

МПД-ОЗ-14

Федеральное государственное бюджетное образовательное учреждение высшего образования «Северо-Осетинская государственная медицинская академия»
Министерства здравоохранения Российской Федерации

Кафедра общественного здоровья, здравоохранения и социально-экономических наук

**МЕДИЦИНСКАЯ СТАТИСТИКА.
ОРГАНИЗАЦИЯ И ЭТАПЫ СТАТИСТИЧЕСКОГО
ИССЛЕДОВАНИЯ**

(учебное пособие для аспирантов медицинских специальностей)

Владикавказ 2019

Аликова З.Р., Бадоева З.А., Магаев К.А. Медицинская статистика. Организация и этапы статистического исследования: учебное пособие для аспирантов медицинских специальностей. Северо-Осетинская государственная медицинская академия. - Владикавказ 2019. - 70 с.

Рецензенты:

Кусова А.Р. – доктор мед.наук, профессор, зав. кафедрой гигиены
ФГБОУ ВО СОГМА Минздрава России

И.Ф. Боциев - канд. тех. наук, доцент кафедры химии и физики
ФГБОУ ВО СОГМА Минздрава России

Учебно-методическое пособие предназначается для подготовки к практическим занятиям, аудиторной и внеаудиторной самостоятельной работе аспирантов различных факультетов медицинских вузов. В пособии представлены методы медицинской статистики, широко применяемые как при изучении показателей здоровья, так и при анализе деятельности медицинских учреждений, системы охраны, укрепления и восстановления здоровья населения.

Пособие направлено на формирование у аспирантов базовых компетенций обработки и анализа информации, выраженной числовыми данными.

Учебный материал охватывает основные вопросы организации и этапы статистического исследования и формирует представление о возможностях и особенностях применения статистических методов для выявления закономерностей развития различных общественных явлений. Даны типовые задачи, образцы их решения, контрольные вопросы, задачи для самостоятельной работы, вопросы тестового контроля.

*Учебно-методическое пособие составлено в соответствии с учебной программой,
Утверждено и рекомендовано к печати Центральным координационным учебно-методическим советом ФГБОУ ВО СОГМА Минздрава России ЦКУМС ФГБОУ ВО СОГМА Минздрава России. (протокол №5 от 23.05.2019 г.).*

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
Глава 1. ОРГАНИЗАЦИЯ И ЭТАПЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ.	6
1.1. Организация статистического исследования.....	6
1.1. Этапы статистического исследования.	7
1.2. I этап статистического исследования. Составление плана и программы исследования.....	8
1.3. II этап. Сбор материала (статистическое наблюдение). ..	16
1.4. III этап. Разработка материала. Статистическая группировка и сводка..	17
1.4.1. Первое свойство статистической совокупности.	22
1.4.2. Второе свойство статистической совокупности.	27
1.4.3. Третье свойство статистической совокупности.	33
1.4.4. Четвертое свойство статистической совокупности.	39
1.4.5. Пятое свойство статистической совокупности.....	49
1.5. IV этап. Статистический анализ изучаемого явления, формулировка выводов.....	59
КОНТРОЛЬНЫЕ ВОПРОСЫ.	63
ВОПРОСЫ ТЕСТОВОГО КОНТРОЛЯ.....	64
ТИПОВЫЕ ЗАДАЧИ.....	66
ЛИТЕРАТУРА(основная, дополнительная).....	69

ВВЕДЕНИЕ

Интенсификация труда медицинских работников в условиях бюджетно-страхового здравоохранения предъявляет повышенные требования к научно-

организационным факторам. В данных условиях возрастает роль медицинской статистики в научной и практической деятельности медицинского учреждения.

В практической и научно-исследовательской деятельности врач, как правило, анализирует результаты своей деятельности не только на индивидуальном, но и на групповом и популяционном уровнях. Это необходимо врачу для подтверждения уровня квалификации, а также в целях дальнейшего усовершенствования и профессиональной специализации. Поэтому умение правильно организовать и провести статистическое исследование необходимо всем врачам различного профиля, руководителям учреждений и органов здравоохранения. Такие знания и умения способствуют повышению качества и эффективности медицинской помощи населению через непрерывную подготовку кадров (важнейший элемент ресурсного обеспечения) и, таким образом, конкурентоспособности лечебно-профилактических учреждений различных форм собственности в условиях рыночной экономики.

Руководители здравоохранения в оперативной и прогностической работе постоянно используют статистические данные. Только квалифицированный анализ статистических данных, оценка событий и соответствующие выводы позволяют принять правильное управленческое решение, способствуют лучшей организации работы, более точному планированию и прогнозированию. Статистика помогает контролировать деятельность учреждения, оперативно управлять им, судить о качестве и эффективности лечебно-профилактической работы. Руководитель при составлении текущих и перспективных планов работы должен основываться на изучении и анализе тенденций и закономерностей развития, как здравоохранения, так и состояния здоровья населения своего района, города, области и т. д.

Традиционная статистическая система в здравоохранении основана на получении данных в виде отчетов, которые составляются в низовых учреждениях и затем суммируются на промежуточных и высших уровнях. Система отчетов имеет не только преимущества (единая программа,

обеспечение сравнимости, показатели объема работы и использования ресурсов, простота и малая стоимость сбора материалов), но и определенные недостатки (малая оперативность, жесткость, негибкая программа, ограниченный набор сведений, неконтролируемые ошибки учета и др.).

Анализ проделанной работы должен осуществляться врачами не только на основании существующей отчетной документации, но и путем специально проводимых выборочных статистических исследований.

Глава 1

ОРГАНИЗАЦИЯ И ЭТАПЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

1.1. Организация статистического исследования

Статистическое исследование (СИ) — это научно организованный по единой программе сбор, сводка и анализ данных (фактов) о социально-экономических, демографических и других явлениях и процессах общественной жизни в государстве с регистрацией их наиболее существенных признаков в учетной документации. Отличительными чертами (спецификой) статистического исследования являются: целенаправленность, организованность, массовость, системность (комплексность), сопоставимость, документированность, контролируемость, практичность.

Статистическое исследование позволяет получить представление о том или ином явлении, изучить его размер, уровень, выявить закономерности. Предметом статистического исследования могут быть здоровье населения, организация медицинской помощи, факторы внешней среды, влияющие на здоровье и т. д.

При проведении статистического исследования могут быть использованы два методических подхода:

- изучение интенсивности явления в среде, распространенности явления, выявление тенденций состояния здоровья населения - проводятся на генеральных совокупностях или достаточно больших по численности выборочных совокупностях, позволяющих получить интенсивные показатели и обоснованно перенести полученные данные на всю генеральную совокупность

- проведение строго спланированных исследований по изучению отдельных факторов без выявления интенсивности явления в среде - проводятся, как правило, на небольших по численности совокупностях с целью выявления новых факторов, изучения неизвестных или малоизвестных причинно-следственных связей.

1.2. Этапы статистического исследования:

Статистическое исследование состоит из последовательно проводимых 4-х этапов:

1. Составление плана и программы исследований.

2. Статистическое наблюдение (сбор материала).
3. Разработка материала, статистическая группировка и сводка.
4. Статистический анализ изучаемого явления, формулировка выводов(вычисление относительных и средних величин и их графическое изображение) (рис.1).



Рис. 1. Этапы статистического исследования.

I этап статистического исследования. Составление программы и плана исследования.

В программу проведения исследования входит определение цели и задач исследования, который ориентирует на конечный результат, составляется план исследования, разрабатывается программа сводки статистического материала и решаются организационные вопросы:

а) цель и задачи исследования должны быть четко сформулированы; цель определяет основное направление исследования и носит, как правило, не только теоретический, но и практический характер, она формулируется ясно, четко, недвусмысленно; для раскрытия поставленной цели определяются задачи исследования:

б) необходимо изучить по данной теме литературу,

в) необходимо разработать организационный план,

г) разработка плана исследования,

д) необходимо составить программу исследования (наблюдения),

е) необходимо составить программу сводки полученных данных.

Организационный план предусматривает определение:

- места (административно-территориальных границ наблюдения),
- времени (конкретных сроков осуществления наблюдения, проведения разработки и анализа материала),

- субъекта исследования (организаторов, исполнителей, методического и организационного руководства, источников финансирования исследования).

Разработка плана исследования включает определение:

- объекта исследования (статистической совокупности);

- объема исследования (сплошное, несплошное);

- видов (текущее, прерывное, периодическое, единовременное);

- способов сбора статистической информации.

Объект исследования в статистике называют **статистической совокупностью**, которую определяют, как группу относительно однородных элементов – единиц, объединенных вместе в границах времени и пространства.

Статистическая совокупность состоит из единиц наблюдения (рис. 2.).



Рис. 2. Структура статистической совокупности.

Различают два вида совокупности — генеральная и выборочная.

Генеральная совокупность — это группа, состоящая из всех относительно однородных элементов в соответствии с поставленной целью.

Выборочная совокупность — отобранная для исследования часть генеральной совокупности и предназначенная для характеристики всей генеральной совокупности. Она должна быть репрезентативна (представительна) по количеству и качеству по отношению к генеральной совокупности.

Репрезентативность количественная основана на законе больших чисел и означает достаточную численность элементов выборочной совокупности, рассчитываемую по специальным формулам и таблицам.

Репрезентативность качественная основана на законе вероятности и означает соответствие (однотипность) признаков, характеризующих элементы выборочной совокупности по отношению к генеральной.

Объем статистической совокупности — это численность элементов совокупности, взятых для исследования.

Сроки и место (территория) проведения исследования — это составление календарного плана выполнения данного исследования по данному этапу на конкретной территории.

Виды наблюдения. Статистические наблюдения подразделяются на виды по следующим признакам:

- по времени регистрации данных;
- по полноте охвата единиц совокупности;
- по способу получения сведений в ходе проведения и характеру его осуществления.

Виды статистического наблюдения по времени регистрации:

- *текущее (непрерывное) наблюдение* - проводится для изучения текущих явлений и процессов. Явление изучается за какой-то отдельный период времени (неделю, квартал, год и т. д.) путем повседневной регистрации явления по мере возникновения каждого случая (учет числа родившихся, умерших, заболевших, выписанных из стационара). Так учитываются быстро меняющиеся явления.

- *прерывное наблюдение* — проводится по мере необходимости, при этом допускаются временные разрывы в регистрации данных:

- *периодическое наблюдение* — проводится через сравнительно равные интервалы времени (перепись населения).

- *единовременное наблюдение* — осуществляется без соблюдения строгой периодичности его проведения. Статистические данные собираются на определенный (критический) момент времени (перепись населения, изучение физического развития детей, профилактические осмотры населения).

Единовременная регистрация отражает состояние явления на момент изучения, используется для изучения медленно меняющихся явлений.

Выбор вида наблюдения по времени определяется целью и задачами исследования (характеристику госпитализированных больных можно получить в результате текущей регистрации выбывших из стационара - *текущее наблюдение* или путем однодневной переписи больных, находящихся в стационаре - *единовременное наблюдение*).

По полноте охвата единиц совокупности различают следующие виды статистического наблюдения:

- *сплошное наблюдение* — представляет собой сбор и получение информации обо всех единицах изучаемой совокупности. Характеризуется высокими материальными и трудовыми затратами, недостаточной оперативностью информации. Проводят с целью установления абсолютных размеров явления (общей численности населения, общего количества родившихся или умерших). Применяется и в тех случаях, когда сведения необходимы для оперативной работы (учет инфекционной заболеваемости, нагрузка врачей и др.)

- *несплошное (выборочное) наблюдение* — основано на принципе случайного отбора единиц изучаемой совокупности, при этом в выборочной совокупности должны быть представлены все типы единиц, имеющих в совокупности. Имеет ряд преимуществ перед сплошным наблюдением: сокращение временных и денежных затрат. Выборочное наблюдение - самый распространенный метод, сводится к исследованию некоторой специально отобранной части единиц наблюдения для характеристики всей генеральной совокупности. Преимуществом этого метода является получение результатов высокой степени надежности, а также значительно более низкая стоимость. В исследовании занято меньшее число исполнителей, Кроме того он требует меньших затрат времени. В медицинской статистике роль и место выборочного метода особенно велики, поскольку медицинские работники имеют дело

обычно только с частью изучаемого явления (изучают группу больных с тем или иным заболеванием, анализируют работу отдельных подразделений).

Методы проведения исследования на выборочной совокупности.

Монографический метод применяется при изучении какого-либо одного объекта, когда из множества объектов избирается один и исследуется с максимальной полнотой с целью показа передового опыта, выявления тенденций развития явления.

Пример: описание новой хирургической технологии.

Метод основного массива применяется при изучении тех объектов, в которых сосредоточено большинство изучаемых явлений. Суть его состоит в том, что из всех единиц наблюдения, входящих в состав данного объекта, избирается их основная часть, характеризующая всю статистическую совокупность.

Пример: на заводе имеется 7 основных цехов, в которых занято 1300 рабочих и два небольших вспомогательных цеха со 100 рабочими. Для наблюдения можно взять только основные цеха и по ним сделать выводы, касающиеся всего завода.

Анкетный метод применяется для сбора статистических сведений с помощью специально разработанных анкет.

Пример: при изучении распространенности желудочно-кишечных заболеваний среди учащихся профессионально-технических училищ города Н. была разработана анкета с перечнем вопросов, интересующих исследователя.

По способу получения сведений в ходе проведения и характеру его осуществления:

- *непосредственное наблюдение* (клинический осмотр больных, проведение лабораторных, инструментальных исследований, антропометрические измерения и т. п.);

- *социологические методы:* метод интервью (очный опрос), анкетирование (заочный опрос - анонимный или неанонимный) и др.;

- *документальное исследование* (выкопировка сведений из учетно-отчетных медицинских документов, сведения официальной статистики учреждений и организаций).

Методы отбора изучаемых явлений и формирования выборочной совокупности.

Существуют следующие методы отбора изучаемых явлений: случайный, механический, гнездовой, направленный, типологический.

Случайный отбор — это отбор, проводимый по жребию (по начальной букве фамилии или по дню рождения и т.п.).

Механический отбор — это отбор, когда у всей совокупности берется для изучения механически отобранная каждая пятая (20 %) или десятая (10 %) единица наблюдения.

Гнездовой (серийный) отбор — когда из генеральной совокупности выбираются не отдельные единицы, а гнезда (серии), которые отбираются путем случайной или механической выборки.

Пример: для изучения заболеваемости сельского населения М-ской области изучается заболеваемость сельского населения одного, наиболее типичного пункта. Результаты распространяются на все сельское население области.

Направленный отбор — это отбор, когда из генеральной совокупности с целью выявления определенных закономерностей отбираются только те единицы наблюдения, которые позволят выявить влияние неизвестных факторов при устранении влияния известных. Пример: при изучении влияния стажа рабочих на травматизм отбираются рабочие одной профессии, одного возраста, одного цеха, одного образовательного уровня.

Типологический отбор — это отбор единиц из заранее сгруппированных однотипных качественных групп. Пример: при изучении закономерности смертности среди городского населения следует сгруппировать изучаемые города по численности населения в них.

Характеристика исполнителей (кадры). Сколько человек и какой квалификации проводят исследование. Пример: исследование по изучению санитарно-гигиенического режима учащихся старших классов средних общеобразовательных школ района проводят два врача и два помощника санитарного врача центра гигиены и эпидемиологии данного административного округа.

Характеристика технического оснащения и требуемых материальных средств:

- лабораторное оборудование и приборы, соответствующие цели исследования;
- канцелярские товары (бумага, бланки); без дополнительных ассигнований.

Программа исследования (наблюдения) - это перечень вопросов, по которым собираются сведения, либо перечень признаков и показателей, подлежащих регистрации.

Программа исследования(наблюдения) включает:

- определение единицы наблюдения;
- перечень вопросов (учетных признаков), подлежащих регистрации в отношении каждой единицы наблюдения
- разработку индивидуального учетного (регистрационного) бланка с перечнем вопросов и признаков, подлежащих учету;
- разработку макетов таблиц, в которые затем вносятся результаты исследования.

На каждую единицу наблюдения заполняется отдельный бланк, он содержит паспортную часть, четко сформулированные, поставленные в определенной последовательности вопросы программы и дату заполнения документа. В качестве учетных бланков могут быть использованы применяемые в практике лечебно-профилактических учреждений учетные медицинские формы.

Источниками получения информации могут служить другие медицинские документы (истории болезни, и индивидуальные карты амбулаторного больного, истории развития ребенка, истории родов), отчетные формы лечебно-профилактических учреждений и др.

Для обеспечения возможности статистической разработки данных из этих документов производят выкопировку сведений на специально разработанные учетные бланки, содержание которых определяется в каждом отдельном случае в соответствии с задачами исследования.

Единица наблюдения — это каждый первичный элемент статистической совокупности.

Единица наблюдения наделена признаками сходства и различия, которые подлежат учету и дальнейшему наблюдению, поэтому эти признаки называются учитываемыми (учетными).

Учитываемые признаки — это признаки, по которым различаются элементы единицы наблюдения в статистической совокупности.

Признаки классифицируются:

- по характеру на:

а) качественные или атрибутивные (описательные) признаки — выражены словесно (пол, наличие вредных привычек, состояние здоровья и т.п.);

б) количественные признаки — выражены числом (возраст, число выкуриваемых сигарет, длительность заболевания, стаж курения и т.п.);

- по роли в совокупности на:

а) факторные признаки, влияющие на изучаемое явление (наличие вредных привычек и стаж курения);

б) результативные признаки, изменяющиеся под влиянием факторных признаков (состояние здоровья, наличие заболевания и т.п.).

Одним из важнейших моментов организационного этапа при составлении программы исследования является уточнение используемых в работе понятий и

терминов. Это обусловлено тем, что исследователи не всегда вкладывают одно и то же содержание в одно и то же понятие, термин.

Программа сбора материала представляет собой последовательное изложение учитываемых признаков — вопросов, на которые необходимо получить ответы при проведении данного исследования. Это может быть специально составленный исследователем опросный лист, анкета, карта. Документ должен иметь четкое название. Вопросы (учитываемые признаки) должны быть четкими, краткими, соответствовать цели и задачам исследования; на каждый вопрос следует предусмотреть варианты ответов. Эти варианты готовых ответов носят название "группировка".

Группировка признаков осуществляется с целью выделения однородных групп для изучения тех или иных закономерностей изучаемого явления.

II этап статистического исследования. Сбор материала.

Задачей этого этапа является регистрация каждой единицы наблюдения, её учетных признаков в строгом соответствии с программой наблюдения и заполнение учетных документов. Перед, и в ходе выполнения этой работы проводится инструктаж (устный или письменный) исполнителей наблюдения, обеспечение их формами регистрации.

Для сбора материала необходимо определить источники получения информации и разработать соответствующие документы. В медико-социальных исследованиях для изучения состояния здоровья (заболеваемости, физического развития, смертности, рождаемости и т.д.) в качестве источника используют информацию из официальной статистики государственных учреждений и организаций, из первичной медицинской документации (истории болезни, индивидуальные карты амбулаторного больного, истории развития ребенка и др.), данные медицинских осмотров. С целью выкопировки сведений из документации составляют выборочную карту, которая при изучении заболеваемости содержит паспортную часть (фамилия, имя и отчество обследуемого, его возраст, пол, место жительства и др.) и перечень вопросов:

даты обращения за медицинской помощью в связи с заболеванием, профилактическим осмотром и др.,

III этап статистического исследования. Разработка материала, статистическая группировка и сводка - начинается с проверки и уточнения числа наблюдений, полноты и правильности полученных сведений, выявлении и устранении ошибок, дубликатов записей и т. д.

Для правильной разработки материала применяется шифровка первичных учетных документов, т. е. обозначение каждого признака и его группы знаком - буквенным или цифровым.

Шифровка - это технический прием, облегчающий и ускоряющий разработку материала, повышающий качество, точность разработки. Шифры - условные обозначения - вырабатываются произвольно. При шифровке диагнозов рекомендуется пользоваться международной номенклатурой и классификацией болезней; при шифровке профессий - словарем профессий. Преимуществом шифровки является то, что при необходимости после окончания основной разработки можно вернуться к материалу для разработки с целью выяснения новых связей и зависимостей. Зашифрованный учетный материал позволяет сделать это легче и быстрее, чем незашифрованный. После проверки проводится группировка признаков.

Группировка — расчленение совокупности изучаемых данных на однородные, типичные группы по наиболее существенным признакам. Группировка может проводиться по качественным и количественным признакам. Выбор группировочного признака зависит от характера изучаемой совокупности и задач исследования.

Группировочный признак — это признак, по которому происходит объединение отдельных единиц совокупности в однородные группы. Выбор группировочных признаков всегда должен быть основан на анализе качественной природы изучаемого явления с тем, чтобы выявить существенные

признаки. Группировочные признаки могут быть выражены количественно и качественно.

Различают типологические и вариационные группировки.

А) Типологическая группировка производится по качественным (описательным, атрибутивным) признакам (пол, профессия, группы болезни)

Б) Вариационная группировка (по количественным признакам) проводится на основании числовых размеров признака (возраст, длительность заболевания, продолжительность лечения и т. д.). Количественная группировка требует решения вопроса о величине группировочного интервала: интервал может быть равным, а в ряде случаев - неравным, даже включать так называемые открытые группы (при группировке по возрасту могут быть определены открытые группы: до 1 года, 50 лет и старше).

При определении числа групп исходят из цели и задач исследования. Необходимо, чтобы группировки могли вскрыть закономерности изучаемого явления. Большое число групп может привести к чрезмерному дроблению материала, ненужной детализации. Малое число групп приводит к затушевыванию характерных черт.

Рекомендуется следующее число групп: до 40 наблюдений - 5-6 групп, 40-60 наблюдений 6-8 групп, 60-100 наблюдений - 7-10 групп и т. д.

Закончив группировку материала, приступают к сводке - *обобщение единичных случаев, полученных в результате статистического исследования, в определенные группы, их подсчет и внесение в макеты таблиц.*

Сводку статистического материала проводят при помощи статистических таблиц. Таблица, не заполненная цифрами, называется макетом.

Статистические таблицы бывают перечневые, хронологические, территориальные. Таблица имеет подлежащее и сказуемое. Статистическое подлежащее обычно размещается по горизонтальным строкам в левой части таблицы и отражает главный, основной признак. Статистическое сказуемое

размещается слева направо по вертикальным графам и отражает дополнительные учетные признаки.

Статистические таблицы делятся на:

- **простые** - представлено числовое распределение материала по одному признаку, составных частей его. Простая таблица содержит обычно простой перечень или итог по всей совокупности изучаемого явления (табл. 1).

- **групповые** - представлено сочетание двух признаков в связи друг с другом (табл. 2).

Таблица (простая) 1

Распределение курящих студентов по факультетам (в абс. числах и в % к итогу)

Наименование факультетов	Всего студентов	
	абсолютное число студентов	в %
1. Лечебный		
2. Медико-профилактический		
3. Фармацевтический и т.д.		
Итого:		100,0

Таблица (сгруппированная) 2

Распределение студентов различных факультетов по полу и возрасту, в котором они выкурили первую сигарету

Наименование факультетов	Пол		Возраст, в котором выкурили первую сигарету			Всего
	муж	жен	до 15 лет	15 - 18 лет	старше 18 лет	
1. Лечебный						
2. Медико-профилактический						
3. Фармацевтический и т.д.						
Итого:						

- **комбинационные** - дается распределение материала по трем и более взаимосвязанным признакам (табл. 3).

При составлении таблиц должны соблюдаться определенные требования:

- каждая таблица должна иметь заголовок, отражающий ее содержание;
- внутри таблицы все графы также должны иметь четкие краткие названия;

- при заполнении таблицы все клетки таблицы должны содержать соответствующие числовые данные. Оставшиеся незаполненными из-за отсутствия данной комбинации клетки таблицы прочеркиваются ("-"), а при отсутствии сведений в клетке проставляется "н. с." или "...";

- после заполнения таблицы в нижней горизонтальном ряду и в последнем справа вертикальном столбце подводятся итоги вертикальных граф и горизонтальных строк;

- таблицы должны иметь единую последовательную нумерацию.

Таблица (комбинационная) 3

Распределение курящих студентов различных факультетов по полу и среднему количеству сигарет (папирос), выкуриваемых в день

Наименование факультетов	Среднее количество сигарет (папирос), выкуриваемых студентами в день									Всего		
	10 и менее			11 - 20			более 20					
	м	ж	оба пола	м	ж	оба пола	м	ж	оба пола	м	ж	оба пола
1. Лечебный												
2. Медико- профилактический												
3. Фармацевтический и т.д.												
Итого:												

В исследованиях, имеющих небольшой объем наблюдений, сводка проводится вручную. Все учетные документы раскладываются на группы в соответствии с шифром признака. Далее проводится подсчет и запись данных в соответствующую клетку таблицы. В настоящее время в проведении сортировки и сводки материала широко используются ЭВМ. Которые позволяют не только отсортировать материал по изучаемым признакам, но выполнить расчеты показателей.

Статистическое обработка материала исследования определяется исходя из основных свойств статистической совокупности:

Групповые свойства статистической совокупности.

Статистическая совокупность как группа изучаемых явлений обладает особыми групповыми свойствами. Эти свойства присущи только статистической совокупности (табл. 4).

Таблица 4

Основные групповые свойства и статистические критерии совокупности

Групповые свойства	Статистические критерии
Распределение признака	Частота, доля (% , ‰, относительные величины)
Средний уровень признака	M_o – мода, M_e – медиана M – средняя арифметическая величина
Разнообразие признака	Lim – лимит; Am – амплитуда σ – среднее квадратическое отклонение
Репрезентативность признака	m_M – средняя ошибка средней арифметической $m_{\%}$ – средняя ошибка относительной величины (%) χ^2 – коэффициент соответствия (хи – квадрат)
Взаимосвязь между признаками	r_{xy} – коэффициент корреляции

К основным групповым свойствам статистической совокупности относятся следующие: распределение признака, его средний уровень, разнообразие и репрезентативность, взаимосвязь между признаками.

Распределение признака – это одно из важных свойств совокупности; оно заключается в том, что элементы совокупности распределяются неодинаково, по величине признака, образуя определенную внутреннюю структуру группы. Анализ структуры позволяет вскрыть важнейшие закономерности, присущие тому или иному явлению.

Средний уровень признака как свойство совокупности дает обобщенную характеристику различным величинам количественного признака, выразив его одним числом.

Средним уровнем характеризуют количественные признаки, т.е. признаки, имеющие числовые значения (рост, масса и др.).

Разнообразие признака указывает, что группа состоит из объектов, различных по величине изучаемого признака.

Репрезентативность признака (представительность, типичность) означает способность выборочных совокупностей отражать свойства генеральных совокупностей. Для измерения этого свойства используются ошибки репрезентативности (m).

Ошибки репрезентативности показывают, насколько результаты выборочного исследования отличаются от результатов, которые могли бы быть получены при сплошном изучении генеральной совокупности.

Взаимосвязь между признаками – это степень зависимости размеров признаков друг от друга.

1.4.1. Первое свойство статистической совокупности - распределение признака в статистической совокупности

Основными величинами, которыми оперирует статистика, являются абсолютные и относительные *величины*, отражающие качественную структуру статистической совокупности и характеризующие распределение признаков (*первое свойство статистической совокупности*).

Абсолютные величины используют очень широко. Они нужны для общей характеристики явления, признака: например, это численность населения в мире, в конкретной стране, в городе, районе и т. д.; при учете редких инфекционных заболеваний (малярия, полиомиелит и др.); число врачей, учреждений здравоохранения, коек и пр.

Однако при сравнении данных возникает необходимость применения *относительных величин* (или показателей), которые получают при

соотношении, сопоставлении двух сравниваемых чисел. Типы распределения признаков в статистической совокупности представлены на рис. 3.



Рис. 3. Типы распределения признаков в статистической совокупности

В целом относительные величины, характеризующие распределение признаков и применяемые в медицине, представлены следующими показателями:

- экстенсивные показатели
- интенсивные показатели

- показатели соотношения
- показатель наглядности
- показатели динамического ряда.

Для удобства сопоставления, обычно перечисленные показатели вычисляют на 100, 1000, 10 000, 100 000, для чего умножают на круглое число (100, 1000, 10 000, 100 000 и т. д.), называемое основанием. В результате полученные коэффициенты приобретают форму "процентов" (%), "промилле" (‰), "продецимилле" (‱) и т. д. Чем реже встречается изучаемое явление, тем больше числовое основание следует выбрать, чтобы не было коэффициентов меньше единицы, которыми неудобно пользоваться. Принято, например, структуру какого-то явления выражать в процентах, а демографические показатели (рождаемость, смертность) — в промилле; распространенность заболеваний — на 100000 населения.

Экстенсивные показатели (коэффициенты) характеризуют распределение явления на его составные части, определяют его внутреннюю структуру, удельный вес части в целом, долю признака в статистической совокупности. При вычислении экстенсивных показателей используют только одну статистическую совокупность и ее состав. Большинство экстенсивных показателей определяют в процентах (%), реже в промилле (‰) и долях единицы. Вычисление осуществляют по формуле:

$$\text{Экстенсивный показатель} = \frac{\text{часть явления}}{\text{явление в целом}} * 100\%$$

*Например, в прошедшем году среди учащихся 1-го класса было зарегистрировано 387 случаев заболеваний желудочно-кишечного тракта, в том числе 224 случая заболеваний гастритом. Удельный вес лиц, страдающих гастритом, составил $57,9\% \left(\frac{224}{387} * 100\% = 57,9\% \right)$.*

В качестве примеров экстенсивных показателей, используемых в здравоохранении, можно назвать структуру заболеваемости, смертности, инвалидности и т. д. Следует помнить, что экстенсивными показателями

следует пользоваться в тех случаях, когда необходимо охарактеризовать статистическую совокупность в данном месте и в данное время. Для сравнения показателей в динамике коэффициенты экстенсивности непригодны. Для этой цели нужно знать численность среды, в которой происходит явление, и пользоваться интенсивными показателями.

Интенсивные показатели (коэффициенты) характеризуют частоту (интенсивность, уровень, распространенность) явления за определенный промежуток времени (чаще за год) в среде, в которой это явление происходит и с которой оно непосредственно связано. При вычислении интенсивных показателей необходимо знание двух статистических совокупностей, одна из которых представляет среду, другая — явление. В демографической и санитарной статистике в качестве среды чаще всего рассматривают население и в расчете на него вычисляют то или иное явление, например, число рождений, число заболеваний, общее число смертей за год на определенную численность населения. Показатель интенсивности рассчитывают на 100, 1000, 10 000 или 100 000) по формуле:

$$\text{Интенсивный показатель} = \frac{\text{явление}}{\text{среда}} * 1000 (\text{‰})$$

Типичными интенсивными показателями являются демографические показатели и коэффициенты заболеваемости, вычисляемые для населения в целом или для отдельных определенных групп людей, например, в одном из регионов страны за год родились живыми 3000 детей, в течение 1-го года умерли 52 ребенка. Показатель младенческой смертности составил 17,3 ‰ ($\frac{52}{3000} * 100\%$).

Показатели соотношения (коэффициенты) характеризуют частоту встречаемости признака в статистической совокупности, его распространенность и применяются при сравнении двух, не связанных между собой статистических совокупностей, но сопоставимых логически и по содержанию. К ним относят такие показатели, как число врачей, число

больничных коек на 1000 населения, количество различных исследований, переливаний крови на 100 (или 1000) больных. По методике вычисления показатели соотношения схожи с показателями интенсивности, но различны по содержанию.

Динамический ряд — это совокупность однородных статистических величин, показывающих изменение какого-либо явления (признака) во времени. Числа, из которых составляется динамический ряд, могут быть представлены абсолютными, средними и относительными величинами. Если динамический ряд состоит из абсолютных величин, то он называется простым, если он составлен из средних или относительных чисел, то такой динамический ряд называется сложным или производным.

Динамические ряды бывают двух видов:

- моментными, состоящими из величин, характеризующих явление на определенный момент времени (на конец года, месяца, декады и т. д.);
- интервальными, состоящими из величин, которые характеризуют явление за определенный промежуток (интервал) времени (за год, месяц и т. п.).

С целью изучения особенностей изучаемого процесса и достижения наглядности в характеристике рассматриваемого явления рассчитывают специальные показатели динамического ряда, среди которых следует назвать абсолютный прирост, темп прироста (убыли), показатель роста (снижение), показатель наглядности (табл. 3.6).

Абсолютный прирост представляет собой разность между последующим и предыдущим уровнем. Он измеряется в тех же единицах, что и уровни динамического ряда.

Темп прироста (убыли) показывает отношение абсолютного прироста (снижения) каждого последующего уровня к пре

Динамический ряд, как правило, позволяет проследить основную закономерность явления, проявляющуюся в последовательном снижении или увеличении показателей динамического ряда.

Анализ динамического ряда включает в себя вычисление ряда показателей, при необходимости проведение выравнивания и графическое изображение.

1.4.2. Второе свойство статистической совокупности — средний уровень признака в статистической совокупности

В медицине и здравоохранении при анализе, кроме распределения, довольно часто используют средние величины, характеризующие физическое развитие пациентов (рост, масса тела, окружность груди и т. д.), данные их обследования (частота дыхания и пульса, артериальное, внутриглазное и внутричерепное давление и т. д.), результаты анализов (содержание гемоглобина, число эритроцитов, лейкоцитов, СОЭ) и др. В каждой совокупности и в данных конкретных условиях один и тот же признак отличается от величины этого же признака в другой совокупности, при наличии других условий. Так, величина пульса, артериального давления, температуры тела, длительность временной нетрудоспособности и другие критерии отличаются (варьируют) даже у больных с одним диагнозом. Иными словами, признаки могут принимать различные числовые значения у разных единиц совокупности, при этом нередко могут повторяться у нескольких единиц наблюдения.

Полученные при исследовании одного и того же признака у единиц наблюдения статистической совокупности абсолютные величины сначала записывают в том порядке, как их получает исследователь, т. е. хаотично.

Каждое числовое значение изучаемого признака называется *вариантой* (V), а числа, показывающие, как часто повторяются одни и те же варианты, называют *частотой* (P).

Ряд вариантов одного и того же признака, расположенных в определенном порядке (по степени возрастания или убывания), с соответствующими им частотами, образуют *вариационный ряд*.

Вариационные ряды бывают простые или несгруппированные, которые составляют, как правило, при малом числе наблюдений (до 30 единиц наблюдения), и сгруппированные, которые составляют при большом числе

наблюдений (более 30 единиц наблюдения).

Обобщенной характеристикой вариационного ряда являются *средние величины*, положительные качества которых заключаются в том, что они характеризуют большую совокупность однородных явлений.

Различают несколько видов средних величин: «Мода, медиана, средняя арифметическая, средняя геометрическая, средняя гармоническая и т. д.

Модой (Мо) называется варианта, встречающаяся с наибольшей частотой.

Медианой (Ме) — варианта, которая делит вариационный ряд пополам и расположена в середине вариационного ряда, если ряд нечетный, и если ряд четный, то определяется как полусумма двух средних вариантов.

Наиболее часто в характеристике вариационного ряда используют *среднюю арифметическую*. Средняя арифметическая, которая рассчитана в вариационном ряду, где каждая варианта встречается только 1 раз, называется *средней арифметической простой*. Ее определяют по формуле:

$$M = \frac{\sum V}{n},$$

где М — средняя арифметическая,

V — варианта изучаемого признака,

n — число наблюдений.

Если в исследуемом ряду одна или несколько вариантов повторяются несколько раз, то вычисляют *среднюю арифметическую взвешенную*, когда учитывается вес каждой варианты в зависимости от частоты ее встречаемости.

Расчет такой средней проводят по формуле: $M = \frac{\sum V \cdot p}{n}$,

где М - средняя арифметическая взвешенная,

V - варианты (числовые значения изучаемого признака), p — частота, с которой встречается одна и та же варианта признака, т. е. сумма вариантов с данным значением признака,

n - число наблюдений, т. е. сумма всех частот или общее число всех вариантов ($\sum p$).

Например, при определении среднего пульса у студентов перед экзаменом

следует сначала вычислить $\Sigma V \cdot p$, а затем среднюю величину, которая составила $M = 76,9$ уд/мин ($2000/26$) (табл. 5).

Таблица 5

Определение среднего пульса у студентов-мужчин 20-22 лет

Пульс у студентов-мужчин (V)	Число студентов (p)	$V \cdot p$
68	1	68
69	1	69
72	4	288
76	5	380
77	4	308
79	7	553
80	1	80
84	2	168
86	1	86
	$\Sigma p = n = 26$	$\Sigma V \cdot p = 2000$

Нередко при большом числе наблюдений для вычисления средней арифметической взвешенной используют сгруппированный вариационный (или разбитый на равные интервалы) ряд. Такой вариационный ряд должен быть непрерывным, варианты, расположенные в определенном порядке (возрастания или убывания), следуют друг за другом (табл. 6).

При группировке вариационного ряда следует учитывать, что интервал выбирает исследователь, величина интервала зависит от цели и задач исследования.

Число групп в сгруппированном вариационном ряду определяют в зависимости от числа наблюдений. При числе наблюдений от 31 до 100 рекомендуется иметь 5-6 групп, от 101 до 300 - от 6 до 8 групп, от 301 до 1000 наблюдений можно использовать от 10 до 15 групп. Расчет интервала (i) производится по формуле (округление в сторону увеличения):

$$i = \frac{V_{max} - V_{min}}{n}$$

где i - величина интервала,

$V_{\max}$ - максимальное значение варианты,

$V_{\min}$ - минимальное значение варианты.

Таблица 6

Определение среднего роста студентов-мужчин 20-22 лет

Рост студентов-мужчин (V), см.	Центральная варианта группы (V_1)	Число студентов (p)	$V_1 \cdot p$
160-164	162	4	648
165-169	167	21	3507
170-174	172	47	8084
175-179	177	68	12036
180-184	182	54	9828
185-189	187	18	3366
		$\Sigma p = n = 212$	$\Sigma V_1 \cdot p = 37469$

Расчет средней взвешенной в сгруппированном (или интервальном) ряду требует определения середины интервала, которую вычисляют как полусумму крайних значений группы.

Пример составления сгруппированного вариационного ряда представлен в табл. 6, а расчет средней величины производят по формуле:

$$M = \frac{\Sigma V_1 \cdot p}{n} = \frac{37469}{212} = 176,7 \text{ см.}$$

Однако при большом числе наблюдений, достаточно протяженном вариационном ряду рекомендуется среднюю взвешенную вычислять по способу моментов (табл. 7).

Таблица 7

Определение среднего роста студентов-мужчин 20-22 лет

Рост студентов-мужчин (V), см.	Центральная варианта группы (V_1)	Число студентов (p)	$a = \frac{V_1 - A}{i}$	$a \cdot p$
160-164	162	4	-3	-12

165-169	167	21	-2	-42
170-174	172	47	-1	-47
175-179	177	68	0	0
180-184	182	54	+1	+54
185-189	187	18	+2	+36
		$\Sigma p = n = 212$		$\Sigma a \cdot p = -11$

Этот способ основан на том, что средняя равна любой произвольно (условно) взятой средней (M_1) за которую чаще всего принимается мода (M_0), плюс среднее отклонение всех вариантов от условно средней (первый момент средней):

$$M = A_i + \frac{\Sigma a \cdot p}{n} i ,$$

где M - средняя арифметическая (взвешенная),

A_1 - условно взятая средняя величина (наиболее чаще встречающееся величина),

i - величина интервала,

a - отклонение между центральными вариантами групп и условной средней величиной, выраженное в интервалах $\left(\frac{V_1 - A_i}{i}\right)$,

p - частота (число раз, с которым встречается одна и та же варианта признака),

n — число наблюдений, т. е. сумма всех частот или общее число всех вариантов (Σp);

$$M = A_i + \frac{\Sigma a \cdot p}{n} i = 177 + \frac{-11 \cdot 5}{212} = 176.74 \text{ см.}$$

Таким образом, средняя взвешенная, вычисленная по способу моментов, составила 176,74 см, что практически совпало с расчетами средней обычным методом - 176,7 см. Однако при вычислении средней по способу моментов используют простые цифры, вычисления менее громоздки, что значительно облегчает и ускоряет расчеты.

Средняя арифметическая (средняя взвешенная) имеет ряд свойств, которые используют в некоторых случаях для упрощения расчета средней и получения ориентировочной величины.

1. Средняя арифметическая занимает срединное положение в строго симметричном вариационном ряду ($M = M_o = M_e$).
2. Средняя арифметическая имеет абстрактный характер и является обобщающей величиной, выявляющей закономерность.
3. Алгебраическая сумма отклонений всех вариантов от средней равна нулю. На этом свойстве основан расчет средней по способу моментов.
4. Если к каждой вариате вариационного ряда прибавить или отнять одно и то же число, то на столько же увеличится или уменьшится средняя арифметическая величина.
5. Если каждую вариату разделить или умножить на одно и то же число, то во столько же раз уменьшится или увеличится средняя арифметическая.

Два последних свойства используют в тех случаях, когда варианты представлены очень малыми или наоборот большими числами.

В медицине и здравоохранении средними величинами оценивают отдельные показатели (параметры физического развития), сравнивая данные лабораторных и других исследований с нормой. Следует учитывать, что средние рассчитывают на большой однородной группе, поскольку нарушение этого принципа приводит к искажению реальных процессов.

Графическое изображение вариационного ряда может быть представлено в виде графиков симметричного асимметричного распределения в зависимости от числа наблюдений и изучаемого признака.

1.4.3. Третье свойство статистической совокупности — разнообразие признака в статистической совокупности

Разнообразие признака характеризует третье свойство статистической совокупности, которое проявляется в том, что в однородных статистических

совокупностях величины количественных признаков различны. Например, при измерении окружности головы 2-летних детей в трех группах были получены данные, представленные в табл. 8.

Таблица 8

Окружность голову у детей 2-летнего возраста в трех группах

Показатель	1-я группа (n = 42)	2-я группа (n = 42)	3-я группа (n = 42)
Окружность головы (V), см.	46	44 45 46 47 48	42 43 44 45 46 47 48 49 50
Число детей (p)	42	1 8 25 6 2	1 2 4 6 14 10 3 - 2

В каждой из сравниваемых групп было одно и то же число наблюдений, равное 42, одна и та же средняя взвешенная (окружность головы), составившая 46 см, но группы были различны по составляющим их вариантам. Если в 1-й группе все дети имели окружность головы 46 см, то во 2-ю группу вошли дети с окружностью головы от 44 до 48 см, а в 3-ю — от 42 до 50 см. В этом проявляется разнообразие, колеблемость признака изучаемой совокупности.

Наибольшее разнообразие признака (окружность головы) представлено в 3-й группе, в которую вошли дети с наименьшими и наибольшими размерами окружности головы. *Статистическими критериями*, характеризующими разнообразие признака, являются:

- лимит ($\lim$);
- амплитуда (A_m);
- среднее квадратическое отклонение (σ);
- коэффициент вариации (C_v).

Лимит ($\lim$) определяется крайними значениями вариант в вариационном ряду ($V_{\max} - V_{\min}$) и составляет для 2-й группы $\lim_2 =$ от 44 до 48 см, для 3-й группы $\lim_3 =$ от 42 до 50 см.

Амплитуда (A_m) равна разности между крайними вариантами ($V_{\max} - V_{\min}$). Для 2-й группы амплитуда составила $A_{m2} = 4$ см, для 3-й - $A_{m3} = 8$ см.

И лимит, и амплитуда свидетельствуют о большем разнообразии признака в 3-й группе (с учетом лишь крайних значений вариант вариационного ряда).

Наиболее полную характеристику разнообразию вариационного ряда дает среднее квадратическое отклонение (σ), которое учитывает разнообразие всех вариант вариационного ряда. Существует два способа расчета этого показателя. Один из них простой и называется среднеарифметическим, для чего используют формулу:

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}},$$

где σ - среднее квадратическое отклонение;

d- разница между вариантой и средней арифметической (V - M);

n - число наблюдений.

Этот способ вычисления среднего квадратического отклонения применяют, если число наблюдений не превышает 30 ($n < 30$), каждая варианта встречается 1 раз ($P = 1$), то среднюю величину вычисляют как простую среднюю арифметическую. В том случае, если варианты имеют различную частоту больше 1, то вычисляют среднее, взвешенное квадратическое отклонение и применяют формулу:

$$\sigma = \sqrt{\frac{\sum d^2 * p}{n-1}},$$

где σ — среднее квадратическое отклонение;

d— разница между вариантой и средней арифметической (V - M);

n — число наблюдений;

p — частота, с которой встречается каждая варианта.

В качестве примера вычисления средней и среднего квадратического отклонения при числе наблюдений $n < 30$ приведем определение пульса у студентов перед сдачей экзамена (табл. 9).

Таблица 9

Определение среднего квадратического отклонения среднего пульса у студентов-мужчин 20-22 лет перед сдачей экзамена

Пульс у студентов-мужчин (V)	Число студентов (p)	$V \cdot p$	$d = (V - M)$	d^2	$d^2 \cdot p$
68	1	68	-8,9	79,21	79,21
69	1	69	-7,9	62,41	62,41
72	4	288	-4,9	24,01	96,04
76	5	380	-0,9	0,81	0,81
77	4	308	0	0	0
79	7	553	+2,1	4,41	30,87
80	1	80	+3,1	9,61	9,61
84	2	168	+7,1	50,41	100,82
86	1	86	+9,1	82,81	82,81
$\Sigma p = n = 26$		$\Sigma V \cdot p = 2000$			$\Sigma d^2 \cdot p = 462,58$

Для вычисления среднего квадратического отклонения необходимо рассчитать отклонение каждой варианты от средней, возвести в квадрат, умножить на частоту встречаемости данной варианты, суммировать произведение и подставить в формулу:

$$\sigma = \sqrt{\frac{\Sigma d^2 \cdot p}{n - 1}} = \sqrt{\frac{462,58}{26 - 1}} = \sqrt{18,50} = 4,3$$

Таким образом, среднее квадратическое отклонение (σ) при обследовании 26 студентов ($n = 26$) и средней величине пульса (M) 76,9 уд/мин. соответствовал 4,3.

В том случае, если число наблюдений превышает 30 и среднюю величину рассчитывают по способу моментов, то и среднее квадратическое отклонение вычисляют по способу моментов по формуле:

$$\sigma = \sqrt{\frac{\Sigma a^2 \cdot p i^2}{n} - \left(\frac{\Sigma a \cdot p i}{n}\right)^2}$$

где $\sqrt{\frac{\sum a \cdot pi}{n}}$ – первый момент средней,

$\sqrt{\frac{\sum a^2 \cdot pi^2}{n}}$ – второй момент средней

a - разница между вариантой и условно средней,

p - частота, с которой встречается каждая варианта,

i - величина интервала,

n - число наблюдений.

Для вычисления среднего квадратического отклонения в таблицу вычисления средней по способу моментов необходимо включить еще одну графу (табл. 10).

Среднее квадратическое отклонение при обследовании (n) 212 студентов и M = 176,74 см. составило:

$$\sigma = \sqrt{\frac{\sum a^2 \cdot pi^2}{n} - \left(\frac{\sum a \cdot pi}{n}\right)^2} = \sqrt{\frac{293 \cdot 5^2}{212} - \left(\frac{-11 \cdot 5}{212}\right)^2} = \sqrt{34,55 + (0,26)^2} = \sqrt{36,62} = 5,88$$

По величине среднего квадратического отклонения можно судить о разнообразии вариационного ряда: чем больше величина σ , тем больше разнообразие, чем меньше значение σ тем меньше разнообразие вариантов и тем более однороден вариационный ряд.

Таблица 10

Определение среднего роста студентов-мужчин 20-22 лет

Рост студентов-мужчин (V), см.	Центральная варианта группы (V ₁)	Число студентов (p)	a $= \frac{V_1 - M}{i}$	a · p	a ² · p
--------------------------------	---	---------------------	----------------------------	-------	--------------------

160-164	162	4	-3	-12	36
165-169	167	21	-2	-42	84
170-174	172	47	-1	-47	47
175-179	177	68	0	0	0
180-184	182	54	+1	+54	54
185-189	187	18	+2	+36	72
		$\Sigma p = n = 212$			$\Sigma a^2 \cdot p = 293$
				$\Sigma a \cdot p = -11$	

Среднее квадратическое отклонение связано со структурой распределения.

При нормальном распределении в пределах $M \pm \sigma$ находится 68,3 %, $M \pm 2\sigma$ - 95,5 %, $M \pm 3\sigma$ - 99,7 % всех вариантов (рис.3).

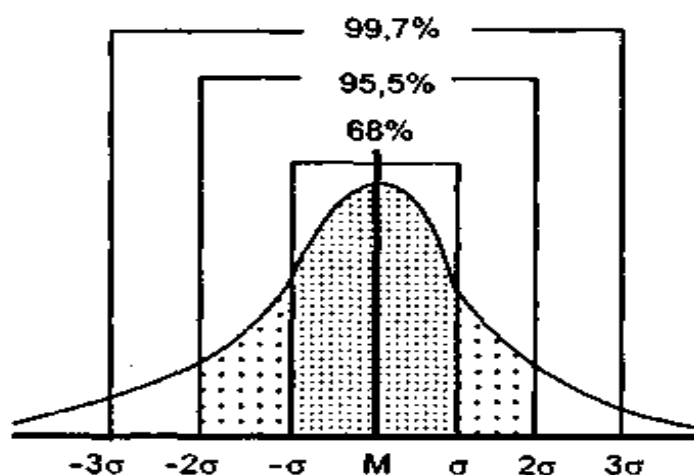


Рис. Связь среднего квадратического отклонения со структурой вариационного ряда.

Иными словами, если 95 % всех вариантов вариационного ряда находится в пределах $M \pm 2\sigma$, то средняя величина характерна для данного ряда и можно говорить о ее представительности для статистической совокупности и не требуется увеличивать число наблюдений. Практическое значение среднего квадратического отклонения состоит в том, что по его величине можно:

- определить структуру вариационного ряда;
- охарактеризовать степень однородности вариационного ряда в

- зависимости от величины a ;
- судить о типичности средней арифметической в зависимости от
- распределения вариант в вариационном ряду;
- оценить отдельные признаки у каждого индивидуума по стандартному отклонению от средней арифметической по формуле:

$$t = \frac{V-M}{\sigma},$$

где t – доверительный интервал,

M – средняя величина;

- определить коэффициент вариации при сравнении степени разнообразия разных признаков в одной совокупности или однородных признаков в разных совокупностях;
- определить достоверность (репрезентативность) результатов исследования.

Коэффициент вариации (C_v) является относительной мерой разнообразия, так как вычисляется по отношению среднего квадратического отклонения (σ) к средней арифметической величине (M), выраженному в процентах (100 %):

$$C_v = \frac{\sigma}{M} * 100\%$$

Коэффициент вариации применяют в том случае, когда необходимо сравнить разнообразие разных признаков в одной совокупности (разные показатели крови) или одного признака в разных совокупностях (масса тела у младенцев и подростков). Для ориентировочной оценки степени разнообразия признака пользуются следующими градациями коэффициента вариации. При величине коэффициента вариации (C_v) больше 20 % отмечают сильное разнообразие, если C_v от 20 до 10 %, то разнообразие среднее, а если C_v меньше 10 %, то разнообразие слабое.

1.4.4. Четвертое свойство статистической совокупности — достоверность признака в статистической совокупности

При изучении генеральной (сплошной) совокупности для ее количественной характеристики достаточно рассчитать M и σ . Однако на практике, как правило, исследование проводят на выборочной совокупности, которая должна быть репрезентативна (достоверна) или представительна для генеральной совокупности. *Репрезентативность* выборочной совокупности оценивают специальными методами отбора, она *означает* представительность в ней всех учитываемых признаков генеральной совокупности.

Оценка *достоверности* результатов исследования базируется на теоретических основах репрезентативности и дает возможность достаточно обоснованно охарактеризовать выявленные закономерности. При определении достоверности результатов

исследования необходимо установить, с какой вероятностью можно перенести результаты изученных признаков выборочной совокупности (части явления) на всю генеральную совокупность (явление в целом). Оценка результатов исследования предусматривает вычисление:

- средней ошибки (m) для средних (M) или относительных (P) величин;
- доверительных границ средних (M) или относительных (P) величин;
- достоверности разности средних (M) или относительных (P) величин по

критерию t .

Среднюю ошибку (m), которая является важнейшей статистической величиной, определяют по формулам:

- для относительной величины (P)

$$m_{\%} = \sqrt{\frac{pq}{n}},$$

- для средней величины (M) при $n \leq 30$

$$m_M = \frac{\sigma}{\sqrt{n-1}},$$

- для средней величины (M) при $n > 30$

$$m_M = \frac{\sigma}{\sqrt{n}},$$

где m – ошибка средней или относительной величины

P – показатель, выраженный в процентах (%), промилле (‰) и т.д.

q – величина, равная $100 - P$ (при вычислениях в процентах, т.е. когда P выражен в процентах), $1000 - P$ (при вычислениях в промиллях)

σ – среднееквадратическое отклонение;

n – число наблюдений.

Величина ошибки средней арифметической прямо пропорциональна степени разнообразия признака в статистической совокупности и обратно пропорциональна квадратному корню из числа наблюдений. Следовательно, уменьшение величины ошибки возможно за счет либо снижения степени разнообразия признака, либо увеличения числа наблюдений. Как правило, в медико-статистических исследованиях обычно используют доверительную вероятность (надежность), равную 95,5 - 99,7 %.

Таким образом, если среднее квадратическое отклонение (σ) характеризует степень разнообразия вариантов в вариационном ряду и необходима для определения ошибки средней, то величина ошибки (m) вместе с доверительным коэффициентом (t) является важным условием определения доверительного интервала, с помощью которого оценивают доверительные границы изучаемого признака.

Например, при вычислении среднего пульса у 26 студентов перед экзаменом ($M = 76,9$ уд/мин) средняя ошибка составила:

$$m = \frac{4,3}{26-1} = 0,86$$

Это означает, что при аналогичных исследованиях при доверительной вероятности 68,3 % средняя величина пульса у студентов перед экзаменом может быть в пределах $76,9 \pm 0,86$, или от 76 до 78 уд/мин.

При более высокой степени доверительной вероятности (95,5 %, 99,7 % и др.) средняя величина будет иметь больший предел варьирования среднего результата, который определяется произведением " $t \cdot m$ ", где доверительный

коэффициент (t) берется из соответствующей таблицы (табл.11).

Таблица 11
Значение критерия Стьюдента (t)

K = n-1	Доверительная вероятность (P)			K = n-1	Доверительная вероятность (P)		
	95,5%	99,7%	99,9%		95,5%	99,7%	99,9%
1	12,70	63,36	637,59	16	2,12	2,92	4,02
2	4,30	9,92	31,60	17	2,11	2,90	3,96
3	3,18	5,84	12,94	18	2,10	2,88	3,92
4	2,78	4,60	8,61	19	2,09	2,86	3,88
5	2,57	4,03	6,86	20	2,09	2,84	3,85
6	2,42	3,71	5,96	21	2,08	2,83	3,82
7	2,36	3,50	5,31	22	2,07	2,82	3,79
8	2,31	3,36	5,04	23	2,07	2,81	3,77
9	2,26	3,25	4,78	24	2,06	2,80	3,75
10	2,23	3,17	4,59	25	2,06	2,79	3,73
11	2,20	3,11	4,44	26	2,06	2,78	3,71
12	2,18	3,06	4,32	27	2,05	2,77	3,69
13	2,16	3,01	4,22	28	2,05	2,76	3,67
14	2,14	2,98	4,14	29	2,04	2,76	3,66
15	2,13	2,95	4,07	30	2,04 – 1,96	2,75 – 2,58	3,64 – 3,29

При обследовании 26 студентов и доверительном коэффициенте $t = 2,06$ (табл. 12), если доверительная вероятность равна 95,5 %, средний пульс у студентов при аналогичных исследованиях составит $76,9 \pm 2,06 \cdot 0,86$ уд/мин, или его величина будет в среднем от 75 до 79 уд/мин.

При большом числе наблюдений ($n > 30$), как правило, доверительный коэффициент (t) берется равным 1 при $P = 68,3 \%$, 2 — при $P = 95,5 \%$ и 3 — при $P = 99,7 \%$. Так, при обследовании 212 студентов-мужчин 22 лет (см. табл. 7) и вычислении среднего роста ($M = 176,74$ см), средней ошибки ($m = 0,4$ см) предел колеблемости средней при повторных аналогичных исследованиях при $P = 68,3 \%$ составит 176—178 см ($176,74 \pm 0,4$ см), а при $P = 95,5 \%$ и $t = 2$ средний рост студентов-мужчин будет в пределах от 175 до 178 см ($176,74 \pm 2 \cdot 0,4$ см).

Таким образом, при $t=1$ вероятность приведенного предела колебаний средней величины (средний пульс, средний рост и т. д.) составит лишь 68,3 %, а 31,7 % полученных средних могут оказаться вне вычисленных пределов. При $t=2,06$ или $t=2,0$ вероятность получения средней в обозначенных пределах при проведении аналогичных исследований составит 95,5 % и лишь 4,5 % результатов окажется вне этих границ. Следовательно, с повышением доверительной вероятности увеличивается ширина доверительного интервала.

Сравнение средних и относительных величин по критерию t .

На практике нередко для того, чтобы сделать вывод об эффективности предлагаемого нового метода лечения или диагностики, приходится сравнить результаты, полученные в исследуемой и контрольной группах. Целью сравнения двух средних (M) или относительных показателей (P) являются оценка существенности их различий, установление их достоверности.

Достоверность разности между двумя средними и относительными величинами определяют по формулам:

$$\text{для средних величин } t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}},$$

$$\text{для относительных величин } t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}}$$

где M_1 и M_2 - средние величины, полученные в двух самостоятельных независимых группах наблюдений (исследуемая и контрольная);

P_1 и P_2 - показатели изучаемого признака, выраженные в процентах;

m_1 и m_2 - средние ошибки средних или относительных величин;

t - доверительный коэффициент.

При $t > 2$ различие между двумя средними величинами существенно и не случайно, т. е. достоверно. Это значит, что в генеральной совокупности

сравниваемые средние величины (или относительные величины) имеют различие и при повторении подобных наблюдений будут получены аналогичные различия. При $t = 2$ надежность такого вывода будет не меньше 95 %. С увеличением критерия достоверности (t) степень надежности различия между средними величинами (или относительными величинами) также повышается, а риск ошибки уменьшается. При $t < 2$ достоверность разности средних величин (или относительных величин) считается недоказанной.

Так, изучение содержания гормонов в крови показало, что уровень эстрадиола у мальчиков в возрасте 7-10 лет составил среди больных эпилепсией $23,5 \pm 3,4$ (45 детей), в контрольной группе $32,4 \pm 2,3$ (53 ребенка). Для определения достоверности различия средних необходимо использовать приведенную ранее формулу и провести следующие расчеты:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{32,4 - 23,5}{\sqrt{2,3^2 + 3,4^2}} = \frac{8,9}{4,1} = 2,17$$

Полученные данные свидетельствуют о том, что различие между двумя средними существенно, достоверно, не случайны, статистически значимы, так как $t = 2,17$ и, следовательно, с достоверностью более 95,5 % можно утверждать о снижении содержания эстрадиола в крови 7—10-летних мальчиков, больных эпилепсией, в сравнении с этим показателем у детей в контрольной группе.

При анализе числа нейроэндокринных нарушений у детей, больных эпилепсией (115 человек), и в группе сравнения (92 человека) было отмечено, что нарушения пубертатного развития среди больных отмечаются в 27,4 %, а в контрольной группе — в 11,2 % случаев. Для определения достоверности различия между двумя показателями распространенности нарушений пубертатного развития сначала следует вычислить ошибки для каждого показателя по соответствующим формулам:

$$m_1 = \pm \sqrt{\frac{P * q}{n}} = \pm \sqrt{\frac{27,4 * 72,6}{115}} \pm \sqrt{17,3} = \pm 4,1\%$$

$$m_2 = \pm \sqrt{\frac{P * q}{n}} = \pm \sqrt{\frac{11,2 * 88,8}{92}} = \pm \sqrt{10,8} = \pm 3,3,$$

Далее производим вычисление по формуле:

$$t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} = \frac{27,4 - 11,2}{\sqrt{4,1^2 + 3,3^2}} = \frac{16,2}{5,3} = 3,05.$$

Полученные данные свидетельствуют о том, что имеющиеся различия в показателях существенны, достоверны, не случайны, статистически значимы, поэтому можно утверждать о большей распространенности нарушений пубертатного развития у детей, больных эпилепсией, по сравнению со здоровыми детьми. Если доверительный коэффициент $t > 3$, то с достоверностью более 99,7 % можно утверждать, что число нарушений пубертатного развития у детей с эпилепсией в 2,4 раза выше, чем у здоровых детей.

Определение необходимого объема наблюдений.

Зная величину ошибки, можно рассчитать необходимый объем наблюдений для проведения выборочного исследования и получения достоверных результатов. Поскольку искомые закономерности можно выявить, исследовав только генеральную совокупность, не следует забывать, что проведение выборочных исследований требует существенно меньше средств. Однако получение достоверных результатов возможно лишь в том случае, если выборочная совокупность является репрезентативной, т. е. представительной по отношению к генеральной совокупности по количеству и по качеству. Научной основой выборочного метода является закон больших чисел и теория вероятности, которые позволяют сформулировать следующие основные положения закона больших чисел:

- при увеличении числа наблюдений данные выборочной совокупности стремятся воспроизвести данные генеральной совокупности;
- при достижении определенного, достаточного количества данные выборочной совокупности воспроизводят данные генеральной совокупности.

Статистические методы позволяют обрабатывать любое число наблюдений

(как было показано ранее), но чтобы застраховать себя от получения недостоверного результата, следует заранее определить необходимый объем наблюдений для получения достоверных результатов. С этой целью используют специальные формулы, полученные при помощи преобразования формул предельных ошибок для средних и относительных величин:

$$tm_{\text{абс.}} = t \frac{\sigma}{\sqrt{n}}, \quad tm_{\%} = t \sqrt{\frac{P \cdot q}{n}},$$

где n — число наблюдений,

t — доверительный коэффициент,

σ — среднее квадратическое отклонение,

p — результирующий признак, выраженный в процентах,

q — величина, равная $100 - p$,

$tm = \Delta$ — предельно допустимая максимальная ошибка с достаточной степенью вероятности.

Решая приведенные равенства путем алгебраических преобразований относительно " n ", получают формулы для определения числа наблюдений, когда выбранный признак берется:

- в абсолютных цифрах $n = \frac{t^2 \sigma^2}{\Delta^2}$,
- в относительных величинах $n = \frac{t^2 \cdot P \cdot q}{\Delta^2}$

Приведенные формулы для случайного повторного отбора определяют оптимальный объем выборки для исследования существующих закономерностей и получения достоверных результатов. Необходимо лишь выбрать признак или признаки для проведения соответствующих вычислений, что в ряде случаев затруднено, поскольку единицы наблюдения в исследуемых совокупностях характеризуются множеством учетных признаков. Так, в комплексных социально-гигиенических исследованиях изучают не только комплекс показателей, свидетельствующих о состоянии здоровья изучаемой группы населения, но и окружающие условия, факторы, воздействующие на

здоровье. Вести расчет по всем исследуемым признакам невозможно, поэтому, как правило, выделяют 2—3 результирующих признака и конкретно по ним рассчитывают необходимый объем наблюдений.

Сведения о критериях, характеризующих разнообразие признаков (σ и P) и нужных для расчета необходимого числа наблюдений, получают:

- из данных литературы;
- если признак выражен абсолютным числом и в литературе данных о нем нет, то используют результаты пробного исследования, объем которого обычно составляет не менее 30 единиц наблюдения;
- если признак выражен относительным числом, то берут значение максимального произведения $pq = 2500$ ($p = q = 50\%$) или $pq = 0,25$ ($p = q = 0,5$), так как не следует пробное исследование проводить и вычислять проценты при числе наблюдений менее 100.

Степень доверительной вероятности при вычислении необходимого объема наблюдений зависит от целей и задач исследования, но обычно составляет не менее 95,5 % (или 0,95), при которой $t = 2$. Величина максимальной ошибки (Δ) выражается в единицах изучаемого признака, характеризует допускаемую неточность искомой величины в сравнении с генеральной совокупностью, задается в разумных пределах в зависимости от целей и задач исследования. Незначительное увеличение Δ ведет к резкому сокращению необходимого объема выборки, напротив, незначительное уменьшение Δ резко его увеличивает.

Например, для изучения состояния здоровья детей первых 3 лет жизни в связи с особенностями состояния здоровья родителей необходимо рассчитать среднее квадратическое отклонение, равное 1,6; число наблюдений для получения достоверного результата. В качестве результирующего признака была взята кратность перенесенных детьми заболеваний на 2-м году жизни. Индивидуальная характеристика заболеваемости детей ранее не изучалась. При проведении пробного исследования, включающего обследование 28 детей 2-го

года жизни, были получены следующие данные:

M - среднее число перенесенных ребенком заболеваний в год составило 3,8;

σ — предельно допустимая ошибка, равная 0,2;

p - доверительная вероятность, равная 95,5 %;

t - доверительный коэффициент, равный 2.

Расчет производят по формуле: $n = \frac{t^2 \sigma^2}{\Delta^2} = \frac{2^2 \cdot 1,62}{0,2^2} = 256$ единиц наблюдения

Вывод: для получения статистически достоверных данных о состоянии здоровья детей первых 3 лет жизни в связи с особенностями здоровья их родителей необходимо обследовать 256 детей.

В качестве другого результирующего признака был взят показатель "индекс здоровья", отражающий удельный вес детей, не болевших на 1-м году жизни. При анализе данных литературы было установлено, что среди детей количество не болевших на 1-м году жизни в Москве составило 9,5 %. При доверительной вероятности 95,5 % и предельно допустимой ошибке 3,5 % необходимое число наблюдений было определено по формуле и составило 280,7:

$$n = \frac{t^2 P q}{\Delta^2} = \frac{2^2 \cdot 9,5 \cdot 90,5}{3,5^2} = 280,7 \text{ единиц наблюдения}$$

где p - результирующий признак, выраженный в процентах и равный 9,5 %;

q- величина, равная 100 — p и составившая 90,5 %;

Δ - предельно допустимая ошибка, равная 3,5 %;

P - доверительная вероятность, равная 95,5 %;

t- доверительный коэффициент, равный 2.

Вывод: для получения статистически достоверных данных о состоянии здоровья детей необходимо обследовать 281 ребенка в возрасте от 0 до 3 лет.

Учитывая результаты определения необходимого объема наблюдений для получения статистически достоверных данных о состоянии здоровья детей первых 3 лет жизни на основании произведенных расчетов по двум

результатирующим признакам (в первом случае 256 единиц наблюдения, а во втором - 281), необходимо обследовать не менее 282 детей данного возраста.

Иногда при определении основного (результатирующего) признака и его предельно допустимой ошибки, особенно при изучении совсем незнакомых совокупностей, возникают трудности, которых можно избежать, если воспользоваться данными табл.12, где значения Δ и σ не определяют заранее, а берут ориентировочное соотношение Δ и σ (Δ/σ). Последнее обозначают как коэффициент точности (K), уровень соотношения выбирают в зависимости от цели и задач исследования в пределах от 0,5 до 0,1.

Как следует из табл.12, необходимый объем выборки для получения устойчивых результатов с достаточной степенью уверенности и точности составляет 400 единиц наблюдения при точности исследования (K) 0,1 и доверительной вероятности (P) 95,4 %, что несколько выше, чем при проведении расчетов.

Таблица 12

Необходимый объем выборки в зависимости от точности исследования и доверительной вероятности

Вид исследования	Желаемая точность исследования $K = \Delta/\sigma$	Доверительная вероятность		
		t = 2 P = 95,5%	t = 2,5 P = 98,6%	t = 3,0 P = 99,7 %
Ориентировочное знакомство	0,5	16	25	36
	0,4	25	39	56
	0,3	44	69	100
Исследование средней точности	0,2	100	156	225
Исследование повышенной точности	0,1	400	625	900

1.4.5.Пятое свойство статистической совокупности — взаимосвязь признаков в статистической совокупности

Пятое свойство статистической совокупности — это определение взаимосвязи между признаками. В природе и обществе все процессы и явления взаимно связаны. Формами проявления количественных связей служат функциональная и корреляционная связи.

Функциональная связь характеризуется тем, что каждому значению одного признака соответствует строго определенное значение другого признака и изменение величины одного признака неизбежно вызывает совершенно определенные изменения величины другого признака. Как правило, функциональная связь характерна для физико-химических явлений и присуща неживой природе: например, с возрастанием скорости увеличивается пройденное расстояние.

При корреляционной связи каждому значению одного признака соответствует несколько значений другого признака, поэтому она проявляется лишь при массовом сопоставлении признаков в качественно однородной совокупности и характерна для социально-гигиенических и медико-биологических процессов. Например, при увеличении роста, как правило, возрастает масса тела человека, хотя могут встречаться высокие люди, но с дефицитом массы тела, и наоборот, при небольшом росте наблюдается избыточная масса тела.

Признаки могут быть качественными и количественными негруппированными величинами (абсолютными и производными). Главным является установление причинных взаимосвязей, подтверждающих зависимость одного явления от другого или от какой-то общей причины. С этой целью определяют коэффициент корреляции, который позволяет оценить характер, силу и достоверность взаимосвязи изучаемых признаков. По направлению связь между явлениями может быть прямой (+) и обратной (-). Если связь между признаками прямая, то с увеличением одного признака увеличивается и другой. Например, чем старше ребенок, тем больше его рост; по мере снижения температуры тела, как правило, частота пульса уменьшается

и т. д. При обратной связи между изучаемыми явлениями с увеличением одного признака другой увеличиваться не будет. Например, с увеличением возраста снижается показатель средней продолжительности предстоящей жизни. В зависимости от численного выражения коэффициента корреляции различают связь слабую (от 0,0 до 0,3), среднюю (от 0,3 до 0,7) и сильную (от 0,7 до 1,0). Фактически, если коэффициент корреляции составит 0,0, то это значит, что связь между явлениями отсутствует, а если коэффициент корреляции будет равен 1,0, то это свидетельствует о сильной и полной (или функциональной) связи между явлениями. Достоверность коэффициента корреляции определяется величиной ошибки и доверительным коэффициентом t . В том случае, если полученный коэффициент корреляции в 3 раза и более превышает свою ошибку, он считается достоверным.

Для вычисления коэффициента корреляции используют методы рангов, или метод Спирмена (r), квадратов, или метод Пирсона (r), корреляционной решетки (η) и множественной корреляции. Наиболее простым методом является вычисление коэффициента корреляции методом рангов (метод Спирмена), но полученный коэффициент дает приближенные результаты. Для вычисления коэффициента корреляции данным методом используют формулу:

$$R_{xy} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)},$$

где d — разность рангов;

1 и 6 — постоянные коэффициенты;

n — число наблюдений сравниваемых пар.

Применение данной формулы предполагает обязательное ранжирование признаков в порядке их возрастания (или убывания)

Определение связи между ростом и массой тела у студентов-мужчин 20-22
лет по методу рангов

Число обследованных	Признаки		Ранги		Разность рангов	
	Рост (x), см.	Масса тела (y), кг.	x	y	-y	d ²
1	169	55	1	1	0	0
2	170	61	2	2	0	0
3	171	62	3	3	0	0
4	174	68	4	4	0	0
5	176	75	5	6,5	-1,5	2,25
6	180	75	6	6,5	-0,5	0,25
7	181	81	7	10,5	-3,5	12,25
8	184	78	8	8	0	0
9	185	71	9,5	5	4,5	20,25
10	185	80	9,5	9	0,5	0,25
11	187	81	11	10,5	0,5	0,25
12	188	82	12	12	0	0
n = 12						$\Sigma = 35,5$

Главным условием является соблюдение определенных правил:

- ранжировать каждую колонку цифр (вариант) нужно независимо от данных других столбцов;
- при ранжировании соблюдать единые требования, состоящие в том, что если в первой колонке вариант ранжирование начали с минимальной величины, то и во второй колонке цифр необходимо сделать также. Например, для определения взаимосвязи роста и массы тела были обследованы 12 студентов-мужчин в возрасте 20—22 года. Полученные данные представлены в табл. 13.

При определении порядкового номера следует учитывать, что при наличии одинаковых вариантов им всем дается среднее значение тех рангов, которые они (варианты) занимают. Так, в данном примере две варианты роста имели одинаковое значение 185 см и занимали по порядку 9-е и 10-е места, при этом каждая варианта получила среднее значение приходящихся на них порядковых (ранговых) мест, равное $2J_{15} = 9,5$. Аналогичным образом рассчитывали ранговые места для массы тела.

Подставляя полученные значения (d и n) в формулу, вычисляем коэффициент корреляции, он равен $+0,876$, что свидетельствует о наличии прямой и сильной зависимости между ростом и массой тела у студентов-мужчин в возрасте 20 - 22 года

$$R_{xy} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 35,5}{12 \cdot 143} = 1 - 0,124 = +0,876$$

Для определения достоверности полученного коэффициента корреляции вычисляют величину ошибки по формуле:

$$m_p = \sqrt{\frac{1 - R^2}{n - 2}}$$

где m_p — средняя ошибка коэффициента корреляции, вычисленного методом рангов;

R — величина коэффициента корреляции, вычисленного методом рангов;

n — число наблюдений.

Величина ошибки коэффициента корреляции ($R = +0,876$).

Таким образом, вычисленный по методу рангов коэффициент корреляции, равный $+0,876 \pm 0,153$, отражает наличие прямой, сильной и достоверной корреляционной зависимости и свидетельствует о том, что в целом с увеличением роста возрастает масса тела.

Метод рангов применяют в тех случаях, когда:

- число наблюдений не больше 30;
- признаки имеют не только количественное, но и качественное выражение (описательного характера);
- ряды распределения имеют открытые варианты (20 лет и более).

Наиболее точным и часто применяемым является метод квадратов или метод Пирсона:

$$r_{xy} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}},$$

где r — коэффициент корреляции, вычисленный методом квадратов

d_x — отклонения вариантов от средней величины ($V_x - M_x$)

d_y - отклонения вариант от средней величины ($V_y - M_y$)

Ошибку для коэффициента корреляции, вычисленного методом квадратов, определяют по формуле:

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

где m_r - ошибка коэффициента корреляции, вычисленного методом квадратов;

n - число наблюдений.

При большом числе наблюдений (более 30 единиц) для определения коэффициента корреляции целесообразно применять метод корреляционной решетки, а при определении связи одновременно между тремя и более признаками используют метод множественной корреляции. При этом коэффициент корреляции (характер, сила и достоверность) оценивают по критериям, описанным ранее.

Метод стандартизации. При сравнении тех или иных общих показателей необходимо соблюдать важнейшее условие, которое заключается в однородности состава сравниваемых совокупностей по признаку, влияющему на величину этих показателей. Так, например, для того чтобы сопоставить летальность в двух больницах и сделать вывод, где она выше, следует прежде всего проанализировать, однороден ли в этих больницах состав больных по нозологическим формам болезни, по возрастному составу больных, по уровню оснащенности оборудованием и медикаментами и т. д.

Безусловно, в больнице, в которой среди госпитализированных больше лиц с тяжелыми хроническими заболеваниями, находящихся в более старших возрастных группах, будет выше общий показатель летальности. Разный состав пациентов в этих больницах делает несопоставимыми общие показатели летальности. Однако в ряде случаев не только в социально-гигиенических клинических исследованиях, а и в практическом здравоохранении нередко невозможно составить однородные группы для сравнения тех или иных

показателей. Это касается, прежде всего, заболеваемости, рождаемости, общей смертности, а также показателей, характеризующих деятельность лечебно-профилактических учреждений в странах, областях, городах, районах внутри города, имеющих разный состав населения по возрасту, полу, исходному состоянию здоровья, условиям жизни и т. д., например, при сравнении деятельности учреждений здравоохранения одним из ведущих показателей является число больных, выписанных из стационара с осложнениями. Так, в одной из больниц города этот показатель составил 3,4, а в другой — 3,7 на 100 больных. Но прежде чем сравнивать эти показатели и делать выводы, следует проанализировать совокупности, из которых они получены. В данном случае необходимо обратить внимание на распределение больных по отделениям. Преобладание терапевтических больных, как правило, способствует увеличению показателя, так как именно в это отделение госпитализируют тяжелобольных, нередко имеющих сочетанную патологию.

Для сравнения данных показателей, вычисленных из неоднородных по своему составу совокупностей, применяют специальный метод, который называется методом стандартизации.

Стандартизация — метод сравнения показателей в двух неоднородных совокупностях на основании расчета условных (стандартизованных) показателей при использовании стандарта.

Рассчитанные при помощи этого метода стандартизованные показатели условны, потому что они, косвенно устраняя влияние того или иного фактора на истинные показатели, указывают, какими были бы эти показатели, если бы влияние данного фактора, мешающего сравнению, отсутствовало.

Следовательно, стандартизованные показатели могут быть использованы только с целью сравнения. Существует несколько способов расчета стандартизованных показателей, среди которых различают прямой, косвенный и обратный методы. Наиболее распространенным является прямой метод.

Суть прямого метода состоит в вычислении общих стандартизованных показателей в сравниваемых совокупностях, которые уравниваются благодаря выбранному стандарту. Предварительно на I этапе вычисляют интенсивные показатели в каждой из сравниваемых совокупностей, по которым судят об истинной частоте изучаемого явления в исследуемых совокупностях. Следующий II этап заключается в выборе стандарта, за который можно принять сумму данных двух рассматриваемых совокупностей, или средний состав обеих групп, или одну из сравниваемых групп, или какую-то третью группу, близкую к сравниваемым данным одной из статистических совокупностей. Главное условие — стандарт должен быть одинаковым для сравниваемых совокупностей.

На III этапе рассчитывают условные величины в каждой группе стандарта и на IV этапе — общие стандартизованные показатели, которые тоже являются условными, гипотетическими. Целесообразно все последовательные расчетные операции по этапам стандартизации оформить в виде таблицы (табл.14).

Таблица 14
Распределение больных, выписанных с осложнениями в больницах
А и Б (данные условные)

Отделение	Больница А		Больница Б		I этап интенсив ные показате ли (на 100 больных)		II этап Выбор стандарта	III этап ожидаемое число больных на стандарт	
	Число больных	Из них с осложне ниями	Число больны х	Из них с осложне ниями	А	Б	1+2 графы	А	Б
Терапевти ческое	500	25	200	12	5,0	6,0	700	35,0	42,0
	400	8	600	12	2,0	2,0	1000	20,0	20,0
Хирурги ческое	100	4	200	10	4,0	5,0	300	12,0	15,0
Гинеколо гическое	1000	37	1000	34	3,7	3,4	2000	67,0	77,0
В целом по больнице						IV этап	100	3,35	3,85

Таким образом, применение прямого метода стандартизации включает последовательное выполнение 5 этапов. На I этапе рассчитывают интенсивные и общие показатели (или средние величины) по всем группам в двух сравниваемых совокупностях. В нашем примере это вычисление частоты осложнений у больных в каждом отделении и в целом в двух больницах. На II этапе определяют стандарт, который в приведенном примере был равен сумме числа больных в каждом отделении. На III этапе рассчитывают ожидаемые величины осложнений в каждом отделении в больницах. На IV этапе определяют стандартизованные показатели в больницах А и Б. Выполнение V этапа, на котором сравнивают интенсивные и стандартизованные показатели в двух больницах, позволяет сделать вывод о величине показателя и степени влияния разного состава сравниваемых совокупностей по стандартизуемому признаку.

Так, в нашем примере (см. табл.14) полученные интенсивные показатели свидетельствуют о том, что частота осложнений выше в больнице А по сравнению с таковой в больнице Б (соответственно 3,7 и 3,4 случая осложнений на 1000 больных). Однако более высокий уровень осложнений в больнице А обусловлен преобладанием в ней терапевтических больных (число терапевтических коек в больнице А 500, а в больнице Б 400).

Такой вывод можно сделать на основании использования метода стандартизации. После проведения стандартизации (т. е. устранения различий в распределении больных в имеющихся отделениях) вычисленные стандартизованные показатели распространенности осложнений оказались ниже в больнице А по сравнению с показателями в больнице Б при одинаковой численности больных в отделениях (соответственно 3,35 и 3,85 ожидаемого случая осложнений на 100 больных). Таким образом, неравномерное распределение больных в отделениях влияет на показатель частоты осложнений в больнице в целом. Возможно, следует провести стандартизацию в отношении и других факторов, воздействующих на частоту осложнений.

Графические изображения. Статистическая обработка полученных данных завершается графическими изображениями, позволяющими дать наглядное представление результатов исследования. Практически в каждом статистическом исследовании применяют графический метод, для правильного использования которого нужно знать основные виды графических изображений и правила их построения.

Графиками в статистике называют условные изображения числовых величин (средних и относительных) в виде различных геометрических образцов (линий, плоских и объемных фигур в виде многоугольников, круга и т. д.). Статистический график дает возможность наглядно оценить характер изучаемого явления, присущие ему закономерности, особенности, тенденции развития, взаимосвязь характеризующих его показателей.

Каждый график состоит из графического образа и вспомогательных элементов. *Графический образ* — это совокупность точек, линий и фигур, с помощью которых изображают статистические данные. Вспомогательные элементы — это общее назначение графика, пояснение условных знаков и смысла графического образа, оси координат, шкалы, числовые сетки и числовые данные, дополняющие или уточняющие изображаемые показатели. Вспомогательные элементы облегчают чтение и толкование графика. Название графика должно быть кратким и точно раскрывать его содержание, располагается обычно под графиком в отличие от таблицы, название которой находится над ней. Пояснительные тексты могут располагаться в пределах графического образа, рядом с ним или выноситься за его пределы. Оси координат с нанесенными на них шкалами и числовые сетки необходимы для построения графика и пользования им. Шкалы могут быть прямо- или криволинейными (круговыми); равномерными (линейными) и неравномерными (например, логарифмическая шкала).

Статистические графики делят по разным признакам: способу построения, характеру графического образа и назначению (содержанию).

По *способу построения* графики делят на диаграммы, картограммы и картодиаграммы. По *характеру графического образа* различают графики точечные, линейные, плоскостные (столбиковые, полосовые или ленточные, квадратные, круговые, секторные, фигурные) и объемные. По *назначению* (содержанию) выделяют графики, изображающие различные относительные величины (структура явления, динамика процесса и др.) или показывающие сравнения в пространстве, размещения по территории, колеблемость вариационных рядов и взаимосвязанных показателей.

Диаграммой называют изображение статистических данных в виде точек, линий, плоскостей, фигур; они могут быть представлены в виде линейных, плоскостных, объемных и фигурных изображений.

Картограмма — графическое изображение статистических величин, представленных на географической карте.

В том случае, если статистические данные изображены в виде диаграммы на географической карте, то такой вид графического изображения называют *картодиаграммой*.

Вид графического изображения выбирают в зависимости от того, какие статистические величины требуется представить наглядно. Так, абсолютные величины (например, численность населения страны, города и т. д.) и интенсивные показатели (например, показатели рождаемости, смертности и т. д.) можно изобразить в виде диаграмм, картограмм и картодиаграмм. Причем в случае изображения динамики процесса пользуются линейными диаграммами, а если необходимо дать характеристику процесса за единый отрезок времени на различных территориях, то следует применять столбиковые диаграммы. Это же относится к использованию объемных или фигурных диаграмм. Для графического изображения экстенсивных показателей необходимо пользоваться внутри столбиковыми и секторными диаграммами.

Таким образом, обязательным правилом является применение графического метода в строгом соответствии графического изображения

имеющимся статистическим величинам и правилам построения диаграмм.

Линейную диаграмму обычно используют для изображения динамики процесса, явления во времени.

Столбиковую диаграмму применяют для иллюстрации однородных, но не связанных между собой интенсивных показателей

Секторную диаграмму применяют для изображения экстенсивных показателей.

Наряду с секторной диаграммой для изображения экстенсивных показателей применяют внутри столбиковую диаграмму, в которой ширина и высота столбика — произвольные.

IV этап. Статистический анализ изучаемого явления, формулировка выводов - ответственный этап исследования, на котором проводится вычисление статистических показателей (частоты, структуры, средних размеров изучаемого явления), дается их графическое изображение, Изучается динамика, тенденции, устанавливаются связи между явлениями. Даются прогнозы и т. д. Анализ предполагает интерпретацию полученных данных, оценку достоверности результатов исследования. В заключение делаются выводы.

В проведении статистического исследования важнейшим элементом является соблюдение строгой последовательности в осуществлении названных этапов.

Результаты исследования могут быть оформлены в виде статьи, отчета, доклада, диссертации и др. Для каждого вида оформления существуют определенные требования, которые должны соблюдаться при литературной обработке результатов статистического исследования.

Результаты медико-статистического исследования внедряются в практику здравоохранения. Возможны различные варианты использования результатов исследования: ознакомление с результатами широкой аудитории медицинских

и научных работников; подготовка инструктивно-методических документов; оформление рационализаторского предложения и другие

По завершении статистического исследования разрабатываются рекомендации и управленческие решения, проводится внедрение результатов исследования в практику, оценивается эффективность.

Ошибки статистического анализа. Наиболее частые ошибки можно сгруппировать следующим образом:

- 1) методические ошибки;
- 2) неправильная оценка показателей;
- 3) логические ошибки формального анализа.

К первой группе — *методическим ошибкам* — относят дефекты программы и плана исследования, которые чаще всего заключаются в следующем:

- неправильное определение единицы наблюдения, в связи с чем формируется неоднородная статистическая совокупность, что не позволяет выявить закономерности и сделать правильные выводы;
- недостаточное число наблюдений ведет к получению недостоверных результатов, поэтому следует рассчитывать необходимый объем наблюдений по нескольким результирующим признакам;
- использование слишком сложных таблиц, содержащих много признаков (более 3—4 статистических сказуемых при одном статистическом подлежащем) и приводящее к дроблению материала и получению малочисленных групп, при этом трудно доказать достоверность полученных результатов и невозможно сделать обоснованные выводы.

Вторая группа ошибок обусловлена недостатками, которые связаны с расчетами, неправильным выбором статистического метода обработки полученного материала и *неправильной оценкой статистических величин*. К ним относятся:

- арифметические ошибки, при этом целесообразно проводить проверку и

перепроверку различных вычислений, особенно когда возникает неожиданный результат;

- недостаточная статистическая обработка данных, когда вывод делается только на основании анализа абсолютных чисел; не составлены динамические ряды, не рассчитаны относительные и средние величины, не доказана их достоверность, не рассчитаны коэффициенты корреляции и т. д.;

- неправильная оценка показателей и прежде всего экстенсивных, для сравнения которых следует использовать формулу сравнения показателей (см. ранее), а вывод о больших или меньших масштабах каких-то явлений или процессов можно и нужно делать только на основании анализа интенсивных показателей;

- сравнение результатов, полученных в качественно неоднородных или искусственно отобранных группах (например, при апробации нового медикаментозного средства для лечения какого-то заболевания среди пациентов с определенными показаниями нельзя сделать вывод, что это средство дает лучший эффект по сравнению с другими препаратами, лечение которыми может быть несколько менее эффективно, но применение их не имеет противопоказаний);

- оценка темпа роста без учета исходного уровня показателей, поскольку известно, что в соответствии со статистической закономерностью, чем ниже исходный уровень каждого явления, тем выше темп роста, и наоборот;

- неиспользование метода стандартизации при анализе показателей в неоднородных статистических совокупностях.

Третья группа ошибок статистического анализа включает в себя *логические ошибки формального анализа*, которые можно сгруппировать следующим образом:

- выводы, которые сделаны на основе простого сравнения цифр без учета качественной характеристики явления, в связи с чем не учитываются причины, способствующие возникновению изучаемого явления, и затруднена разработка

мероприятий, направленных на воздействие, управление изучаемыми явлениями, процессами;

- выводы, сделанные по принципу "после этого, значит вследствие этого", что неверно (например, у пациента с гипертонической болезнью обострение наступило после ссоры с соседями по квартире, но это не означает, что ссора с соседями явилась причиной обострения, поскольку и проживание в коммунальной квартире, и наличие самого заболевания нельзя не учитывать в профилактической работе с данными пациентом).

Таким образом, статистический анализ — это не только анализ цифр и явлений, но и в значительной мере искусство специалиста, умение выделить из ряда последовательных событий ведущие, установить достоверную связь между ними, наметить пути воздействия.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Организация статистического исследования.
2. Этапы статистического исследования.
3. I этап статистического исследования. Цель и задачи исследования.
4. Организационный план статистического исследования. Разработка плана исследования.
5. Программа статистического исследования.
6. Определение единицы наблюдения и составление программы сбора материала.
7. II этап статистического исследования. Сбор материала (статистическое наблюдение). Виды статистического наблюдения.
8. Виды статистического наблюдения по времени регистрации.
9. Виды статистического наблюдения по полноте охвата единиц совокупности.
10. III этап статистического исследования. Разработка материала, статистическая группировка и сводка.
11. Первое свойство статистической совокупности
12. Второе свойство статистической совокупности
13. Третье свойство статистической совокупности
14. Четвертое свойство статистической совокупности
15. Пятое свойство статистической совокупности
16. Статистические таблицы. Составные части таблицы.
17. Виды статистических таблиц.
18. IV этап статистического исследования. Статистический анализ изучаемого явления, формулировка выводов.

ВОПРОСЫ ТЕСТОВОГО КОНТРОЛЯ

1. В каких случаях применяются статистические методы исследования?

- а) при изучении явления на различных этапах развития
- б) для точного определения показателей здоровья населения, качества оказания медицинской помощи и профилактической работы

Ответы: а)

2. Что включает в себя программа исследования?

- а) единицу наблюдения
- б) объект исследования
- в) статистические документы
- г) объем и место проведения исследования
- д) группировки и составление макетов таблиц

Ответы: а) в) д)

3. Что включает в себя план исследования?

- а) метод исследования
- б) единицу наблюдения
- в) временной период
- г) руководство и финансирование
- д) группировки и составление макетов таблиц
- е) объем и место проведения исследования

Ответы: а) в) г) е)

4. Из каких пунктов состоит план статистического исследования?

- а) тема исследования
- б) цель исследования
- в) объект исследования
- г) единица наблюдения
- д) методы исследования

Ответы: а) б) в) д)

5. Что включает в себя разработка статистического материала?

- а) контроль и шифровка статистического материала
- б) составление плана исследования
- в) вычисление статистических показателей и создание графических изображений
- г) группировки и составление макетов таблиц

Ответы: а) в)

6. Сколько этапов статистического исследования вы знаете?

- а) 2
- б) 4
- в) 5

Ответы: в)

7. Назовите этапы статистического исследования:

- а) составление программы исследования
- б) анализ, выводы и заключение
- в) составление плана и программы исследования
- г) разработка статистического материала
- д) сбор статистического материала

Ответы: б) в) г) д)

8. Что может быть объектом социально-гигиенического исследования?

- а) возрастно-половые группы
- б) заболеваемость
- в) семьи
- г) группы людей, объединенных по социальным, профессиональным и территориальным признакам
- д) группы населения, сформированные по состоянию здоровья

Ответы: а) в) г) д)

9. Какие виды группировок Вы знаете?

- а) количественные
- б) большие
- в) малые
- г) качественные

Ответы: а) г)

10. Какие виды статистических таблиц Вы знаете?

- а) простые
- б) комплексные
- в) комбинационные
- г) групповые
- д) контрольные

Ответы: а) в) г)

11. Назовите виды статистического наблюдения по объему:

- а) текущее
- б) генеральное
- в) сплошное
- г) постоянное
- д) выборочное

Ответы: в) д)

12. Назовите виды статистического наблюдения по времени:

- а) текущее
- б) генеральное
- в) сплошное
- г) постоянное
- д) выборочное

Ответы: а)

ТИПОВЫЕ ЗАДАЧИ

Задача-эталон

Для разработки комплексного плана оздоровительных мероприятий для студентов медицинского вуза главным врачом студенческой поликлиники совместно с представителями студенческого профсоюзного комитета вуза проведено изучение влияния факторов риска на распространенность болезней органов пищеварения (БОП) у студентов.

Решение

Цель исследования. Разработать мероприятия по снижению болезней органов пищеварения (БОП) у студентов медицинского вуза.

Задачи исследования.

- Изучить распространенность различных болезней органов пищеварения (БОП) у студентов медицинского вуза.
- Определить факторы риска возникновения БОП.
- Разработать предложения для администрации вуза

Программа исследования:

Единица наблюдения — студент с диагнозом БОП, обучающийся в медицинском вузе на данном факультете.

Атрибутивные признаки: пол, диагноз, характер питания.

Количественные признаки: возраст, длительность заболевания, интервал между приемами пищи, число приемов пищи в день.

Результативные признаки: наличие заболевания системы органов пищеварения.

Факторные признаки: пол, возраст, характер питания и др.

Программа сбора материала (анкета, заполненная студентом)

1. ФИО
2. Курс: 1,2,3,4,5,6
3. Факультет: лечебный (1), медико-профилактический (2), фармацевтический (3)

4. Возраст: до 20 лет включительно — (1), 21-22 — (2), 23-24 — (3), 25 и более (4)
5. Пол: муж (1), жен (2)
6. Сколько раз в течение дня Вы принимаете пищу? Один — (1), два — (2), три и более (3)
7. Прием пищи состоит из бутербродов без чая (1), бутербродов с чаем (2), полного обеда (3), другого (4) (укажите)
8. Каков интервал между приемами пищи: до 1 ч (1), 1-2 ч (2), 3-4ч (3), 5 ч и более (4)
9. Предусмотрено ли в расписании занятий время на обед: (да - (1), нет - (2)
10. Имеете ли Вы заболевание системы органов пищеварения: да - (1), нет - (2)
11. Если Вы ответили "да", то укажите диагноз: _____
12. Длительность заболевания: до 1 года — (1), 2-3 года — (2), 4-5 лет — (3), 6 лет и более — (4)

И другие вопросы в соответствии с целью и задачами исследования.

Программа разработки материала

Типологическая группировка: группировка студентов по факультетам, полу, по диагнозу заболевания.

Вариационная группировка: группировка по длительности заболевания (до 1 года, 2-3 года, 4-5 лет, 6 лет и более), интервал между приемами пищи (до 1 ч, 1-2 ч, 3-4 ч, 5 ч и более).

Макеты статистических таблиц.

Простая таблица.

Таблица 1.

Распределение студентов, имеющих заболевания системы органов пищеварения по нозологическим формам (в % к итогу)

Заболевание	Всего студентов	
	абсолютное число студентов	в %
1. Гастрит		
2. Язвенная болезнь желудка		
3. Язвенная болезнь 12-перстной кишки		
4. Прочие		
Итого:		100,0

Групповая таблица.

Таблица 2.

Распределение студентов, имеющих заболевания системы органов пищеварения по полу и возрасту (в % к итогу)

Заболевание	Пол		Возраст			Всего
	муж	жен	до 15 лет	15 - 18 лет	старше 18 лет	
1. Гастрит						
2. Язвенная болезнь желудка						
3. Язвенная болезнь 12-перстной кишки						
4. Прочие						
Итого:						

Комбинационная таблица

Таблица 3.

Распределение студентов, имеющих заболевания системы органов пищеварения, по факультетам и полу (в % к итогу)

Заболевание	Лечебный			Медико-профилактический			Фармацевтический			Всего		
	м	ж	оба пола	м	ж	оба пола	м	ж	оба пола	м	ж	оба пола
1. Гастрит												
2. Язвенная болезнь желудка												
3. Язвенная болезнь 12-ти перстной кишки												
4. Прочие												
Итого:												

План исследования:

- Объект исследования — студент медицинского вуза, обучающиеся в данном -медицинском вузе на данном факультете.
- Объем статистической совокупности: достаточное число наблюдений.
- Совокупность: выборочная, репрезентативная по качеству и количеству.
- Сроки проведения исследования: 6 февраля — 6 июня текущего года.
- Методы сбора материала: анкетирование, выкопировка из медицинских документов студенческой поликлиники.

ЛИТЕРАТУРА.

Основная литература:

1. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения. Под ред. чл.-корр. РАМН, проф. В.З.Кучеренко. М., "Гэотар-Медиа", 2007, учебное пособие для вузов.
2. Кобринский Б.А., Зарубина Т.В. Медицинская информатика: Учебник. М: изд. "Академия", 2009.
3. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения: Учебное пособие для практических занятий. /Под ред. В.З.Кучеренко. – М.:ГЭОТАР-МЕД, 2004. – 192 с.

Дополнительная литература

4. Лисицын Ю.П. Общественное здоровье и здравоохранение. Учебник для вузов. — М.: ГЭОТАР-МЕД, 2007. — 512 с.
5. Медик В.А., Юрьев В.К. Курс лекций по общественному здоровью и здравоохранению: Часть 1. Общественное здоровье. — М.: Медицина, 2003. — 368 с.
6. Кучеренко В.З., Агарков Н.М. и др. Социальная гигиена и организация здравоохранения (Учебное пособие) — Москва, 2000. — 432 с.
7. Ниворожкина Л. И. Статистическое наблюдение // Ниворожкина Л. И. Теория статистики — Ростов на Дону: «Мини Тайп», «Феникс», 2005. — С. 30—32.

