.е. x | $$x∙f$$ |
| До 50 | 5 | 45 | 225 |
| 50 - 60 | 15 | 55 | 825 |
| 60 - 70 | 20 | 65 | 1300 |
| 70 - 80 | 30 | 75 | 2250 |
| 80- 90 | 16 | 85 | 1360 |
| 90 и более | 14 | 95 | 1330 |
| Итого | 100 | - | 7290 |

 **Средняя арифметическая и ее свойства.** Рассмотрим основные свойства средней арифметической.

 *Первое свойство*. Нулевое свойство средней величины заключается в том, что сумма отклонений вариант от их средней арифметической величины равна нулю.



 *Второе свойство***.** Если все варианты уменьшить (или увеличить)на одно и то же постоянное число, то средняя арифметическая из этих вариант уменьшится (или увеличится) на то же самое число.



 *Пример.* Пусть заработная плата каждого работника фирмы увеличилась за некоторый период на 15000 д.е. Тогда средняя заработная плата всех работников фирмы увеличилась также на 15000 д.е.

 *Третье свойство***.** Если все варианты одинаково увеличить (или уменьшить) в одно и то же число раз, то средняя арифметическая увеличится (или уменьшится) во столько же раз.



 *Пример.* Так, если бы заработная плата каждого работника фирмы увеличилась на 10%, то и средняя заработная плата всех работников фирмы увеличилась бы на 10%.

 *Четвертое свойство***.** Если же все веса средней одинаково увеличить (или уменьшить) в несколько раз, средняя арифметическая не изменится. Увеличение всех весов в несколько раз приводит к тому, что во столько же одновременно увеличится и числитель, и знаменатель дроби (средней арифметической), поэтому значение дроби не изменяется.

 Для упрощения расчётов средней идут по пути уменьшения значений вариантов и частот. Наибольшее упрощение достигается, когда в качестве А выбирается значение одного из центральных вариантов, обладающего наибольшей частотой, в качестве I – величина интервала. Величина А называется началом отсчёта, поэтому такой метод вычисления средней называется «способом отсчёта от условного нуля» или «*способом моментов*».

$\overbar{x}$=$m\_{1}i+A, где m\_{1}=\frac{\sum\_{}^{}(\frac{x-A}{i})∙f}{\sum\_{}^{}f}$

**Средняя гармоническая**. Средняя гармоническая применяется в тех случаях, когда частоты (веса) не приводятся непосредственно, а входят сомножителями в один из имеющихся показателей.

*Пример*. Автомобиль доставил товары в три магазина фирмы, которые удалены от головного предприятия на одинаковое расстояние. Так, до первого магазина, расположенного на шоссейной дороге, автомобиль прошел путь со скоростью 50 км/час, до второго, по проселочной дороге, — 40 км/час, а в третьем случае автомобилю пришлось полпути пройти через лесной массив, и скорость движения составила только 30 км/час.

Требуется определить среднюю скорость движения автомобиля.

На первый взгляд представляется, что средняя скорость движения может быть определена по формуле простой арифметической:

$\overbar{Va}$*=* $\frac{50+40+30}{3}=\frac{120}{3}=40 км/ч$

В подобных случаях нужно применить формулу средней гармонической простой (не взвешенной):



или в сокращенном виде

где *XГ—* средняя гармоническая;

1 1 1 1

—; —; —... — — числа, обратные заданным вариантам.

*х х1 х2 хп*

**Простая гармоническая средняя есть отношение числа вариант к сумме обратных значений этих вариант.**

Для нашего примера будем иметь:



В нашем примере средняя арифметическая *(*$\overbar{V}а$*)* оказалась больше средней гармонической VГ, при этом абсолютная ошибка завышения составляет 2 км/час(38 – 40), а относительная — 5%

( 2 ∙100 / 40)

**Средняя геометрическая.** Средняя геометрическая применяется в тех случаях, когда индивидуальные значения признака представляют собой, как правило, относительные величины динамики, полученные как произведение цепных коэффициентов роста. Средняя геометрическая рассчитывается как корень степени n из произведений вариантов признака x:

$\overbar{x}$ = = 

где n – число вариант; Π – знак произведения.

**Средняя хронологическая—** это средний уровень ряда динамики т.е. средняя, исчисленная по совокупности значений показателя в разные моменты или периоды времени.В зависимости от вида ряда динамики применяются различные способы ее расчета, а именно расчет: средней хронологической интервального ряда; средней хронологической моментного ряда.

Средней хронологической интервального ряда является средняя величина из уровней интервального ряда динамикии исчисляется по формуле:

$\overbar{у}$ = 

 где $\overbar{у}$— средний уровень ряда;

 *у —* уровень ряда динамики;

 *п —* число членов ряда.

Средней хронологической моментного ряда является средняя величина из уровней моментного ряда динамики.

При равных промежутках времени меж-
ду датами, на которые имеются данные, и равномерном изменении размера показателя между датами средняя хронологическая моментного ряда обычно исчисляется по формуле:

y =

n — число всех членов ряда;

*у -—* средний уровень.

 Если периоды времени, отделяющие одну дату от другой, не равны между собой, то расчет средней хронологической моментного ряда производится по формуле средней взвешенной арифметической, в качестве весов которой принимаются отрезки времени между датами, т.е. по формуле:

 =

где *Т —* время, в течение которого данный уровень ряда *(у)* оставался без изменения.

Например, По состоянию на 01.01.2012 года численность работников локомотивного депо составила 1220 человек. 5 января на работу было принято 2 человека, 16 января – уволено 3 человека, 29 января - принято 4 человека. Среднесписочная численность работников за январь месяц составит 1221 человек.



**Структурные средние.** Особым видом средних величин являются *структурные средние*. Они применяются для изучения внутреннего строения и структуры рядов распределения значений признака. К таким показателям относятся мода и медиана.

*Мода* (М0)– это есть варианта, у которой частота (вес) в ряду распределения наибольшая.

Вычисление моды для интервальных рядов с равными интервалами производится по формуле:

М0 = х0 + ri,

где хо – начало модального интервала; ri – величина интервала;

m2 – частота модального интервала;

m1 – частота интервала, предшествующего модальному;

m2 – частота интервала, следующего за модальным.

*Медиана*  (Ме) – варианта, находящаяся в середине ранжированного (упорядоченного) ряда распределения. Для ее определения достаточно расположить в порядке возрастания или убывания все варианты. Серединная варианта и будет являться медианой. Медиана делит ряд на две равные части – со значениями признака меньше медианы и со значениями признака больше медианы. Если число значений в распределённом ряду чётное, то медиана равна средней из двух вариант, находящихся в середине ряда.

Расчет медианы интервального ряда производится по формуле:

М = хМе + ri

где хМе – начало (нижняя граница) медианного интервала;

ri – величина интервала;

- сумма накопленных частот ряда;

mn – накопленная частота вариант, предшествующих медианному;

mMe – частота медианного интервала.

Соотношение моды, медианы и средней арифметической указывает на характер распределения признака в совокупности, позволяет его асимметрию Если М0 < Ме<, то имеет место правосторонняя асимметрия, при < Ме< М0 следует сделать вывод о левосторонней асимметрии ряда.

 *Вариация* – это различие в значениях какого-либо признака у разных единиц данной совокупности в один и тот же период или момент времени.

 К показателям вариации относятся: размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации.

 *Размах вариации* R – представляет собой разность между максимальным xmax и минимальным xmin значениями признака.

$$R = x\_{max} - x\_{min}.$$

 *Среднее линейное отклонение:*

для не сгруппированных значений

$$\overbar{d}= \frac{\sum\_{}^{}\left|x\_{i}-\overbar{x\_{}}\right|}{n};$$

где n – число членов ряда;

для сгруппированных значений,

$$\overbar{d}= \frac{\sum\_{}^{}\left|x\_{i}-\overbar{x\_{}}\right|f\_{i}}{\sum\_{}^{}f\_{i}};$$

где Σf – сумма частот вариационного ряда.

*Дисперсия* признака представляет собой средний квадрат отклонений вариантов от их средней величины, она вычисляется по формулам простой и взвешенной дисперсий ( в зависимости от исходных данных):

*простая дисперсия*

для не сгруппированных значений $σ^{2}= \frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}}{n};$

для сгруппированных значений $σ^{2}= \frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}};$

Расчёты по этим формулам можно упростить, если использовать **свойства дисперсии.** *Первое* – если все значения признака уменьшить или увеличить на одну и ту же величину А , то дисперсия от этого не измениться;

*Второе* – если все значения признака уменьшить или увеличить в одно и то же число раз ( i ) , то дисперсия соответственно уменьшится или увеличится в i2 раз.

 Используя второе свойство дисперсии, разделив все варианты на величину интервала, получим следующую формулу вычисления дисперсии в вариационных рядах с равными интервалами по способу моментов:

$σ^{2}=i^{2}(m\_{2}-m\_{1}^{2})$*,*

где i - величина интервала;

момент второго порядка;

$$m\_{2}= \frac{\sum\_{}^{}\left(\frac{\overbar{x\_{i}}-C }{i}\right)^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}},$$

квадрат момента первого порядка.

$$m\_{1}= \frac{\sum\_{}^{}\left(\frac{\overbar{x\_{i}}-C }{i}\right)^{}∙f\_{i}}{\sum\_{}^{}f\_{i}}$$

 *Среднее квадратическое отклонение*:

для несгруппированных значений $ σ=\sqrt{\frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}}{n}};$

для сгруппированных значений $ σ=\sqrt{\frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}}};$

 Среднее квадратическое отклонение – это обобщающая характеристика размеров вариации признака в совокупности; оно показывает на сколько в среднем отклоняются конкретные варианты от их среднего значения. Среднее квадратическое отклонение является абсолютной мерой колеблемости признака и выражается в тех же единицах, что и варианты.

 **Дисперсия альтернативного признака**. Среди варьирующих признаков встречаются такие, вариация которых проявляется в том, что у одних единиц совокупности они встречаются , а у других нет. Обозначим: 1 – наличие интересующего нас признака; 0 – его отсутствие; p – доля единиц, обладающих данным признаком; q - доля единиц, не обладающих данным признаком;

p + q = 1.Отсюда q = 1 – p.

 Исчислим среднее значение альтернативного признака и его дисперсию.

среднее значение альтернативного признака$\overbar{x}= \frac{1∙p+0∙q}{p+q}=p,$

дисперсия альтернативного признака$σ\_{p}^{2}$=$\frac{\left(1-p\right)^{2}p+\left(0-p\right)^{2}q}{p+q} =q^{2}p+p^{2}q=pq\left(q+p\right)=pq,$ Чем меньше значение дисперсии и среднего квадратического отклонения, тем однороднее совокупность и тем более типичной будет средняя величина.

 *Коэффициент вариации* представляет собой выраженное в процентах отношение среднего квадратического отклонения к средней арифметической:

$$V= \frac{σ}{\overbar{x}}∙100$$

 Коэффициент вариации используют не только для сравнительной оценки вариации единиц совокупности, но и как характеристику однородности совокупности. Совокупность считается количественно однородной, если коэффициент вариации не превышает 33%.

 **Правило сложения дисперсий.** Вариация признака обусловлена различными факторами, некоторые из этих факторов можно выделить, если статистическую совокупность разбить на группы по какому-либо признаку. Тогда, наряду с изучением вариации признака по всей совокупности в целом, становится возможным изучить вариацию для каждой из составляющих её групп, а также и между этими группами. Если совокупность расчленена на группы по одному фактору, изучение вариации осуществляется посредством исчисления и анализа трёх видов дисперсий: *общей, межгрупповой и внутригрупповой.*

 *Общая дисперсия* σ2  измеряет вариацию признака по всей совокупности под влиянием всех факторов, обусловивших эту вариацию

для несгруппированных значений $σ^{2}= \frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}}{n};$

для сгруппированных значений $σ^{2}= \frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}};$

*Межгрупповая дисперсия* δ2 характеризует систематическую вариацию результативного порядка, обусловленную влиянием признака-фактора, положенного в основание - группировки. Она равна среднему квадрату отклонений групповых (частных) средних xiот общей средней.

$$δ^{2}= \frac{\sum\_{}^{}\left(\overbar{x\_{i}}-\overbar{x\_{}}\right)^{2}}{n}; δ^{2}= \frac{\sum\_{}^{}\left(\overbar{x\_{i}}-\overbar{x\_{}}\right)^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}};$$

 *Внутригрупповая (частная) дисперсия* σ i 2  отражает случайную вариацию, т.е. часть вариации, обусловленную влиянием неучтённых факторов и не зависящую от признака-фактора, положенного в основание группировки. Она равна среднему квадрату отклонений отдельных значений признака внутри группы х от средней арифметической этой группы и может быть исчислена как простая дисперсия или как взвешенная дисперсия по формулам:

$$σ\_{i}^{2}= \frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}}{n}; σ\_{i}^{2}= \frac{\sum\_{}^{}\left(x\_{i}-\overbar{x\_{}}\right)^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}};$$

На величину частных, внутригрупповых дисперсий не влияет групповой признак. Поэтому, чтобы получить представление об общей вариации признака под влиянием случайных причин, следует рассчитать среднюю из внутригрупповых или частных дисперсий, которая будет равна:

$$\overbar{σ^{2}}= \frac{\sum\_{}^{}σ\_{i}^{2}}{n}; \overbar{σ^{2}}= \frac{\sum\_{}^{}σ\_{i}^{2}∙f\_{i}}{\sum\_{}^{}f\_{i}}; $$

Согласно *правилу сложения дисперсий* общая дисперсия равна сумме средней из внутригрупповых и межгрупповых дисперсий:

$$σ^{2}= \overbar{σ^{2}}+δ^{2}.$$

**3. Выборочное наблюдение**

 Наблюдение организуется таким образом, что эта часть отобранных единиц представляет всю совокупность.

 *Вся изучаемая совокупность, из которой производится отбор некоторого числа единиц для выборочного наблюдения*, *называется* **генеральной совокупностью.**

*Часть генеральной совокупности, подлежащая выборочному обследованию, называется* **выборочной совокупностью.**

 Численность (объем) генеральной совокупности обозначим буквой N, а численность выборочной совокупности обозначим буквой n. При выборочном наблюдении обычно ставят две задачи: определение среднего размера изучаемого признака и определение доли изучаемого признака в данной совокупности. При статистических исследованиях (сплошных и выборочных) возникают ошибки двух видов: регистрации и репрезентативности.

 *Ошибки регистрации* могут иметь *случайный* (непреднамеренный) и *систематический* (тенденциозный) характер. *Случайные ошибки* обычно уравновешивают друг друга, поскольку не имеют преимущественного направления в сторону преувеличения или преуменьшения значения изучаемого показателя. Систематические ошибки направлены в одну сторону вследствие преднамеренного нарушения правил отбора. Их можно избежать при правильной организации и проведении наблюдения.

 *Ошибки репрезентативности* присущи только выборочному наблюдению и возникают в силу того, что выборочная совокупность не полностью воспроизводит генеральную. Они представляют собой расхождение между значениями показателей, полученных при выборке, и значениями показателей этих же величин, которые были бы получены при проведённом с одинаковой степенью точности сплошном наблюдении, т.е. между величинами выборочных и соответствующих генеральных показателей.

 Исчисленные обобщающие характеристики в генеральной совокупности называются генеральными: $\overbar{x }$- генеральная средняя, *а*- генеральное среднее квадратическое отклонение, Р – генеральная доля, полученная как отношение числа М единиц, обладающих данным признаком, ко всей численности N генеральной совокупности,



 Исчисленные обобщающие характеристики в выборочной совокупности называются выборочными: - выборочная средняя,  - выборочное среднее квадратическое отклонение, w – выборочная доля (частость) – отношение числа m единиц выборочной совокупности, обладающих данным признаком, ко всей численности n выборочной совокупности, т.е.



**Пример.** В табл. 5 приведены данные испытания крепости нити. Требуется определить среднее квадратическое отклонение крепости нити выборочной совокупности от ее генеральной.

Таблица 5 - **Данные испытания 100 одиночных нитей на крепость**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Крепостьнити, вграммах | Числопроб | Средняя крепость нити в граммах | xf | х-х0 |  |  |
| А | 1 | 2 | 3 | 4 | 5 | 6 |
|  80-100100-120120-140140-160160-180180-200200-220220-240240-260260-280280-300 | 358101826128532 | 90110130150170190210230250270290 | 2705501040150030604940252018401250810580 | -100-80-60-40-20020406080100 | -5-4-3-2-1012345 | 758072401801232454850 |
| Итого |  100 |  - |  18360 |  -  |  | 472 |

К – число, кратное величине интервала.

 Средняя выборочная (средняя крепость нити выборочной совокупности) равна 183,6 грамма



Среднее квадратическое отклонение равно 43,0 граммам:



 Поскольку проведение сплошного наблюдения заменяется выборочным, а исчисление средней генеральной заменяется исчислением средней выборочной, важно установить, насколько полученная средняя выборочная является характерной для данной генеральной совокупности и представляет ли она среднюю генеральную , т.е.

 Чем меньше величина отклонения , тем точнее выборочная средняя воспроизводит генеральную среднюю. Величина этого отклонения и определяет степень точности выборочного наблюдения.

*Ошибки выборочного наблюдения*, которые иначе называют ошибками *репрезентативности*, возникают вследствие специфики самого метода и именно потому, что обследуется не вся совокупность, а лишь его часть, отобранная в случайном порядке.

 Определение средней величины этих ошибок и их возможных границ, а следовательно, определение достоверности данных выборочного наблюдения, является основной задачей теории выборочного исследования.

 Средняя величина ошибки  для выборочной совокупности, отобранной в случайном порядке



где - средняя ошибка выборки;

 - среднее квадратическое отклонение;

n – численность выборочной совокупности.

 Применяя эту формулу, получим следующую величину средней ошибки для нашего примера



 Величина средней ошибки выборки зависит, прежде всего, от показателей колеблемости значений признаков в выборочной совокупности. Степень колеблемости значений признаков определяется средним квадратическим отклонением . Чем меньше величина среднего квадратического отклонения, тем меньше величина средней ошибки при той же численности выборки.

 Величина средней ошибки зависит от численности выборки. Увеличивая или уменьшая объем выборки n, можно регулировать величину ошибки . Чем больше единиц будет охвачено выборочным наблюдением, тем меньше будет величина ошибки, так как тем точнее будет представлена генеральная совокупность. Полученная величина ошибки  характеризует среднее отклонение средней выборочной от средней генеральной.

 *Величина пределов конкретной ошибки зависит от степени вероятности, с которой измеряется ошибка выборки.*

 Ошибка выборки, исчисленная с заданной степенью вероятности, представляет предельную ошибку выборки.

 Если через  обозначим предельную ошибку, частное от деления  на  приравняем к t, тогда можно записать , отсюда  а так как  то 

 Величина предельной ошибки зависит от величины средней ошибки и коэффициента t. Коэффициент зависит от степени вероятности, с которой производится выборочное наблюдение. При вероятности 0,683 значении t равно 1. При вероятности 0,954 значении t равно 2. При вероятности 0,997 значении t равно 3.

 Величину вероятности для различных значений t можно определить на основе теоремы Лапласа. На практике пользуются готовыми таблицами значений этой функции, вычисленных для различных значений t. С увеличением значения t вероятность Р быстро приближается к единице, так что практически обычно ограничиваются значениями t, не превышающими 2-3 единицы.

 Уже при значении t, равном 3, вероятность очень близка к единице. Это означает, что если бы из одной и той же генеральной совокупности было произведено большое число случайных выборок одинаковой численности, то в среднем на 1000 выборок приходилось бы 997 таких, в которых отклонение выборочной средней от генеральной не превышало 3 , и только в трех выборках отклонение могло бы выйти за эти пределы.

 Указывая вероятные пределы случайной ошибки выборки, мы тем самым указываем и те пределы, за которые не выйдет характеристика генеральной совокупности.

 Определим для нашего примера, в каких границах должна заключаться средняя крепость нити в генеральной совокупности, с вероятностью 0,997.

 Средняя ошибка равна  т.е. 

 Предельная ошибка при заданной степени вероятности А равна 3, т.е.



 При проведении выборочного наблюдения часто возникает необходимость предварительного определения *численности выборочной* *совокупности*. Предположим, что мы хотим получить ошибку выборки вдвое меньшую, чем мы получили, т.е. ставим определенные условия: величина должна быть равна 2,15 вместо 4,30. Чтобы добиться уменьшения ошибки вдвое, нужно увеличить число наблюдений. Но на какое количество? Формула средней ошибки выборки позволяет ответить на этот вопрос:

 или  а отсюда  или  т.е. при сокращении ошибки вдвое численность выборки должна быть увеличена в четыре раза, при сокращении второе объем выборки должен быть увеличен в девять раз и т.д. Следовательно, чтобы получить среднюю ошибку выборки, для нашего примера равную 2,15, нужно подвергнуть наблюдению не 100, а 400 нитей.

 Для определения доли, изучаемого признака пользуются формулой средней ошибки выборки, которая имеет следующий вид:



 где Р – доля единиц, обладающих данным признаком в генеральной совокупности. Но этот показатель неизвестен, и его как раз нужно определить на основе выборочного наблюдения. Поэтому величина Р заменяется частостью  :



 Допустим, что нужно установить для нашего примера долю нитей, имеющих крепость 190 граммов и больше. Частость (доля данного признака в выборочной совокупности) равна 0,56 

 Отсюда средняя ошибка для доли:



 При заданной степени вероятности (0,997) предельная ошибка доли равна:



Пределы генеральной доли определяем по формуле:

Р = 

Отсюда Р = 0,56 0,14868.

 **Виды отбора единиц в выборочную совокупность** По виду различают индивидуальный, групповой и комбинированный отбор. При *индивидуальном отборе* в выборочную совокупность отбираются отдельные единицы генеральной совокупности; при *групповом отборе* – качественно однородные группы или серии изучаемых единиц; *комбинированный отбор* предполагает сочетание первого и второго видов.

 **Случайный отбор.** Единицы для наблюдения отбираются из всей совокупности. При этом каждая единица генеральной совокупности имеет равную возможность попасть в выборочную совокупность.

 По *методу* отбора различают *повторную* и *бесповторную выборки*.

 При *повторной выборке* общая численность единиц генеральной совокупности в процессе выборки остаётся неизменной. Ту или иную единицу, попавшую в выборку, после регистрации снова возвращают в генеральную совокупность, и она сохраняет равную возможность со всеми прочими единицами при повторном отборе единиц вновь попасть в выборку. Такая выборка встречается редко. Чаще пользуются бесповторной выборкой.

*Бесповторным* называется такой отбор, когда отобранная единица не возвращается обратно в генеральную совокупность. Следовательно, численность генеральной совокупности с каждой отобранной единице сокращается.

 Способом отбора определяется конкретный механизм или процедура выборки единиц из генеральной совокупности. По степени охвата единиц совокупности различают *большие* и *малые* (менее 30 наблюдений) выборки,

наибольшее распространение получили следующие виды выборки: собственно-случайная, механическая, типическая, серийная, комбинированная.

 Собственно-случайный отбор представляет собой отбор единиц из всей генеральной совокупности посредством жеребьёвки или какого-либо иного подобного способа, например, с помощью таблицы случайных чисел.

**Механический отбор.** При механическом отборе также применяется принцип случайного отбора. При этом из генеральной совокупности отбирается определенное число единиц через определенный интервал. При таком способе отбора генеральную совокупность механически разбивают на равные группы, число которых равно численности выборочной совокупности.

Если при случайном отборе возникает лишь возможность попадания в выборку представителей всех тех состояний, которыми характеризуется изучаемый признак общей совокупности, то механический отбор направлен на то, чтобы действительно обеспечить попадание в выборку таких представителей. При механическом отборе средняя ошибка выборки определяется по тем же формулам, как и при повторном случайном отборе.

**Типический отбор** используется в тех случаях, когда все единицы генеральной совокупности можно разбить на несколько типических групп. Типический отбор предполагает выборку единиц из каждой типической группы собственно-случайным или механическим способом.

Отбор единиц в типическую выборку может быть организован либо пропорционально объему типических групп, либо пропорционально внутригрупповой дифференциации признака.

Отобранное по каждой группе количество единиц является частной выборочной совокупностью ni. Для каждой такой выборочной совокупности можно установить средний размер изучаемого признака и среднее квадратическое отклонение , которое характеризует внутригрупповую колеблемость признака в пределах своей группы. Этот показатель можно обобщить для всей совокупности в целом, т.е. найти показатель внутригрупповой колеблемости признака для всех групп совокупности, вместе взятых, .

Чтобы получить общую выборочную среднюю для всех обследованных групп$\overbar{ x}$, надо из частных выборочных средних$\overbar{x}$, вывести среднюю арифметическую взвешенную, причем в качестве весов можно взять или общую численность каждой группы, или численность выборки в каждой группе. Результат будет одинаковым, так как количество обследуемых единиц распределяется по группам пропорционально их удельному весу в общей совокупности.

Ошибка выборки при типическом отборе определяется по той же формуле, что и при случайном отборе, однако вместо общей дисперсии признака  в этой формуле участвует средняя дисперсия из внутригрупповых .

Чтобы получить общую выборочную среднюю для всех обследованных групп , вывести среднюю арифметическую взвешенную, причем в качестве весов можно взять или общую численность каждой группы, или численность выборки в каждой группе. Результат будет одинаковым, так как количество обследуемых единиц распределяется по группам пропорционально их удельному весу в общей совокупности.

**Серийная выборка.** Иногда в практике выборочного наблюдения производят отбор целых групп единиц и внутри отобранных групп подвергают наблюдению все единицы без исключения. Для отбора серий применяют либо случайную выборку, либо механический отбор. Такая выборка называется *серийной.*

Серийный отбор имеет большое практическое значение, так как легче организовать отбор и изучение нескольких серий единиц, чем сотен отдельных единиц. Но серийный отбор оказывается менее точным в смысле репрезентативности изучаемых показателей, чем другие способы отбора.

Средняя ошибка серийной выборки исчисляется по формулам:

При *повторном отборе* серий:

для средней ; для доли .

При *бесповторном отборе* серий:

для средней  для доли 

где S2 – межсерийный квадрат отклонений;

R - число серий в генеральной совокупности;

S - число отобранных серий;

S- доля данного признака в среднем по всем обследованным сериям.

 Необходимая численность выборки (n) определяется на основе формулы предельной ошибки выборки. Так, если выборка повторная, то n при собственно-случайном и механическом отборах определяется из формулы

 ,

где t – коэффициент доверия, вычисляемый по таблицам в зависимости от вероятности.

 Чтобы найти n, возведем обе части уравнения в квадрат и получим:

, откуда , а 

Объем выработки при исчислении доли определяется по этой же формуле, только вместо  берется , т.е. .

 При бесповторном отборе численность выборки определяется из формулы

.

 Возведя обе части уравнения в квадрат, получим:

****

После преобразований имеем:



Отсюда 

 Аналогично исчисляется объем выборки и при определении доли, только вместо  берется 

 *Пример*. Объём генеральной совокупности составляет 2500 единиц. Требуется определить необходимый объем собственно-случайной выборки для повторного и бесповторного отборов для того, чтобы с вероятностью 0,954 предельная ошибка выборки при определении средней не превышало 20 единиц при среднем квадратическом отклонении 300 единиц.

 По условию задачи N=2500, t=2, =20 и =300.

 Для бесповторного отбора:



 Таким образом, чтобы с вероятностью 0,954 получить предельную ошибку выборки не более 20 ед. при среднем квадратическом отклонении 300 ед., необходимо отобрать из 2500 ед. при повторном отборе 900, а при бесповторном – 662. Как видно, при прочих равных условиях объем выборки при бесповторном отборе меньше, чем при повторном.

 Так же исчисляется объем выборки и при определении доли. Определим по данным нашего примера, сколько нужно отобрать единиц для выборочного наблюдения, чтобы ошибка доли с вероятностью 0,954 не превышала 3% ( =0,03) при удельном весе рассматриваемых единиц в выборке, равном 80% (=0,8), N=2500.

 Для повторного отбора:



 Для бесповторного отбора:



 При определении необходимой численности выборки d2 и p(w) генеральной и выборочной совокупностей неизвестны, причем  и w выборочной совокупности могут быть получены в результате проведения выборочного наблюдения. А без них нельзя установить необходимую численность выборки. В таких случаях фактическое значение дисперсии заменяют приближенным, полученным в результате проведения аналогичного выборочного наблюдения или пробного для ориентировочного суждения о ее размерах. Если признак альтернативный, то исходят из того, что ω = 0,5, а произведение $ω(1- ω) = 0,5∙ 0,5 = 0,25$. Вообще, при определении выборочных данных для вычисления необходимой численности выборки исходят из максимально возможных значений.

 Рассмотрим другой пример выборки при исчислении выборочной доли для бесповторного отбора.

**Пример.**Предполагается, что партия деталей содержит 8% бракованных. Необходимо определить нужный объем выборки, чтобы с вероятностью 0,954 можно установить долю брака с погрешностью не более 2%. Исследуемая партия 5 тыс. деталей.

Объем выборки при исчислении выборочной доли для бесповторного отбора определяется по формуле:



 По условию задачи t=2, доля бракованных деталей =0,08, . Предельная ошибка доли по условию равна =0,02, а N=5000.

 Подставляем эти данные в формулу:



 Чтобы с вероятностью 0,954 можно было утверждать, что предельная ошибка доли брака не превышает 2% необходимо из 5000 деталей отобрать 642.

 Распространение выборочных результатов на генеральную совокупность:

 *способ прямого пересчета и способ коэффициентов.*

*Способ прямого пересчета состоит в том, что средняя величина признака, найденная посредством выборки, умножается на число единиц генеральной совокупности.*

 Например, необходимо определить средний процент брака в партии деталей, состоящей из 10000 деталей. Для выборочного наблюдения в случайном порядке было отобрано 900 деталей. Анализ качества отобранных деталей показал, что средний процент брака в данной совокупности составил 1,5%. Среднее квадратическое отклонение равно 0,3%. Максимальная ошибка выборочного наблюдения с вероятностью 0,997 равна 0,3%.

 Таким образом, средний процент брака в генеральной совокупности находится в пределах 1,5%0,3%, т.е. колеблется от 1,2% до 1,8%.

 Имея данные об общей величине партии, определяем общее количество бракованных деталей, которое будет колебаться в пределах 1,8-1,2% от 10000, или 180-120 единиц. Можно пределы не указывать, а пользоваться средней выборочной как генеральной средней. Тогда среднее количество бракованных банок в генеральной совокупности составит 1,5% от 10000, т.е. 150 единиц.

 Второй способ, или *способ коэффициентов*, применяется тогда, когда выборочное обследование проводится в целях проверки данных сплошного наблюдения.

 Сущность этого метода заключается в том, что на основании сопоставления данных сплошного и данных выборочного наблюдений устанавливают процент расхождений, который и служит коэффициентом поправки, налагаемой на данные сплошного наблюдения.