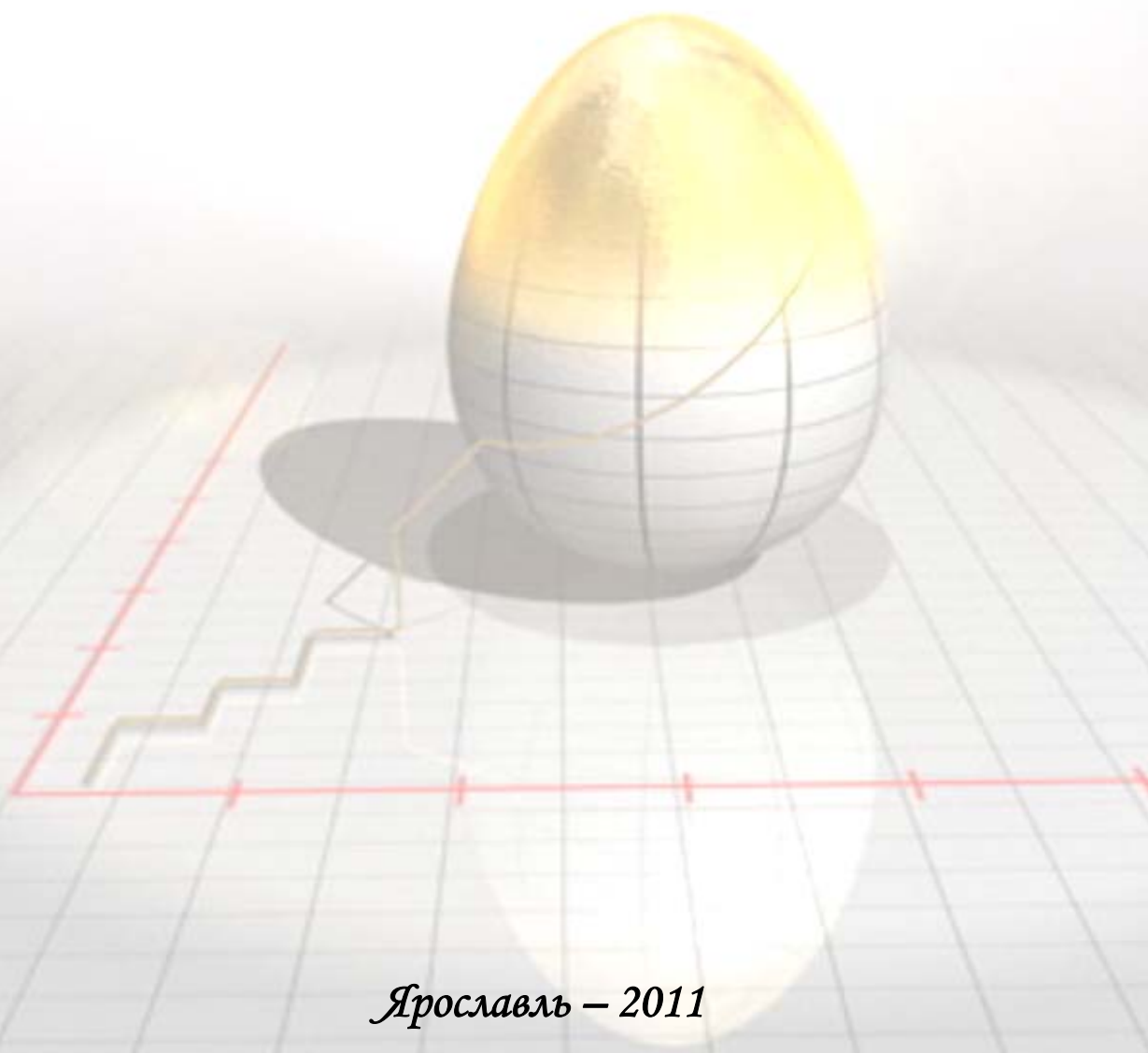


*Е.М. Спиридонова*

---

*Теоретические основы*  
*образовательной статистики*



*Ярославль – 2011*

Автор:

***Спиридонова Елена Михайловна,***

доктор экономических наук, доцент.

Рецензенты:

***Завьялов Федор Николаевич,***

доктор экономических наук, профессор,  
заведующий кафедрой мировой экономики и статистики  
Ярославского государственного университета им. П.Г. Демидова

***Потехин Николай Владимирович,***

руководитель ЦИТ ИРО.

**Спиридонова Е.М.** Теоретические основы образовательной статистики. /  
Институт развития образования. Ярославль, 2011. – 50 с.

Учебное пособие предназначено для слушателей курсов повышения квалификации соответствующей тематики и для работников образовательных учреждений, занимающихся сбором и анализом информации.

В пособии изложены основные понятия статистики, описана методология статистического исследования, даны теоретические основы и методики расчета важнейших статистических показателей, таких как показатели структуры, динамики, средние величины, показатели вариации и т.п.; приведены примеры, связанные с деятельностью образовательных учреждений.

© *Е.М. Спиридонова*

## Содержание:

---

<b>ЛЕКЦИЯ 1. ОСНОВНЫЕ ПОНЯТИЯ СТАТИСТИКИ.....</b>	<b>4</b>
1.1. Понятие и развитие статистики. ....	4
1.2. Определение науки статистики.....	6
1.3. Статистические категории и методология.....	8
<b>ЛЕКЦИЯ 2. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ. ....</b>	<b>12</b>
2.1. Понятие статистического наблюдения и его виды.....	12
2.2. Программа и этапы статистического наблюдения.....	15
2.3. Контроль данных наблюдения. ....	19
<b>ЛЕКЦИЯ 3. ПЕРВИЧНАЯ ОБРАБОТКА МАТЕРИАЛОВ НАБЛЮДЕНИЯ. ....</b>	<b>22</b>
3.1. Сводка и ее виды. ....	22
3.2. Группировка и ее виды. ....	23
3.3. Основные принципы построения группировок.....	26
3.4. Ряды распределения, их виды и графическое изображение.....	28
3.5. Статистические таблицы и правила их оформления. ....	31
<b>ЛЕКЦИЯ 4. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ. ....</b>	<b>33</b>
4.1. Абсолютные величины и их виды. ....	33
4.2. Относительные величины и их виды.....	35
<b>ЛЕКЦИЯ 5. СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАЦИИ.....</b>	<b>41</b>
5.1. Средние величины и их виды. ....	41
5.2. Средние степенные.....	42
5.3. Средние порядковые.....	45
5.4. Показатели вариации. ....	46
<b>РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА.....</b>	<b>50</b>

# Лекция 1.

## Основные понятия статистики.

---

### 1.1. Понятие и развитие статистики.

В широком смысле **термин «статистика»** обычно трактуется как:

- сфера практической деятельности, направленная на сбор, обработку и публикацию данных (статистические органы);
- или отрасль науки (и соответствующая учебная дисциплина).

В узком смысле этот термин также употребляется как:

- совокупность цифровых данных, характеризующих явление или объект;
- или в качестве названия конкретного показателя (F-статистика, t-статистика, статистика Дарбина-Уотсона и др.)

Термин «статистика» происходит от группы латино-итальянских слов:

- status – состояние, положение вещей,
- stato, statista – государство, его знаток.

#### Краткая историческая справка.

Начало статистической практики относится примерно к тому времени, когда возникло государство. Имеются сведения об элементарном счете населения и земель, проводившемся несколько тысячелетий назад: в Китае – в XXII в до н.э., в Древнем Египте, в Персии, в Империи инков. В Древнем Риме с целью налогообложения проводились имущественные цензы (учет земель) и переписи населения. Греческий философ Аристотель (384-322 гг. до н.э.) составил описания 157 городов и государств своего времени. В эпоху феодализма первые упоминания статистического учета относятся к концу IX в.: известно о проведенной Карлом Великим инвентаризации королевских имений и переписи населения, способного носить оружие. Средневековье оставило также уникальный памятник истории – «Книгу страшного суда» (1061 г.) – свод материалов всеобщей переписи населения Англии и его имущества (240 тыс. дворов). С образованием централизованных государств и особенно – в эпоху развития капитализма – объем применения статистического учета значительно расширился. Переписи населения начали проводиться регулярно. Возникали элементарные формы статистического учета и в других областях общественной жизни.

**Как наука** статистика появилась значительно позднее.

Ее истоки находятся в так называемой «политической арифметике», созданной в конце XVII в. благодаря трудам англичан У. Петти (1623-1687), который был, по определению К. Маркса, «... в некотором роде изобретателем статистики...», и Дж. Граунта (1620-1674), впервые отметившего закономерности в движении населения. В то время статистика еще не отделилась от политической экономии и других социально-экономических дисциплин.

Другой исторической дисциплиной, предшествовавшей современной статистике, было «государствоведение», которое начало формироваться одновременно с политической арифметикой в трудах немецкого ученого Г. Конринга (1606-1681) и получило особое развитие в Германии в XVIII в., а несколько позднее и в России. В рамках государствоведения статистика рассматривалась как одно целое вместе с географией, этнографией, юридическими сведениями и т.д.

**Термин «статистика»** в научный оборот в 1746 году ввел немецкий ученый Г. Ахенваль (1719-1772), предложив заменить название курса «Государствоведение», преподававшегося в университетах Германии, на «Статистику», и положив тем самым начало развитию статистики как науки и учебной дисциплины.

Со временем - в XIX в. - описательное направление постепенно утрачивает свое значение, но его заслугой является создание табличных и графических методов исследования, а также – собственно определение науки статистики. В этот период развитие статистики все больше определяется прогрессом математики, и в первую очередь – Теории вероятностей.

К середине XIX в. бельгийский статистик А. Кетле (1796-1874) и его последователи доказали наличие закономерностей в статистических рядах. Кетле принадлежит заслуга систематического использования математических методов в обработке статистических данных. Кроме того, активно занимаясь статистической практикой, причем, не только на родине, в Бельгии, но и в масштабах всей Европы, Кетле выработал правила переписей населения и обосновал регулярность их проведения в развитых странах. По его инициативе проводились международные статистические конгрессы, а в 1885 г. был основан Международный статистический институт, существующий и поныне как профессиональная ассоциация статистиков, имеющая консультативный статус при ЭКОСОС и ЮНЕСКО.

Во второй половине XIX в. и начале XX в. происходило интенсивное развитие статистики. Этому способствовало проведение различного рода периодических переписей и обследований, которые собирали богатейший материал о различных объектах обследования (предприятиях, хозяйствах, отдельных людях). В это же время совершенствовались органы государственной статистики, в первую очередь те, которые осуществляли переписи. Тогда же начала формироваться и специальная научная дисциплина - «математическая статистика», являющаяся, по сути, частью математики.

Таким образом, можно констатировать, что у истоков статистической науки стояли две школы: немецкая, отдающая предпочтение описательным методам, и английская школа «политических арифметиков».

Представители описательной школы (Г. Конринг, Г. Ахенваль, А. Л. Шленцер) своей задачей считали описание достопримечательностей государства: территории, населения, климата, политического устройства, вероисповедания, торговли и т.п. – без анализа закономерностей и связей между явлениями.

Представители школы политических арифметиков (У. Петти, Дж. Граунт, Э. Галлей) своей главной задачей считали выявление на основе большого числа наблюдений различных закономерностей и взаимосвязей в изучаемых явлениях.

Каждая школа развивалась своим путем, используя свои методы в исследованиях, но предмет изучения у них был общий – государство, общество и, в частности, массовые явления и процессы, происходящие в нем. Статистика сформировалась как наука в результате синтеза государствоведения и политической арифметики, причем от последней она взяла больше, поскольку одна из задач статистики в настоящее время - выявление различного рода закономерностей и тенденций развития.

Однако представители этих двух школ не дошли до теоретического обобщения практики учетно-статистических работ, до создания **теории статистики**. Эта задача была решена А. Кетле (и его последователями), который дал определение предмета статистики, раскрыл суть ее методов. Под влиянием его идей и возникло третье направление статистической науки – математико-статистическое, которое в дальнейшем получило свое развитие в работах таких известных ученых как: англичане К. Пирсон, В. Госсет, Р. Фишер, М. Дж. Кендэл, Ф. Гальтон, Ф. Эджворт, итальянец К. Джини, русские – П.Л. Чебышев, А.А. Марков, А.М. Ляпунов, С.Н. Бернштейн, А.Я. Хинчин, А.Н. Колмогоров, В.И. Романовский, Ю. Янсон, В.Я. Буняковский, А.И. Чупров, А.А. Чупров и др.

Отметим, что и в настоящее время статистика продолжает активно развиваться, особенно в части методологии регрессионного, динамического и многомерного статистического анализа, создаются «новые» научные дисциплины на стыке статистики и других наук (например – эконометрика). Это обусловлено ростом практических потребностей в статистическом анализе в связи с развитием математики, физики, экономики, других социальных и естественных наук, а главное – стремительным прогрессом возможностей вычислительной техники и средств связи.

---

## **1.2. Определение науки статистики.**

**Статистика** – общественная наука, количественно отображающая качества и свойства массовых процессов и явлений в конкретных условиях места и времени.

В общем случае, объектом изучения статистики являются массовые явления и процессы любой природы, в том числе – общественные и социально-экономические. Предметом статистики выступает количественная сторона (размеры, соотношения, закономерности, взаимосвязи, тенденции и т.п.) изучаемых явлений и совокупностей объектов.

Отметим, что существуют и другие точки зрения на предмет статистики и, в связи с этим, – различные ее определения. Так, например, есть мнение, что статистика не имеет собственного предмета изучения, это наука о методе.

Статистика представляет собой целый комплекс научных дисциплин:

- **Теория статистики** занимается определением понятийного аппарата и категорий статистической науки, выработкой общих принципов проведения исследования и правил сбора данных, разработкой универсальных методов обработки и анализа информации, т.е. – общей методологией статистического исследования массовых процессов и явлений.

- **Экономическая статистика** занимается исследованием и количественной оценкой экономических явлений и процессов, разрабатывает специальные обобщающие экономические показатели (производства продукции, рабочей силы и рабочего времени и т.п.).

Внутри Экономической статистики часто выделяют микро- и макроуровень. В первом случае речь идет о расчете показателей отдельных предприятий и организаций, а во втором – о макроэкономических показателях государства в целом (таких как ВВП, ВНП, НД, темпы экономического роста и др.), в т.ч. и для международных сопоставлений. Содержание этих понятий и совершенствование методик их расчета – предмет изучения специальной дисциплины – *макроэкономической статистики*.

- *Социальная статистика* занимается исследованиями и разработкой обобщающих показателей в различных областях общественной жизни, таких как показатели уровня и качества жизни, занятости и безработицы, показатели социальной сферы, в т.ч. образования, здравоохранения и др.

При изучении социальную и экономическую статистику часто объединяют в один курс *Социально-экономической статистики*.

- Система *отраслевых статистик* изучает процессы и явления в соответствующих отраслях и сферах социально-экономической и культурной жизни общества. Существуют (и изучаются) отдельные дисциплины:

- Статистика промышленности (по видам экономической деятельности),
- Статистика сельского хозяйства,
- Статистика строительства,
- Статистика финансов,
- Статистика образования и т.п.

Как научная дисциплина, носящая общественный характер, статистика тесно связана с другими науками и сферами человеческой деятельности.

Методологическими основами статистики являются диалектика и экономическая теория. В статистике применяются общефилософские понятия и методы научного познания (индукция, дедукция, анализ, синтез, сравнение, аналогия), диалектические категории (система, структура, развитие и др.), и законы развития (в диалектическом материализме – законы единства и борьбы противоположностей, перехода количественных изменений в качественные, и отрицания отрицания). Экономическую теорию (и другие социальные науки) статистика использует для определения содержания основных социально-экономических категорий, чтобы описывать их с помощью соответствующих статистических показателей.

Статистика широко применяет математический аппарат, т.е. свои методические основы статистика берет в различных математических дисциплинах. В первую очередь это Теория вероятностей и математическая статистика, а также – Линейная алгебра, Дискретная математика, Численные методы. Для решения отдельных задач также используются Методы оптимизации, Линейное программирование, Теория игр и др.

Информационные основы статистики – оперативно-технический и бухгалтерский учет, результаты которых представляются всеми предприятиями и организациями в виде статистической отчетности, а также – данные отдельных исследований, мониторингов, опросов и т.п.

В свою очередь статистика обеспечивает аналитической информацией органы управления различных уровней: от администрации Президента и кабинета Министров до руководителей отдельных предприятий и организаций всех сфер деятельности. Статистическая информация используется и другими науками, такими, как экономическая география, история и др.; основные показатели доступны средствам массовой информации и населению. Статистические методы проведения исследований и расчета обобщающих показателей активно применяются в маркетинге, социологии, психологии, медицине и др.

Таким образом, главная задача статистики, как сферы деятельности, – предоставление полной и достоверной информации обо всех процессах и явлениях, происходящих в социально-экономической системе страны (в т.ч. – и в региональном разрезе).

В общем случае, задачами любого статистического исследования, для решения которых и создавалась наука статистика, являются:

- изучение основных характеристик и общих свойств единиц совокупности,
- определение размеров и внутренней структуры явления,
- отражение достигнутого уровня развития (на момент или дату) и/или произошедших за определенный период изменений,
- выявление имеющихся взаимосвязей и зависимостей,
- описание закономерностей развития и построение прогнозов;

Эти цели достигаются с помощью расчета и анализа соответствующих статистических показателей.

---

### **1.3. Статистические категории и методология.**

При изучении количественной стороны массовых явлений и процессов статистика оперирует специфичными понятиями, относящимися к основным категориям статистической науки: признак, совокупность, показатель.

□ **Признак** – характерное для объекта качество или свойство, которое можно наблюдать и/или измерить.

Например, признаками образовательного учреждения являются его тип, численность учащихся и т.п.

#### Виды признаков.

➤ По степени значимости:

1. **Существенные**, основные – определяют главное содержание явления, изучаются в первую очередь и подробно.
2. **Второстепенные**, несущественные – отражают дополнительные характеристики, рассматриваются как «прочие».



➤ По виду («природе») явления:

1. **Атрибутивные** (качественные). Отдельные значения признака имеют существенные качественные различия. Как правило, выражены словесно. Например, школы: начальные, средние, основные.
2. **Количественные**. Значения различаются по величине. Выражены цифрами. Пример: школы по численности учащихся.

➤ По порядку определения:

1. **Первичные**. Лежат в основе программы сбора первичной информации.
2. **Вторичные**. Получаются в процессе обработки и анализа данных. Например, средняя заработная плата работников учреждения.

Статистика имеет дело с **варьирующими**, т.е. меняющимися в определенных пределах, признаками (изучать «постоянные» характеристики – нет смысла). По количеству групп в них особо выделяют **альтернативные** признаки, которые по своей природе могут быть и качественными и количественными, – это признаки, которые могут принимать одно из двух противоположных значений (например, студенты – «успешно сдавшие сессию» или «имеющие задолженность хотя бы по одному предмету»); для их обработки применяются специальные методы.

□ **Совокупность** – множество объектов, лиц, явлений, имеющих один или несколько общих признаков и различающихся между собой по другим признакам. Например, совокупность общеобразовательных школ города (один и тот же тип учреждений, но могут иметь разную специализацию).

**Единица совокупности** – наименьшая неделимая часть совокупности, обладающая всеми ее основными свойствами.

**Объем совокупности** – количество единиц в ней.

Виды совокупностей.

➤ По объему включаемых единиц:

1. **Бесконечные**. В статистической практике не рассматриваются.
2. **Большие**. Наиболее характерны для описания массовых явлений.
3. **Малые** - объемом менее 20-50 единиц. Требуют особых методов расчета некоторых характеристик.
4. **Единичные**. Встречаются в монографических исследованиях.

➤ По степени однородности:

1. **Однородные** – если вся совокупность находится под воздействием одних и тех же наиболее существенных признаков.
2. **Разнородные** (неоднородные) – в противном случае.

Совокупность может быть однородной по одним признакам и разнородной по другим. Так, например, в целом однородная по признакам «профессия» и «образование» совокупность школьных педагогов окажется разнородной по признакам «возраст» и «стаж».

□ **Статистический показатель** – количественная характеристика изучаемого объекта, процесса, явления в конкретных условиях места и времени.

#### Виды показателей.

➤ По охвату совокупности:

1. **Индивидуальные** – характеризуют отдельную единицу совокупности (например, конкретную школу);
2. **Сводные (общие)** – характеризуют группу или всю совокупность единиц (школы района, области).

➤ По способу определения:

1. **Оценочные.** Характеризуют первичные признаки, т.е. их получают непосредственно в ходе наблюдения, обычно – в виде абсолютных величин (размеры, количество и т.п.).
2. **Аналитические.** Получаются в результате дальнейшего анализа и расчетов, часто – в виде относительных и средних величин.

**Система статистических показателей** – совокупность (определенный «набор») взаимосвязанных показателей, всестороннее отражающих различные аспекты изучаемого явления. В силу разных причин «набор» показателей со временем может меняться.

В статистике Российской Федерации наиболее распространенным является иерархический принцип построения системы показателей:

- по территории (по народному хозяйству) РФ в целом – *макро*-уровень;
- по регионам (по видам экономической деятельности, министерствам, департаментам, ведомствам) – *мезо*-уровень;
- по отдельным населенным пунктам (предприятиям, организациям) – *микро*-уровень.

Применяются и другие принципы построения системы, а степень детализации и количество уровней могут варьироваться.

Для изучения предмета статистики разработаны и применяются специальные приемы, образующие в своей совокупности методологию статистики. В ее основе, как уже отмечалось, лежит диалектический метод познания, согласно которому общественные явления рассматриваются в развитии, взаимной связи и причинной обусловленности, что нашло отражение в конкретных статистических методах.

**Статистическая методология** – система приемов, методов, способов и правил статистического исследования массовых (социально-экономических) явлений, обеспечивающих сбор данных, их обработку и анализ.

Принято выделять три стадии статистического исследования:

1. **Статистическое наблюдение** – планомерный, специально организованный сбор данных об анализируемом объекте, процессе, явлении,

т.е. – регистрация признаков у всех единиц изучаемой совокупности. На этом этапе крайне важно обеспечить полноту и достоверность собранной информации.

2. **Первичная обработка результатов наблюдения.** Осуществляется сводка и группировка данных, строятся ряды распределения и простейшие графики, данные заносятся в таблицы. Т.е. на этом этапе данные систематизируются и группируются, тем самым осуществляется переход от описания отдельных единиц к описанию выделенных групп и объекта исследования в целом.

3. **Расчет и анализ обобщающих показателей.** Применяются разнообразные математико-статистические методы для обобщения и последующего анализа данных: рассчитываются средние значения и показатели вариации, выявляются взаимосвязи, определяется динамика и т.п. На этом этапе важно правильно выбрать показатели (согласно целям анализа) и методику их расчета (учитывая особенности используемых данных). Результаты анализа, как правило, излагают в текстовой форме, также используя табличные формы и графические иллюстрации.

На каждой из вышеописанных стадий статистического исследования применяются соответствующие статистические методы:

1. Метод массовых наблюдений.
2. Метод группировок. Графический и табличный методы.
3. Методы анализа с помощью обобщающих показателей.

В целом, статистическая методология направлена на изучение количественных закономерностей, проявляющихся в структуре, динамике и взаимосвязи явлений; она позволяет исследовать всю совокупность фактов, изобразить процесс в целом, учесть разнообразие форм явления, проанализировать закономерности и тенденции развития, выявить причинно-следственные связи и формы зависимостей. При этом статистические методы достаточно универсальны и могут применяться в исследованиях любого рода.

---

## Лекция 2.

### Статистическое наблюдение.

---

#### 2.1. Понятие статистического наблюдения и его виды.

**Статистическое наблюдение** – специально организованный сбор сведений о входящих в состав изучаемой совокупности единицах.

Необходимо зарегистрировать признаки каждой единицы объекта наблюдения и занести полученные данные в определенные «документы», сформировав тем самым первичную статистическую информацию, которая впоследствии будет систематизирована и должным образом обработана.

К статистическому наблюдению предъявляются следующие требования:

- наблюдение должно иметь научно-практическую ценность;
- собираемая информация должна быть полной и достоверной;
- наблюдение должно проводиться по заранее разработанной программе.

#### Виды статистического наблюдения.

➤ По охвату совокупности выделяют:

1). **Сплошное** наблюдение.

В этом случае обследуются все, без исключения, единицы.

Результаты такого наблюдения всегда точные и достоверные, но его проведение требует больших материальных и трудовых затрат, и, как правило, - весьма продолжительно по времени.

Примеры: перепись населения, отчетность предприятий и организаций; мониторинг образовательных учреждений (в рамках области).

2). Виды несплошного наблюдения:

а). **Метод основного массива.**

Обследуется большая часть единиц, где признак наиболее ярко выражен.

Такое наблюдение, как правило, дает достаточно точные результаты, но зачастую возникает проблема выделения «основного массива».

Примеры: голосование на выборах (лица старше 18 лет); наблюдение за ценами на городских рынках производится в нескольких наиболее крупных городах области (в них проживает более 80% городского населения области), малые города из этого наблюдения исключаются.

б). **Выборочное** наблюдение.

Обследуется некоторое количество или определенная доля единиц, которые отбирают случайным или каким-то специальным образом (например: каждый десятый из общего списка).

Этот вид наблюдения – наиболее оперативный и наименее затратный из всех, но при расчете итоговых показателей, распространяемых на всю совокупность, надо учитывать ошибки репрезентативности (возникающие из-за разницы выборки и всей совокупности единиц). Более того, если выборка окажется неудачной, то рассчитанные на ее основе результирующие показатели могут оказаться весьма далекими от реальных значений.

в). **Монографическое** исследование.

Обследуются одна или несколько наиболее типичных единиц.

Такие наблюдения, как правило, достаточно точные, они интересны и богаты разнообразной информацией. Проблема здесь чаще всего заключается в выявлении «типичных» единиц, обычно для этого требуется провести предварительное исследование.

➤ По времени проведения выделяют:

1). **Текущее** наблюдение.

Это наблюдение проводится постоянно.

Пример: отчетность предприятий и организаций; мониторинг ОУ.

2). **Периодическое** наблюдение.

Проводится через определенные (часто - равные) промежутки времени.

Примеры: перепись населения проводится раз в 10 лет; сбор информации об отчисленных учащихся – каждые полгода (в октябре и апреле).

3). **Единоновременное** наблюдение.

Такие исследования проводятся со специальными целями и могут не повторяться.

Некоторые наблюдения в зависимости от степени их востребованности могут переходить из одной группы в другую: единовременное может стать периодическим, периодическое – текущим, и наоборот.

➤ По источникам получения первичной информации можно выделить следующие способы:

1). **Документальный** – основан на использовании различных документов первичного учета, ведущегося на предприятиях и в организациях.

Самый характерный пример – статистическая отчетность.

2). **Непосредственное наблюдение**. Регистрация производится на основе непосредственного осмотра, подсчета, замеров и т.п. Наиболее точный способ.

Примеры: ревизия, инвентаризация; регистрация цен и т.п.

3). **Опрос**. Основной источник данных – сведения, полученные от опрашиваемых лиц. Опросы могут проводиться разными способами (интервьюирование, анкетирование и др.); степень достоверности полученной информации зависит от конкретного способа проведения опроса.

➤ В целом, способов и форм организации наблюдений достаточно много, но их можно объединить в две принципиально различающиеся группы:

1). **С непосредственным участием регистратора:**

- Непосредственное наблюдение [см. выше].

- **Экспедиционный способ.** Специально обученные регистраторы производят опрос и со слов опрашиваемых заполняют бланки.

Так, например, проводится перепись населения.

- Устный опрос, интервью и т.п.

Такие способы проведения наблюдения дают достаточно точную и достоверную информацию, но требуют больших трудовых и временных затрат.

## 2). **Без участия регистратора:**

- Стат.отчетность [с.м. выше – документальный способ].
- **Корреспондентский способ.** Организации, проводящие наблюдения на регулярной основе, могут создать сеть корреспондентов – из других организаций или специально подобранных лиц, давших согласие периодически заполнять бланки и высылать их обратно в установленные сроки.

Этот способ сбора информации достаточно широко использовался в Советском Союзе при изучении спроса населения и конъюнктуры торговли, и, по нашему мнению, учитывая современный уровень развития и распространения информационно-коммуникационных технологий, имеет хорошие перспективы для его использования в маркетинговых и т.п. исследованиях.

- **Саморегистрация.** Бланки заполняются самими опрашиваемыми. Задача организации, проводящей наблюдение, – раздать бланки опрашиваемым, проинструктировать их, собрать заполненные бланки.

Этот метод используется при бюджетных обследованиях населения.

- Письменный опрос, анкетирование и т.п.

Все эти способы, как правило, малозатратны (необходимы средства только на печать и рассылку бланков или анкет), но при этом часто встречаются ошибки (в т.ч. и умышленные), связанные с неправильным пониманием вопросов или нежеланием отвечать на них.

Кроме того, если речь идет об анкетировании, то может возникнуть проблема малой возвратности анкет, вследствие чего результаты анализа могут оказаться некорректными. Тем не менее, этот метод очень популярен при проведении разного рода социологических опросов, например, среди читателей для определения интереса к публикуемым материалам. В настоящее время все популярнее становятся опросы в сети Интернет. Необходимо заранее продумать, как заинтересовать потенциальных респондентов ответить на вопросы (и вернуть заполненную анкету исследователям), например, разыграть среди ответивших приз, привлекательный для данной целевой аудитории. Другой способ заключается в том, чтобы значительно увеличить число распространяемых анкет; необходимое их количество можно определить, исходя из процента возвратности в аналогичных, ранее проведенных исследованиях.



## 2.2. Программа и этапы статистического наблюдения.

В процессе статистического наблюдения принято выделять три стадии:

### I. Подготовительная стадия.

Как правило, эта стадия – самая продолжительная по времени. На стадии подготовки наблюдения необходимо решить все методологические и организационные вопросы, разработать программу и инструментарий наблюдения.

### II. Непосредственное проведение наблюдения.

Может занимать от одного дня до нескольких недель. На этом этапе производится сбор информации об объекте исследования: проведение опроса, «замер» показателей, сбор заполненных анкет, форм отчетности и т.п.

### III. Контроль данных наблюдения.

На этом этапе проводится предварительный анализ достоверности собранной информации, выявляются и по возможности исправляются ошибки, допущенные респондентами при заполнении анкет и форм отчетности.

Следует более подробно рассмотреть первую, подготовительную, стадию. В общем виде в ней выделяют следующие этапы:

#### 1. Сначала решают **методологические вопросы** исследования:

а). Определяют цели и задачи исследования.

б). В соответствии с целями выделяют субъект, объект и предмет исследования.

Пример: Департамент образования области проводит мониторинг материально-технической базы образовательных учреждений (ОУ) всех видов (кроме финансируемых из федерального бюджета). Здесь: субъект – Департамент образования области, объект – образовательные учреждения области (вся их совокупность, кроме федеральных), предмет – состояние материально-технической базы ОУ.

в). Формируется исследуемая совокупность, определяются ее границы и объем. При необходимости вводится ценз (ограничительные признаки).

Например: исследуются результаты ЕГЭ по школам. Совокупность – множество всех школ области, единица совокупности – одна, отдельно взятая, школа; объем совокупности – общее количество школ в области. Цензом при этом может быть наличие выпускных классов в соответствующем году.

г). В соответствии с целями и задачами определяется вид наблюдения.

Т.е. необходимо заранее определить, каким должно быть наблюдение по охвату совокупности (сплошным или выборочным), по периодичности, способу проведения, форме отчетности и т.п. [*Виды наблюдений подробно рассмотрены в предыдущем вопросе.*]

Так, например, проводимый Департаментом образования мониторинг МТБ ОУ Ярославской области является сплошным (в рамках области), периодическим (2 раза в год), корреспондирующим (администрация ОУ самостоятельно заполняет электронные формы отчетности) наблюдением.

д). Вводятся временные параметры наблюдения:

- **критический момент** – момент времени (конкретная дата), на который фиксируется состояние объекта (достигнутый им уровень развития);
- **период наблюдения (период регистрации)** – за этот (отчетный) период происходят определенные события и изменения, которые отражаются в отчетности.

Например, предоставляемые общеобразовательными учреждениями формы федерального статистического наблюдения № ОШ-1 «Сведения об учреждении» и № 83-РИК «Сведения о численности и составе работников учреждения» заполняются по состоянию на 20 сентября (критический момент), а форма № ОШ-2 «Сведения о финансировании и расходах учреждения» – за период с 1 января по 31 декабря, т.е. за календарный год (период наблюдения).

## 2. Разрабатывается программа и инструментарий наблюдения.

а). **Программа наблюдения** – перечень вопросов, на которые должны быть получены ответы от каждой единицы совокупности. Конкретное содержание программы определяется особенностями объекта и целями исследования.

Программа может быть:

- обширной, т.е. затрагивающей большое число вопросов, что позволяет получить разнообразные сведения и более глубоко изучить объект наблюдения;
- узкой, т.е. – с ограниченным кругом вопросов, имеющих, как правило, сугубо практическое значение;
- или – специальной, т.е. ориентированной на учет каких-то специфических особенностей объекта наблюдения.

Программу мониторинга образовательных учреждений Ярославской области, например, можно считать достаточно обширной – в ней содержатся вопросы по материально-технической базе учреждений, по персоналу, контингенту детей и др. Однако, иногда возникает необходимость проведения дополнительных специальных наблюдений для удовлетворения поступивших от Департамента образования или руководства области «новых» запросов.

Отметим, что разработка программы – одна из важнейших проблем исследования, в решении которой должны принимать участие и практики и аналитики; от ее качества зависит ценность собранной информации. В общем случае программа наблюдения должна содержать все основные признаки объекта, а также – те признаки, которые могут понадобиться при проведении дальнейшего, более глубокого анализа. В то же время в программу не следует включать второстепенные признаки, которые могут затруднить обработку,



увеличив объем «ненужных» сведений. Т.о., следует придерживаться принципа «необходимой полноты» собираемой информации. Кроме того, в программу рекомендуется включать «вопросы контрольного характера», позволяющие обеспечить проверку информации на этапе заполнения формуляров.

б). На основании составленной программы разрабатывают основной и вспомогательный инструментарий наблюдения.

□ К **основному инструментарию** относится статистический формуляр и инструкция к его заполнению.

➤ **Статистический формуляр** – особым образом сформированный бланк, в котором отражается перечень вопросов программы наблюдения.

Бланк формуляра может иметь любой вид: форма отчетности, опросный лист, таблица и т.п., на бумажном носителе или в электронном виде. Бланки могут быть индивидуальными (на каждую единицу совокупности – отдельный бланк) или списочными (для каждой единицы – отдельная строка). Независимо от вида, формуляр – носитель первичной информации.

В общем виде статистический формуляр обязательно содержит:

- Адресно-именную часть:
  - полное название проводимого наблюдения, его место и время;
  - сведения об организации, проводящей исследование;
  - сведения о подотчетной единице (об «источнике» информации).

В формуляре могут также присутствовать подписи ответственных лиц, с указанием их телефонов и режимов работы (если это формы отчетности), или, наоборот, отсутствовать информация о конкретном опрашиваемом (в случае анонимного анкетирования).

- Основную часть:
  - показатели, характеризующие единицу наблюдения,
  - и вопросы, на которые нужно получить ответы.

Вопросы в анкете могут быть сформулированы как:

- закрытые – содержат все возможные варианты ответа, из которых надо выбрать один, соответствующий ситуации;
  - открытые – предполагают ответы в свободной форме;
- Такие вопросы не слишком удобны для дальнейшей обработки, но дают наиболее полную картину мнений респондентов.
- дихотомические – с ответами строго «да» или «нет»;
  - шкальные – где ответы упорядочены согласно какому-то принципу (от большего к меньшему, от худшего к лучшему и т.п.);
  - поливариантные – где можно выбрать сразу несколько вариантов ответа;
  - рейтинговые – ответы (все или несколько) надо упорядочить или проранжировать в порядке предпочтения.

Показатели в формуляре могут быть представлены в виде открытых позиций, куда надо вписать конкретное число (что предпочтительнее), или в виде некой шкалы, где надо выбрать соответствующий интервал.

Общие требования к составлению формуляра:

- ✓ Объем формуляра (количество вопросов) должен быть не слишком большим, чтобы респонденты могли заполнить его за приемлемое время.
- ✓ Отдельные элементы программы выстраиваются в логическом порядке: от простых вопросов к сложным, расчетные показатели после первичных и т.п.
- ✓ Вопросы должны иметь четкую, понятную формулировку и возможность однозначного ответа (или необходимо давать разъяснения и «подсказки»).
- ✓ В анкетах желательно представить наиболее популярные и ожидаемые ответы на вопросы, оставив строку «другое» для отличающихся вариантов, тем самым повышая удобство и скорость их заполнения.
- ✓ Что касается графического оформления, то здесь важно соблюсти баланс между размером формуляра (скажем, одна страница формата А4) и размером используемого шрифта (он должен быть хорошо читаемым). Также необходимо следить за тем, чтобы вопросы и варианты ответов располагались рядом, на одной странице, чтобы оставалось достаточно места для внесения цифр и «своих» ответов и т.п.

➤ **Инструкция** определяет порядок и правила заполнения формуляра. В ней приводятся:

- разъяснения используемых терминов,
- методики расчета показателей,
- примеры расчета в типичных и особо сложных ситуациях,
- указание на связи между показателями
- и возможности их взаимопроверки.

Инструкция может быть оформлена отдельной брошюрой или - в кратком виде – совмещена с формуляром, например, в виде указания формулы расчета показателя или характера соотношений между взаимосвязанными вопросами.

□ **Вспомогательный инструментарий** (в зависимости от конкретного вида наблюдения) может включать:

- образцы форм первичного учета и бланков отчетности,
- перечни переписных участков,
- схемы маршрутов движения регистраторов по участкам,
- списки регистраторов с указанием их режима работы,
- списки организаций-корреспондентов, занимающихся сбором данных, с указанием их адресов, телефонов, графиков работы,
- списки операторов и программистов, отвечающих за внесение информации в базы данных и написание обрабатывающих программ,
- сметную документацию и т.д. и т.п.

3. После того, как решены все программно-методологические вопросы, приступают к решению **организационных вопросов** исследования:

а). Определяется общий порядок проведения наблюдения, приема и сдачи его материалов.

б). Составляется подробный, поэтапный план наблюдения, с указанием сроков и ответственных лиц.

в). Размножаются и рассылаются формуляры (и инструкции).

г). При необходимости проводится набор регистраторов, осуществляется их учеба или инструктаж.

д). В некоторых случаях (например, при переписи населения) проводится агитационно-разъяснительная кампания в средствах массовой информации.

е). Составляются подробные сметы затрат на проведение наблюдения: определяются суммы на закупку необходимой техники и канцелярских принадлежностей, на печать бланков и их рассылку, на доставку и охрану заполненных бланков, на оплату труда регистраторов и программистов, на рекламу и т.д. Здесь важно учесть все, даже самые незначительные, расходы.

ж). Изыскиваются источники финансирования (из бюджета или за счет заинтересованных организаций).

Отметим, что, к сожалению, не всегда удастся найти достаточный объем финансирования; тогда - придется либо сократить программу наблюдения, либо отложить его проведение на некоторое время.

---

### **2.3. Контроль данных наблюдения.**

Даже если наблюдение тщательно подготовлено, имеет детально проработанную программу и четкие инструкции, в процессе его проведения могут иметь место различного рода ошибки, негативно влияющие на достоверность полученных результатов.

**Ошибка наблюдения** – расхождение зарегистрированных и фактических значений признаков у отдельных единиц совокупности.

Строгой классификации ошибок наблюдения не существует, но можно выделить следующие их виды:

1). **Ошибки регистрации** – возникают из-за неверного определения значений признака или в результате ошибочной записи. Могут быть:

а). **Непреднамеренные**, неумышленные – случайные опiski и опечатки, ошибки округления, неверная трактовка вопроса.

Случайные ошибки, как правило, не оказывают существенного влияния на результаты наблюдений с большим числом единиц, т.к. с равной вероятностью могут завышать или занижать истинное значение, а потому – в процессе сводки они взаимопогашаются. Кроме того, такие ошибки обычно поддаются простейшему контролю и могут быть исправлены.

Во избежание ошибок, связанных с неверным пониманием вопроса, следует больше внимания уделять разработке формуляра и инструкции, качеству подготовки регистраторов и т.п.

б). **Преднамеренные**, умышленные – сознательное искажение фактов. Например, занижение доходов и прибыли с целью получения льгот или

уклонения от налогообложения. Или наоборот – завышение социального статуса в социологических опросах.

Такие «ошибки» чаще всего возникают в наблюдениях, не предполагающих какой-либо серьезной проверки (анонимное анкетирование, Интернет-опросы и т.п.), но могут встречаться и в документальных видах наблюдения, например – в отчетности предприятий и организаций. Одним из способов борьбы с подобными «ошибками» в отчетности является усиление юридической ответственности: обычно – административной (штрафы), а в некоторых случаях – вплоть до уголовной.

**2). Ошибки организации наблюдения.** Имеется в виду, что в ходе организации наблюдения могли быть некорректно выбраны место и/или время его проведения, неудачно отобрана исследуемая совокупность, неверно определены (на подготовительной стадии) формы и способы сбора материала.

Причем, эти «ошибки» тоже могут быть:

- непреднамеренными – случайными или под давлением изменившихся обстоятельств,
- или преднамеренными – работа «по заказу», с целью получения заведомо оговоренного результата.

Данные таких исследований не могут считаться достоверными, полученные результаты, скорее всего, будут искажены.

**3). Ошибки репрезентативности** – присутствуют в любом выборочном наблюдении; от них нельзя избавиться, но можно уменьшить их размер, либо увеличивая численность выборки, либо применяя специальные способы отбора единиц.

**4). «Машинные» ошибки** – ошибки обрабатывающих программ (в т.ч. и алгоритмические), а также – различные технические сбои, в результате которых часть информации может быть утеряна или заперчена.

Перед тем, как использовать программы для реальных расчетов, их необходимо тщательно протестировать на правильность расчета и отображения итоговых показателей, а в процессе введения информации в электронные таблицы и базы данных следует соблюдать основные правила работы с ними. В частности, рекомендуется подключить частое автоматическое сохранение файлов, в которые вводится новые данные, запретить несанкционированное изменение уже введенной информации, создавать резервные копии файлов с исходными первичными показателями и промежуточными результатами.

С целью повышения точности полученной информации, процесс статистического наблюдения обычно завершается этапом контроля, осуществляемого по следующим направлениям:

- полнота охвата совокупности, что особенно актуально для сплошных и документально-отчетных видов наблюдения;
- качество и полнота заполнения бланков (часть из них могут оказаться непригодными для обработки);

- правильность представленных первичных данных – выявляются ошибки регистрации, которые могут быть исправлены.

Для проверки полученной первичной информации используются следующие способы контроля:

1). **Логический** – сопоставление ответов на вопросы в рамках одного наблюдения, сравнение с уже известными результатами предыдущих исследований, с плановыми и нормативными значениями, с минимумами, максимумами и т.п.

2). **Арифметический** – проверка правильности расчетных процедур по взаимосвязанным показателям, пересчет сумм, средних, процентов и т.п.

3). **Контрольная проверка** (повторное контрольное исследование). Обычно применяется к наблюдениям-опросам, где предполагается наличие систематических ошибок завышения или занижения показателей.

По окончании основного исследования часть уже обследованных единиц подвергается документальной проверке, по итогам которой рассчитываются «поправочные коэффициенты», характеризующие соотношение точных, проверенных значений показателей со значениями, указанными самими опрашиваемыми. Предполагается, что остальные – непроверенные – единицы «исказили» свои данные примерно в той же пропорции, поэтому показатели основного исследования, скорректированные (умноженные) на «поправочные коэффициенты», считаются более реалистичными.

---

## Лекция 3.

### Первичная обработка материалов наблюдения.

---

#### 3.1. Сводка и ее виды.

**Статистическая сводка** – комплекс последовательных действий по первичной обработке материалов статистического наблюдения, включающий подсчет итоговых показателей по совокупности или ее частям, систематизацию данных и занесение их в специально разработанные таблицы.

Сводка – это следующий за статистическим наблюдением этап статистического исследования, главная задача которого упорядочить собранную информацию и на этой основе дать сводную (общую) характеристику изучаемой совокупности.

##### Виды сводки.

##### 1. По степени детализации:

- а) **простая** – подсчет общих итогов по совокупности (чаще всего – элементарное суммирование);
- б) **сложная** – когда сначала производится группировка единиц изучаемой совокупности по одному или нескольким признакам, а затем подсчитываются итоги по группам.

##### 2. По месту проведения:

- а) **централизованная** – если сбор и обработка информации производится в одном месте (неким «центральным органом»);
- б) **децентрализованная** – данные собираются (и проходят первичную обработку) на местах и поступают в «центр» в обработанном (агрегированном) виде.

Примером централизованной сводки может служить обработка результатов голосования на выборах, а децентрализованной – отчетность предприятий и организаций, а также процесс мониторинга образовательных учреждений области.

##### 3. По способу подсчета:

- а) **ручная** – результаты подсчитываются вручную (самый «яркий» пример – подсчет результатов голосования на местах);
- б) **автоматизированная (машинная)** – информация заносится в базы данных и электронные таблицы, итоги подсчитываются с помощью специальных программных средств.

### 3.2. Группировка и ее виды.

Полученные в результате простой сводки итоговые показатели дают слишком общее представление об изучаемой совокупности; этого обычно недостаточно для последующего анализа. Для более точной характеристики совокупности рекомендуется выделить в ней группы единиц и подсчитать групповые итоги, что в дальнейшем поможет проанализировать структуру явления и выявить взаимосвязи показателей.

**Группировка** – метод, при котором изучаемая совокупность разбивается на однородные части по одному или нескольким существенным признакам. Группировка создает основу для последующей сложной сводки, а также – для расчета и анализа обобщающих показателей, характеризующих совокупность.

**Группировочный признак** – признак, положенный в основу группировки, – должен отвечать следующим требованиям:

- его выбор должен быть обоснован социально-экономической сущностью явления (и/или – логически, математически, статистически);
- им должны обладать не менее 80% единиц совокупности.

Например, если изучается качество жизни населения, то в основу группировки разумно положить доходный признак, разбив население на группы по уровню среднедушевых доходов.

Не следует путать группировку с классификацией! **Классификация** – систематизированное распределение объектов по классам, группам, разрядам на основании их сходства или различия. В основу классификации всегда кладется существенный качественный признак, тогда как группировки часто формируются по количественным признакам. Кроме того, классификация – устойчива во времени, а группировка производится для целей конкретного исследования.

#### Виды группировки.

1. По функциональному назначению:

**а). Типологическая** – совокупность делится на однородные группы по типам явлений. Основой типологической группировки, как правило, является атрибутивный (качественный) признак, выраженный словесно.

Примеры: группировка населения по полу; группировка образовательных учреждений по типу местности и т.п.

**б). Вариационная (структурная)** – однородная совокупность делится на группы по варьирующему признаку, тем самым характеризуется внутренняя структура явления. Если построить вариационные группировки за ряд периодов, то можно оценить структурные сдвиги в составе совокупности. Группировочный признак, как правило, – количественный.

Примеры: группировка населения по возрасту; группировка школ по численности учащихся и т.п.

**в). Смешанная** – когда в основе деления лежат одновременно и атрибутивный и количественный признаки.

Примеры:

➤ В статистике РФ при расчетах величины прожиточного минимума используется следующая половозрастная группировка населения:

- дети 0-6 лет,
- дети 7-15 лет,
- мужчины трудоспособного возраста,
- женщины трудоспособного возраста,
- пенсионеры.

➤ В Ярославской области при оценке качества образования используется группировка образовательных учреждений (ОУ) по типу местности, виду ОУ и численности учащихся, в результате чего выделено 8 «кластеров» ОУ:

- кластер № 1 – ОУ инновационного типа, школы с углубленным изучением предметов, лицеи, гимназии;
- кластер № 2 – вечерние (сменные) общеобразовательные школы;
- кластер № 3 – спецкоррекционные ОУ, учебно-воспитательные ОУ, школы-интернаты;
- кластер № 4 – крупные (более 400 учащихся) городские начальные, средние и основные общеобразовательные школы (НОШ, СОШ и ООШ);
- кластер № 5 – малые (до 399 учащихся) городские НОШ, СОШ и ООШ;
- кластер № 6 – крупные (более 200 учащихся) сельские и поселковые НОШ, СОШ и ООШ;
- кластер № 7 – средние (от 31 до 199 учащихся) сельские и поселковые НОШ, СОШ и ООШ;
- кластер № 8 – малые (до 30 учащихся) сельские и поселковые НОШ, СОШ и ООШ.

**г). Аналитическая (факторная)** – позволяет выявить взаимосвязь между изучаемым явлением и группировочным признаком. Применяется для исследования связи между различными показателями, характеризующими совокупность. При этом один из них рассматривается как результативный, а другой – факторный – кладется в основу группировки. Изменение факторного признака обуславливает и изменение результативного.

Если оба признака количественные, то по аналитической группировке можно судить о направлении связи между рассматриваемыми показателями. Группы по факторному признаку упорядочиваются по возрастанию, и в каждой вычисляется среднее значение результативного признака. Если получившиеся средние значения результирующего показателя в целом возрастают, то можно судить о наличии прямой связи, если убывают – об обратной. Если четкой тенденции в значениях результирующих показателей нет, говорят, что связь не выявлена.

Например, так можно выявить есть ли связь между заполняемостью классов и успеваемостью школьников, между стажем или квалификацией учителя и результатами сдачи ЕГЭ и т.п.



## 2. По количеству группировочных признаков:

а) **Простые** – если в основу группировки положен один, наиболее важный признак;

б) **Комбинационные** – если в основе группировки 2 (и более) признака, взятых в комбинации. Группы первого признака разбиваются по группам второго (те, в свою очередь, могут разбиваться по группам третьего и т.д.), в каждой подгруппе рассчитывается среднее значение результирующего показателя.

На основании комбинационной группировки можно выявить оптимальное сочетание признаков (комбинацию с наилучшим значением результирующего показателя). Однако если группировочных признаков или групп в них будет слишком много, то тенденция может и не выявиться.

Например, рассматривается, как влияет вид образовательного учреждения (ОУ) и тип местности на результаты сдачи ЕГЭ в целом по области. Строится комбинационная группировка следующего вида:

Вид ОУ	Тип местности	% успешно сдавших ЕГЭ
Спец. школы, гимназии, лицеи	городские	$Y_1$
	сельские	$Y_2$
Обычные школы	городские	$Y_3$
	сельские	$Y_4$
Вечерние школы	городские	$Y_5$
	сельские	$Y_6$

Очевидно, что в данном примере «наилучшим сочетанием признаков» окажется та их комбинация, где показатель  $Y_i$  будет максимальным (предположительно, это будет  $Y_1$ , т.е. – городские спец. школы).

в) **Многомерные** – учитывают большое количество признаков, причем не «последовательно» (как комбинационная группировка), а одновременно весь их набор, образующий многомерное признаковое пространство. Эта задача может быть решена методами кластерного анализа или многомерного шкалирования.

## 3. По порядку определения:

а) **Первичные** – их получают непосредственно в ходе наблюдения, на основе первичной статистической информации;

б) **Вторичные** – получают на основе уже имеющихся группировок. Их используют в случае несопоставимости результатов первичных группировок и рассчитывают путем укрупнения, разделения или пересчета первичных интервалов.

Следует также помнить, что с каждым последующим пересчетом информативность группировки обычно снижается. Поэтому необходимо заранее продумывать вопросы сопоставимости сгруппированных данных с результатами аналогичных или предыдущих наблюдений.

### 3.3. Основные принципы построения группировок.

Этапы построения группировки:

**I.** Выбирается группировочный признак.

Как уже отмечалось, в основу группировки может быть положен как атрибутивный (качественный по своей сути), так и количественный признак (такая группировка называется **количественной**).

**II.** Определяется число групп.

При группировке по атрибутивному признаку количество (и конкретное наполнение) групп жестко задается числом возможных значений признака.

В случае количественной группировки часто встает вопрос о выборе числа групп и определении величины интервалов. При этом важно учесть не только диапазон изменения значений признака, но и объем исследуемой совокупности: чем он больше, тем больше групп следует выделить.

Если нет каких-то специальных условий, то при определении количества групп для достаточно больших совокупностей можно ориентироваться на расчет по формуле Стерджесса:

$$k = \log_2 N + 1 = 3,322 \cdot \lg N + 1 = 1,443 \cdot \ln N + 1,$$

где  $k$  – количество групп,  $N$  – объем совокупности.

**III.** Для количественной группировки определяются интервалы.

**Интервал** группировки – диапазон значений варьирующего признака, разница между максимальным и минимальным значениями признака в группе.

Виды интервалов.

**1.** По величине:

**а). Равные** – используются, если изменение признака по группам совокупности происходит достаточно равномерно.

В этом случае величина интервалов определяется по простой формуле:

$$l = \frac{x_{\max} - x_{\min}}{k},$$

где  $l$  – величина интервала,  $x_{\max}$  и  $x_{\min}$  – максимальное и минимальное значения признака в совокупности,  $k$  – число групп.

Полученное по данной формуле значение для удобства работы принято округлять (до десятых, целых, десятков, сотен и т.д.).

**б). Неравные** – используются, если признак изменяется крайне неравномерно и/или в больших пределах.

**в). Специализированные** – задаются сутью явления и отражают качественное своеобразие групп.

Например, группировка населения по возрасту для расчета показателей трудоспособности:

- до 16 лет (население «дорабочего» возраста);
- женщины 16-54 лет и мужчины 16-59 лет (население трудоспособного возраста);
- женщины 55 лет и старше, мужчины 60 лет и старше (население «послерабочего» возраста).

2. По количеству включаемых единиц:

а). **Равные** – используются для абсолютно равномерных распределений, на практике встречаются крайне редко.

б). **Неравные** – включают разное количество единиц.

Обычно группировки формируются таким образом, чтобы интервалы были равными по величине, а количество единиц в них – какое получится.

3. По виду чисел:

а). **из непрерывных чисел**

б). **из дискретных чисел**

Рассмотрим, например возрастные группы педагогических работников:

- моложе 25
- 25 – 35
- 35 – 45
- 45 – 55
- старше 55

- моложе 25
- 25 – 34
- 35 – 44
- 45 – 54
- 55 и старше

Признак «возраст» рассматривается как:

некая непрерывная величина.

«целое число исполнившихся лет».

Интервалы между собой:

не различаются.

различаются на **единицу масштаба**.

Границы интервалов:

по умолчанию – одна включается в интервал, другая – нет.

обе включаются в интервал.

Величина (длина) интервала:

$$l = x_v - x_n$$

$$l = x_v - x_n + e_m$$

где  $l$  – величина интервала,  $x_v$  и  $x_n$  – верхняя и нижняя границы интервала,  $e$  – единица масштаба (задается сутью явления или точностью измерений).

В приведенном выше примере с возрастными группами педагогов:

$$l = 35 - 25 = 10 \text{ (лет)}$$

$$l = 34 - 25 + 1 = 10 \text{ (лет)}$$

Середина интервала:

$$x_c = \frac{x_v + x_n}{2}$$

где  $x_c$  – середина интервала,  $x_v$  и  $x_n$  – верхняя и нижняя границы интервала.

В приведенном выше примере с возрастными группами педагогов:

$$x_c = (35 + 25) / 2 = 30$$

$$x_c = (34 + 25) / 2 = 29,5$$

4. По наличию границ:

а). **Закрытые** – имеют обе границы.

б). **Открытые** – если отсутствует нижняя или верхняя граница.

Открытыми часто оказываются первый (нет нижней границы) и/или последний (нет верхней границы) интервалы. Проводя группировку, следует помнить, что в такие интервалы должно попадать не более 10% единиц совокупности.

При определении величины и середины открытых интервалов (что может потребоваться для расчета средних и показателей вариации) руководствуются следующим правилом: если остальные интервалы примерно равны по величине, то открытые интервалы приравнивают по величине ближайшим закрытым. В противном случае, недостающие границы определяются экспертным путем.

### 3.4. Ряды распределения, их виды и графическое изображение.

**Ряд распределения** – упорядоченное распределение единиц изучаемой совокупности по группам (значениям) варьирующего признака.

Строго говоря, ряд распределения может быть и атрибутивным – т.е. построенным по качественному признаку, но обычно рассматриваются вариационные ряды, представляющие собой упорядоченное по возрастанию (или убыванию) отображение результатов количественной группировки.

Кроме того, ряд распределения можно рассматривать как более удобную и компактную форму записи первичного ряда данных.

Например, известны экзаменационные оценки школьников в порядке их получения: 4, 3, 4, 2, 4, 5, 3, 5, 3, 4. Это «первичный ряд», в котором повторяются одни и те же значения признака (всего их 4 – «2», «3», «4» и «5»).

В виде «ряда распределения» эти данные будут выглядеть следующим образом:

Оценка	Количество, чел.
2	1
3	3
4	4
5	2
Итого:	10

Общий вид ряда распределения:

Признак $x$	Частота $f$
$x_1$	$f_1$
$x_2$	$f_2$
...	...
$x_n$	$f_n$
Итого:	$\Sigma f$

**Частота  $f$**  может быть представлена как количество или доля единиц, обладающих данным значением признака. Если  $f_i$  – доли, то  $\Sigma f = 1$  (или 100%).

Виды (вариационных) рядов распределения.

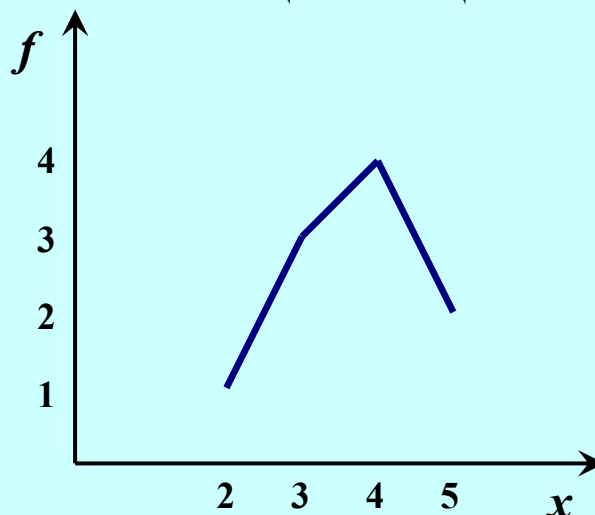
1. По виду (количественного) признака  $x$ :

а). **Дискретные** – если признак принимает конкретные числовые (часто – целочисленные) значения.

Графическим изображением дискретного ряда является **полигон**.

Пример: распределение школьников по экзаменационным оценкам.

Оценка ( $x$ )	Количество, чел. ( $f$ )
2	1
3	3
4	4
5	2
Итого:	10

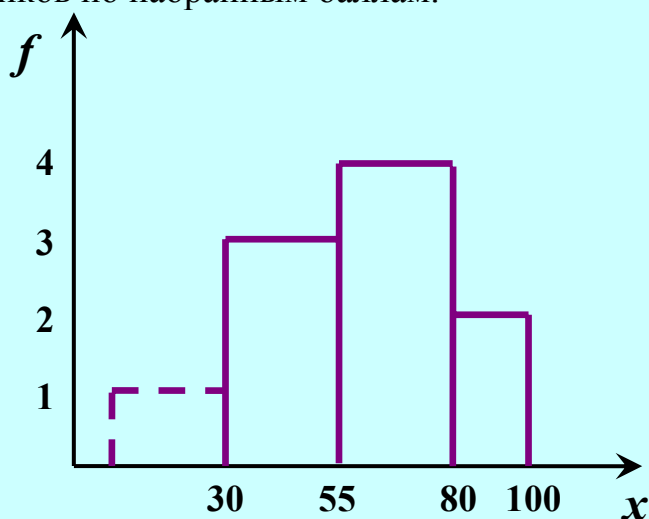


б). **Интервальные** – если признак может принимать любые значения в границах выделенных интервалов.

Графическим изображением интервального ряда является **гистограмма**.

Пример: распределение школьников по набранным баллам.

Баллы ( $x$ )	Количество, чел. ( $f$ )
< 30	1
30 – 55	3
55 – 80	4
> 80	2
Итого:	10



2. По виду частот:

а). **Простые** – если частоты представлены как количество или доля единиц (как в двух вышеприведенных примерах).

**б). Кумулятивные** – получаются из простых последовательным сложением частот. Требуются для последующих расчетов, в частности – для определения медианы.

Если ряд упорядочен по возрастанию, то расчет кумулятивных (или – накопленных) частот осуществляется сверху-вниз (по принципу «менее чем»). Последняя кумулятивная частота будет равна сумме обычных частот.

Если вернуться к примеру с экзаменационными оценками школьников, то кумулятивные частоты будут рассчитаны следующим образом:

Оценка (значения признака $x$ )	Количество, чел. (обычные частоты $f$ )	Кумулятивные частоты ( $S$ )
2	1	1
3	3	1+3=4
4	4	1+3+4=8
5	2	1+3+4+2=10
Итого:	10	

### 3. Особый вид ряда распределения – *ранжированный ряд*.

Значениям признака, упорядоченным по возрастанию, в соответствие ставятся ранги. Если несколько единиц имеют одинаковое значение признака, то им присваивается среднее арифметическое соответствующих рангов. Последний ранг, в общем случае, соответствует объему совокупности.

Ранжирование может иметь как самостоятельное значение («расстановка мест»), так и быть составной частью последующих расчетов.

Графическое изображение ранжированного ряда называется *огива*. В отличие от других графиков, здесь значения признака  $x$  откладываются по оси ординат, а по оси абсцисс – ранги. Огива имеет вид строго возрастающей ломаной кривой.

Пусть, например, известны результаты сдачи ЕГЭ семи школьников.

Фамилия школьника	А	Б	В	Г	Д	Е	Ж
Балл ( $x$ )	48	38	27	38	71	38	48
Ранг ( $p$ )	5,5	3	1	3	7	3	5,5

Первый ранг присваивается школьнику В – с минимальным значением балла – 27. Следующее по возрастанию значение балла – 38 – сразу у трех школьников – Б, Г и Е, которым, по порядку, следовало бы присвоить второй, третий и четвертый ранги. Значит, каждому из них будет присвоено среднее арифметическое указанных рангов, а именно:  $(2+3+4) / 3 = 3$ . Аналогичная ситуация и со следующим значением балла – школьники А и Ж «поделят» пятый и шестой ранги. Последний, седьмой ранг, получит школьник Д, набравший максимальный балл – 71.

### 3.5. Статистические таблицы и правила их оформления.

**Статистическая таблица** – одна из наиболее удобных, наглядных и эффективных форм отображения статистических показателей.

С помощью статистических таблиц осуществляется представление результатов наблюдения, сводки и группировки; в них дается сводная характеристика совокупности, подводятся один или несколько итогов. Табличный метод отображения информации распространен и в других областях знаний и сферах деятельности.

Строение статистической таблицы:

#### Заголовок таблицы

A	Б	В	1	2	3	4	5	6
	<b>Подлежащее</b>			–	<b>Сказуемое</b>			
				×				
				::				

#### Примечание

- В **заголовке** таблицы отражается сущность явления, его место и время.
- **Подлежащее** таблицы содержит перечень характеризующих явление признаков (строки таблицы), а также – их коды, шифры и единицы измерения.
- **Сказуемое** таблицы содержит систему показателей, описывающих признаки из подлежащего. Таблица может иметь достаточно сложную разработку сказуемого (столбцы с общими шапками).
- В **примечании** могут быть приведены ссылки на источники информации, пояснения и комментарии к отдельным показателям, методики их расчета и т.п.

Правила оформления таблиц:

1. Таблицы в тексте должны быть пронумерованы. Нумерация может быть сквозной (*Таблица 1, ..., Таблица N*) или по главам (*Таблица 1.1, Таблица 1.2* и т.д.), приводится перед заголовком и выравнивается по правому краю.
2. В таблице допускаются только общепринятые сокращения терминов и единиц измерения.
3. Общие для ряда показателей слова следует вынести в соответствующие заголовки строк и/или столбцов.

4. Если все показатели в таблице измерены в одних и тех же единицах и/или по отношению к одному и тому же периоду, то эти единицы и период выносятся в заголовок таблицы (*Заголовок таблицы, %, 2000=100%*).

5. Показатели внутри таблицы следует располагать в логическом (сначала – оценочные, потом – расчетные) и в строгом хронологическом (от более ранних к более поздним, или наоборот) порядке.

6. Все однотипные показатели в таблице приводятся с одинаковой степенью точности и округления. Коэффициенты принято округлять до трех знаков после запятой, проценты – до одного.

7. Громоздкую таблицу, не уместящуюся на листе, лучше разбить на несколько маленьких, по группам показателей.

8. В случае необходимости переноса таблицы на следующий лист руководствуются следующими соображениями:

а). Если шапка таблицы небольшая, в одну-две строки, – имеет смысл переносить вместе с шапкой.

б). Если же шапка таблицы имеет сложную разработку и занимает много места, то на первой странице столбцы нумеруются (сразу под шапкой), и на следующие страницы переносится только строка нумерации. Причем, столбцы сказуемого рекомендуется нумеровать заглавными русскими или латинскими буквами, а столбцы сказуемого – арабскими цифрами; двойная нумерация столбцов не допускается.

9. Для соответствия размера таблицы форме листа допускается подлежащее и сказуемое менять местами, т.е. строки таблицы станут столбцами, а столбцы – строками.

10. В таблицах используются специальные символы:

а). Если показателя не существует по смыслу явления, то в соответствующей клеточке таблицы ставится «крестик» (×).

б). Если показатель, как таковой, существует, но в данном случае его нет (или он равен 0), то ставится «прочерк» (–).

в). Если показатель очень мал по сравнению с другими, т.е. равен 0 точностью до принятых в таблице округлений, то ставится «ноль» с соответствующим количеством нулей после запятой (0,000).

г). Если показатель по какой-то причине неизвестен, то либо пишется «нет сведений» (если есть место для такой надписи), либо ставятся четыре точки – в линию (....) или «квадратиком» (::).

Правильное оформление таблицы, так же как и тщательная разработка ее макета, способствует скорости восприятия удобству дальнейшей обработки информации.



## Лекция 4.

### Абсолютные и относительные величины.

---

#### 4.1. Абсолютные величины и их виды.

Обобщающие показатели, характеризующие статистическую совокупность, могут быть представлены в виде:

- абсолютных величин,
- относительных величин,
- средних величин.

Полученные в результате статистического наблюдения показатели чаще всего бывают представлены в виде абсолютных величин, которые, в свою очередь, служат основой для последующих расчетов и анализа.

**Абсолютная величина** – обобщающий статистический показатель, характеризующий численность совокупности или объем изучаемого признака в конкретных условиях места и времени, выраженный в соответствующих единицах измерения.

Например, такие показатели из отчетности образовательных учреждений (ОУ), как: площадь помещений ОУ (форма ОШ-1), численность работников ОУ (форма 83-РИК), объемы финансирования и расходы ОУ (форма ОШ-2) и т.п.

#### Виды абсолютных величин.

➤ По охвату совокупности:

3. **Индивидуальные** – характеризуют отдельную единицу совокупности (например, конкретную школу);

4. **Суммарные**:

- а. групповые – характеризуют группу единиц (школы района),
- б. общие - характеризуют всю совокупность (все школы области).

➤ По порядку определения:

1. **Первичные** – получают непосредственно в результате наблюдения (путем замеров и подсчетов);

2. **Вторичные** – получаются из первичных в результате их суммирования (или вычитания).

➤ По характеру явления:

1. **Моментные** – характеризуют состояние объекта на определенный момент времени (например, численность работников на 1 января);

2. **Интервальные** – характеризуют показатели объекта за определенный период (фонд оплаты труда за отчетный год).

➤ По форме учета (единицам измерения):

1. **Натуральные** – измеренные в каких-либо физических единицах:
  - а. простые (например, общая площадь помещений – в м<sup>2</sup>, количество персонала – в чел.);
  - б. составные (расход электроэнергии - в кВт\*ч, трудоемкость учебной программы – в человеко-часах).

Показатели, измеренные чел.-дн., чел-час. и т.п., иногда выделяют в особую группу – «трудовые», обладающую теми же свойствами, что и натуральные.

Достоинства натуральных величин: сопоставимы во времени и по территориям.

Недостатки: несопоставимы по разным видам продукции, работ и услуг.

2. **Условно-натуральные** – предназначены для сопоставления однородной продукции (однотипных работ, услуг). Получают путем пересчета всех видов продукции по какому-то одному виду, принятому за эталон. Пересчет может вестись как по физическим характеристикам (мощность, КПД, доля полезных свойств), так и по экономическим (себестоимость, материалоемкость, трудоемкость и т.п.).

Условно-натуральные величины широко используются в анализе производственной деятельности для получения суммарных показателей по всем видам продукции (работ, услуг) в натуральном эквиваленте.

### **Пример.**

Пусть учреждение, расположенное в сельской местности и имеющее собственную (твердотопливную) котельную, сожгло за зиму 2 т дров, 1,5 т каменного угля и 1 т топливных брикетов. Требуется определить общее количество использованного топлива.

### Решение:

В данном случае нельзя определить общий объем топлива простым суммированием (2+1,5+1=4,5 т), т.к. использовались разные виды топлива, имеющие разную теплотворную способность.

Для учета суммарного потребления всех видов органического топлива используется так называемое «условное топливо». В России за «условное топливо» принят каменный уголь, теплота сгорания 1 кг которого равна 29,3 МДж (7000 ккал). Для других видов топлива нужно использовать коэффициенты пересчета (калорийные эквиваленты). Известно также, что удельная теплота сгорания дров 10,2 МДж/кг, а топливных брикетов – 17,5 МДж/кг.

Тогда, коэффициенты пересчета:

$$К.Э. \text{ (для дров)} = \frac{10,2 \text{ МДж / кг}}{29,3 \text{ МДж / кг}} = 0,35 \quad К.Э. \text{ (для брикетов)} = \frac{17,5 \text{ МДж / кг}}{29,3 \text{ МДж / кг}} = 0,60$$

И, следовательно, общий объем потребленного учреждением топлива составит (2\*0,35+1,5+1\*0,6)=2,8 усл.т.

3. **Стоимостные** – измеренные в денежном выражении (в руб.).

Достоинства стоимостных показателей: сопоставимы по разным видам продукции и работ (через цену продукции или ставку оплаты).

Недостатки: несопоставимы за длительные промежутки времени, т.к. подвержены влиянию инфляции, изменению курсов валют и т.п. (в этих случаях необходимо производить пересчет «в твердых ценах» или с поправкой на индекс цен).

---

## 4.2. Относительные величины и их виды.

**Относительная величина** – обобщающий статистический показатель, характеризующий меру соотношения двух сопоставляемых величин.

Общий вид:  $O = \frac{A}{B} \cdot 10^k$

При расчете относительных величин важно, чтобы сравниваемые величины (**A** и **B**) были сопоставимы, и между ними существовала реальная взаимосвязь.

### Виды относительных величин.

➤ По масштабу сопоставления:

- $k=0$  – доля;
- $k=1$  – дециль;
- $k=2$  – процент (%);
- $k=3$  – промилле (‰);
- $k=4$  – продецимилле (‱).

➤ По порядку определения:

1. **Первичные** – если **A** и **B** – абсолютные величины;
2. **Вторичные** – если **A** и **B** – относительные величины.

### **Пример.**

Пусть в классе **A** успешность сдачи (на «4» и «5») ЕГЭ по математике составляет 55%, а в классе **B** – 50%. Насколько класс **A** опережает класс **B** по данному показателю?

Решение:

Расчет соотношения производится по формуле:  $\frac{55\%}{50\%} \cdot 100\% = 110\%$ .

Следовательно, класс **A** опережает класс **B** на 10%.

Если посчитать «абсолютную разницу» относительных величин, выраженных в процентах, то эта разница измеряется в процентных пунктах. Другими словами, класс **A** опережает класс **B** на  $(55-50)=5$  п.п.

➤ По функциональному назначению можно выделить следующие группы относительных величин:

### 1. Показатели плана (и норм).

а. **Показатель планового задания:**

$$П.З. = \frac{\text{планируемый на следующий период уровень}}{\text{фактический уровень базисного периода}}$$

Характеризует перспективы развития.

б. **Показатель выполнения плана:**

$$В.П. = \frac{\text{фактический уровень}}{\text{плановый уровень}}$$

Характеризует степень реализации плана. Если  $В.П. > 1$  (100%), план перевыполнен; если  $В.П. < 1$  (100%), план недовыполнен.

Если план направлен на снижение уровня показателя (например, если речь идет о снижении затрат или себестоимости), то показатель выполнения плана следует считать, как обратный:

$$В.П.^* = \frac{\text{плановый уровень}}{\text{фактический уровень}}$$

#### Пример.

Пусть в школе внедряется новая программа обучения, на которую необходимо набрать 50 школьников. При этом планируется снизить недельную нагрузку на 5%. Фактически было набрано 52 ученика, а нагрузка снизилась на 3%. Рассчитать показатели выполнения плана.

Решение:

Выполнение плана по набору учеников рассчитывается по формуле:

$$В.П. (\text{по набору}) = \frac{\text{факт}}{\text{план}} = \frac{52}{50} \cdot 100\% = 104\%.$$

Т.е. план по набору перевыполнен на 4%.

Выполнении плана по снижению недельной нагрузки рассчитывается по обратной формуле:  $В.П.^* (\text{по } \downarrow \text{нагрузки}) = \frac{\text{план}}{\text{факт}} = \frac{95\% \cdot T}{97\% \cdot T} \cdot 100\% = 97,9\%$

Т.о. план по снижению нагрузки оказался недовыполненным на 2,1%.

### 2. Показатели динамики.

Характеризуют развитие (динамику) явления.

Рассматривается значение некоего показателя  $Y$  за  $n$  периодов. Вводится нумерация значений показателя по периодам:  $Y_1, Y_2, \dots, Y_n$ . Тогда:

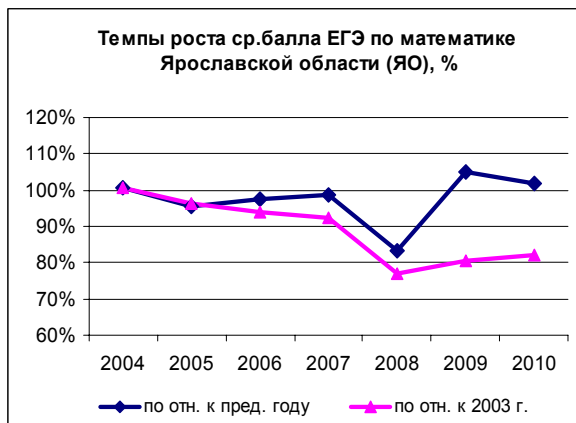
а. **Базисные показатели динамики:**  $Дб_i = \frac{Y_i}{Y_0},$

где  $Y_0$  – период, принятый за базу сравнения (обычно – самый первый).

б. *Цепные показатели динамики:*

$$Ду_i = \frac{Y_i}{Y_{i-1}}$$

Относительные величины динамики также называют **темпами роста**, и их часто переводят в проценты. Графически они могут быть представлены как обычный график (ломаная кривая) или как столбиковая диаграмма.



Для анализа динамики показателя полезно знать и средние значения его темпов роста (убыли) и прироста, т.е. на сколько процентов увеличивается (уменьшается) значение показателя в среднем за период.

**Средний темп роста** показателя определяется по формуле:  $\overline{Tp} = \sqrt[n-1]{\frac{Y_n}{Y_1}} \cdot 100\%$ ;

а **средний темп прироста** равен  $(\overline{Tp} - 100\%)$ .

### Пример.

Известно, что средний балл ЕГЭ по математике выпускников колледжа составил: в 2008 г. – 51,65, в 2009 г. – 54,34, в 2010 г. – 58,13. Рассчитать показатели динамики.

### Решение:

Результаты расчетов представлены в таблице:

Год	Базисные показатели динамики по сравнению с 2008 г., %	Цепные показатели динамики по сравнению с предыдущим г., %
• 2008	100%	::
• 2009	$\frac{54,34}{51,65} \cdot 100\% = 105,2\%$	$\frac{54,34}{51,65} \cdot 100\% = 105,2\%$
• 2010	$\frac{58,13}{51,65} \cdot 100\% = 112,5\%$	$\frac{58,13}{54,34} \cdot 100\% = 107,0\%$

Средний балл ЕГЭ по математике в колледже ежегодно увеличивался. За два года это средний балл возрос на 12,5%. Причем, увеличение шло нарастающими темпами: в 2009 г. прирост составил 5,2%, в 2010 г. – 7%.

Среднее значение темпа роста:  $\sqrt[3-1]{\frac{58,13}{51,65}} \cdot 100\% = \sqrt{1,125} \cdot 100\% = 106,1\%$ , т.е. на протяжении исследуемого периода показатель увеличивался на 6,1% в среднем за год.

### с. *Показатели (коэффициенты) опережения.*

Рассматриваются значения двух разных показателей (или одного показателя, но по двум разным территориям)  $Y$  и  $Z$  за  $n$  периодов. Для них рассчитываются относительные величины динамики (базисные и/или цепные).

Тогда базисные и цепные показатели опережения:  $Об_i = \frac{Дб(Y)_i}{Дб(Z)_i}$ ,  $Оц_i = \frac{Дц(Y)_i}{Дц(Z)_i}$ .

Коэффициенты опережения характеризуют, насколько рост (убыль) одного показателя идет быстрее (медленнее) динамики другого показателя.

Например, насколько темп роста доходов опережает темпы роста цен, или насколько темпы снижения среднеобластного балла ЕГЭ по математике опережают темпы снижения среднего балла по России.

## 3. Показатели внутреннего строения.

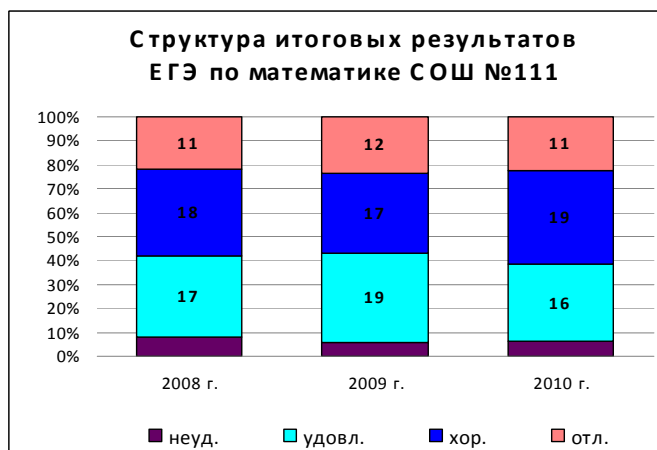
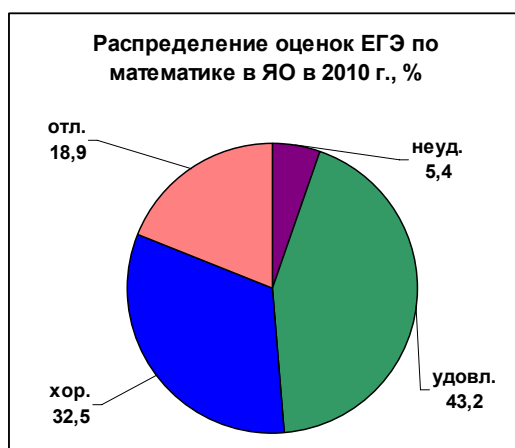
Рассматриваются отдельные составные части объекта или явления.

а. *Показатели структуры* – характеризуют удельный вес (долю) части совокупности в ее общем объеме. Измеряются в долях и процентах.

Доля  $\omega_i = \frac{\text{показатель части совокупности}}{\text{показатель всей совокупности}}$ ;  $\sum_i \omega_i = 1$  (или 100%).

Если в результате округлений сумма показателей структуры не равна 100%, следует скорректировать самый большой показатель.

Графически показатели структуры принято отображать на круговой диаграмме. Если требуется сравнить структуру нескольких однотипных объектов или одного, но за несколько периодов, то удобнее использовать нормированную гистограмму.



б. *Показатели координации* – характеризуют соотношение частей совокупности между собой и обычно задаются как количество единиц одной группы, приходящихся на определенное количество (1, 10, 100, 1000 и т.д.) единиц другой. Например, сколько мужчин приходится на 1000 женщин в населении области, сколько мужчин на 100 женщин среди школьных педагогов и т.п. В качестве базы сравнения, как правило, выбирается самая значимая часть совокупности (с наибольшей долей).

$$K_i = \frac{\text{показатель части совокупности}}{\text{показатель части, принятой за базу сравнения}} \cdot N_0, N_0 = \{1, 10, 100, 1000 \dots\}$$

### Пример.

На 1 июля 2010 г. в учреждениях дополнительного образования области работало 3500 человек, из них: административного персонала (АП) – 420, педагогического (ПП) – 1800, учебно-вспомогательного (УВП) – 320 и обслуживающего (ОП) – 960.

Рассчитать показатели структуры и координации (в пересчете на 100 педагогов).

### Решение:

Результаты расчетов представлены в таблице:

Категория персонала	Структура персонала учреждений доп.образования, %	Показатели координации в расчете на 100 педагогов, чел.
• АП	$\frac{420}{3500} \cdot 100\% = 12,0\%$	$\frac{420}{1800} \cdot 100 \approx 23$
• ПП	$\frac{1800}{3500} \cdot 100\% = 51,5\%$	
• УВП	$\frac{320}{3500} \cdot 100\% = 9,1\%$	$\frac{320}{1800} \cdot 100 \approx 18$
• ОП	$\frac{960}{3500} \cdot 100\% = 27,4\%$	$\frac{960}{1800} \cdot 100 \approx 53$
Итого:	100,0%	×

Таким образом, в структуре работников учреждений дополнительного образования педагоги составляют более половины – 51,5%. Вторая по численности группа – обслуживающий персонал – 27,4%. Администрация составляет 12% от общей численности персонала учреждений. На 100 педагогов приходится 23 административных работника, 18 представителей учебно-вспомогательного и 53 – обслуживающего персонала.

## 4. Показатели дифференциации.

Характеризуют различие (дифференциацию) значений исследуемого признака у отдельных единиц совокупности и позволяют оценить степень этих различий. Измеряются в долях (в размах) или процентах.

Рассматриваются значения признака  $X$  у различных единиц совокупности, и сопоставляются с неким значением  $X^0$ , принятым за базу

сравнения:  $D_i = \frac{X_i}{X^0}$ .

В качестве  $X^0$  могут выступать:

- некое «эталонное» или наиболее удобное значение признака;
- одно из реально встречающихся в данной совокупности значений признака, чаще всего – минимальное или максимальное;
- одна из средних величин, рассчитанных по совокупности.



Здесь могут применяться не только обычные средние (простая или взвешенная), но и «средняя прогрессивная» (средняя из тех единиц, что лучше средней) и «средняя регрессивная» (средняя из тех, что хуже средней).

Показатели дифференциации применяются для выделения «лучших» единиц и расчета резервов роста (снижения) показателя. Их также можно применять для группировки единиц по соотношению со средними (прогрессивными и регрессивными) значениями признака.

## 5. Показатели развития (интенсивности).

Характеризуют степень распространения одного явления в среде другого и определяются сущностью социально-экономических явлений.

До сих пор рассматривались относительные величины, где в числителе и знаменателе стояли одноименные показатели (простейшие из них, где сопоставляются два объекта, часто называют **показателями сравнения**). Однако, рамки применимости общей формулы для расчета относительных величин куда шире: делимые друг на друга показатели *A* и *B* могут иметь различную природу (и, соответственно, – разные единицы измерения). Таким образом, полученные в результате расчета относительные показатели могут быть величинами именованными (имеющими единицы измерения) или выражаться в процентах или промилле.

В качестве примера можно привести следующие показатели:

- плотность населения – общая численность населения делится на площадь территории (чел. на км<sup>2</sup>);
- среднедушевой ВВП – значение ВВП делится на численность населения (руб. на чел.);
- коэффициенты рождаемости и смертности – численность родившихся или умерших делится на численность населения (измеряется в ‰);
- наполняемость классов – общее количество учеников делится на количество классов (чел.);
- нагрузка на одного учителя – количество учеников делится на количество педагогов (чел.); и др.

Приведенная выше классификация относительных величин по функциональному назначению не является «строгой», могут выделяться и другие виды и группы показателей.





## Лекция 5.

### Средние величины и показатели вариации.

#### 5.1. Средние величины и их виды.

**Средняя величина** (среднее значение) – обобщенная количественная характеристика варьирующего признака единиц совокупности. Средняя величина относится к единице совокупности, но абстрагируется от индивидуальных особенностей отдельных единиц.

Например, показатель средней заработной платы рассчитывается делением фонда оплаты труда на численность работающих и представляет собой обобщающую количественную характеристику уровня оплаты труда в конкретной организации (в данном виде деятельности). Используя значения средней заработной платы, можно достаточно объективно сравнивать оплату труда в различных организациях (видах деятельности).

Также можно сравнивать школы с разной численностью учащихся по среднему баллу ЕГЭ по определенному предмету, который будет характеризовать некий усредненный уровень знаний выпускников школ по данному предмету, тогда как индивидуальные результаты ЕГЭ отдельных учеников каждой школы могут сильно варьироваться между собой.

Однако следует помнить, что средняя величина может получиться «абстрактной» (известное выражение - «средняя температура по больнице»), т.е. ни одна из единиц не имеет значения, равного или близкого среднему. Это обычно происходит, когда изучаемый признак сильно варьируется по разным частям совокупности (например, заработная плата - по категориям персонала учреждения).

Таким образом, для правильного расчета и дальнейшего анализа средних величин важно, чтобы расчет производился по однородной совокупности с достаточно большим числом единиц. При выборе методики расчета также следует учитывать качественное содержание усредняемого признака и особенности исходных данных.

#### Виды средних величин:

**I. Средние аналитические** – рассчитываются по формулам. Выделяют:

1. **Средние степенные** – в их основе лежит степенная функция.

К этой группе относятся наиболее «популярные» виды средних: **средняя арифметическая, средняя геометрическая, средняя гармоническая, средняя квадратическая** и др.

2. Другие средние аналитические.

Например, при расчете средних для явлений, имеющих ярко выраженный сезонный характер, используются тригонометрические функции.

**II. Средние порядковые (средние позиционные, структурные средние) –** определяются по месту расположения в упорядоченной совокупности.

---

## 5.2. Средние степенные.

Общий вид средней степенной: 
$$\bar{x} = \sqrt[n]{\frac{\sum_{i=1}^n x_i^\alpha}{n}} = \sqrt[n]{\frac{x_1^\alpha + x_2^\alpha + \dots + x_n^\alpha}{n}},$$

где  $\alpha$  – степень средней,  $n$  – объем совокупности,  $x_i$  – индивидуальные значения признака.

Это формула «простой средней», используется для совокупностей небольшого объема, представленных в виде первичного ряда данных.

Совокупности большого объема могут быть представлены в виде ряда распределения, с соответствующими индивидуальным значениям признака частотами  $f_i$ . В этом случае удобнее использовать «взвешенную» формулу, в которой учитывается, что варианты значений признака встречаются разное количество раз (имеют разный «вес»).

Общий вид средней степенной взвешенной:

$$\bar{x} = \sqrt[n]{\frac{\sum_{i=1}^n x_i^\alpha f_i}{\sum_{i=1}^n f_i}} = \sqrt[n]{\frac{x_1^\alpha f_1 + x_2^\alpha f_2 + \dots + x_n^\alpha f_n}{f_1 + f_2 + \dots + f_n}}$$

В практике статистического анализа наиболее часто используются: средняя арифметическая ( $\alpha = 1$ ), средняя гармоническая ( $\alpha = -1$ ) и средняя геометрическая.

### □ Средняя арифметическая.

Наиболее распространенный вид средней. Используется при расчете средней из абсолютных и средних величин.

Простая средняя арифметическая: 
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Взвешенная: 
$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{f_1 + f_2 + \dots + f_n}$$

📌 Если средняя рассчитывается по интервальному ряду, то в качестве  $x_i$  берутся середины интервалов, рассчитываемые как полусуммы верхних и нижних границ. В результате полученная средняя окажется не точной, степень ее приближения к фактическому среднему значению зависит от равномерности распределения единиц внутри интервалов.

### □ Средняя гармоническая.

Используется в тех же случаях, что и средняя арифметическая, если исходные данные не позволяют рассчитать среднюю по формуле средней арифметической (когда известен числитель логической формулы, а знаменатель требует дополнительных расчетов).

Простая средняя гармоническая: 
$$\bar{x} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

Взвешенная: 
$$\bar{x} = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n \frac{f_i}{x_i}} = \frac{f_1 + f_2 + \dots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}}$$

#### Пример.

Известны следующие данные об оплате труда сотрудников ОУ:

Категории персонала	Численность, чел.	Фонд оплаты труда, тыс. руб.	Средняя заработная плата, тыс. руб.
• АП	4	88	22
• ПП	17	170	10
• УВП и ОП	9	72	8
Итого:	30	330	×

Определить среднюю заработную плату сотрудников ОУ.

#### Решение:

В данном примере рассчитать среднюю заработную плату можно тремя разными способами.

1. Т.к. известны значения фонда оплаты труда и численности сотрудников, то можно воспользоваться *логической* формулой:

$$\text{средняя заработная плата} = \frac{\text{фонд оплаты труда}}{\text{численность работающих}} = \frac{330}{30} = 11 \text{ (тыс.руб.)}$$

2. Предположим, что значение фонда оплаты труда – неизвестно, а известны только значения средней заработной платы (качественный признак  $x$ ) и численность (количественный признак  $f$ ) по категориям персонала. Тогда следует использовать формулу *средней арифметической* взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} = \frac{22 \cdot 4 + 10 \cdot 17 + 8 \cdot 9}{4 + 17 + 9} = \frac{330}{30} = 11 \text{ (тыс.руб.)}$$

3. Пусть теперь неизвестна численность работников, а значения фонда оплаты труда ( $F$ ) и средней заработной платы ( $x$ ) по категориям – известны. Тогда следует использовать формулу *средней гармонической* взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^n F_i}{\sum_{i=1}^n x_i} = \frac{88 + 170 + 72}{22 + 10 + 8} = \frac{330}{30} = 11 \text{ (тыс. руб.)}$$

Таким образом, средняя заработная плата в ОУ составляет 11 тыс. руб.

### Средняя геометрическая.

Используется при расчете средней из относительных величин (темпов роста, индексов и т.п.).

Простая средняя геометрическая:  $\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$

### Пример.

Известны следующие показатели официальной статистики дневных общеобразовательных учреждений (ООУ) Ярославской области\*:

На начало учебного года:	Число дневных ООУ	Численность обучающихся ООУ, тыс.чел.	Показатели динамики по отношению к предыдущему году:	
			числа ООУ	численности обучающихся
• 2005–2006	530	122,2	×	×
• 2006–2007	511	114,7	0,964	0,939
• 2007–2008	498	108,9	0,975	0,949
• 2008–2009	490	106,9	0,984	0,982
• 2009–2010	470	106,6	0,959	0,997

Определить средние темпы убыли числа ООУ и численности обучающихся, и сравнить полученные средние между собой.

### Решение:

Требуется рассчитать средние значения для относительных величин - показателей динамики (темпов убыли), для чего необходимо использовать

формулу средней геометрической:  $\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i}$ , где  $x_i$  – показатели динамики.

Средний темп роста (убыли) числа общеобразовательных учреждений:  $\bar{x} = \sqrt[4]{0,964 \cdot 0,975 \cdot 0,984 \cdot 0,959} = \sqrt[4]{0,887} = 0,970$ ; средний темп роста (убыли) численности учащихся:  $\bar{x} = \sqrt[4]{0,939 \cdot 0,949 \cdot 0,982 \cdot 0,997} = \sqrt[4]{0,872} = 0,966$ .

Таким образом, на протяжении исследуемого периода численность учащихся снижается в среднем на 3,4% в среднем за год, что обуславливается главным образом демографическими причинами. Число школ при этом тоже

\* Регионы России. Социально-экономические показатели. 2010: Стат. сб. / Росстат – М., 2010. (табл. 6.4 - 6.5)

сокращается, но немного медленнее, – на 3,0% в среднем за год, и в основном за счет закрытия малокомплектных школ.

Следует отметить, что при использовании приведенных в § 4.2 формул для расчета среднего темпа роста по исходным показателям, получатся те же результаты:

$$\overline{Tr} = \sqrt[n-1]{\frac{Y_n}{Y_1}} \cdot 100\%: \quad \overline{Tr} \text{ (для числа ООУ)} = \sqrt[5-1]{\frac{470}{530}} \cdot 100\% = \sqrt[4]{0,887} \cdot 100\% = 97,0\%$$

### 5.3. Средние порядковые.

Используются наряду со средними степенными для более полной характеристики распределения изучаемого признака в совокупности. Наиболее важные и часто упоминаемые средние порядковые – мода и медиана.

□ **Мода** (модальное значение) – наиболее часто встречающееся значение признака в совокупности.

Мода, как статистический показатель, обладает свойством минимизации количества ошибок и активно применяется в статистическом анализе. Например, в анализе спроса и потребления – для определения наиболее распространенного размера одежды и обуви. Большое значение в статистике уровня и качества жизни население имеет понятие модального дохода, а в анализе результатов ЕГЭ обязательно рассчитывается модальный балл – балл, который набрали наибольшее количество сдававших.

Определение моды производится:

- в дискретном ряду – согласно смыслу определения: мода – значение признака, которому соответствует наибольшая частота;
- в интервальном ряду – по формуле:  $Mo = x_{mo} + l_{mo} \cdot \frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})}$ ,

где  $x_{mo}$  – нижняя граница модального интервала,  $l_{mo}$  – его величина,  $f_{mo}$  – его частота,  $f_{mo-1}$  – частота предыдущего,  $f_{mo+1}$  – частота следующего интервалов.

Модальный интервал – интервал с наибольшей частотой.

□ **Медиана** (медианное значение) – центральное значение признака в упорядоченном ряду, т.е. у половины единиц совокупности значение признака больше медианного, у половины – меньше.

Медиана, как статистический показатель, обладает свойством минимизации суммы ошибок и часто используется, например, в анализе доходов населения. В анализе результатов ЕГЭ помимо среднего и модального вычисляется также и медианный балл, позволяющий более точно определить характер распределения оценок.

Определение медианы производится строго по упорядоченному (по возрастанию) ряду распределения:

- если ряд – дискретный, то медиана – значение признака, которому соответствует кумулятивная частота, равная половине суммы обычных частот;

- в интервальном ряду – по формуле:  $Me = x_{me} + l_{me} \cdot \frac{\frac{1}{2} \sum f_i - S_{me-1}}{f_{me}}$ ,

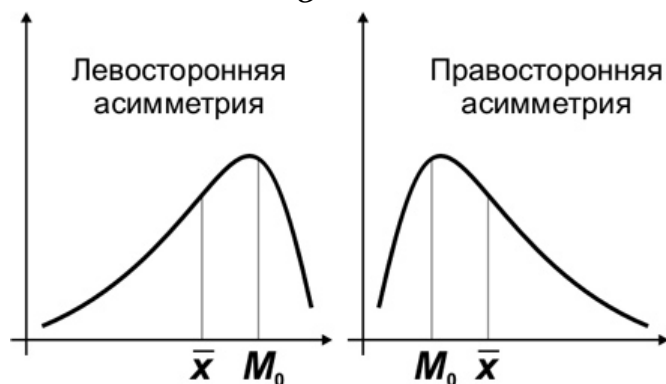
где  $x_{me}$  – нижняя граница медианного интервала,  $l_{me}$  – его величина,  $f_{me}$  – его частота,  $S_{me-1}$  – кумулятивная частота предыдущего интервала,  $\sum f_i$  – сумма обычных частот.

Медианным интервалом будет тот, в котором в первом по счету сверху кумулятивная частота превысит половину суммы обычных частот:  $S_i > \frac{1}{2} \sum f_i$ .

Следует помнить, что модальный и медианный интервалы определяются независимо друг от друга, несмотря на то, что на практике они часто совпадают, что происходит, если распределение относительно симметрично.

Среднюю арифметическую, моду и медиану называют показателями центра распределения. По их соотношению можно судить о симметричности или несимметричности распределения. Обычно (но не всегда!) эти соотношения таковы:  $\bar{x} \leq Me \leq Mo$  или  $\bar{x} \geq Me \geq Mo$ . Если все три показателя достаточно близки друг к другу:  $\bar{x} \approx Me \approx Mo$ , то можно говорить о симметричном характере распределения. В противном случае следует рассчитать **показатель асимметрии**. Его упрощенная

формула:  $As = \frac{\bar{x} - Mo}{\sigma}$ , где  $\sigma$  – среднее квадратическое отклонение (см. § 5.4).



Если  $As > 0$ , то асимметрия – правосторонняя; если  $As < 0$ , то – левосторонняя.

Распределение доходов в обществе, например, всегда имеет правостороннюю асимметрию (т.е. у большинства доходы меньше среднего).

## 5.4. Показатели вариации.

**Вариация** – это различие индивидуальных значений признака у отдельных единиц совокупности.

В научной литературе могут использоваться синонимы термина «вариация»: разброс (значений), колеблемость (признака), рассеивание и др.

Необходимость расчета показателей вариации обуславливается тем, что совокупности, имеющие одинаковые или очень близкие средние, могут сильно различаться по степени вариации. Чем меньше различия значений признака в совокупности, тем совокупность однороднее, а значение средней – точнее.

Таким образом, анализ показателей вариации имеет важное практическое значение: с его помощью можно определить пределы колеблемости признака, степень однородности совокупности, «надежность» средней и т.п.

Для более полной характеристики и сравнительного анализа совокупностей используют несколько – абсолютных и относительных – показателей вариации. Наиболее популярные из них представлены ниже.

#### □ Абсолютные показатели вариации:

1. **Размах вариации** – характеризует пределы колеблемости признака и определяется разницей между наибольшим и наименьшим значениями признака в совокупности:  $R = x_{\max} - x_{\min}$

2. **Среднее линейное отклонение** – характеризует насколько в среднем индивидуальные значения признака отличаются от среднего значения, и определяется как средняя арифметическая из абсолютных отклонений.

$$\text{Простая формула: } d = \frac{\sum |x_i - \bar{x}|}{n} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

$$\text{Взвешенная формула: } d = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i} = \frac{|x_1 - \bar{x}| \cdot f_1 + \dots + |x_n - \bar{x}| \cdot f_n}{f_1 + \dots + f_n}$$

☞ Если показатели вариации рассчитываются по интервальному ряду, то, как и в случае с расчетом средней, в качестве  $x_i$  необходимо брать середины интервалов.

Среднее линейное отклонение имеет определенный практический смысл: в интервал  $(\bar{x} - d, \bar{x} + d)$  попадает около половины всех единиц совокупности.

3. **Дисперсия** – наиболее популярный в математической статистике показатель вариации, характеризующий меру рассеивания значения признака.

$$\text{Простая формула: } D = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

$$\text{Взвешенная: } D = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{(x_1 - \bar{x})^2 \cdot f_1 + \dots + (x_n - \bar{x})^2 \cdot f_n}{f_1 + \dots + f_n}$$

$$\text{Дисперсию также можно рассчитать по формуле: } D = \overline{x^2} - (\bar{x})^2 = \frac{\sum x_i^2}{n} - \left( \frac{\sum x_i}{n} \right)^2.$$

4. **Среднее квадратическое (среднеквадратичное, стандартное) отклонение** – так же, как и линейное, характеризует усредненную величину отличий индивидуальных значений признака от среднего значения, но рассчитывается как средняя квадратическая из отклонений, тем самым придавая большую значимость большим отличиям.

Если дисперсия уже рассчитана, то среднеквадратическое отклонение проще всего рассчитать как корень из дисперсии:  $\sigma = \sqrt{D}$ .

Или использовать простую или взвешенную формулы соответственно:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}, \quad \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}.$$

Как правило, среднее квадратическое отклонение больше линейного, и в интервал  $(\bar{x} - \sigma, \bar{x} + \sigma)$  попадает около двух третей всех единиц совокупности.

Следует также помнить, что размах вариации  $R$ , среднее линейное отклонение  $d$  и среднее квадратическое отклонение  $\sigma$  – величины именованные, измеряются в тех же единицах, что и изучаемый признак  $x$ ; а дисперсия  $D$  – величина безразмерная, единицы измерения для нее не пишутся.

### □ Относительные показатели вариации.

**1. Относительное линейное отклонение** – предназначается для сравнения разных совокупностей по степени вариации изучаемого признака и определяется соотношением среднего линейного отклонения и средней.

$$k_d = \frac{d}{\bar{x}} \cdot 100\%$$

Чем относительное линейное отклонение  $k_d$  меньше, тем совокупность однороднее.

**2. Коэффициент вариации** – характеризует степень однородности совокупности, соизмеряя среднеквадратическое отклонение со средней.

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

Если коэффициент вариации  $v \leq 33,3\%$ , то совокупность считается однородной, в противном случае – разнородной.

### Пример.

Известны результаты ЕГЭ по математике двух выпускных классов школы: в классе А средний балл составил 46,5 при стандартном отклонении 16,3 балла, а в классе Б – 42,0 и 15,7 балла соответственно. В каком классе результаты ЕГЭ оказались более однородными?

#### Решение:

Для сравнения разных (в т.ч. и по объему) совокупностей следует использовать относительные показатели вариации.

В данном примере можно рассчитать коэффициенты вариации  $v$ :

$$v (\text{для класса А}) = \frac{16,3}{46,5} \cdot 100\% = 35,1\%; \quad v (\text{для класса Б}) = \frac{15,7}{42,0} \cdot 100\% = 37,4\%.$$

Получается, что оба класса разнородны по результатам ЕГЭ, но класс А – чуть менее разнороден, чем класс Б. Таким образом, класс А более успешно сдал ЕГЭ, чем класс Б, – и средний балл выше (на 10,7%) и совокупность учеников более однородна по данному показателю.



### Показатели вариации атрибутивных признаков.

Если рассматриваемый признак – атрибутивный или качественный по своей природе, то рассчитать его среднее значение невозможно и приведенные выше формулы показателей вариации теряют смысл. Однако и в этом случае можно сравнить совокупности по степени однородности, для чего следует применить **анализ вариации долей** (для оценки разницы между отдельными частями совокупности).

Пусть в совокупности выделено  $k$  групп, и  $w_i$  – доля  $i$ -ой группы. Тогда:

средняя доля  $\bar{w} = \frac{\sum_{i=1}^k w_i}{k}$ , дисперсия долей  $D_w = \overline{w^2} - (\bar{w})^2 = \frac{\sum w_i^2}{k} - \frac{1}{k^2}$ , а показатель

вариации долей  $v_w = \frac{\sqrt{D_w}}{\bar{w}} = \sqrt{k \sum w_i^2 - 1}$ . Чем  $v_w$  больше, тем совокупность однороднее.

#### Пример.

Известны итоговые результаты ЕГЭ по физике для Ярославской области:

Уровень достижений:	Доля выпускников, %	
	2009 г.	2010 г.
• ниже минимального	1,4	2,1
• минимальный	7,5	9,7
• низкий	18,9	25,6
• удовлетворительный	33,7	24,8
• хороший	23,9	19,7
• отличный	14,6	18,1

Определить, в каком году результаты более однородны.

Решение:

«Уровень достижений» в данном случае не является количественным показателем (он не выражен конкретной оценкой или баллом). Поэтому следует использовать показатели вариации долей:

$$v_w^{09} = \sqrt{k \sum w_i^2 - 1} = \sqrt{6 \cdot (0,014^2 + 0,075^2 + \dots + 0,146^2) - 1} = 0,633$$

$$v_w^{10} = \sqrt{k \sum w_i^2 - 1} = \sqrt{6 \cdot (0,021^2 + 0,097^2 + \dots + 0,181^2) - 1} = 0,501$$

Следовательно, совокупность сдававших физику выпускников 2009 г. оказалась более однородной, чем в 2010 г.

## Рекомендуемая литература.

---

- ✓ Мониторинг качества, статистика и социология образования: учебный курс / М.Л. Агранович, О.Я. Дымарская, Е.С. Заир-Бек, А.Г. Каспржак, О.Н. Кожевникова. – М.: Университетская книга, 2006. – 188 с.
- ✓ Писарева С.А., Иванов С.А., Пискунова Е.В. Мониторинг, статистика, социология в деятельности образовательного учреждения. Учебное пособие. – М.: НФПК, ООО «Миралл», 2005. – 184 с.
- ✓ Теория статистики. Учебник для вузов. / Под ред. Г.Л. Громыко. 2-е изд., перераб. и доп. – М.: Инфра-М, 2006, 2009, 2011. – 476 с.