

СТАТИСТИКА

*Методические указания к самостоятельной работе
для студентов бакалавриата направления 38.03.01*

САНКТ-ПЕТЕРБУРГ
2023

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования
Санкт-Петербургский горный университет

Кафедра экономики, организации и управления

СТАТИСТИКА

*Методические указания к самостоятельной работе
для студентов бакалавриата направления 38.03.01*

САНКТ-ПЕТЕРБУРГ
2023

УДК 330.4 (073)

СТАТИСТИКА: Методические указания к самостоятельной работе / Санкт-Петербургский горный университет. Сост. *Н.В. Василенко*. СПб, 2023. 60 с.

В методических указаниях содержатся материалы для самостоятельной работы по основным теоретическим разделам общей статистики, а также оценочные средства для подготовки к текущему контролю и экзамену по дисциплине «Статистика».

Предназначены для студентов направления подготовки 38.03.01 «Экономика».

Научный редактор проф. *А.Е Черепорвицын*.

Рецензент к.э.н. доцент *О.А. Пономарева* (ФГБОУ ВО «Санкт-Петербургский политехнический университет Петра Великого»)

© Санкт-Петербургский
горный университет, 2023

ВВЕДЕНИЕ

Овладение статистическим инструментарием в соответствии в ФГОС 3++ рассматривается как неотъемлемая часть подготовки бакалавров по направлению 38.03.01 – Экономика. Основной дисциплиной, решающей эту задачу, является «Статистика».

Цель изучения дисциплины «Статистика»:

– формирование у будущих бакалавров теоретических знаний и практических навыков о статистических методах и инструментарии в области сбора, анализа и интерпретации результатов исследования социально-экономических процессов и явлений на предприятиях, в отрасли и народном хозяйстве.

Основные задачи дисциплины:

– знать основные источники статистической информации с целью осуществления сбора, анализа и обработки данных, необходимых для решения профессиональных задач;

– умение использовать систему статистических показателей и процедур с целью обработки экономических данных в соответствии с поставленной задачей, проанализировать результаты расчетов и обосновать полученные выводы;

– овладение навыками выбора и применения инструментальных средств для обработки экономических данных в соответствии с поставленной задачей, анализа результаты расчетов и обоснования полученные выводы.

Самостоятельная работа обучающихся по дисциплине «Статистика» направлена на углубление и закрепление знаний, полученных на лекциях и других занятиях, выработку навыков самостоятельного активного приобретения новых, дополнительных знаний, подготовку к предстоящим учебным занятиям и промежуточному контролю в форме экзамена.

В первой части методических рекомендаций в сжатой форме изложены основные теоретические сведения, охватывающие основные темы по дисциплине «Статистика».

Предлагаемые студентам материалы разделены на девять разделов (табл.1).

Таблица 1

Тематика и трудоемкость самостоятельной работы

№ п/п	Тематика практических занятий	Трудоемкость в ак. часах
1	Методические основы статистики	4
2	Общие сведения о статистическом наблюдении	4
3	Сводка и группировка статистических данных	10
4	Обобщающие статистические показатели	6
5	Анализ вариационных рядов	10
6	Выборочные наблюдения	10
7	Статистическое изучение корреляционных зависимостей	10
8	Ряды динамики	8
9	Экономические индексы	10
	Итого:	72, в том числе 36 – подготовка к экзамену

Во второй части методических рекомендаций размещены вопросы для текущего контроля самостоятельной работы, программа экзамена и примерный перечень вопросов к экзамену.

Использование предлагаемых методических рекомендаций поможет студентам не только успешно освоить пройденный материал и подготовиться к экзамену, но и осмыслить возможности применения статистического инструментария для проведения самостоятельных научных исследований.

1. МАТЕРИАЛЫ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПО РАЗДЕЛАМ

1.1. Методические основы статистики

Изучаемые вопросы

1. Понятие статистики.
2. Краткий обзор развития статистики как науки.
3. Предмет и метод статистики.
4. Теория статистики как научная (учебная) дисциплина

Основные понятия

Термин «статистика» (от лат. *stato* – государство и *status* – положение вещей, политическое состояние) в научный обиход ввел немецкий ученый Готфрид Ахенваль (1719-1772), представитель описательной школы государственоведения. В настоящее время термин «статистика» употребляется в нескольких значениях:

- совокупность сведений (фактов) о разных явлениях в той или иной стране или ее регионах;
- процесс получения сведений с последующей их обработкой, то есть практический вид экономической деятельности, выполняемый специальными статистическими службами, в том числе Федеральная служба государственной статистики (ФСГС <http://www.gks.ru>);
- параметр ряда случайных величин, получаемый по определенному алгоритму из результатов индивидуальных наблюдений;
- наука, изучающая с количественной стороны массовые явления и их закономерности.

Развитие статистики как науки вначале шло по двум направлениям:

- *государствоведение* или *описательная школа*, считавшая основной задачей описание достопримечательностей государства: территории, населения, климата, политического устройства, вероисповедания, ведения хозяйства, торговли, благосостояния государства и граждан и т.п. без анализа закономерностей и взаимосвязей между явлениями. Основатель школы: немецкий ученый *Герман Конринг* (1606-1681). Основные представители: *Готфрид Ахенваль* (1719-1772), *Август Людвиг Шлёцер* (1735-1809);

– *политическая арифметика*, фокусирующаяся на выявлении на основе большого числа наблюдений различных закономерностей и взаимосвязей в изучаемых явлениях. Основатель школы: *Уильям Петти* (1623-1687). Основные представители: *Джон Граунт* (1620-1674), *Эдмунд Галлей* (1656-1742).

В России последователями «школы государственоведения» были *Иван Кириллович Кирилов* (1689-1737), *Василий Никитич Татищев* (1686-1750), *Михаил Васильевич Ломоносов* (1711-1765), *Карл Федорович Герман* (1767-1838), *Константин Иванович Арсеньев* (1789-1865) и др. В научной постановке статистического дела в России большая заслуга принадлежит Русскому географическому обществу (РГО), основанному в 1845 г. и имевшему в своем составе отделение статистики, впоследствии преобразованное в отделение экономической географии.

Задача теоретического обобщения практики учетно-статистических работ и создания теории статистики была решена в XIX в. бельгийским ученым *Адольфом Кетле* (1796-1874), который дал определение предмета статистики (массовые явления, связанные с жизнью общества, государства) и раскрыл суть методов статистики.

Под влиянием идей Кетле в XIX в. возникло третье направление статистической науки – *математико-статистическое*. Основные представители: *Фрэнсис Гальтон* (1822-1911), предложивший алгоритм вычисления коэффициента корреляции; *Карл Пирсон* (1857-1936), разработавший теорию корреляции, критерии согласия, алгоритмы принятия решений и оценки параметров, например хи-квадрат; *Одни Дж. Юл* (1871-1951), один из основателей анализа временных рядов; *Артур Бóули* (1869-1957), суммировавший все передовые математические методы того времени, предложенные Курно, Джевансом, Парето, Эджуортом, Маршаллом, Пигу; *Уильям Гóссет* (1876-1937), более известный под своим псевдонимом Стьюдент (Student) благодаря своим работам по исследованию т. н. распределения Стьюдента; *Рóналд Фишер* (1890- 1962), ставший продолжателем классических работ и методов Карла Пирсона; *Мóрис Кéндалл* (1907-1983), давший имя одному из коэффициентов ранго-

вой корреляции; *Коррадо Джини* (1884-1965), давший имя одному из коэффициентов концентрации признака и др.

Бурное развития статистической науки и практики во второй половина XIX — начало XX в. в России связана с деятельностью представителей т.н. *академической статистики*. К ним следует отнести:

– представителей математической школы Петербургского университета: *Пафнутия Львовича Чебышёва* (1821-1894), сформулировавшего закон больших чисел; *Андрея Андреевича Маркова* (1856—1922), создавшего т. н. схему цепей Маркова; *Александра Михайловича Ляпунова* (1857-1918), заложившего теоретические основы в практику выборочного наблюдения в статистике;

– собственно статистиков *Юрия Эдуардовича Янсона* (1835—1893), придававшему большое значение вопросам организации наблюдения и группировкам, а также выявлению причинно-следственных связей; *Александра Ивановича Чупрова* (1842-1908) и *Александра Александровича Чупрова* (1874—1926), развивавших т. н. стохастическую (вероятностную) статистику и др.

Статистика, как и любая другая наука, имеет свой предмет и метод исследования.

Изучаемые статистикой массовые явления в виде множества однокачественных единиц с отличающимися индивидуальными признаками называют *статистическими совокупностями*.

Элементы, множество которых образует изучаемую статистическую совокупность, называют *единицами совокупности*. Каждая единица совокупности может быть охарактеризована разного рода *качественными* (атрибутивными) и *количественными признаками*. Если определенный признак имеет разные значения у отдельных единиц совокупности, то говорят, что он варьирует или имеет некоторую *вариацию*.

Предметом статистики являются различные статистические совокупности, исследование которых связано с количественной характеристикой и выявлением присущих им закономерностей в конкретных условиях места и времени.

Закономерность, выявленная на основе массового наблюдения, т.е. проявившаяся в большой массе данных через преодоление

свойственной ее единичным элементам случайности, называется **статистической закономерностью**.

Массовое наблюдение, группировка и сводка его результатов, вычисление и анализ обобщающих показателей — все это вместе составляет специфический **метод статистики**.

Изучение статистических совокупностей связано с решением таких основных задач, как:

- получение итоговых данных по совокупности;
- определение структуры совокупности и соотношения отдельных ее частей;
- изучение особенности распределения единиц совокупности по отдельным признакам;
- определение средней величины того или иного количественного показателя и его вариации;
- выявление взаимосвязи между отдельными показателями (признаками);
- изучение динамики отдельных показателей (как единичных, так и агрегированных);
- оценка выборочных данных и пр.

1.2. Общие сведения о статистическом наблюдении

Изучаемые вопросы

1. Статистическое наблюдение как первый этап статистического исследования.
2. Организационные формы статистического наблюдения.
3. Виды и способы статистического наблюдения.
4. Программно-методологические и организационные вопросы статистического наблюдения.
5. Ошибки статистического наблюдения и контроль данных наблюдения.

Основные понятия

Любое законченное статистическое исследование проходит **три этапа**:

- первый этап – **статистическое наблюдение**, предполагающий научно организованный сбор сведений, заключающийся в регистрации тех или иных фактов, признаков, относящихся к единицам изучаемой совокупности;

– второй этап – **сводка**, заключающаяся в обработке собранных первичных данных о каждой единице совокупности, включающая их группировку, обобщение и оформление сведенных данных в таблицах;

– третий этап – **научный анализ исследуемых явлений**, подразумевающий расчет на основе итоговых данных сводки различных обобщающих показателей в виде средних и относительных величин, выявление определенных закономерностей в распределениях, динамике показателей и т.п.

Поскольку **статистическое наблюдение** является начальным этапом, то оно во многом определяет успех всей работы.

Организационные формы статистического наблюдения:

- статистическая отчетность;
- специально организованные статистические обследования (наблюдения);
- регистры.

Виды статистического наблюдения:

1) *по непрерывности* – текущее (непрерывное) и прерывное (единовременное или периодическое);

2) *по охвату единиц наблюдаемого объекта* – сплошное и несплошное наблюдение.

Различают следующие **виды несплошного наблюдения**:

- 1) наблюдение основного массива;
- 2) анкетное;
- 3) выборочное;
- 4) монографическое.

Способы статистического наблюдения:

1) *по источникам информации* – непосредственное наблюдение, документальное и опрос;

2) *по организации сбора информации* – экспедиционный; саморегистрация; корреспондентский; явочный.

Программно-методологические вопросы на подготовительном этапе статистического наблюдения включают:

- определение цели, объекта и единицы наблюдения;
- разработку программы наблюдения и статистического формуляра, содержащего ее.

При определении *объекта наблюдения* должны быть строго проведены его границы во времени, пространстве и материальной сущности.

Под *программой статистического наблюдения* понимается перечень тех признаков, которыми каждая единица наблюдения должна быть охарактеризована. Другими словами, это перечень вопросов, на которые в процессе наблюдения должны быть получены ответы.

Вопросы программы и ответы на них фиксируются в особых *статистических формулярах*, которые могут именоваться по-разному: переписной лист, бланк, форма и др. Статистические формуляры бывают двух видов:

– индивидуальные (карточные), которые заводятся на каждую единицу наблюдения отдельно, т.е. в каждом формуляре содержатся сведения лишь об одной единице наблюдения.

– списочные, когда один формуляр (список) составляется на несколько единиц.

Организационные вопросы на подготовительном этапе статистического наблюдения предполагают определение:

– субъекта наблюдения;

– места и времени наблюдения;

– организационной формы, вида и способа наблюдения.

В ходе статистического наблюдения могут возникнуть погрешности, именуемые ошибками. ***Ошибки регистрации***, которые могут возникнуть как при сплошном, так и при несплошном наблюдении, можно разделить на преднамеренные и непреднамеренные. Непреднамеренные, в свою очередь, могут носить случайный и систематический характер.

Случайные ошибки наблюдения, возникающие по вине отвечающего или по вине регистраторов в результате опечаток, оговорок, незнания и т.п., при большом числе наблюдений они взаимопогашаются, нейтрализуются.

Непреднамеренные систематические ошибки возникают главным образом при опросе за счет округлений количественных показателей (округление возраста, стажа работы, дохода и т.п.) или

за счет неточностей измерительных приборов при непосредственном наблюдении.

Преднамеренные ошибки возникают в силу сознательного стремления лиц, дающих сведения, уменьшить или увеличить величину того или иного показателя, как правило, в одном направлении и потому опасны.

Проверка правильности зафиксированных в статистических формулярах сведений должна проводиться посредством *логического* и *арифметического контроля*.

При несплошном наблюдении могут возникнуть *ошибки репрезентативности*, под которыми понимают расхождения между показателями несплошного и сплошного наблюдения. *Случайные ошибки репрезентативности* при выборочном наблюдении неизбежны, но поддаются учету специальными методами. *Систематические ошибки репрезентативности*, как правило, возникают при неправильной организации выборки.

1.3. Сводка и группировка статистических данных

Изучаемые вопросы

1. Понятие сводки и группировки.
2. Виды группировок.
3. Число групп и типы интервалов.
4. Вторичные группировки.
5. Статистические таблицы: структура и правила составления.

Основные понятия

Сводка данных, полученных в результате статистического наблюдения, состоит в систематизации, обработке и получении общих и групповых итогов, а также расчете производных показателей (средних и относительных величин).

По способу организации различают следующие *виды сводки*:

- *централизованную*, при которой все данные наблюдения сосредотачиваются в одном центре, где они обрабатываются;
- *децентрализованную*, когда первичные статистические материалы разрабатываются на уровне административных районов, итоги сводятся на уровне субъектов Российской Федерации, а затем на уровне государства в целом.

На практике имеет место сочетание децентрализованной и централизованной сводки.

Группировкой называется разбиение единиц статистической совокупности на группы, однородные по какому-либо одному или нескольким признакам. Группировка позволяет систематизировать данные статистического наблюдения, которые в результате преобразуются в упорядоченную статистическую информацию, пригодную для дальнейшего статистического анализа.

Каждая единица исследуемой совокупности обладает рядом свойств, или признаков. Отдельные значения, которые может принимать тот или иной варьирующий признак, называются его вариантами.

По характеру вариантов признаки делятся на:

- *атрибутивные*, если варианты значений признака не выражаются числами, например, пол, возраст, место нахождения и т.д.;
- *количественные*, если варианты значений выражаются в виде чисел, например, объем производства, количество детей в семье.

Признаки, на основе которых получена группировка, называются *группировочными*. Если группировка получена по количественному признаку, она называется *количественной*, если качественному – *качественной*.

От качественных следует отличать *типологические* группировки, поскольку некоторые типы явлений могут быть выделены и по количественному признаку. Например, группировка предприятий на малые, средние и крупные проводится по таким количественным признакам, как численность персонала, объем продукции, стоимость основных фондов, причем для разных видов деятельности значение этих признаков различно. Другим примером типологических группировок служит деление населения на такие группы, как молодежь, лица среднего возраста и др.

Пространственные группировки создаются по географическому признаку, при этом в основу группировок могут быть положены существующее административно-территориальное деление, природно-климатические зоны, части света и т.д. Данные, сгруппированные по *территориальному* признаку, представляют важный

информационный массив как для анализа явлений в пределах отдельных территорий, так и для сопоставления одних и тех же явлений (например, уровни цен и доходов, показатели рождаемости и смертности и др.) на различных территориях.

Для исследования зависимости между явлениями используют *аналитические* группировки. При их построении можно установить взаимосвязь между двумя признаками и более. Факторными называются признаки, под воздействием которых изменяются результативные признаки. В случае если изучается влияние на результат нескольких факторов, используют *многофакторную аналитическую* группировку.

Данные любой группировки соответствуют определенному моменту времени или периоду, так как с течением времени изменяется как численность совокупности, так и численность и соотношение отдельных ее групп.

При систематизации статистических данных большую роль играют *классификации*, под которыми понимают *устойчивое разбиение на группы проводится на основе свойств и различий элементов совокупности*. В отличие от классификации группировка проводится обычно для целей конкретной обследования.

Число групп при выполнении группировки зависит от того, является ли признак дискретной или непрерывной величиной.

В совокупности, где варьирующий признак носит *дискретный характер* и может принимать *ограниченное число значений*, количество групп, как правило, равно количеству возможных значений. Примером такой группировки служит распределение семей одного из городов по числу детей.

Если варьирующий признак является *непрерывной величиной* или *дискретной величиной, которая может принимать очень большое число значений* (например, численность работников на предприятии может изменяться от одного до нескольких тысяч), то число групп зависит от степени вариации данного признака, а также от объема изучаемой совокупности.

В общем случае на вопрос о том, насколько групп следует разбить совокупность, нет стандартного, однозначного ответа.

Если распределение признака в границах его вариации достаточно равномерно или близко к нормальному, диапазон колебаний признака разбивают на равные интервалы, длину которых определяют по формуле:

$$h = \frac{x_{max} - x_{min}}{k} \quad (1),$$

где x_{max} – максимальное значение признака в совокупности,
 x_{min} – минимальное значение признака в совокупности,
 k – количество групп.

Число групп может быть задано на основе опыта предыдущих обследований либо можно использовать формулу Стерджесса для определения оптимального числа *групп с равными интервалами*:

$$k = 1 + 3,222 \lg N \quad (2),$$

где N – число единиц в совокупности.

Полученное значение следует округлить для облегчения расчетов до 1000 руб. Процедуру округления при расчете интервала проводят всегда. Трехзначное, четырехзначное или большее число округляют до ближайшего числа, кратного 50 или 100. Если число имеет два знака до запятой и несколько знаков после запятой, его округляют до целого, если один знак до запятой и несколько знаков после запятой – до десятых и т.д.

Часто значения варьирующего признака распределены таким образом, что при использовании равного интервала для образования групп излишне увеличивается их количество, при этом многие группы будут малочисленными. В этих условиях совокупность разбивают на *группы с неравными интервалами*. Примером такой группировки может служить распределение населения по размеру среднедушевого дохода.

При образовании групп с неравными интервалами группировочного признака могут использоваться два *вида интервалов*:

- *закрытые*, у которых указаны обе границы;
- *открытые*, имеющие одну границу (верхнюю у первого интервала и нижнюю у последнего интервала).

Для расчета показателей статистической совокупности открытые интервалы необходимо «закрывать».

В целях статистического исследования часто приходится пользоваться данными, относящимися к различным периодам, сопоставлять информацию по отдельным отраслям, регионам, странам, опираясь на уже сгруппированные данные, причем сгруппированными, как правило, на разной основе.

Операция перегруппировки, т.е. образование новых групп на базе ранее созданных группировок, называется вторичной группировкой.

В ходе укрупнения интервалов количественных группировок некоторые из них целиком войдут во вновь образованные интервалы. Другие интервалы придется разбивать на части согласно новым границам. При этом в новом интервале число единиц признака будет пропорционально части старого интервала, которая попадает в соответствующий новый интервал.

Для совокупностей, сгруппированных не только по количественным, но и по качественным признакам, например, при сопоставлении данных разных стран, сначала разрабатывают ключи перехода от национальной классификации к международной, затем на этой основе проводят перегруппировку данных национальной статистики.

Результаты сводки и группировки статистических данных оформляются в *статистические таблицы*, представляющие собой форму наиболее краткого и рационального изложения цифровых данных об изучаемой статистической совокупности. Незаполненная цифрами статистическая таблица называется *макетом*.

Каждая статистическая таблица содержит *подлежащее* и *сказуемое*. *Подлежащим* таблицы называется объект, отдельные единицы или его части (группы), которые характеризуются соответствующими показателями. *Сказуемым* называются показатели, которые характеризуют подлежащее. Подлежащее таблицы обычно составляет название ее строк, сказуемое – название колонок.

По характеру подлежащего различают *простые, групповые* и *комбинационные* таблицы.

Разработка сказуемого статистической таблицы может быть *простой* и *сложной*.

1.4. Обобщающие статистические показатели

Изучаемые вопросы

1. Обобщающие статистические показатели.
2. Абсолютные величины.
3. Понятие и виды относительных величин. Относительные величины структуры, динамики, сравнения, интенсивности и координации.
4. Понятие и виды средних величин: средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая.

Основные понятия

В результате сводки данных статистического наблюдения получают различные показатели, одни из которых характеризуют совокупность в целом, другие – отдельные ее части. Под *статистическим показателем* понимаются обобщающая количественная характеристика изучаемого объекта или его свойства.

В зависимости от методов расчета обобщающие показатели могут быть *абсолютными*, *относительными* или *средними величинами*.

При этом любой статистический показатель должен быть точно определенным, что выдвигает ряд требований к его наименованию. В нем должны быть указаны:

- статистическая структура (среднее значение, сумма, процент к итогу и т.д.);
- содержание (население, инвестиции, объем добычи и т.д.);
- позиция в классификации, совокупность объектов (обрабатывающая промышленность России, предприятия угольной промышленности Кузнецкого бассейна и т.п.);
- единица измерения (человек, тонна, километр и др.);
- время (на начало года, за 2010—2020 гг. и т.п.);
- специальные уточнения (в рыночных ценах 2020 г. и пр.).

Абсолютные обобщающие показатели – это число единиц по совокупности в целом или по ее отдельным группам, которое получают *в результате суммирования* зарегистрированных значений признаков первичного статистического материала.

Абсолютные величины как обобщающие показатели могут характеризовать:

– *численность* совокупности, например, численность экономически активного населения, количество предприятий различных форм собственности и т.д.;

– *объем* признаков совокупности, например, размер инвестиций, затраты на рабочую силу и т.д.

Абсолютный обобщающий показатель может иметь следующие виды *единиц измерения*:

– *натуральные* в том числе простые (тонна, штука, квадратный и кубический метр, километр и т.д.) и сложные, представляющие собой комбинацию двух величин (тонно-километр, киловатт-час, и др.), а также измеренные в единицах труда (человеко-дней или человеко-часов);

– *условно-натуральные*, применяемые для получения абсолютных обобщающих показателей, когда отдельные группы слагаемых, входящие в совокупность, не поддаются непосредственному суммированию, например, различные виды топлива соизмеряют по условному топливу с теплотворной способностью 7000 ккал/кг, продукты химической промышленности, руды металлов – по содержанию полезного вещества и т.д.;

– *стоимостные*, позволяющие соизмерять величины в денежной форме.

Относительная обобщающая величина представляет собой результат деления (сравнения) двух величин. В числителе дроби стоит величина, которую сравнивают, в знаменателе – величина, с которой сравнивают.

Последняя называется *базой* (или *основанием*) сравнения. В случае если основание принимается за 100, относительная величина выражается в *процентах* (%), если за 1000 – в *промилле* (‰), если за 10000 – в *продецимилле* (‱).

Различают относительные величины:

– *структуры*, получаемые путем деления численности каждой группы, входящей в совокупность, на численность всей совокупности; показывающие удельный вес каждой группы в общей численности совокупности; дающие возможность сопоставлять

структуры одной и той же совокупности в различные моменты времени или сопоставлять структуры совокупностей, объем которых различен, например, доля городского населения в общем населении страны;

– *динамики*, являющиеся результатом сопоставления уровней одного и того же явления, относящихся к различным периодам или моментам времени, требующие соблюдения сопоставимости данных, например, рост добычи нефти в 2020 г. по сравнению с 2019 г.;

– *сравнения*, получаемые в результате сопоставления одноименных абсолютных показателей, относящихся к разным совокупностям, например, можно сравнить уровень потребления в расчете на душу населения жителями Ленинградской и Новгородской областей во втором квартале текущего года и т.д.;

– *интенсивности*, получаемые при сопоставлении:

либо разноименных признаков одной совокупности, например, коэффициент рождаемости (число родившихся в расчете на 1000 человек населения), уровень занятости (отношение числа занятых к численности экономически активного населения), выражаемые обычно в процентах, промилле и т.п.;

либо объектов двух связанных между собой совокупностей, единицы измерения которых определяются показателями, на основе которых они рассчитаны, например, плотность населения (число людей, приходящихся на 1 км² территории), фондоотдача (стоимость продукции, произведенной на 1 руб. основных фондов) и т.п.;

– *координации*, получаемые как соотношение между частями одного целого, например, соотношение числа мужчин и женщин, отношение численности неработающих лиц к численности занятого населения, отношение стоимости импортных продуктов питания к стоимости отечественного продовольствия и т.п.

Средняя величина – это обобщающая характеристика множества индивидуальных значений некоторого количественного признака. Средняя, являясь функцией множества индивидуальных значений, представляет одним значением всю совокупность и отражает то общее, что присуще всем ее единицам. В среднем значении отклонения, характерные для индивидуальных значений, погашаются.

Различают **два типа** средних величин:

– средняя *типическая*, которая характеризует качественно однородные группы изучаемого явления;

– средняя *системная*, являющаяся обобщающей характеристикой не только разнородных пространственных, например, средняя величина ВВП на душу населения в РФ, но и динамических систем, например, среднегодовая или среднемесячная температура воздуха той или иной местности.

В статистике используются различные **формы средних величин**: средняя *арифметическая*; средняя *гармоническая*; средняя *геометрическая*; средняя *квадратическая*. Выбор той или иной формы средней зависит от содержания определяемого признака и конкретных данных, по которым ее приходится вычислять.

Указанные средние величины могут быть вычислены в двух ситуациях:

– если каждый вариант признака в данной статистической совокупности встречается только один раз, может быть вычислена *простая средняя*,

– если варианты признака в данной статистической совокупности повторяются различное число раз, может быть вычислена *взвешенная средняя*. При этом число повторений вариантов называется *частотой* или *статистическим весом*.

1.5. Анализ вариационных рядов

Изучаемые вопросы

1. Основные характеристики вариационного ряда: частота, частность, плотность и др.

2. Графическое изображение вариационных рядов: полигон, гистограмма, кумулята и огива. Кривая Лоренца.

3. Показатели среднего уровня вариационного ряда. Мода. Медиана. Показатели дифференциации и концентрации вариационного ряда.

4. Показатели вариации признаков.

5. Межгрупповая и внутригрупповая дисперсии. Эмпирический коэффициент детерминации и эмпирическое корреляционное отношение.

Основные понятия

Ряды распределения, построенные по количественному признаку, называются *вариационными*.

По своей конструкции вариационный ряд состоит из двух столбцов (граф): один столбец – это значения варьирующего признака (x – варианты), другой – частоты (m – абсолютное число случаев данного варианта) или частоты (w – относительная доля каждой частоты в общей сумме частот).

Вариационные ряды по способу построения бывают **двух видов**: дискретные и интервальные. Интервальные вариационные ряды могут иметь интервалы равной и неравной длины.

Для построения *дискретного ряда* в первом столбце перечисляются отдельные значения признака, во втором указывается их частота.

Для построения *интервального ряда* определяют прежде всего число групп (7-10), на которые следует разбить всю совокупность. Группы должны содержать по меньшей мере три статистические единицы, а доля отдельной единицы в таком агрегате не должна превышать 85% всего агрегата.

Открытые интервалы закрывают. Для этого длина первого интервала приравнивается к длине второго, а длина предпоследнего – к длине последнего.

Накопленная частота показывает число единиц совокупности, у которых значение варианта не больше данного. Накопленная частота для данного варианта или для верхней границы данного интервала получается суммированием (накапливанием) частот всех предшествующих интервалов, включая данный.

Если вместо абсолютных частот использовать частоты, то аналогично получим *накопленные частоты*. Ряд частоты обычно применяют, когда совокупность очень велика. Кроме того, они позволяют сравнивать распределения по одному и тому же признаку в разных по численности совокупностях.

Абсолютная плотность распределения – это частота, приходящаяся на единицу длины интервала, может быть выражена в долях или процентах и рассчитана по формуле (3):

$$p_i = m_i/h_i \quad (3),$$

где m_i – частота, а h_i – длина i -го интервала.

Относительная плотность распределения – это частость, приходящаяся на единицу длины интервала, может быть выражена в долях или процентах и рассчитана по формуле (4):

$$p_i = w_i/h_i \quad (4),$$

где w_i – частота, а h_i – длина i -го интервала.

Плотность распределения используется в рядах с неравными интервалами для расчета такой характеристики, как мода, или для графического изображения вариационного ряда в виде гистограммы.

Графически вариационный ряд можно изобразить, как и любой ряд значений аргумента и функции, используя прямоугольную систему координат и строя точки с координатами (x_1, m_1) , (x_2, m_2) , ..., (x_n, m_n) или (x_i, w_i) . Если затем последовательно соединить полученные точки отрезками прямой, а из первой и последней точки опустить перпендикуляры на ось X, получим замкнутую фигуру в виде многоугольника, которая называется *полигоном* (рис. 1).

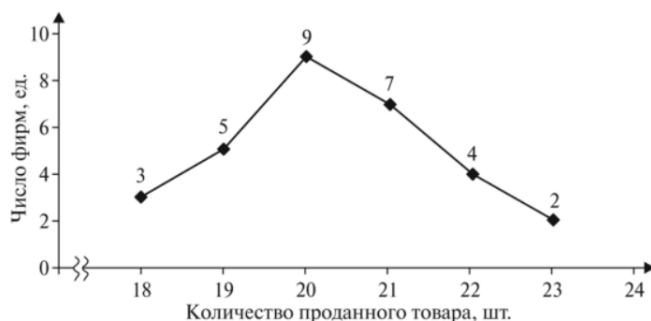


Рис. 1. Полигон распределения фирм по количеству проданного товара

Интервальный вариационный ряд изображают в виде *гистограммы*. Для *интервального ряда с равными интервалами* на оси x откладывают отрезки, равные длине интервала. На этих отрезках, как на основаниях, строят прямоугольники, высота которых пропорциональна частоте или частости (рис. 2).

Для *интервального ряда с неравными интервалами* на оси ординат откладывают плотности распределения, так как в этом слу-

чае именно плотность дает представление о заполненности каждого интервала.

Количество работников

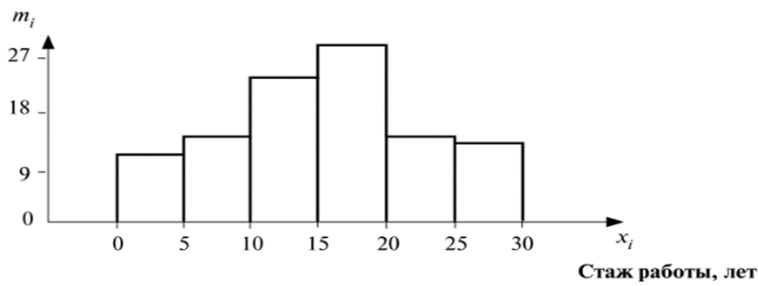


Рис. 2. Гистограмма распределения работников по стажу работы

Площадь всей гистограммы численно равна сумме частот, или численности единиц в совокупности (если на оси ординат отложить частоты).

Вариационный ряд можно представить графически в виде кривой накопленных частот (или частостей), которую называют **кумулятой**. При этом на оси x откладывают варианты или верхние границы интервалов, а на *оси* y – соответствующие накопленные частоты (или частости) (рис.3).

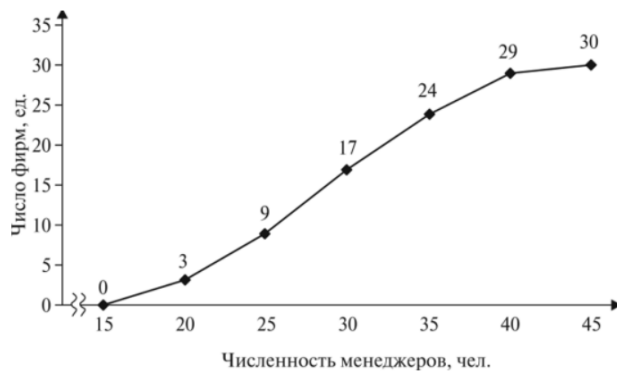


Рис. 3. Кумулята распределения фирм по численности менеджеров.

Если значения x (варианты) откладывать на оси y , а накопленные частоты (или частоты) – на оси x , то построенная на них кумулятивная кривая называется **огивой**.

Если при исследовании вариационных рядов нас интересует параллельное изменение нарастающих долей единиц совокупности и нарастающих долей значений признака в общем объеме используют так называемую **кривую Лоренца**. Она строится на квадрате 100×100 , где по оси абсцисс откладывают накопленные частоты, характеризующие распределение единиц совокупности, а по оси ординат – накопленные доли значений признака в общем объеме.

Для характеристики центра распределения вариационного ряда применяются показатели, получившие название **средних величин**. Наиболее часто применяются:

– *средняя арифметическая простая*:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_i x_i \quad (5),$$

где x_i – значение i -го варианта, а n – количество наблюдений.

– *средняя арифметическая взвешенная*:

$$\bar{x} = \frac{x_1 * m_1 + x_2 * m_2 + \dots + x_n * m_n}{n} = \frac{\sum_i x_i * m_i}{\sum_i m_i} \quad (6),$$

или

$$\bar{x} = \frac{x_1 * w_1 + x_2 * w_2 + \dots + x_n * w_n}{n} = \frac{\sum_i x_i * w_i}{\sum_i w_i} \quad (7),$$

где m_i – частота i -го варианта, а w_i – частота i -го варианта.

– *средняя гармоническая простая*:

$$\bar{x} = \frac{n}{\sum_i \frac{1}{x_i}} \quad (8),$$

– *средняя гармоническая взвешенная*:

$$\bar{x} = \frac{V_1 * x_1 + V_2 * x_2 + \dots + V_n * x_n}{V_1 + V_2 + \dots + V_n} = \frac{\sum_i V_i x_i}{\sum_i V_i} \quad (9),$$

где V_i – частота i -го варианта.

Среднюю гармоническую применяют в тех случаях, когда изучаемые показатели связаны как x и $1/x$, например затраты вре-

мени на единицу продукции и выработка продукции в единицу времени.

Мода – это значение признака, которое чаще всего встречается в вариационном ряду. Моду наиболее удобно применять при изучении *рядов с неопределенными границами*.

Для **дискретного ряда** мода находится непосредственно по определению – от наибольшего значения частоты.

Для **интервального ряда с равными интервалами** сначала определяется модальный интервал $[x_{k-1} - x_k]$, которому соответствуют максимальная частота m_k или частота w_k . Затем мода вычисляется по формуле

$$M_0 = x_{k-1} + h_k \frac{m_k - m_{k-1}}{(m_k - m_{k-1}) + (m_k - m_{k+1})} \quad (10),$$

где x_{k-1} – нижняя граница, а h_k – длина модального интервала.

Для **ряда с неравными интервалами** в формуле моды вместо частот m_{k-1} , m_k , m_{k+1} следует взять плотности распределения y_{k-1} , y_k , y_{k+1} .

Графически моду определяют по гистограмме распределения. Для этого выбирают самый высокий прямоугольник, который и является модальным, далее верхнюю правую вершину модального прямоугольника соединяют с верхней правой вершиной предшествующего прямоугольника, а верхнюю левую вершину модального прямоугольника – с верхней левой вершиной последующего прямоугольника. Абсцисса точки пересечения этих отрезков и будет модой распределения.

Медианой называют такое значение признака, которое приходится на середину ранжированного ряда. Таким образом, в ранжированном ряду распределения одна половина ряда имеет значения признака больше медианы, другая – меньше медианы.

В **дискретном ряду** медиана находится непосредственно по определению на основе накопленных частот.

Для **интервального ряда** сначала при помощи накопленных частот определяется медианный интервал $[x_{k-1} - x_k]$. Затем медиана вычисляется по формуле

$$M_e = x_{k-1} + h_k \frac{1/2 \sum_i m_i - F_{k-1}}{m_k} \quad (11),$$

где F_{k-1} – накопленная частота интервала, предшествующего медианному.

Если вместо частот использовать частоты, медиану можно найти по формуле (12):

$$M_e = x_{k-1} + h_k \frac{1/2 \sum_i w_i - P_{k-1}}{w_k} \quad (12),$$

где P_{k-1} – накопленная плотность интервала, предшествующего медианному.

Графически медиану определяют с помощью кумуляты.

К показателям **дифференциации** признака в вариационном ряду относятся квартили и децили, расширяющие понятие медианы.

К показателям **концентрации** признака в вариационном ряду относятся коэффициенты Джини, Герфиндаля, Лоренца и др.

Для анализа вариационных рядов большое значение имеет оценка рассеяния значений признака. К **показателям вариации** относятся:

– *размах вариации* (13):

$$R = x_{max} + x_{min} \quad (13),$$

– *среднее линейное отклонение* (14):

$$\bar{d} = \frac{\sum_i |x_i - \bar{x}| * m_i}{\sum_i m_i} \quad (14),$$

– *дисперсия* (15):

$$D = \sigma^2 = \frac{\sum_i (x_i - \bar{x})^2 * m_i}{\sum_i m_i} = \overline{x^2} - (\bar{x})^2 \quad (15),$$

– *среднее квадратическое отклонение* (16):

$$\sigma = \sqrt{\frac{\sum_i (x_i - \bar{x})^2 * m_i}{\sum_i m_i}} \quad (16),$$

– *коэффициент вариации* (17):

$$V = \frac{\bar{d}}{\bar{x}} \quad \text{либо} \quad V = \frac{\sigma}{\bar{x}} \quad (17).$$

Если статистическая совокупность разбита на группы по какому-либо признаку и для этих групп известны (или могут быть найдены) средний уровень и дисперсия, то нередко при объединении частных групп в совокупность требуется оценить вариации показателей объединенной совокупности на основе показателей отдельных частных групп. При этом необходимо учитывать, что вариация признака в целом по совокупности зависит как от вариации признака

внутри каждой группы, так и от вариации групповых средних, т.е. от межгрупповой вариации признака. Другими словами, общую дисперсию, характеризующую вариацию признака под влиянием всех факторов, можно получить как сумму ее составляющих – **межгрупповой и внутригрупповой дисперсий**. Это так называемое **правило сложения дисперсий**.

Правило сложения дисперсий позволяет выявить зависимость результатов от определяющих факторов при помощи:

– *эмпирического коэффициента детерминации* (18):

$$\eta_{\text{эмп}}^2 = \frac{\delta^2}{\sigma_{\text{общ}}^2} \quad (18),$$

– *эмпирического корреляционного отношения*, получаемого путем извлечения квадратного корня из эмпирического коэффициента детерминации.

Величина эмпирического корреляционного отношения изменяется от 0 до 1. Если связь отсутствует, то $\eta_{\text{эмп}} = 0$. В этом случае дисперсия групповых средних равна нулю и межгрупповой вариации нет, то есть группировочный признак не влияет на вариацию исследуемого признака. Чем больше значение корреляционного отношения приближается к единице, тем влияние группировочного признака сильнее.

1.6. Выборочные наблюдения

Изучаемые вопросы

1. Понятие и этапы выборочного наблюдения.
2. Виды случайного отбора: случайный повторный и бесповторный, систематический (механический), типический (стратифицированный, расслоенный), серийный (гнездовой).
3. Ошибки выборки при разных видах отбора.
4. Определение необходимой численности выборки при разных способах отбора.
5. Особенности малой выборки.

Основные понятия

Вся совокупность единиц, из которой осуществляется отбор, называется *генеральной совокупностью* (с количеством единиц N и средним значением признака \bar{x}), а единицы, отобранные для непосредственного наблюдения, представляют собой *выборочную сово-*

купность, или просто *выборку* (с количеством единиц n и средним значением признака \bar{x}).

Отбор из генеральной совокупности проводится таким образом, чтобы на основе выборки можно было получить достаточно точное представление об основных параметрах совокупности в целом.

Выборочный метод имеет **ряд преимуществ**:

- значительная экономия материальных и финансовых ресурсов при проведении статистического наблюдения;
- высокая достоверность получаемых данных;
- возможность провести исследование в случае, когда сплошное наблюдение связано с уничтожением или порчей обследуемых единиц.

Основные этапы выборочного наблюдения:

- 1) определение цели, задач и составление программы наблюдения;
- 2) анализ информационных источников, используемых для выделения генеральной совокупности объектов наблюдения (основы выборки);
- 3) формирование генеральной совокупности для проведения выборочного обследования;
- 4) разработка методологии формирования выборочной совокупности, включающей выбор *способа отбора*, определение необходимого *объема выборки*, этапов отбора единиц из генеральной совокупности, планирование и проведение пробной выборки;
- 5) формирование выборки;
- 6) сбор данных на основе разработанной программы;
- 7) анализ полученных результатов и расчет основных характеристик выборочной совокупности;
- 8) расчет *ошибок выборки* и распространение ее результатов на генеральную совокупность.

Различают следующие **виды отбора**:

- **случайный повторный** отбор, при котором каждая единица, отобранная в случайном порядке из генеральной совокупности, после проведения наблюдения возвращается в эту совокупность и может быть вновь подвергнута обследованию;

– **случайный бесповторный** отбор, при котором обследованные единицы в генеральную совокупность не возвращаются;

– **систематический (механический) отбор**, основанный на предварительном упорядочении генеральной совокупности на основе процента отбора и шага отбора;

– **типический (стратифицированный, расслоенный) отбор**, при котором обследуемая совокупность предварительно разбивается на типически однородные группы и выбор осуществляется из каждой такой группы механическим или собственно случайным способом;

– **серийный (гнездовой) отбор**, при котором генеральная совокупность разбита на группы еще до начала выборочного обследования и из генеральной совокупности можно отбирать не отдельные единицы, а целые их серии (с численность S) и обследовать в рамках каждой серии все попавшие в нее единицы (численностью s);

– **многоступенчатый отбор**, при котором на первом этапе из совокупности отбираются укрупненные единицы (серии), а затем без проведения наблюдения за всеми единицами в рамках серии осуществляется собственно случайный или механический отбор единиц из каждой отобранной серии;

– **многофазный отбор**, при котором из единиц совокупности, отобранных на первом этапе, осуществляется подвыборка в целях изучения дополнительных характеристик обследуемой совокупности.

Распространяя результаты выборочного обследования на генеральную совокупность, следует иметь в виду, что между характеристиками генеральной и выборочной совокупности возможно расхождение, обусловленное тем, что обследуется не вся совокупность, а лишь ее часть. Для оценки таких расхождений применяют разные показатели ошибки.

Средняя ошибка выборки (μ) – это среднее квадратическое отклонение всех возможных значений выборочной средней от генеральной средней, т.е. от своего математического ожидания.

Формулы средней ошибки выборки для разных видов отбора систематизированы в табл. 2.

Таблица 2

Формулы средней ошибки выборки при разных видах отбора

Оцениваемый параметр	Повторный отбор	Бесповторный отбор
Собственно случайный и механический отбор		
Средняя	$\mu = \sqrt{\frac{\sigma^2}{n}}$	$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$
Доля	$\mu = \sqrt{\frac{w(1-w)}{n}}$	$\mu = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Типический отбор		
<i>Пропорциональное распределение единиц выборочной совокупности по группам</i>		
Средняя	$\mu = \sqrt{\frac{\sigma_i^2}{n}}$	$\mu = \sqrt{\frac{\sigma_i^2}{n} \left(1 - \frac{n}{N}\right)}$
Доля	$\mu = \sqrt{\frac{w_i(1-w_i)}{n}}$	$\mu = \sqrt{\frac{w_i(1-w_i)}{n} \left(1 - \frac{n}{N}\right)}$
<i>Оптимальное распределение единиц выборочной совокупности по группам</i>		
Средняя	$\mu = \frac{1}{N} \sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i}}$	$\mu = \frac{1}{N} \sqrt{\sum \left[\frac{\sigma_i^2 N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right) \right]}$
Доля	$\mu = \frac{1}{N} \sqrt{\sum \frac{w_i(1-w_i)N_i^2}{n_i}}$	$\mu = \frac{1}{N} \sqrt{\sum \left[\frac{w_i(1-w_i)N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right) \right]}$
Серийный отбор		
Средняя	$\mu = \sqrt{\frac{\delta_x^2}{s}}$	$\mu = \sqrt{\frac{\delta_x^2}{s} * \frac{S-s}{S-1}}$
Доля	$\mu = \sqrt{\frac{\delta_w^2}{s}}$	$\mu = \sqrt{\frac{\delta_w^2}{s} * \frac{S-s}{S-1}}$

где w – доля единиц в выборке, обладающих изучаемым признаком; $(1-w)$ – доля единиц в выборке, не обладающих изучаемым признаком; δ_x^2 – межсерийная дисперсия при равновеликих сериях; δ_w^2 – межсерийная дисперсия для доли.

Предельной ошибкой выборки (Δ) называется отклонение выборочной характеристики от генеральной, рассчитанной в долях средней ошибки с заданной вероятностью (19):

$$\Delta = t * \mu \quad (19),$$

где t – коэффициент доверия, зависящий от вероятности, с которой определяется ошибка выборки.

Опираясь на теоремы теории вероятности можно показать, что вероятность того, что отклонение средней в генеральной совокупности от выборочной средней не выйдет за пределы 2μ (т.е. $t = 2$), равна 0,954, а вероятность того, что оно не выйдет за пределы 3μ , - соответственно 0,997. На практике доверительная вероятность принимается чаще всего на уровне 0,95 и 0,99 с соответствующими коэффициентами доверия 1,96 и 2,58 соответственно.

Относительная предельная ошибка ($\Delta_{отн}$) определяется по формуле (20):

$$\Delta_{отн} = \frac{\Delta}{\bar{x}} \quad \text{или} \quad \Delta_{отн} = \frac{\Delta}{w} \quad (20),$$

где \bar{x} – среднее значение признака по выборке, w – доля единиц в выборке, обладающих изучаемым признаком.

При любом способе отбора предельная ошибка выборки обратно пропорциональна числу обследованных единиц. Чтобы уменьшить ошибку выборки, необходимо увеличить ее объем, но при этом возрастут и затраты на проведение обследования.

Формулы **необходимой численности выборки** для разных видов отбора систематизированы в табл. 3.

Таблица 3

Формулы необходимой численности выборки при разных видах отбора

Оцениваемый параметр	Повторный отбор	Бесповторный отбор
Собственно случайный и механический отбор		
Средняя	$n = \frac{t^2 \sigma^2}{\Delta^2}$	$n = \frac{N t^2 \sigma^2}{N \Delta^2 + t^2 \sigma^2}$
Доля	$n = \frac{t^2 p q}{\Delta^2}$	$n = \frac{N t^2 p q}{N \Delta^2 + t^2 p q}$
Типический отбор		
Средняя	$n = \frac{t^2 \overline{\sigma_i^2}}{\Delta^2}$	$n = \frac{N t^2 \overline{\sigma_i^2}}{N \Delta^2 + t^2 \overline{\sigma_i^2}}$
Доля	$n = \frac{t^2 \overline{p_i q_i}}{\Delta^2}$	$n = \frac{N t^2 \overline{p_i q_i}}{N \Delta^2 + t^2 \overline{p_i q_i}}$

Серийный отбор		
Средняя	$s = \frac{t^2 \delta_{\bar{x}}^2}{\Delta^2}$	$s = \frac{St^2 \delta_{\bar{x}}^2}{(S-1)\Delta^2 + t^2 \delta_{\bar{x}}^2}$
Доля	$s = \frac{t^2 \delta_p^2}{\Delta^2}$	$s = \frac{St^2 \delta_p^2}{(S-1)\Delta^2 + t^2 \delta_p^2}$

где p – доля единиц в серии, обладающих изучаемым признаком; q – доля единиц в серии, не обладающих изучаемым признаком; $\delta_{\bar{x}}^2$ – межсерийная дисперсия при равновеликих сериях; δ_p^2 – межсерийная дисперсия для доли.

На практике иногда приходится отбирать из генеральной совокупности небольшое число единиц. Как правило, **выборка считается малой**, если обследуется не более 30 единиц.

Средняя ошибка малой выборки при собственно случайном или механическом повторном отборе рассчитывается по формуле (21):

$$\mu = \sqrt{\frac{\sigma^2}{n-1}} \quad (21).$$

Для определения предельной ошибки следует использовать вероятности, соответствующие не нормальному распределению, а распределению Стьюдента.

1.7. Статистическое изучение корреляционных зависимостей

Изучаемые вопросы

1. Понятие и методы выявления корреляционной связи.
2. Изучение связи между качественными признаками на основе таблиц сопряженности.
3. Показатели тесноты связи между качественными и количественными признаками.
4. Коэффициенты корреляции рангов. Коэффициент конкордации.

Основные понятия

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

Соответственно, первые называют *факторными*, а вторые – *результативными*.

Следует различать два **вида связи между признаками**:

– *функциональную (детерминированную)*, когда определенному значению переменной x строго соответствует одно или несколько значений другой переменной y и с изменением значения x значение y меняется строго определенно;

– *стохастическую (статистическую)*, обнаруживаемую только при массовом наблюдении.

Частным случаем стохастической связи является **корреляционная**, под которой понимают связь, проявляющаяся при большом числе наблюдений в виде определенной зависимости между средним значением результативного признака и признаками-факторами. Корреляционную связь условно можно рассматривать как своего рода функциональную связь средней величины одного признака (результативного) со значением другого (или других).

Если рассматривается связь средней величины результативного показателя y с одним признаком-фактором x , корреляция называется *парной*, а если факторных признаков два и более (x_1, x_2, \dots, x_n) – *множественной*. При изучении множественной корреляции под *частной корреляцией* понимается зависимость между результативным показателем y и одним из факторных признаков x_i , когда влияние на них остальных факторов, учитываемых на фиксированном уровне, устранено. По характеру изменений x и y в парной корреляции различают *прямую* и *обратную* корреляцию.

Изучение корреляционных связей сводится к решению следующих задач:

– выявление наличия (или отсутствия) корреляционной связи между изучаемыми признаками;

– измерение тесноты связи между двумя (и более) признаками с помощью специальных коэффициентов (*корреляционный анализ*);

– построение и верификация уравнения регрессии (*регрессионный анализ*).

Теория корреляции начала разрабатываться во второй половине XIX в. и особенно расцвета достигла в XX в. Основополож-

никами теории корреляции являются английские биометрики Ф. Гальтон и К. Пирсон. В России их идеи получили развитие в трудах А.А. Чупрова.

Для выявления наличия и характера корреляционной связи в статистике используются следующие **методы**:

– параллельное рассмотрение значений x и y в каждой из n единиц, как правило, сопровождаемое расчетом коэффициента Фехнера, ранговых коэффициентов корреляции или линейного коэффициента корреляции;

– графический метод, предполагающий построение *эмпирической линии регрессии* или *корреляционного поля*;

– метод *аналитических группировок* и *корреляционных таблиц*, дополненный расчетом эмпирического корреляционного отношения (см. п 1.5);

– расчет коэффициентов корреляции.

Коэффициент Фехнера основан на сравнении поведения отклонения индивидуальных значений каждого признака (x и y) от своей средней величины. При этом во внимание принимаются знаки отклонений. Если совпадение знаков обозначить символом С, а несовпадений Н, то коэффициент Фехнера можно рассчитать по формуле (22):

$$K_{\Phi} = \frac{\sum C - \sum H}{\sum C + \sum H} \quad (22).$$

Коэффициент Фехнера принимает значения в диапазоне от 0 (связи нет) до ± 1 связь прямая (+1) или обратная (-1). Поскольку коэффициент Фехнера зависит только от знаков, он характеризует не столько тесноту связи, сколько ее наличие и направление.

Корреляционную зависимость для наглядности можно изобразить графически. Для этого, имея n взаимосвязанных пар значений x и y , пользуясь прямоугольной системой координат, каждую такую пару изображают в виде точки на плоскости с координатами x и y . Соединяя последовательно нанесенные точки, получают ломаную линию, именуемую *эмпирической линией регрессии* (рис. 4).

При большом числе наблюдений для выявления корреляционной связи между двумя количественными показателями x и y удобнее пользоваться **методом группировок**. Чтобы выявить наличие корреляционной связи между двумя признаками, проводится

группировка единиц совокупности по факторному признаку x и для каждой выделенной группы рассчитывается среднее значение результативного признака \bar{y}_i . Если результативный признак y зависит от факторного x , то в изменении среднего значения результативного признака \bar{y}_i будет прослеживаться определенная закономерность.

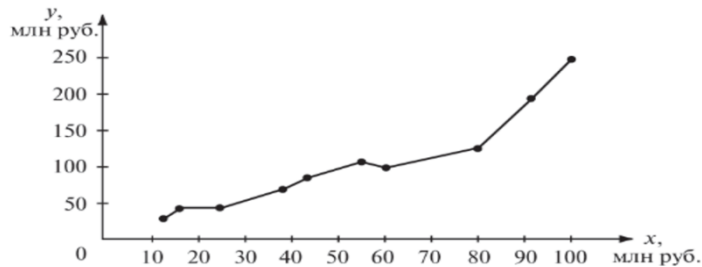


Рис. 4. Эмпирическая линия регрессии

Результаты группировки единиц совокупности могут быть в виде таблицы, в которой приведено комбинационное распределение единиц совокупности по двум признакам. Такие таблицы называют **таблицами взаимной сопряженности**. Если в таблице оба признака, по которым дано распределение единиц совокупности, количественные, то такая таблица взаимной сопряженности называется *корреляционной*.

Корреляционная таблица строится по типу «шахматной»: в подлежащем таблицы выделяются группы по факторному признаку x , а в сказуемом — по результативному y или наоборот, а в клетках таблицы на пересечении x и y показано число случаев совпадения каждого значения x с соответствующим значением y (табл. 4).

По мере увеличения значений x групповые средние значений y тоже увеличиваются от группы к группе, что позволяет сделать вывод о том, что между x и y существует корреляционная связь.

Таблица 4

Корреляционная таблица

Значение признака x_j	Значение признака y_i				Итого (число единиц) $f_x = f_j$	Среднее значение по группам \bar{y}_j
	5	10	15	20		

Продолжение таблицы 4

1	1	3	-	-	4	8,75
3	2	3	7	-	12	12,08
5	-	3	9	4	16	15,31
7	-	-	5	3	8	16,87
Итого (число единиц) $f_y = f_i$	3	9	21	7	$\sum f = 40$	14,00

О наличии и направлении связи можно судить и по «внешнему виду» таблицы. Если частоты расположены в клетках таблицы беспорядочно, то это свидетельствует либо об отсутствии связи между группировочными признаками, либо об их незначительной зависимости. Если же частоты образуют своего рода эллипс, то это свидетельствует о наличии зависимости между x и y , близкой к линейной. Расположение по диагонали из верхнего левого угла в нижний правый свидетельствует о прямой линейной зависимости, а из нижнего левого угла в верхний правый – об обратной.

Корреляционное поле представляет, по существу, ту же корреляционную таблицу, в клетках которой вместо чисел (частот) представлено соответствующее число точек (рис. 5).

Простейшая форма таблицы взаимной сопряженности – таблица «четырёх полей» (четырёхклеточная). В ней по каждому признаку выделяется только две группы, чаще всего по альтернативному принципу («да» – «нет», «хорошо» – «плохо» и т.д.) (табл.5).

Таблица 5

Таблица «четырёх полей»

Группа лиц	Число лиц		
	Заболевших гриппом	Не заболевших гриппом	Итого
Сделавших прививку	30 (a)	270 (b)	300
Не сделавших прививку	120(c)	80 (d)	200
Итого	150	350	500

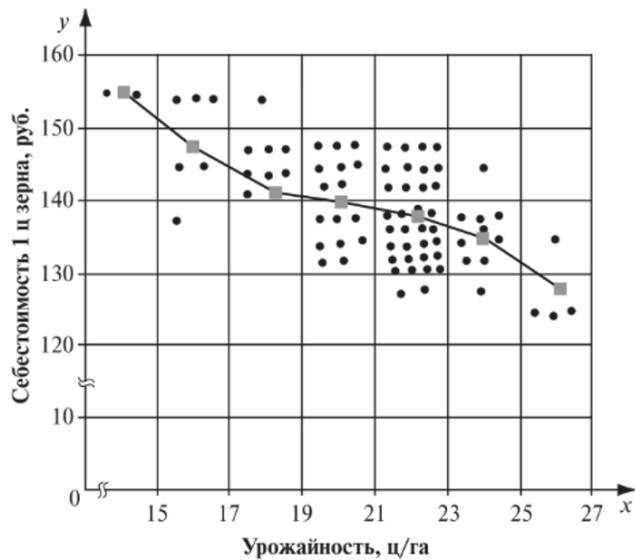


Рис. 5. Поле корреляции и линия средних значений

При случайном распределении распределение частот в каждой строке (или графе) таблицы соответствует (пропорционально) распределению частот в итоговой строке (или графе). Поэтому именно так рассчитываются теоретические частоты (частоты) по строкам (или графам), обозначенные в таблице 5 буквами (a – d).

Критерий Пирсона основан на сопоставлении эмпирических и теоретических частот и может быть определен по формуле (23):

$$\chi^2 = \sum_j \sum_i \frac{(f_{ij} - f'_{ij})^2}{f'_{ij}} \quad (23),$$

где f_{ij} и f'_{ij} – эмпирические и теоретические частоты по группам.

Фактическое значение χ^2 сравнивают с табличным (критическим, пороговым), определяемым по таблице для заданного уровня значимости α (обычно 0,05 или 0,01) и числа степеней свободы $\nu = (k_1 - 1)(k_2 - 1)$, где k_1 и k_2 – число групп по первому и второму признаку группировки. Если фактическое значение больше табличного, наличие корреляционной связи подтверждается.

Для измерения тесноты связи между группировочными признаками в таблицах взаимной сопряженности могут быть использованы такие показатели, как

– коэффициент ассоциации (24):

$$K_{ac} = \frac{ad-bc}{ad+bc} \quad (24);$$

– коэффициент контингенции (25):

$$K_{\text{конт}} = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (25);$$

– коэффициент взаимной сопряженности Пирсона (26):

$$K_{\Pi} = \sqrt{\frac{\chi^2}{\chi^2+n}} \quad (26);$$

– коэффициент взаимной сопряженности Чупрова (27):

$$K_{\text{ч}} = \sqrt{\frac{\chi^2}{n\sqrt{(k_1-1)(k_2-1)}}} \quad (27).$$

Связь считается достаточно значительной и подтвержденной, если $|K_{ac}| > 0,5$ или $|K_{\text{конт}}| > 0,3$.

В качестве показателей тесноты связи между количественными признаками, кроме упоминавшегося ранее коэффициента Фехнера, наиболее часто используются линейный коэффициент корреляции, коэффициенты корреляции рангов Спирмэна и Кендэла, а также коэффициент конкордации, позволяющий проверить тесноту связи между несколькими признаками.

1.8. Ряды динамики

Изучаемые вопросы

1. Понятие и виды рядов динамики.
2. Сопоставимость уровней и смыкание рядов динамики.
3. Основные показатели изменения уровней ряда: абсолютные приросты, темпы роста и прироста.
4. Исчисление средних показателей в рядах динамики. Средний уровень для интервальных и моментных рядов.
5. Средний абсолютный прирост, средние коэффициенты и темпы роста., средний темп прироста.

Основные понятия

Ряд динамики представляет собой числовые значения определенного статистического показателя в последовательные моменты или периоды времени.

Числовые значения того или иного статистического показателя, составляющие ряд динамики, называют *уровнями ряда* и обычно обозначают через y . Первый член ряда y_0 (или y_1) называют *начальным уровнем*, а последний y_n – *конечным*. Моменты или периоды времени, к которым относятся уровни, обозначают через t .

В зависимости от вида показателей уровней ряда различают *ряды абсолютных, относительных и средних величин (показателей)*.

Уровни рядов динамики могут относиться к определенным моментам времени (датам) или же периодам (интервалам). Поэтому в статистике различают:

– *моментные ряды*, уровни которых характеризуют значение показателя (явления) по состоянию на определенные моменты времени (дату).

– *интервальные ряды*, уровни которых характеризуют значение показателя, достигнутое за определенный период (интервал) времени.

Уровни интервальных рядов абсолютных величин можно дробить и складывать (суммировать). Суммируя уровни интервальных рядов абсолютных величин, можно строить ряды с нарастающим итогом.

Уровни анализируемого ряда динамики должны быть сопоставимыми. Возможные **причины несопоставимости**:

- изменение методологии учета или расчета показателей;
- изменение даты учета;
- изменение границ территории, к которой отнесены те или иные показатели;
- различная продолжительность периодов, к которым относятся уровни.

Для достижения сопоставимости уровней ряда их подвергают обработке следующими основными методами:

– *смыканием*, под которым понимают объединение в один ряд (более длинный) двух или нескольких рядов, уровни которых исчислены по разным методологиям или в разных границах;

– *приведением к одному основанию*, подразумевающим переход от абсолютных показателей к относительным.

Уровни любого ряда – это результат взаимодействия разных факторов, одни из которых действуют длительно, другие кратковременно. Поэтому, чтобы сделать правильные выводы о закономерностях развития того или иного показателя, надо суметь отделить главную тенденцию изменения (*тренд*) от колебаний, вызванных случайными кратковременными причинами.

Абсолютный прирост (абсолютное изменение) уровней рассчитывается как разность между двумя уровнями ряда.

В зависимости от базы сравнения различают:

– *цепные* приросты (28):

$$\text{ц}\Delta y = y_i - y_{i-1} \quad (28);$$

– *базисные* приросты (29):

$$\text{б}\Delta y = y_i - y_0 \quad (29).$$

Если же абсолютные приросты от периода к периоду возрастают (или убывают), то уровни изменяются ускоренно (или замедленно) и можно рассчитать *показатель ускорения* как разность между двумя смежными цепными абсолютными приростами (30):

$$\Delta_{\Delta} = \Delta_i - \Delta_{i-1} \quad (30).$$

Темп роста (изменения) – относительный показатель, рассчитываемый как процентное отношение двух уровней ряда и выражаемый в виде безразмерного коэффициента k_p (если база сравнения принимается за единицу), и как T_p в процентах (если база сравнения принимается за 100 единиц).

В зависимости от базы сравнения различают:

– *цепные* коэффициенты роста, когда каждый уровень сопоставляется с уровнем предыдущего периода (31):

$$\text{ц}k_p = y_i / y_{i-1} \quad (31);$$

– *базисные* коэффициенты роста, когда все уровни сопоставляются с уровнем одного какого-то периода, принятого за базу сравнения (32):

$$\text{б}k_p = y_i / y_{i-0} \quad (32).$$

Между цепными и базисными коэффициентами роста существует связь, позволяющая при необходимости переходить от цепных к базисным и наоборот, в частности:

- произведение цепных коэффициентов роста равно базисному;
- результат деления двух базисных коэффициентов равен цепному (промежуточному).

Темп прироста (снижения) – относительный показатель, показывающий, на сколько процентов данный уровень больше (или меньше) другого, принимаемого за базу сравнения и рассчитываемый двумя способами:

- путем вычитания 100% из темпа роста (33):

$$T_{\text{пр}} = T_p - 100\% \quad (33);$$

- как процентное отношение абсолютного прироста к тому уровню, по сравнению с которым рассчитан абсолютный прирост (34):

$$T_{\text{пр}} = \frac{\Delta y}{y_{i-1}} * 100\% \quad (34).$$

Абсолютное значение 1% прироста (α) – отношение абсолютного прироста уровня к темпу прироста (за соответствующий период) (35):

$$\alpha = \frac{\Delta y}{T_{\text{пр}}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} * 100\%} \quad (35).$$

Обобщенной характеристикой динамического ряда может служить прежде всего **средний уровень ряда \bar{y}** . Поскольку средняя величина в данном случае рассчитывается из меняющихся во времени показателей, то она называется *средней хронологической*.

В *интервальном ряду абсолютных величин с равными периодами (интервалами)* средний уровень рассчитывается как средняя арифметическая простая из уровней ряда (36):

$$\bar{y} = \frac{\sum y_i}{n} \quad (36).$$

Для *моментных рядов с равными промежутками* средний уровень может быть рассчитан как средняя хронологическая (37):

$$\bar{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{i=2}^{n-1} y_i}{n-1} \quad (37).$$

В случае *неравных промежутков* между датами среднюю хронологическую для *моментного ряда* можно рассчитать как среднюю арифметическую из средних значений уровней на каждую пару моментов, взвешенных по величине расстояний (отрезков времени) между датами (38):

$$\bar{y} = \frac{\sum(y_i + y_{i+1})t_i}{2\sum t_i} \quad (38).$$

Средний абсолютный прирост (изменение) уровней рассчитывается:

– как средняя арифметическая простая из отдельных цепных приростов (39):

$$\overline{\Delta_y} = \frac{\sum(\Delta y)_i}{n} \quad (39);$$

– на основе накопленного абсолютного прироста за n периодов (40):

$$\overline{\Delta_y} = \frac{y_n - y_0}{n} \quad (40);$$

Средний темп роста рассчитывается как:

– средняя геометрическая из цепных темпов роста (41):

$$\bar{k} = \sqrt[n]{k_1 k_2 \dots k_n} \quad \text{или} \quad \overline{T_p} = \sqrt[n]{k_1 k_2 \dots k_n} * 100\% \quad (41),$$

где k_i – цепной темп роста i -го периода;

– через базисный темп роста (42):

$$\bar{k} = \sqrt[n]{\frac{y_n}{y_0}} \quad \text{или} \quad \overline{T_p} = \sqrt[n]{\frac{y_n}{y_0}} * 100\% \quad (42).$$

Поскольку величина среднего темпа роста, рассчитанный по формулам (41) и (42), зависит от значений крайних уровней ряда, следует тщательно проанализировать ряд и в случае необходимости «длинные» и неодинаковые по характеру изменения периоды разбить на более однородные части (с похожей динамикой уровней), для которых расчет средних темпов роста будет иметь смысл.

Средние темпы прироста рассчитываются на основе средних темпов роста путем вычитания из последних 100% (43):

$$\overline{T_{пр}} = \overline{T_p} - 100\% \quad (43).$$

1.9. Экономические индексы

Изучаемые вопросы

1. Общее понятие и виды индексов. Цепные и базисные индексы.
2. Агрегатные индексы: стоимости, физического объема и цен. Индексы Ласпейреса, Пааше и Фишера.
3. Средние индексы из индивидуальных индексов.
4. Индексы переменного, фиксированного составов, структурных сдвигов.

Основные понятия

В статистике под **индексом** (от лат. index – показатель указатель) от понимается *относительная величина, характеризующая соотношение значений определенного показателя во времени, пространстве, а также сравнение фактических данных с планом или иным нормативом*. Показатель, изменение которого характеризуется индексом, называют *индексируемой величиной*.

В зависимости от **базы сравнения** индексы можно подразделить на:

- *динамические*, отражающие изменение явления во времени;
- *территориальные*, используемые для пространственных, межрегиональных сопоставлений различных показателей.

По **степени охвата элементов** совокупности индексы делятся на

- *индивидуальные*, обозначаемые символом i , характеризующие относительное изменение отдельного единичного элемента сложной совокупности, например, индекс объема одного определенного продукта (44):

$$i_q = \frac{q_1}{q_0} \quad (44);$$

или индекс цены одного определенного продукта (45):

$$i_p = \frac{p_1}{p_0} \quad (45);$$

- *общие* (сводные, агрегатные), обозначаемые символом I и характеризующие относительное изменение экономического показателя за счет изменения индексируемой величины (показателя) в целом по сложной совокупности, отдельные элементы которой несоизмеримы в физических единицах, например, индекс стоимости (46):

$$I_{pq} = \frac{\sum q_1 p_1}{\sum q_0 p_0} \quad (46),$$

где q_1 и q_0 – объем продукции различных видов в отчетном и базовом периодах соответственно;

p_1 и p_0 – цены продукции различных видов в отчетном и базовом периодах соответственно.

С помощью общих индексов характеризуется изменение цен на потребительские товары, изменение уровня жизни населения, развитие производства отдельных отраслей и экономики в целом и многое другое.

Построение общих индексов, относящихся к агрегированным совокупностям, связано с определенными проблемами, решение которых и составляет суть *индексного метода*. Общие индексы позволяют, с одной стороны, обобщать изменения индексируемой величины у отдельных единиц (элементов) в целом по совокупности и, с другой стороны, определять (измерять) влияние изменения отдельных факторов на изменение резульативного показателя явления в целом.

Индексы можно подразделить на индексы:

- *количественных* показателей (индекс физического объема производства, индекс продаж акций и т.п.);
- *качественных* показателей (индекс цен, индекс себестоимости, индекс урожайности, индекс заработной платы и др.);
- *результативных* показателей (индекс стоимости продукции, объема товарооборота, фонда заработной платы, валового сбора и др.).

Общий индекс, полученный путем сопоставления итоговых показателей, количественно выражающих сложное явление в отчетном и базисном периодах с помощью соизмерителей, называют **агрегатным**.

Агрегатный индекс стоимости I_{pq} , рассчитываемый при помощи формулы (46) показывает относительное изменение стоимости продукции как за счет изменения цен, так и за счет изменения объема отдельных товаров.

Общий индекс, исчисленный как отношение стоимости продукции двух периодов в одних и тех же ценах, называют **агрегатным индексом физического объема** I_q (47):

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \quad \text{или} \quad I_q = \frac{\sum q_1 p}{\sum q_0 p} \quad (47),$$

Заметим, что суммы в числителе и знаменателе формулы (47) имеют вполне реальный смысл:

$\sum q_0 p_0$ – стоимость продукции базисного периода в ценах базисного периода;

$\sum q_1 p_0$ – стоимость продукции отчетного периода в ценах базисного периода.

Отличительной особенностью любого агрегатного индекса является то, что и в числителе, и в знаменателе данного индекса имеется сумма произведений двух показателей, один из которых меняется, т.е. выступает в роли *индексируемой* величины, а второй остается неизменным, т.е. выступает в роли *соизмерителя* (или весов).

Разность между числителем и знаменателем агрегатного индекса характеризует изменение в абсолютном выражении сложного (результативного) показателя за счет изменения индексируемой величины.

Агрегатная формула общего индекса цен была впервые предложена в 1864 г. немецким ученым Э. Ласпейресом. Он предлагал строить агрегатный индекс цен, приняв в качестве весов продукцию базисного периода, т.е. q_0 (48):

$$I_p^L = \frac{\sum q_0 p_1}{\sum q_0 p_0} \quad (48).$$

В таком виде, т.е. построенный по продукции базисного периода, этот индекс известен как *индекс цен Ласпейреса*.

В 1874 г. другой немецкий ученый, Г. Пааше, предложил строить агрегатный индекс цен по продукции текущего периода q_1 (49):

$$I_p^П = \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (49).$$

Такой индекс, т.е. построенный по продукции текущего периода, известен как *индекс цен Пааше*.

При анализе динамики потребительских цен индекс Ласпейреса, как правило, больше индекса Пааше. Оно обусловлено тем, что в случае роста потребительских цен изменяется структура потребления, т.е. население переключается на потребление более дешевых товаров, доля которых в отчетном периоде увеличивается.

Индекс цен Ласпейреса удобен для оперативной (недельной, месячной, квартальной) информации об изменении цен на определенный фиксированный набор товаров, когда пересчет каждый раз на текущий набор (количество) товаров сопряжен с большими затратами труда и времени. По формуле Ласпейреса, в частности, рассчитывают индекс потребительских цен (ИПЦ), а также другие индексы цен, наблюдение за динамикой которых осуществляется в оперативном режиме.

Индексу цен Пааше отдается предпочтение, когда индекс цен рассматривается в системе с индексом стоимости и индексом физического объема (50):

$$I_{pq} = I_p * I_q \quad (50).$$

Кроме того, при расчете индекса цен по формуле Пааше, вычитая из числителя знаменатель, легко определить в абсолютном выражении сумму потерь (или прибыли) за счет изменения цен на продукцию отчетного (текущего) периода.

В начале XX в. американский экономист И. Фишер предложил вместо формул индексов цен Ласпейреса и Пааше использовать среднюю геометрическую из них, т.е. корень квадратный из произведения индексов цен Ласпейреса и Пааше (51):

$$I_p^\Phi = \sqrt{\frac{\sum q_0 p_1}{\sum q_0 p_0} * \frac{\sum q_1 p_1}{\sum q_1 p_0}} \quad (51).$$

Однако индекс Фишера из-за его формальности и трудности экономической интерпретации используется редко, в основном при территориальных сопоставлениях.

Любой общий индекс по форме может быть исчислен двояко: как агрегатный и как *средний из индивидуальных*. Основываясь на необходимости выполнения тождества между агрегатным индексом и средним арифметическим и средним гармоническим индексами **физического объема**, путем подстановок можно получить формулы для расчета общий индекс физического объема в форме:

– *среднего арифметического индекса* (52):

$$\bar{I}_q = \frac{\sum i_q q_1 p_0}{\sum q_0 p_0} \quad (52);$$

– *среднего гармонического индекса* (53):

$$\bar{I}_q = \frac{\sum q_1 p_0}{\sum \frac{q_1 p_0}{i_q}} \quad (53).$$

При решении конкретных задач выбор той или иной формы среднего индекса определяется прежде всего тем, какие исходные данные имеются в распоряжении исследователя.

Применительно к **индексам цен** возможны два варианта взвешивания и для среднего арифметического, и для среднего гармонического индексов, в зависимости оттого, по отношению к какому агрегатному индексу рассматривается их тождество: к индексу Ласпейреса или Пааше (табл. 6).

Таблица 6

Формулы средних индексов для агрегатного индекса цен

База расчета	Средний арифметический индекс	Средний гармонический индекс
Индекс Ласпейреса	$\bar{I}_p^L = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}$	$\bar{I}_p^H = \frac{\sum q_1 p_0}{\sum \frac{q_1 p_0}{i_q}}$
Индекс Пааше	$\bar{I}_p^P = \frac{\sum i_q q_0 p_1}{\sum q_0 p_1}$	$\bar{I}_p^H = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_q}}$

Средний арифметический индекс цен с весами $p_0 q_0$, т.е. тождественный агрегатному индексу цен Ласпейреса, в статистической практике используется в основном при расчете сводного индекса потребительских цен (ИПЦ), публикуемого ежемесячно по фиксированному набору товаров и услуг. В этом индексе (ИПЦ) в качестве весов для индексов цен по отдельным группам товаров (продовольственным, непродовольственным, платным услугам) принимается структура потребительских расходов домашних хозяйств в базисном периоде по указанным группам товаров и услуг, т.е. их доля.

Будучи сводной характеристикой качественного показателя, средняя величина складывается как под влиянием значений показателя у индивидуальных элементов (единиц), из которых состоит объект, так и под влиянием соотношения их весов («структуры» объекта).

В результате получим три различных индекса:

– **индекс переменного состава**, отражающий динамику среднего значения показателя (для однородной совокупности) как за счет *изменения индексируемой величины x* у отдельных элементов (частей целого), так и за счет *изменения весов f* , по которым взвешиваются отдельные значения (54):

$$I_{п.с.} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} \quad (54);$$

– **индекс фиксированного состава**, отражающий динамику среднего значения показателя лишь за счет изменения *индексируемой величины x* , при фиксировании весов на уровне, как правило, отчетного периода f_1 , то есть исключает влияние изменения структуры (состава) совокупности на динамику средних величин (55):

$$I_{ф.с.} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} \quad (55);$$

– **индекс структурных сдвигов**, отражающий динамику среднего значения показателя лишь за счет изменения *весов f* при фиксировании индексируемой величины на уровне базисного периода x_0 (56):

$$I_{стр.} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} \quad (56).$$

Нетрудно заметить, что индекс переменного состава есть произведение индекса фиксированного состава на индекс структурных сдвигов.

Если известны данные за несколько периодов, по ним может быть построен ряд индексов: либо с постоянной для всех базой сравнения (*базисных*), либо с переменной (*цепных*). В табл. 7 приведены в качестве примера цепные и базисные индивидуальные индексы цен и физической объема.

Таблица 7

Цепные и базисные индивидуальные индексы цен и физической объема

Название индивидуального индекса	Цепные индексы	Базисные индексы
Индекс цен	$\frac{p_1}{p_0}; \frac{p_2}{p_1}; \dots; \frac{p_t}{p_{t-1}}$	$\frac{p_1}{p_0}; \frac{p_2}{p_0}; \dots; \frac{p_t}{p_0}$
Индекс физического объема	$\frac{q_1}{q_0}; \frac{q_2}{q_1}; \dots; \frac{q_t}{q_{t-1}}$	$\frac{q_1}{q_0}; \frac{q_2}{q_0}; \dots; \frac{q_t}{q_0}$

Между цепными и базисными индивидуальными индексами существует определенная взаимосвязь, что позволяет переходить от одного вида индексов к другому:

- перемножая последовательно цепные индексы, можно получить базисные индексы;
- отношение двух последовательных базисных индексов дает цепной индекс.

Цепные и базисные индексы могут быть построены и для общих индексов. При этом последние могут иметь *постоянные и переменные веса*. Для общих (агрегатных) индексов переход от цепных индексов к базисным строго математически возможен лишь для индексов с постоянными весами.

2. ОЦЕНОЧНЫЕ СРЕДСТВА САМОСТОЯТЕЛЬНОЙ РАБОТЫ

2.1. Вопросы для текущего контроля самостоятельной работы

Раздел 1. Методические основы статистики

1. Информация как статистическая категория.
2. Этапы развития статистики как науки.
3. Статистическая совокупность, ее структура и особенности.
4. Статистическая закономерность и массовые явления.
5. Этапы статистического исследования.

Раздел 2. Общие сведения о статистическом наблюдении

1. Назначение плана статистического наблюдения, его основные элементы.
2. Организационные формы статистического наблюдения.
3. Текущее и непрерывное статистическое наблюдение.
4. Способы статистического наблюдения.
5. Отражение признаков единиц наблюдаемой совокупности в программе статистического наблюдения.

Раздел 3. Сводка и группировка статистических данных

1. Признаки, по которым формируется однородная совокупность.
2. Взаимосвязь сводки и группировки.
3. Назначение статистической группировки, ее основные задачи.
4. Особенности различных видов группировок.
5. Виды статистических таблиц и действия с ними.

Раздел 4. Обобщающие статистические показатели

1. Назначение статистических обобщающих показателей и требования к их качеству.
2. Характеристики абсолютных величин.
3. Правила перехода к условно-натуральному измерению абсолютной величины.

4. Принципы формирования научно-обоснованных статистических средних.

5. Оценка качественной сущности одномерной совокупности с помощью разных видов средних.

Раздел 5. Анализ вариационных рядов

1. Показатели интервальных вариационных рядов.

2. Отличия в построении статистических и математических графиков.

3. Понятия и способы расчета моды и медианы.

4. Понятие показателей вариации, формулы их расчета.

5. Направлениям проведения дисперсионного анализа в экономической статистике.

Раздел 6. Выборочные наблюдения

1. Достоинства и недостатки метода выборочного наблюдения.

2. Виды отбора: случайный, механический, типический, серийный.

3. Предпосылки для выбора того или иного способа отбора.

4. Определение необходимой численности выборки при разных способах отбора.

5. Особенности проведения выборочного наблюдения в условиях малой выборки.

Раздел 7. Статистическое изучение корреляционных зависимостей

1. Статистические и корреляционные зависимости.

2. Способы выявления корреляционной зависимости между количественными признаками.

3. Изучение связи между качественными признаками на основе таблиц сопряженности.

4. Особенности использования линейного коэффициенты корреляции.

5. Ранговые коэффициенты корреляции Спирмена и Кендэла.

Раздел 8. Ряды динамики

1. Различия между динамическими и вариационными рядами.
2. Методы обеспечения сопоставимости динамических рядов.
3. Показатели, оценивающие динамику явления (процесса).
4. Определение среднего уровня ряда для интервальных и моментных рядов.
5. Средние коэффициенты и темпы роста. Средний темп прироста.

Раздел 9. Экономические индексы

1. Классические агрегатные индексы количества и качества.
2. Методы построения системы взаимосвязанных агрегатных индексов.
3. Построение индексов переменного состава прямым и обратным методом.
4. Различия абсолютных разностей качественного и количественного признаков.
5. Метод абсолютных разностей и его использование при оценке влияния отдельных факторов на изменение общего признака.

2.2. Программа экзамена по дисциплине

Раздел 1. Методические основы статистики

Понятие статистики как науки. Задачи описательной школы статистики (Готфрид Ахенваль), школы «политической арифметики» (Уильям Петти), математическо-статистического направления (Карл Пирсон).

Статистическая совокупность: понятие, основные черты. Качественные (атрибутивные) и количественные статистические признаки.

Предмет теории статистики (общей статистики), социально-экономической статистики и отраслевых статистик.

Раздел 2. Общие сведения о статистическом наблюдении

Этапы статистического наблюдения. Сводка. Научный анализ статистической информации. Организационные формы стати-

стического наблюдения: регистр, статистическая отчетность, специально организованное статистическое наблюдение.

Виды статистического наблюдения: текущее и прерывное; сплошное и несплошное; одновременное и повторяющееся; наблюдение основного массива и монографическое. Виды статистического наблюдения по источникам информации (документальный) и организации сбора информации (экспедиционный, саморегистрации).

Ошибки регистрации: случайные, непреднамеренные систематические и преднамеренные: понятие и условия возникновения.

Раздел 3. Сводка и группировка статистических данных

3.1 Сводка и группировка

Понятие сводки. Централизованная, децентрализованная и смешанная сводка.

Понятие и основные черты группировки: типизирующие признаки как основа, соответствие определенному моменту или периоду времени, влияние группировочного признака на число образуемых групп.

Качественные и количественные группировки. Различия между качественными и типологическими группировками. Пространственные и аналитические группировки. Различия между классификацией и группировкой.

Определение количества групп при использовании признаков, имеющих дискретный и непрерывный характер. Формула длины интервала при равномерном близком к нормальному распределении. Формула Стерджесса.

3.2. Статистические таблицы

Макет, подлежащее и сказуемое статистической таблицы. Простая, сложная и комбинированная таблица (с точки зрения подлежащего). Простая и сложная разработка сказуемого в статистической таблице.

Раздел 4. Обобщающие статистические показатели

4.1. Абсолютные и относительные величины

Сущность и примеры абсолютных обобщающих показателей. Единицы измерения абсолютного обобщающего показателя: простые и сложные натуральные, стоимостные.

Виды относительных величин: структуры, динамики, сравнения, интенсивности, координации. Единицы измерения относительного обобщающего показателя: доли, проценты (%), промилли (‰), продецимилли (‱).

4.2. Средние величины

Средняя типическая величина как характеристика качественно однородных групп изучаемого явления. Средняя системная величина как характеристика качественно неоднородных пространственных и динамических статистических совокупностей.

Формула средней степенной (взвешенной) и ее перевод в среднюю арифметическую (взвешенную), среднюю гармоническую (взвешенную), среднюю геометрическую (взвешенную), среднюю квадратическую (взвешенную). Средняя арифметической взвешенной методом отсчета от условного нуля (формула). Общая средняя арифметическая из групповых средних (формула).

Раздел 5. Анализ вариационных рядов

5.1. Построение и графическое изображение вариационных рядов

Дискретные ряды. Частота и частость. Накопленная частота. Кумулята. Полигон.

Интервальные ряды с равными и неравными интервалами. Длина интервала. Абсолютная (из частоты) и относительная (из частости) плотность распределения. Гистограмма. Кумулята и огива. Построение кривой Лоренца.

5.2. Показатели среднего уровня вариационного ряда

Мода: понятие, алгоритм и формулы для аналитического определения значения для дискретных и интервальных рядов с равными и не равными интервалами. Определение моды графически (по гистограмме).

Медиана: понятие, алгоритм и формулы для аналитического определения значения для дискретных и интервальных рядов на основе накопленных частот или частостей. Определение моды графически (по кумуляте).

5.3. Показатели дифференциации и концентрации

Квартили и децили как показатели дифференциации вариационного ряда. Алгоритм аналитического определения значения

первого дециля по накопленным частотам и накопленным частотам. Децильный коэффициент.

Коэффициенты Джини, Херфиндаля-Хиршмана, Лоренца: назначение (показатели концентрации), диапазон принимаемых значений.

5.4. Показатели вариации признаков

Размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, дисперсия, коэффициент вариации: сущность, формулы.

Правило сложения дисперсий. Общая дисперсия, межгрупповая дисперсия: сущность. Эмпирическое корреляционное отношение: правило определения тесноты связи между признаками (очень слабая; слабая; умеренная; заметная; заметная; тесная; весьма тесная; функциональная).

Раздел 6. Выборочные наблюдения

Понятие генеральной и выборочной совокупностей (выборки). Преимущества метода выборочного наблюдения (экономия ресурсов, высокая достоверность получаемых данных, снижение вероятности появления или необнаружения ошибок регистрации). Зависимость точность оценок, полученных на основе выборочного метода, от числа обследованных единиц.

Случайный повторный и бесповторный отбор: понятие, средняя ошибка выборки и средняя ошибка выборки для доли (формулы).

Систематический (механический), типический, гнездовой отбор: понятие, общая схема проведения. Формулы общей, межгрупповой и частной внутригрупповой дисперсии. Средняя ошибка выборки при повторном (бесповторном) отборе при пропорциональном (оптимальном) распределении единиц выборочной совокупности (формулы). Численность выборки при повторном механическом, типическом, серийном отборе.

Раздел 7. Статистическое изучение корреляционных зависимостей

Понятие (жестко) детерминированной, статистической, корреляционной и функциональной типов связей между признаками.

Задачи корреляционного, регрессионного и корреляционно-регрессионного видов анализа.

Коэффициенты Фехнера и Спирмена: диапазон значений, интерпретация типа связи между признаками.

Коэффициент ассоциации, коэффициент контингенции, критерий Пирсона, коэффициент Пирсона для подтверждения наличия / отсутствия связи между признаками на основе таблиц сопряженности (формулы).

Раздел 8. Ряды динамики

Ряды динамики абсолютных, относительных и средних величин (примеры). Интервальные и моментные ряды (примеры). Смыкание, приведение к одному основанию и аналитическое выравнивание рядов динамики: понятие.

Цепные и базисные абсолютные приросты, коэффициенты (темпы) роста, темпы прироста (формулы).

Средний уровень интервального ряда абсолютных величин с равными периодами, моментного ряда абсолютных величин с равными (неравными) промежутками времени (формулы). Средний абсолютный прирост уровней ряда, средний темп роста абсолютных уровней ряда, средний темп роста для уровней ряда, приведенных к одной и той же базе (формулы).

Методы укрупненных интервалов, скользящей средней и аналитического выравнивания для выявления основной тенденции ряда (сущность).

Раздел 9. Экономические индексы

Динамические и территориальные индексы: сущность, примеры. Индивидуальные и сводные: сущность, примеры.

Агрегатный индекс физического объема: индексируемая величина, соизмерители. Агрегатные индексы цен Ласпейреса и Пааше: индексируемая величина, соизмерители, формула. Агрегатные индексы цен Фишера (формула).

Индекс переменного, фиксированного составов, структурных сдвигов (сущность).

Цепные и базисные индексы: понятие. Взаимосвязь между цепными и базисным и индексами.

2.3. Примерный перечень вопросов к экзамену

1. Статистическая совокупность: понятие, основные черты. Качественные (атрибутивные) и количественные статистические признаки.
2. Статистическое наблюдение: понятие и основные этапы.
3. Статистическое наблюдение: виды и организационные формы.
4. Сводка статистических данных: понятие, виды.
5. Группировка статистических данных: понятие, виды.
6. Определение количества групп при использовании признаков, имеющих дискретный и непрерывный характер.
7. Статистические таблицы: виды, правила формирования.
8. Абсолютные обобщающие статистические показатели: понятие, назначение, единицы измерения.
9. Относительные статистические показатели: структуры, динамики, сравнения, интенсивности, координации (понятие, назначение, единицы измерения).
10. Средняя типическая и средняя системная (понятие, назначение, единицы измерения, примеры).
11. Построение дискретных вариационных рядов: общий вид, частота, частость.
12. Построение непрерывных вариационных рядов: общий вид, виды интервалов, длина интервала, частость, плотность.
13. Графическое изображение вариационных рядов: полигон, гистограмма.
14. Накопленные частота и частость. Кумулята и огива.
15. Кривая Лоренца: назначение, правила построения, интерпретация.
16. Мода: понятие, алгоритмы для аналитического и графического (по гистограмме) определения значения для дискретных и интервальных рядов.
17. Медиана: понятие, алгоритмы для аналитического и графического (по кумуляте) определения значения для дискретных и интервальных рядов.

18. Показатели дифференциации вариационного ряда: квантили и децили (понятие, алгоритм расчета, интерпретация). Децильный коэффициент.

19. Показатели концентрации вариационного ряда (коэффициенты Джини, Херфиндаля-Хиршмана, Лоренца): назначение, формулы для расчета, диапазон принимаемых значений, интерпретация.

20. Показатели вариации признаков: размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, дисперсия, коэффициент вариации (сущность, формулы, интерпретация).

21. Правило сложения дисперсий: формулы, сущность, применение. Эмпирическое корреляционное отношение.

22. Метод выборочного наблюдения: сущность, преимущества, ограничения применимости.

23. Типы связей между признаками. Задачи корреляционного, регрессионного и корреляционно-регрессионного видов анализа.

24. Выявление связи между качественными признаками на основе таблиц сопряженности. Критерий Пирсона.

25. Применение коэффициентов ассоциации и контингенции для определения тесноты связи между качественными признаками.

26. Применение коэффициентов Пирсона и Чупрова для определения тесноты связи между качественными признаками.

27. Применение линейного коэффициента корреляции для определения тесноты связи между количественными признаками.

28. Коэффициент корреляции рангов Фехнера: алгоритм расчета, диапазон значений, интерпретация типа связи между признаками.

29. Коэффициент корреляции рангов Спирмена: алгоритм расчета, диапазон значений, интерпретация типа связи между признаками.

30. Ряды динамики: понятие и виды. Смыкание, приведение к одному основанию и аналитическое выравнивание рядов динамики.

31. Основные показатели изменения уровней ряда: абсолютные приросты, темпы роста, темпы прироста (снижения) уровней. Цепные и базисные абсолютные приросты

32. Средний уровень динамического ряда (для интервальных и моментных рядов).

33. Средний абсолютный прирост (изменение), средние темпы (коэффициенты) роста и средние темпы прироста уровней динамического ряда.

34. Экономические индексы: сущность, виды, примеры.

35. Агрегатный индекс физического объема: сущность (индексируемая величина, соизмерители), область применения.

36. Агрегатные индексы цен Ласпейреса, Пааше и Фишера: сущность (индексируемая величина, соизмерители), область применения.

37. Средний индекс физического объема.

38. Средний индекс цен.

39. Индекс переменного, фиксированного составов, структурных сдвигов (на примере индекса цен).

40. Цепные и базисные индексы: понятие, взаимосвязь.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Основной:

1. Теория статистики: Учебник / под ред. проф. Г.Л. Громыко. М.: ИНФРА-М, 2017. 476 с. Режим доступа: <http://znanium.com/catalog.php?bookinfo=851544>
2. Глинский, В.В., Ионин, В.Г. и др. Статистика: учебник / В.В. Глинский, В.Г. Ионин, Л.К. Серга [и др.]; под ред. В.Г. Ионина. М.: ИНФРА-М, 2020. 355 с. Режим доступа: <http://znanium.com/catalog.php?bookinfo=552459>
3. Годин, А.М. Статистика : Учебник для бакалавров / А.М. Годин. М.: Издательско-торговая корпорация «Дашков и К°», 2020. 410 с. Режим доступа: <https://znanium.com/catalog/product/1093663>
4. Гужова, О.А. Статистика в управлении социально-экономическими процессами: Учеб. пособие / О.А. Гужова, Ю.А. Токарев. М.: ИНФРА-М, 2017. 172 с. Режим доступа: <http://znanium.com/catalog.php?bookinfo=556718>
5. Сергеева, И.И. Статистика : Учебник / И.И. Сергеева, Т.А. Чекулина, С.А. Тимофеева. М.: ФОРУМ : ИНФРА-М, 2020. 304 с. Режим доступа: <https://znanium.com/catalog/product/1141798>

Дополнительный:

1. Громыко, Г.Л. Теория статистики: Практикум / Г.Л. Громыко. М.: ИНФРА-М, 2017. 238 с. Режим доступа: <http://znanium.com/catalog.php?bookinfo=671371>
2. Бурова, О.А. Статистика : сборник задач: Учебное пособие / О.А. Бурова М.: МИСИ-МГСУ, 2017. 127 с. Режим доступа: <https://znanium.com/catalog/product/971683>
3. Гужова, О.А. Статистика в управлении социально-экономическими процессами : Учеб. пособие / О.А. Гужова, Ю.А. Токарев. М.: ИНФРА-М, 2017. 172 с. Режим доступа <https://znanium.com/catalog/product/556718>
4. Ефимова, М.Р. Практикум по общей теории статистики: Учеб. пособие / М.Р. Ефимова, О.И. Ганченко, Е.В. Петрова. М.: Финансы и статистика, 2014. 368 с. Режим доступа: <https://e-lanbook.com/book/91212>
5. Экономическая статистика. Практикум : Учеб. пособие / Ю.Н. Иванов, Г.Л. Громыко, А.Н. Воробьев [и др.] ; под ред. Ю.Н. Иванова. М.: ИНФРА-М, 2017. 176 с. Режим доступа: <https://znanium.com/catalog/product/760303>

СОДЕРЖАНИЕ

Введение.....	3
1. Материалы для самостоятельной работы по разделам	5
1.1. Методические основы статистики.....	5
1.2. Общие сведения о статистическом наблюдении.....	8
1.3. Сводка и группировка статистических данных.....	11
1.4. Обобщающие статистические показатели	16
1.5. Анализ вариационных рядов.....	19
1.6. Выборочные наблюдения	26
1.7. Статистическое изучение корреляционных зависимостей	31
1.8. Ряды динамики	37
1.9. Экономические индексы.....	42
2. Оценочные средства самостоятельной работы.....	49
2.1. Вопросы для текущего контроля самостоятельной работы...	49
2.2. Программа экзамена по дисциплине	51
2.3. Примерный перечень вопросов к экзамену	56
Библиографический список.....	59

СТАТИСТИКА

***Методические указания к самостоятельной работе
для студентов бакалавриата направления 38.03.01***

Сост. *Н.В. Василенко*

Печатается с оригинал-макета, подготовленного кафедрой
экономики, организации и управления

Ответственный за выпуск *Н.В. Василенко*

Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 03.02.2023. Формат 60×84/16.
Усл. печ. л. 3,5. Усл.кр.-отт. 3,5. Уч.-изд.л. 3,2. Тираж 50 экз. Заказ 61.

Санкт-Петербургский горный университет
РИЦ Санкт-Петербургского горного университета
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2