

# ОСНОВЫ СТАТИСТИКИ И ОРГАНИЗАЦИИ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

## 2.1. СТАТИСТИКА КАК САМОСТОЯТЕЛЬНАЯ ОБЩЕСТВЕННАЯ НАУКА

Современный этап общественного развития характеризуется широким использованием статистики в различных областях науки, культуры, образования, промышленного и сельскохозяйственного производства, строительства. Трудно определить сферу человеческой деятельности, где бы не применялась статистика. Это в полной мере относится к медицине, биологии и здравоохранению.

**Статистика** — самостоятельная общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной.

Статистику принято считать мощным орудием социального познания. Используя результаты количественных измерений изучаемых явлений, статистика позволяет определить причинно-следственные связи и устанавливать складывающиеся на различных этапах общественного развития закономерности. Количественные изменения, накапливаясь, способны вызвать коренные качественные изменения, вследствие чего старое качество исчезает и возникает новое.

При изучении возрастного состава населения страны определяется доля, которую занимает каждый возраст в общей численности населения, т.е. дается количественная характеристика его возрастной структуры. Выявляются и качественные особенности состава населения — удельный вес граждан, достигших трудоспособного возраста, численность населения в школьном возрасте, количество дошкольников и т.д. Таким образом, при исследовании количественной стороны явлений углубляются знания о свойствах исследуемого явления, составляющих его качество.

Медико-биологический раздел статистики включает:

- теоретические и методологические основы;
- статистические показатели:
  - общественного здоровья;
  - здравоохранения;
  - научно-практических исследований.

Предмет статистики применительно к медицине, биологии, здравоохранению:

- теоретические и методологические основы;
- количественная характеристика общественного здоровья и здравоохранения;
- закономерности результатов научно-практических исследований.

Статистика языком цифр характеризует размеры и количественные соотношения (структуру, темпы развития и т.п.) явлений общественной жизни и проявляющиеся в них закономерности.

Одна из особенностей статистики — то, что при изучении количественной стороны общественных явлений и процессов она всегда отображает качественные особенности исследуемых явлений, т.е. изучает количество в неразрывной связи, единстве с качеством. Качество в научно-философском понимании — свойство, присущее предмету или явлению, которое отличает данный предмет или явление от других. Качество — это то, что делает предметы и явления определенными.

Медико-биологический раздел статистики обеспечивает:

- мониторинг показателей здоровья населения и его отдельных контингентов;
- установление причинно-следственных связей между явлениями в медицине, биологии, здравоохранении и других сферах жизнедеятельности, влияющих на здоровье человека;
- обоснование необходимости оздоровительных мероприятий, их реализации с оценкой эффективности;
- изучение основных количественных и качественных показателей деятельности системы здравоохранения с целью совершенствования ее организации и управления;
- использование результатов научно-практических исследований на доказательной основе;
- разработку исходных данных для экономики здравоохранения.

Следовательно, совершенствование знаний в области теоретических и методологических основ статистики, мониторинг общественного здоровья, обоснование необходимости оздоровительных мероприятий, их реализации с оценкой эффективности, оптимизация организации и управления системой здравоохранения на доказательной основе — необходимые компоненты в повышении квалификации специалистов в сфере общественного здоровья и здравоохранения.

## 2.2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МЕДИКО-БИОЛОГИЧЕСКОЙ СТАТИСТИКИ

Для изучения любого явления необходимо выбрать **объект исследования, единицу наблюдения и ее учетные признаки**, определить **статистическую совокупность и необходимый объем наблюдений**. Объем выборочного исследования должен быть достаточным для получения статистически значимых результатов.

Теоретическое обоснование выборочного метода строится на основе теории вероятностей и закона больших чисел. Суть теории вероятности заключается в том, что появление любого признака рассматривается как случайное событие.

Например, несмотря на случайность заболеваний гриппом, эпидемии возникают с определенной закономерностью и, как правило, в холодное время года, так что эпидемиологи могут с большой вероятностью прогнозировать распространенность этого вида заболевания среди населения.

В последнее время особый интерес медицинской общественности вызывают результаты исследований, полученные на принципах доказательной медицины, основу которой составляют теория вероятностей и закон больших чисел, а также современные методы обработки информации.

**Вероятность** — это мера возможности возникновения случайных событий в конкретных условиях.

В биологии и медицине мы чаще всего имеем дело с выборочными совокупностями.

Вероятность наступления какого-либо события ( $P$ ) определяется отношением наступивших событий ( $m$ ) к числу всех возможных случаев ( $n$ ):

$$P = m/n.$$

Альтернатива вероятности наступившего события — вероятность его отсутствия ( $q$ ).

$$q = (n - m)/n = 1 - m/n = 1 - P; q = 1 - P;$$

$$P + q = 1.$$

Вероятность наступления события находится в пределах от 0 до 1. Соответственно, чем  $P$  ближе к 1, тем выше вероятность наступления того или иного события и чем больше  $P$  приближается к 0, тем меньше возможность его наступления.

Теория вероятностей позволяет вывести **закон больших чисел**.

В соответствии с законом больших чисел:

- с увеличением числа наблюдений результаты выборочного исследования стремятся воспроизвести закономерности генеральной совокупности;
- достаточный объем наблюдений позволяет установить закономерности, которые не удастся обнаружить при небольшом числе наблюдений.

Для иллюстрации сказанного в математике применяют упрощенные схемы моделирования этих процессов с подбрасыванием монет или использованием ящиков с цветными шарами.

Доля, полученная на выборочной совокупности ( $P_1$ ), при большой выборке весьма близка к доле, которую составляет явление в генеральной совокупности ( $P$ ). При большом числе наблюдений вероятность этого несовпадения настолько мала, что практически с ней можно не считаться.

Теорией статистики установлено, что при большой выборке ( $n > 30$ ) с вероятностью, равной 95%, можно утверждать, что разность долей, полученных из этой выборки ( $P_1$ ), и генеральной совокупности ( $P$ ), будет составлять  $2m$ ; с вероятностью, равной 99,7%, можно утверждать, что разность этих долей ( $P_1 - P$ ) не превысит  $3m$ .

Числа 1, 2, 3 и другие, на которые умножают ошибку репрезентативности ( $m$ ), носят название доверительных коэффициентов, и обозначают их буквой  $t$ . С увеличением  $t$  возрастает степень вероятности, с которой можно утверждать, что разность долей, полученных из выборки и генеральной совокупности будет находиться в пределах:  $\Delta = tm$ , где  $\Delta$  — предельная ошибка, допустимая для данного исследования. Предельная ошибка ( $\Delta$ ) может быть с положительным и отрицательным знаком ( $\pm \Delta$ ). Следовательно,  $P + P_1 \pm \Delta$ .

Пользуясь законом больших чисел, увеличивая объем выборки, можно регулировать размер предельной ошибки, доводя ее до минимума. При планировании исследования используют формулы, основывающиеся на законе больших чисел, по которому рассчитывают необходимую численность ( $n$ ) выборки. Для этого надо знать, с какой точностью, в зависимости от задач исследования, необходимо получить результаты, т.е. иметь представление о допустимой для данного исследования ошибке ( $\Delta$ ).

Основные и наиболее общие положения теории вероятностей и закона больших чисел разработаны отечественными учеными — математиками П.Л. Чебышевым, А.М. Ляпуновым, А.А. Марковым. Дальнейшая разработка теории вероятностей произведена советским математиком А.Н. Колмогоровым. **Теорема Чебышева** формулируется следующим образом: с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом числе независимых наблюдений средняя величина изучаемого признака, полученная на основе выборки, будет сколь угодно мало отличаться от средней величины изучаемого признака во всей генеральной совокупности.

Знание теоретических основ статистики позволяет разрабатывать и совершенствовать технологии, обеспечивающие инновационные направления изучения закономерностей, существующих в окружающем мире и обществе, в частности, при исследовании процессов, формулирующих здоровье населения.

В процессе изучения влияния общественных факторов используется множество объектов исследования, информация о которых представляется в виде статистической совокупности.

Комплексное изучение объектов исследования обеспечивается, как уже отмечалось, использованием системного анализа и системного подхода.

**Системный анализ** — методология комплексного изучения системы (объекта, службы, процесса) как единого целого, ее целей, функций, структуры, организации, выявление положительных и отрицательных сторон системы и разработки мероприятий для ее коррекции на основе использования различных наук, математических методов и вычислительной техники.

**Системный подход** предполагает использование совокупности методологических принципов для изучения объектов, явлений, процессов как целостных систем, а также разработку комплекса мероприятий и рекомендаций для их совершенствования. Системный подход помогает лицу, принимающему решение, выбрать последовательность действий путем общего изучения стоящей перед ним проблемы, определения цели, нахождения вариантов решения и сравнения последних под углом зрения соответствующих им результатов.

С позиций системного подхода необходимо рассматривать медицинский объект не только как сложную систему, но и как составную часть (подсистему) другой системы, взаимодействующую со многими другими внешними системами. Например, плохая работа стационара отразится на деятельности многих взаимосвязанных медицинских и немедицинских организаций и наоборот. Развитие специализированной медицинской помощи означает концентрацию тяжелых больных в специализированных организациях, и априори можно ожидать ухудшения показателей деятельности организации, если оно не имеет более квалифицированного персонала и не оснащено современным оборудованием и т.д.

Неблагоприятные демографические процессы в ряде регионов страны влияют на показатели здоровья населения и деятельности служб здравоохранения, что требует для нейтрализации этого явления (влияния) дополнительных средств здравоохранения.

Системный подход предполагает, что при ограниченности средств увеличение ресурсов в одном секторе здравоохранения означает их уменьшение в других, а увеличение трудовых ресурсов в производственной сфере означает их уменьшение в непроизводственной.

Системный подход к изучению проблем управления здравоохранением предусматривает, что если, несмотря на все принятые меры, нет существенного прогресса в решении проблемы, то данная проблема является межотраслевой и ее решение в значительной степени зависит от других взаимосвязанных отраслей народного хозяйства, природных систем и т.д.

Рассмотрение проблем здравоохранения как многофакторных комплексов позволяет выделить группу неуправляемых (природные, географические, климатические и т.д.) и управляемых факторов, зависящих от пациента, врача, медицинского учреждения, вышестоящих органов здравоохранения, внешних систем.

В целях рационального использования тех или иных методов обработки полученной информации необходимо учитывать распределение признака в исследуемой совокупности и его типичности для изучаемого явления.

### **2.3. РАСПРЕДЕЛЕНИЕ ПРИЗНАКА В СТАТИСТИЧЕСКОЙ СОВОКУПНОСТИ**

Статистическая совокупность состоит из элементов, имеющих различные значения изучаемого признака. Количество элементов (единиц наблюдения) в совокупности варьирует. Величина, свидетельствующая о количестве единиц наблюдения с одинаковой величиной признака, называется частотой. Наличие в совокупности элементов с определенным значением признака выражается мерой вероятности, что позволяет с помощью теории вероятностей определить закономерности распределения признака изучаемого явления.

Характер распределения возможно установить только на достаточном объеме наблюдений. Знание характера распределения признака в совокупности

способствует рациональному выбору статистических критериев для оценки результатов исследования.

Распределения, которые наблюдаются в медицинских, в том числе и в социально-гигиенических исследованиях, довольно разнообразны по своему характеру.

Различают основные типы распределений: альтернативное, нормальное (симметричное) и асимметричное (правостороннее, левостороннее, двугорбое — бимодальное и т.д.). На рис. 2.1 показаны наиболее распространенные в практике типы распределения.

В социально-гигиенических исследованиях очень популярен альтернативный тип распределения. Такое распределение имеет только два противоположных значения признака (да, нет). Например, исход лечения состоит только из двух противоположных градаций: числа умерших и числа выживших; по признаку доношенности при рождении распределение состоит также только из двух групп: числа доношенных и числа недоношенных; по признаку успеваемости студентов распределение состоит из числа успевающих и числа неуспевающих и т.д.

Часто встречаются явления, которые распределяются по асимметричному типу. При правостороннем распределении наибольшее число случаев наблюдения скапливается не на уровне середины ряда, а сдвигается в сторону меньшего значения признака. В этом случае  $M > Me > Mo$ . Если наибольшее число случаев наблюдения сдвинуто в сторону большего значения признака, наблюдается левосторонняя асимметрия ( $M < Me < Mo$ ). Если наибольшее



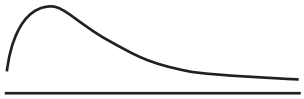


Тип распределения		Графическое изображение	Примеры
I. Альтернативный (качественные признаки)			Исход лечения: выжившие, умершие
II. Нормальный (количественные признаки)			Распределение по росту
III. Асимметричный (количественные признаки)	1. Правосторонний		Число детей в семье
	2. Левосторонний		Кратность прививок
	3. Двугорбый (бимодальный)		Неоднородная группа (рост в группе, состоящей из мальчиков и девочек)

Рис. 2.1. Типы распределения статистической совокупности

число случаев наблюдения скапливается по концам ряда, отмечается двугорбое бимодальное распределение.

Правосторонняя асимметрия характерна для распределения такого признака, как число детей в семье или кратность случаев временной утраты трудоспособности. Как известно, в большинстве семей имеется 1–2 ребенка. С увеличением числа детей в семьях соответственно уменьшается число семей. Если проанализировать ряд по кратности случаев нетрудоспособности в связи с заболеванием в течение года, то он будет иметь вид правосторонней асимметрии, так как основная масса работающих имеет минимальное число случаев временной нетрудоспособности — 1–2 (т.е. значительное число болеющих скапливается у наименьшей градации данного признака).

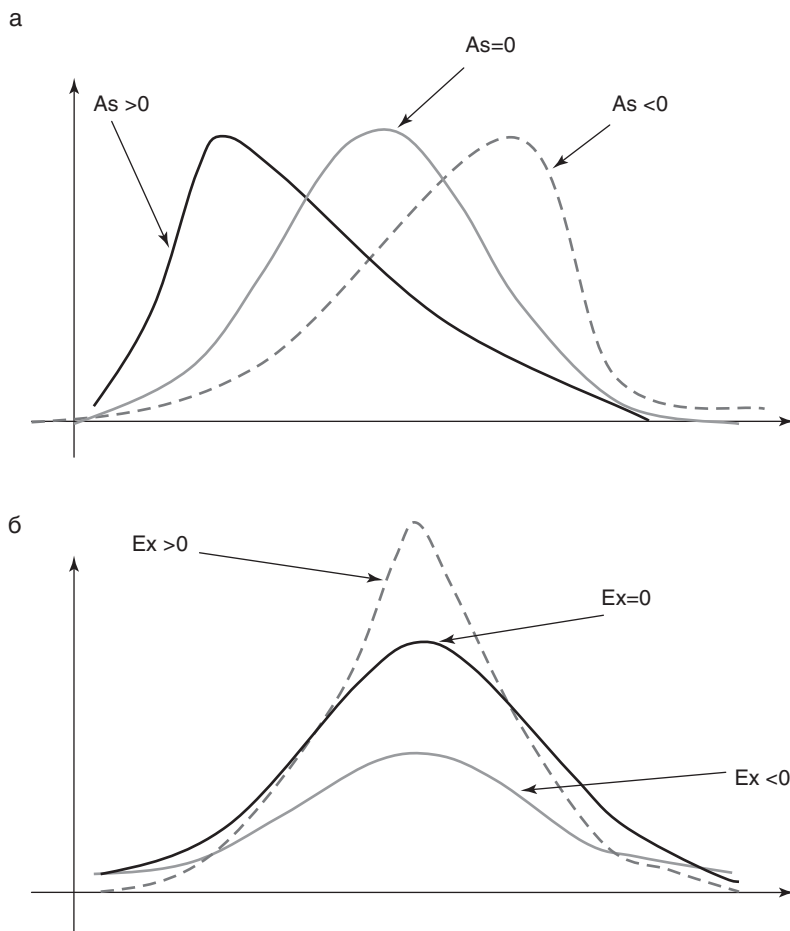
Реже встречается распределение нормальное (симметричное). Обычно нормальное распределение наблюдается при построении рядов количественных признаков: роста, массы тела, артериального давления, сроков госпитализации и др.

При нормальном типе распределения число случаев наблюдений с различной величиной признака располагается симметрично по отношению к середине ряда: от меньшего значения признака к большему его значению, наибольшее число случаев наблюдений приходится на середину ряда. При нормальном распределении признака в совокупности значения моды, медианы и средней арифметической величины равны ( $M = Me = Mo$ ). Если признаки имеют только положительные значения, можно оценить соответствие эмпирического распределения нормальному с помощью среднего квадратического отклонения. Если среднее квадратическое отклонение меньше половины среднего ( $S < M/2$ ), распределение можно считать нормальным, так как симметричность — одна из основных характеристик нормального распределения.

Двугорбое (бимодальное) распределение имеет две вершины. Как правило, такой ряд нуждается в дополнительном анализе. Двугорбый тип распределения указывает, что совокупность неоднородна. Например, если включить в совокупность мальчиков и девочек и измерить их рост, то полученное распределение будет бимодальным.

При альтернативном типе распределения признака в совокупностях специальных методов его определения не требуется. Обработка данных, полученных на альтернативных совокупностях, предполагает вычисление относительных величин, оценку статистической значимости различий между ними и установление других закономерностей.

Определить, отличается ли эмпирическое распределение от нормального, можно с помощью коэффициента асимметрии и эксцесса. Чем более выражена асимметрия ( $As$ ), тем больше значение коэффициента. Знак величины коэффициента асимметрии связан с направлением асимметрии. Если распределение вытянуто в сторону отрицательных значений, коэффициент асимметрии положительный ( $As > 0$ ), если распределение вытянуто в сторону положительных значений, коэффициент асимметрии отрицательный ( $As < 0$ ; рис. 2.2, а).



**Рис. 2.2.** Определение характера распределения признака с помощью коэффициентов эксцесса и асимметрии

Количественное описание островершинности дает коэффициент эксцесса ( $Ex$ ). Если эксцесс больше нуля, распределение принято считать островершинным, в противоположном случае — туповершинным (рис. 2.2, б).

Таким образом, если асимметрия и эксцесс значительно отличаются от нуля, можно считать, что распределение не подчиняется закону нормального распределения (рис. 2.2, б).

В эпоху компьютерных технологий рассчитать эти коэффициенты — несложная задача. В доступной для любого пользователя программе Excel имеется встроенная опция «Анализ данных», которая позволяет получить значения искомых показателей.

С целью определения отличия изучаемого распределения от нормального с помощью коэффициентов асимметрии и эксцесса применяют обычный в биометрии метод сравнения коэффициентов с их ошибками репрезентативности.



Если показатели эксцесса и асимметрии превышают свою ошибку более чем в три раза, можно говорить об отличии эмпирических распределений от нормального.

С помощью компьютерных технологий характер распределения признака возможно определить несколькими способами: построить гистограмму распределения; наглядно оценить, насколько диаграмма близка к графику нормального распределения; проверить нулевую гипотезу о том, что эмпирическое распределение соответствует закону нормального распределения и альтернативную гипотезу о том, что эмпирическое распределение не соответствует закону нормального распределения.

Проверить гипотезы можно с помощью критериев:

- Колмогорова—Смирнова. Условие: среднее значение и среднее квадратическое отклонение признака известны заранее, а не вычисляются по выборке;
- Лиллиефорса. Условие: среднее значение и среднее квадратическое отклонение признака вычислены по выборке;
- Шапиро—Уилка. Условие: среднее значение и среднее квадратическое отклонение признака неизвестны. Этот критерий — наиболее мощный и самый строгий из перечисленных.

Если полученное значение  $P$  для используемого критерия больше критического уровня статистической значимости (0,05), то эмпирическое распределение приближенно считают нормальным.

## 2.4. ПОДГОТОВКА И ПРОВЕДЕНИЕ НАУЧНОГО ИССЛЕДОВАНИЯ

Любое научное исследование должно быть соответствующим образом продумано и организовано. Исследование проводится поэтапно. Подготовительный этап включает определение проблемы, выбор темы и формулировку рабочей гипотезы. На этом этапе следует определить цель и задачи исследования, научную новизну, теоретическую и практическую значимость.

На первом этапе исследователю необходимо составить план и программу. Второй этап включает сбор материала. Третий этап — статистическая обработка собранного материала, на котором проводится сводка и группировка. На четвертом этапе данные подвергаются анализу, делаются соответствующие выводы и даются рекомендации, результаты исследования внедряются в практику (рис. 2.3).

**Программа** представляет общий стратегический замысел исследования, а **план** включает перечень вопросов организационного характера.

Научно-исследовательская работа начинается с определения проблемы. **Проблема исследования** — некая противоречивая ситуация, требующая разрешения. Проблема возникает тогда, когда имеющихся знаний в соответствующей области недостаточно, а новые знания отсутствуют. Исследователь должен показать, где проходит грань между знанием и незнанием с точки зрения предмета исследования. Формулируя проблему, он должен ответить на вопрос, что предстоит изучить из ранее не изученного.



Рис. 2.3. Этапы статистического исследования

Для того чтобы выявить проблему, определить тему будущей научной работы, необходимо изучить имеющуюся литературу по данному кругу вопросов. Изучение литературы позволит понять, что уже разработано по данной проблеме, что частично разработано, что вообще никогда не рассматривалось.

**Тема** научного исследования — это направление, в котором рассматривается проблема. Тема должна быть направлена на решение актуальных для здравоохранения и медицины задач и проблем. В обосновании **актуальности** определяется уровень изученности избранной проблемы в медицине и здравоохранении, указывается степень ее новизны для современной науки, подчеркивается ее связь с важными аспектами медико-социальных проблем современности, решению которых может способствовать ее исследование.

Тема исследования актуальна, если ее научное решение удовлетворяет потребности медицинской практики и заполняет пробел в науке, не располагающей в настоящее время средствами для решения этой актуальной задачи. Другими словами, исследование имеет **научную новизну, теоретическую и практическую значимость**.

**Научная новизна** характеризуется полученными в процессе исследования результатами. Научная новизна — признак, наличие которого дает исследователю право на использование понятия «впервые» при представлении полученных результатов.

Под **практической значимостью** понимают то, как результаты научного исследования могут быть использованы в медицинской практике. Практическая значимость может состоять в разработке системы профилактики и лечения,

методики диагностики отдельных заболеваний, в разработке медико-социальных рекомендаций и т.д. **Теоретическая значимость** заключается в получении нового научного знания.

Описание практической значимости следует начинать со слов «Обоснована эффективность использования...», «Результаты, полученные в ходе исследования... могут быть основой для...», «Предложенная методика позволяет...», «Обоснована необходимость изучения... для...»

Изучение литературных источников позволяет исследователю не только определить тему научной работы, но и приступить к выработке научной гипотезы. **Гипотеза** — это предположение о фактах, связях, функционировании и развитии биосоциальных и медицинских явлений, не имеющих эмпирического или логического обоснования либо недостаточно обоснованных. Гипотеза должна быть проверяемой, содержать предположение, соответствовать фактам. Гипотеза не должна противоречить логике.

При формулировке гипотез можно использовать словесные конструкции типа: «если..., то...»; «так..., как...»; «при условии, что...»

Существует несколько видов гипотез. **Рабочая гипотеза** должна отражать основное содержание исследования. Частным следствием рабочей гипотезы становятся **эмпирические гипотезы**. Они конкретизируют общие гипотезы, выдвинутые в исследовании. С учетом того, что статистический анализ — неотъемлемая часть научно-исследовательской работы, формулируются **статистические гипотезы**. Статистические гипотезы используют для проверки статистической значимости различий между группами сравнения, при оценке характера распределения признака в статистической совокупности. Проверке подвергаются статистические гипотезы: нулевая ( $H_0$ ) — о различии и альтернативная ( $H_1$ ) — о сходстве групп.

Выражение общего замысла исследования — его цель. **Цель** — это обоснованное представление об итоговых результатах научной работы. Определение цели в итоге позволяет исследователю окончательно определиться с темой научной работы. Цель исследования в сфере здравоохранения и общественного здоровья заключается в установлении влияния медико-социальных факторов на общественное здоровье с последующей разработкой системы корригирующих мероприятий и оценкой их эффективности.

Для достижения цели исследования необходимо определить задачи, которые должны отвечать на вопрос, как достичь цели. Задачи должны содержать в себе конкретные статистические гипотезы, проверка которых будет осуществлена выбранными методами статистического анализа. Важно так выстроить последовательность задач, чтобы можно было определить «маршрут» научного поиска, его логику и структуру. Задачи исследования предусматривают:

- формирование методики;
- изучение закономерностей того или иного явления и факторов, их определяющих;
- установление причинно-следственных связей между факторами и закономерностями явления;
- разработку системы корригирующих мероприятий и их внедрение;
- оценку эффективности реализованных мероприятий.

Для формулировки задач исследования рекомендуется использовать такие слова, как «определить», «исследовать», «изучить», «установить», «выявить», «разработать», «оптимизировать», «решить», «выяснить», «провести», «обосновать», «оценить».

Программа сбора информации предполагает разработку **первичной документации**. Это специально разработанные карты (бланки, анкеты) (рис. 2.4) для углубленного изучения проблем в сфере охраны и укрепления здоровья населения, улучшения организации и качества медицинской помощи, эффективного использования ресурсов, решения других проблем здравоохранения и общественного здоровья. В них содержится перечень вопросов, позволяющих собрать информацию об изучаемых признаках, подлежащих регистрации в процессе наблюдения. В заголовке каждой карты (бланка, анкеты) указывают ее предназначение. Далее следуют вопросы (перечисляются изучаемые признаки), требующие ответа.

Заполнение карты может осуществляться с использованием специально разработанного шифра, что обеспечивает ее формализацию и оптимизирует обработку информации.

Программа обработки информации предполагает формирование макетов таблиц. Их заполнение проводится на последующих этапах исследования. Принцип построения таблиц предусматривает обеспечение логичности и компактности представленной в них информации.

**Таблица** должна иметь номер, четкое и краткое название. В ней различают подлежащее и сказуемое. Подлежащее таблицы характеризует основной признак изучаемого явления и располагается на горизонтальных строках таблицы. Сказуемое располагается в вертикальных графах таблицы. Таблица должна содержать итоги по графам и строкам. С использованием итоговых данных таблиц на последующих этапах исследования рассчитывают показатели. Таблицы, за исключением таблиц приложений, следует нумеровать арабскими

<b>КАРТА</b> <b>изучения заболеваемости с временной утратой</b> <b>трудоспособности металлургов</b>	
Ф.И.О. _____	пол _____
возраст _____ образование _____ место работы _____	
_____ специальность _____	общий стаж _____
профессиональный стаж _____	диагноз _____
Дата выдачи больничного листка _____	
Дата выписки на работу _____	
Социально-трудовые условия _____	
Социально-бытовая характеристика _____	
_____	
Дата заполнения карты _____	

**Рис. 2.4.** Образец учетного документа

цифрами сквозной нумерацией. Макеты таблиц должны отражать основное направление исследования и анализа его результатов. Различают простые, групповые и комбинационные таблицы.

Простой считается таблица, в которой представлены данные с учетом лишь одного признака (табл. 2.1).

**Таблица 2.1.** Распределение больных в зависимости от специфических условий производства

Металлургический завод	Число больных
Завод «А»	
Завод «Б»	
Завод «В»	
<b>Итого</b>	

Групповые таблицы содержат информацию по двум признакам: один из них представлен в виде сказуемого, другой в виде подлежащего (табл. 2.2).

**Таблица 2.2.** Характеристика заболеваемости с временной нетрудоспособностью в зависимости от возраста и пола

Возраст, лет	Мужчины	Женщины	Всего
До 30			
31–40			
41 и старше			
<b>Итого</b>			

Комбинационные таблицы располагают информацией, позволяющей учесть влияние трех и более факторов на величину изучаемого явления. Подлежащее такой таблицы характеризуется наличием сказуемого в виде комбинации нескольких признаков (табл. 2.3).

**Таблица 2.3.** Распределение заболеваемости с временной утратой трудоспособности в зависимости от производства, пола и возраста работающих

Металлургический завод	Мужчины						Женщины						Всего
	возраст, лет												
	до 20	20–29	30–39	40–49	50–59	60 и старше	до 20	20–29	30–39	40–49	50–59	60 и старше	
Завод «А»													
Завод «Б»													
<b>Итого</b>													

В соответствии с государственной системой стандартизации Российской Федерации (ГОСТ 1.5-02), при переносе части таблицы на ту же или другие страницы название помещают только над первой частью таблицы. Цифровой материал, как правило, оформляют в виде таблиц в соответствии с рис. 2.5.



Рис. 2.5. Оформление таблицы

Программа анализа материала основана на свойствах статистической совокупности: распределение признаков, средний уровень признаков, вариативность, репрезентативность и взаимосвязь признаков (корреляции и ассоциации).

Оценить распределение признаков можно абсолютными числами и относительными показателями.

**Средний уровень** признаков описывается средними величинами (мода, медиана, средняя геометрическая, средняя арифметическая). Какую именно среднюю величину применить в каждом конкретном случае, исследователь решает, предварительно оценив характер распределения признаков в совокупности.

**Вариативность**, или разнообразие, признаков характеризуют размах (амплитуда —  $A$ ), интерпроцентильный размах, интерквартильный размах и среднее квадратическое отклонение. Для принятия решения об использовании конкретной меры разнообразия оценивают соответствие распределения признака закону нормального распределения.

**Репрезентативность** — представительность признаков, т.е. выборочная совокупность, должна обладать всеми свойствами генеральной совокупности. Меры репрезентативности — ошибка показателя и доверительные интервалы.

Еще одно свойство статистической совокупности — это **взаимосвязь признаков**. Установить наличие взаимосвязи количественных или порядковых признаков — значит рассчитать **коэффициент корреляции**, который покажет, в какой степени изменение значения одного признака сопровождается изменениями значения другого признака. **Метод Пирсона** применяется для признаков, характер распределения которых соответствует закону нормального распределения. Для других распределений используются **непараметрические методы** оценки корреляции. Описать взаимосвязь качественных признаков или ассоциацию можно, используя методы Спирмена, Кендалла, Гамма.

Перечисленные методы подходят для выявления линейной взаимосвязи. Это такая взаимосвязь, при которой увеличение (уменьшение) средних значений одного признака приводит к увеличению (уменьшению) средних значений

другого. Однако встречается и другой тип взаимосвязи между явлениями. Например, исследователь проверяет гипотезу о влиянии продолжительности рабочей смены на производительность труда. Производительность труда с начала рабочей смены постепенно повышается, затем продолжительное время держится на высоком уровне, а к концу смены начинает снижаться. В данном случае можно говорить о криволинейной зависимости, которая коэффициентами линейной корреляции не улавливается. Один из способов оценки корреляционной связи в этом случае — применение корреляционного отношения.

Таким образом, в программе анализа необходимо указать, какие методы применяли для оценки характера распределения признака в совокупности и для проверки статистических гипотез, какие параметры распределения использовали для описания признаков, явлений.

**План исследования** включает перечень организационно-методических вопросов. В соответствии с планом исследователь определяет, что исследовать, в каком направлении, сколько единиц наблюдения взять для исследования, где проводить исследование и т.д.

При выборе места исследования необходимо учитывать типичность медицинских организаций, предприятий, районов, городов, территорий. В таком случае исследование будет представлять ценность, а именно давать возможность использования результатов научного труда на практике, распространения полученных данных на другие подобные организации, территории и т.п.

Под **объектом наблюдения** понимается статистическая совокупность, состоящая из отдельных предметов или явлений (единиц наблюдения), взятых вместе в единых границах времени и пространства. Объект исследования должен иметь четко установленный формат, которым определяются место (территория), сроки, объем и единица наблюдения. Объектами исследований в области организации здравоохранения и общественного здоровья могут быть медицинские организации, население областей, краев, городов и районов, ресурсы системы здравоохранения и т.д.

**Тему** научной работы определяет предмет исследования. **Предмет** — это свойства, особенности, процессы объекта исследования, которые следует изучить. В качестве изучаемого явления можно рассматривать общественное здоровье населения, организацию медицинской помощи, ресурсное обеспечение отрасли и т.п.

В качестве объекта исследования может быть представлено население определенной территории, а предметом в данном случае будет заболеваемость населения изучаемой территории. В одном и том же объекте может быть несколько предметов исследования. Границы между предметом и объектом условны и подвижны. Например, изучение качества жизни, связанного со здоровьем, — предмет исследования для практикующего врача, а для организатора здравоохранения — объект исследования. Точное определение предмета избавляет исследователя от безнадежных попыток «объять необъятное» и позволяет представить новую информацию об объекте исследования.

**Единица наблюдения** — первичный элемент статистической совокупности, наделенный всеми признаками, подлежащими учету и регистрации. Если



объект исследования — работники угольных предприятий, то один работник — единица наблюдения. Если же объект исследования — здравоохранение определенной территории, единицей наблюдения в исследовании будет каждая конкретная медицинская организация.

Общее число единиц наблюдения в исследовании характеризует его **объем**. Если исследование является несплошным, объем выборки должен быть достаточным для получения статистически значимых результатов, т.е. исследование должно иметь необходимую статистическую мощность.

**Статистическая мощность** (чувствительность) исследования — это вероятность того, что при проверке статистической гипотезы исследование с данными объемами выборок выявит статистически значимое, реально существующее различие между выборками. Иными словами, это вероятность объективного отклонения нулевой гипотезы в случае, когда на самом деле верна альтернативная.

Статистическая мощность критерия равна  $1 - \beta$ , где  $\beta$  — вероятность ошибки второго рода (т.е. вероятность ошибочного принятия нулевой гипотезы), допускаемой в исследовании. Если  $\beta$  принимается равной 0,2 (20%), то статистическая мощность исследования равна  $1 - 0,2 = 0,8$  (80%). Чем больше статистическая мощность исследования, тем больше уверенность в том, что имеющиеся в реальности различия между выборками выявлены в ходе исследования.

Для обеспечения большой статистической мощности требуются большие объемы наблюдений. В настоящее время считается, что если в ходе исследования невозможно обеспечить сбор необходимого числа наблюдений, такое исследование должно быть запрещено по этическим соображениям. Для преодоления проблемы большого числа наблюдений организуются многоцентровые исследования, в которых участвуют несколько организаций, иногда из разных государств.

Обычно выбирают статистическую мощность исследования более 80%. Необходимое число единиц наблюдения определяют на основе данных, полученных при проведении пробного (пилотного) исследования, с использованием соответствующих формул, либо таблиц, либо номограмм.

Определить объем совокупности выборочного исследования можно также при помощи специальных таблиц. В пособии для врачей А.М. Меркова и Л.Е. Полякова «Санитарная статистика» приведена таблица «Число наблюдений, необходимое для того, чтобы ошибка в 19 случаях из 20 не превысила заданного предела» (табл. 2.4).

**Таблица 2.4.** Число наблюдений, необходимое для того, чтобы ошибка в 19 случаях из 20 не превысила заданного предела

При величине показателя, %	Предел ошибки, %						При величине показателя, %
	1	2	3	4	6	10	
1	2	3	4	5	6	7	8
1	400	100	45	26	17	5	99



Окончание табл. 2.4

При величине показателя, %	Предел ошибки, %						При величине показателя, %
	1	2	3	4	6	10	
1	2	3	4	5	6	7	8
2	800	200	90	50	32	9	98
3	1200	300	130	74	48	13	97
4	1300	400	120	100	62	16	98
5	1900	500	210	120	77	20	95
6	2300	600	250	140	90	24	94
7	2600	650	290	160	110	27	93
8	3000	740	330	190	120	30	92
9	3300	800	370	210	130	34	91
10	3600	900	400	230	150	37	90
15	5100	1300	570	320	210	52	85
20	6400	1600	710	400	260	65	80
25	7500	1900	830	470	300	76	75
30	8400	2100	930	530	340	85	70
35	9100	2300	1010	570	370	92	65
40	9600	2400	1070	600	390	97	60
45	9900	2500	1100	620	400	100	55
50	10000	2500	1110	630	400	100	50

**Пример.** В хирургическом отделении больницы за год было прооперировано 384 чел., у 64 больных в послеоперационном периоде возникли осложнения. Требуется найти частоту возникновения осложнений.

Вычисляем интенсивный показатель.

$$x = (100 \times 64) / 384 = 16,7.$$

Достаточным ли был объем наблюдения (384 чел.)?

Находим в таблице величину показателя в процентах (15–20). Предельная ошибка, которая нас устроит, — 4%. На пересечении находим число наблюдений, необходимое для того, чтобы ошибка в 19 случаях из 20 не превысила заданного предела. Это число — 320–400 чел. Следовательно, объем наблюдения, равный 384 пациентам, достаточен для получения статистически значимых результатов.

Необходимое число наблюдений для относительного показателя, полученного по выборке, можно определить при помощи преобразования формулы предельной ошибки выборки ( $\Delta$ ):

$$\Delta = \pm tm = \pm t \times \sqrt{\frac{p(100-p)}{n}},$$

где  $t$  — доверительный коэффициент,  $m$  — ошибка относительной величины,  $p$  — показатель,  $n$  — число наблюдений.

Решая приведенное равенство относительно  $n$ , получаем формулу для определения необходимого числа наблюдений:

$$n = \frac{t^2 \times p \times (100 - p)}{\Delta^2}.$$

Используя данные рассмотренного ранее примера, проведем проверку достаточности объема наблюдений выборочной совокупности:

$t$  — доверительный коэффициент; при  $P = 95,5\%$  равен 2;  $p = 16,7$ ;  $\Delta = 5\%$  (задает сам исследователь).

$$n = \frac{2^2 \times 16,7 \times (100 - 16,7)}{5^2} = 222.$$

Следовательно, необходимый объем наблюдений выборочной совокупности равен 222.

Необходимое число наблюдений для выборочного среднего можно определить при помощи преобразования формулы предельной ошибки выборки ( $\Delta$ ):

$$\Delta = \pm tm = \frac{t \times \sigma^2}{n^2},$$

где  $t$  — доверительный коэффициент,  $m$  — ошибка средней арифметической,  $\sigma$  — среднеквадратическое отклонение,  $n$  — число наблюдений.

Решая приведенное равенство относительно  $n$ , получим формулу для определения необходимого числа наблюдений:

$$n = \frac{t^2 \times \sigma^2}{\Delta^2}.$$

Проведем проверку достаточности объема наблюдений выборочной совокупности.

**Пример.** Пусть  $t$  — доверительный коэффициент, при  $P = 95,5\%$  равен 2;  $\sigma$  — составляет  $\pm 1,8$ ;  $\Delta = 0,5$  (задает сам исследователь)

$$n = \frac{2^2 \times 1,8^2}{0,5^2} = 51,8 \approx 52$$

Следовательно, необходимый объем наблюдений выборочной совокупности равен 52.

Оценить необходимый объем наблюдений и мощность критерия можно, используя номограмму Альтмана (рис. 2.6).

**Пример.** Объект исследования — две группы пациентов, пролеченных двумя разными способами лечения. Для проверки гипотез о равенстве средних в группах использовали непарный критерий Стьюдента ( $t$ ). Необходимо оценить эффективность нового метода лечения. Показатель, который будет оцениваться — число дней, на протяжении которых сохранялись симптомы заболевания.

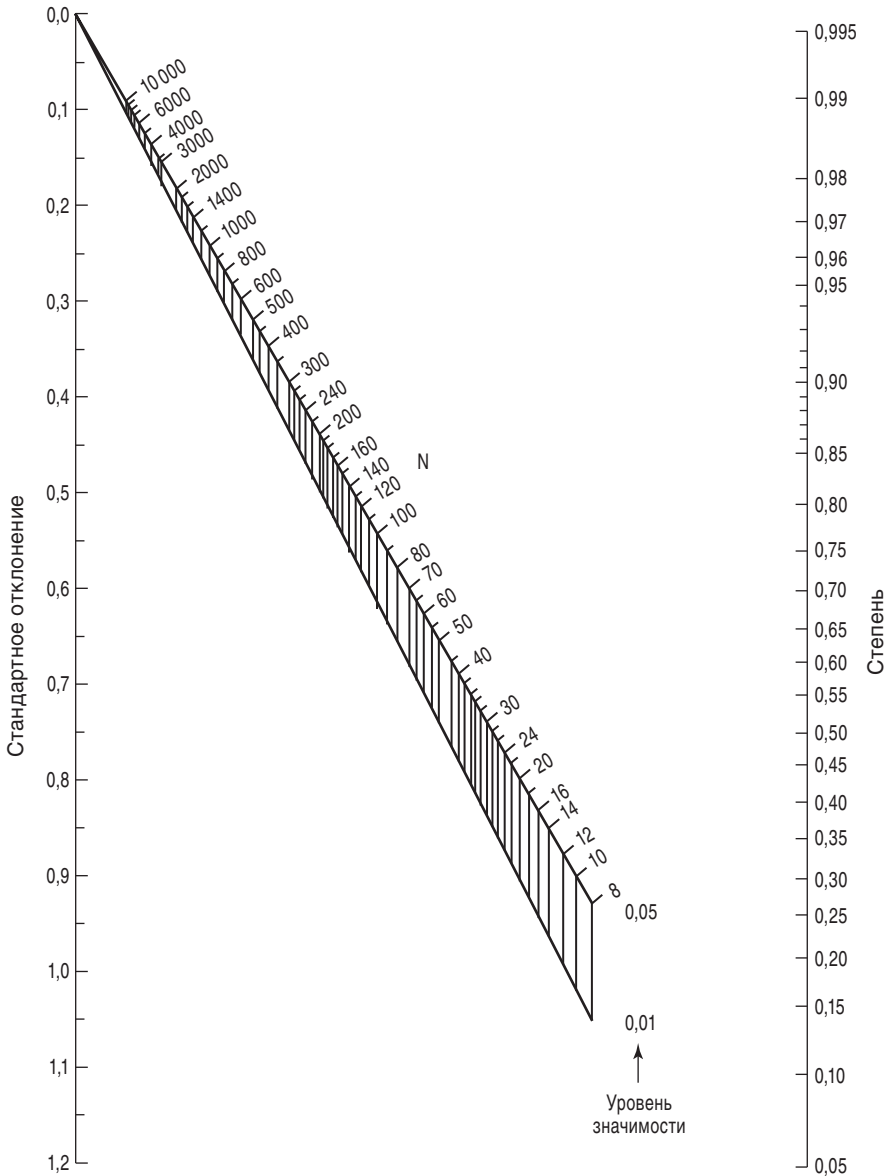


Рис. 2.6. Номограмма Альтмана для определения объема выборки

Необходимо ответить на вопрос, сколько пациентов включить в исследование, чтобы иметь 80% мощность обнаружения 3-дневной разницы в длительности сохранения симптомов между группами пациентов, пролеченных разными методами лечения, при 5% уровне значимости. Предположим, что стандартное отклонение длительности сохранения симптомов заболевания составляет 6 дней. Стандартизованная разница — это соотношение  $\delta/\sigma$ ; где  $\delta$  —

наименьшая разность в средних, которая клинически значима;  $\sigma$  — принятое стандартное отклонение наблюдений, одинаковое в каждой из двух групп (по данным литературы или пилотного исследования). Следовательно,  $3/6 = 0,5$ .

По номограмме Альтмана определим достаточный объем наблюдений. Соединив линией стандартизованную разницу и мощность, получим необходимый объем наблюдений, равный 120 пациентам.

Таким образом, чтобы иметь 80% шанс обнаружения разницы средних в 3 дня ( $\sigma = 6$ ) при 5% уровне значимости с применением непарного критерия Стьюдента ( $t$ ), нужно взять 60 пациентов в группу наблюдения и 60 пациентов в группу сравнения.

Для непарного критерия Стьюдента и критерия  $\chi^2$  Пирсона для расчета объема выборки можно применить формулу Лера:  $16/(\text{стандартизованная разность})^2 = 16/(0,50)^2 = 64$ . Числитель 16 относится к мощности 80%, а числитель 21 — к мощности 90%.

Вывод: для мощности 80% и двухстороннем уровне значимости 0,05 потребуется взять по 64 пациента в каждую группу.

Номограмму Альтмана применяют также для оценки мощности критерия при данном объеме выборки. Если обнаружится, что мощность критерия недостаточна (менее 80%) для получения статистически значимых различий, нужно выбирать другой критерий.

**Единицы наблюдения**, отобранные для исследования, характеризуются признаками, которые принято называть учетными. Объективный выбор единицы наблюдения и соответствующих **учетных признаков** — ответственный момент, определяющий весь ход научного исследования. В одном исследовании может быть несколько единиц наблюдения. Если одной из задач исследования предполагается изучение заболеваемости, то за единицу наблюдения принимается случай заболевания. К учетным признакам относятся диагноз, течение заболевания (острое, хроническое), длительность и т.д. Если изучение заболеваемости предусматривает полицевой учет, то единицей наблюдения будет человек, а учетными признаками при регистрации станут пол, возраст, профессия и т.д.

Анализ данных, характеризующих тот или иной признак, необходим для того, чтобы определить параметры описания признака и методы аналитической статистики для выявления различного рода закономерностей.

Признаки в совокупности могут быть выражены количественными и качественными характеристиками.

**Количественные данные** (числовые) могут быть непрерывными, если они представлены дробными числами, и дискретными. К **непрерывным** количественным данным относят рост, массу тела, возраст, стаж и т.д. Данные, представленные целыми числами, например, число студентов в группе, число пролеченных больных, считаются **дискретными**.

**Качественные данные** (категориальные) можно описать словами, но не числами. Они подразделяются на номинальные и порядковые. К номинальным данным можно отнести диагноз, профессию, условия проживания. Номинальные данные, если они имеют лишь два возможных значения (пол, исход заболевания), относят к бинарным (дихотомическим). Порядковые данные — это качественные

данные, которые отражают степень выраженности, какой-либо характеристики объекта наблюдения (группа инвалидности, стадия заболевания).

Учетные признаки по роли в статистической совокупности могут быть факторными и результативными. **Результативный признак** меняет свое значение под действием другого признака, оказывающего на него влияние. Этот признак будет **факторным**. Например, число лет курения — факторный признак, формирование сердечно-сосудистой патологии — результативный.

В плане исследования необходимо предусмотреть продолжительность календарного периода, за который будет осуществлен сбор данных об объекте, время сбора данных. Этот период зависит от сложности объекта наблюдения, численности объекта исследования и масштаба исследования.

По времени исследование может быть **текущим** и **единовременным**. Перепись населения — классический пример единовременного наблюдения, так как проводится на определенный момент времени (критический момент — 01.01.20...).

Регистрация рождений, смертей, заболеваний на определенной территории в течение года — пример текущего наблюдения.

**Второй этап исследования — сбор материала** (проведение наблюдения). Для сбора и регистрации информации используют первичную документацию, формы которой разрабатывают на первом этапе исследования.

Данные для исследования получают:

- путем непосредственного наблюдения, когда они регистрируются непосредственно исследователем на месте их формирования;
- методом выкопировки (извлечения); в этом случае источник информации представлен учетно-отчетной документацией медицинских организаций. Для сбора информации могут использоваться учетные формы, утвержденные государственными статистическими органами;
- анамнестическим методом, при котором регистрируются ответы респондента;
- анкетным способом, в ходе которого респонденты самостоятельно заполняют анкеты. Этот способ эффективен, когда сами респонденты заинтересованы в анкетировании;
- способом интервьюирования, при котором респондентам задают вопросы. Ответы фиксируют в анкете, вопроснике, карте.

Наблюдение может быть сплошным, когда охватывается вся **генеральная совокупность**, т.е. предполагается учет 100% единиц изучаемого объекта, и частичным — исследование части совокупности, но с обязательным распространением результатов на всю совокупность.

Под **статистической совокупностью** понимают отдельные элементы (единицы наблюдения), взятые для изучения какого-либо явления с учетом определенных признаков. Примеры статистической совокупности: население города, района, родившиеся или умершие, заболевшие и другие контингенты населения. Статистическая совокупность может рассматриваться как генеральная или как выборочная, и от этого зависит интерпретация результатов исследования.

**Пример.** Если в целях изучения распространенности туберкулеза на планете обследовать все население земного шара, то получились бы данные на основе генеральной совокупности. Провести такое исследование не предоставляется возможным. В качестве генеральной совокупности может рассматриваться население какого-либо города или села. Генеральная совокупность определяется не только территориальными границами. Ее формирование возможно с учетом иных признаков: возраст, пол, профессия, а также их сочетанием. Таким образом, в зависимости от цели и задач исследования определяется формат генеральной совокупности.

Зачастую невозможно, а иногда и нецелесообразно проводить углубленный анализ всех единиц генеральной совокупности. Именно поэтому исследование ограничивают определенной частью единиц — **выборочной совокупностью**, обладающей всеми ее свойствами. Такая совокупность должна быть репрезентативной, т.е. представленной с точки зрения содержания составляющих элементов, их соотношения и закономерностей, присущих генеральной совокупности.

**Выборочный метод** — наиболее разработанный и распространенный метод, при котором отбор единиц наблюдения из генеральной совокупности осуществляется так, чтобы у каждой единицы были равные шансы попасть в выборку, что именуется **рандомизацией**. Различают следующие способы отбора единиц наблюдения в выборочную совокупность: механический, типологический, серийный, случайный, парно-сопряженный и др.

**Случайный отбор** осуществляется методом жеребьевки. Используются таблицы случайных чисел. Все единицы генеральной совокупности нумеруются. На одинаковых карточках пишутся номера, методом жеребьевки осуществляется отбор единиц для исследования.

**Механический отбор** производится с учетом начальной буквы фамилии, номера истории болезни или других признаков. Для осуществления отбора все единицы совокупности располагаются в определенном порядке. Если речь идет о выборе историй болезней, то можно взять каждую пятую, шестую или десятую историю болезни. Если для обследования приглашают пациентов, то лучше использовать ту или иную первую букву фамилии.

При **типологическом отборе** выборка формируется из типичных групп всей генеральной совокупности, в которых методом случайной или механической выборки отбираются единицы для исследования, что обеспечивает сохранение идентичности соотношения составляющих компонентов выборки и генеральной совокупности.

**Серийный отбор** предполагает формирование серий (гнезд) в генеральной совокупности, каждая из которых имеет равную возможность быть включенной в выборочную совокупность. В выбранных таким способом сериях обследованию подлежат все единицы наблюдения.

Определенный интерес представляет **парно-сопряженный отбор**, или метод уравновешенных групп (копи-пар). Суть этого метода заключается в том, что используются группы наблюдения (опытные) и сравнения (контрольные), в которых подбираются единицы наблюдения с одним, двумя или несколькими однородными признаками.

Иногда в целях углубленного изучения закономерностей тех или иных явлений на фоне уже изученных используют **метод направленного отбора**. Для определения эффективности внедрения стационарозамещающих технологий необходимо методом направленного отбора взять для исследования медицинские организации, где работа по внедрению ресурсосберегающих технологий поставлена на должном уровне, и сравнить с организациями, где такое практикуется недостаточно или вовсе отсутствует.

Одним из видов направленного отбора считается **когортный метод**. Когортной принято называть статистическую совокупность, которая состоит из относительно однородных элементов, объединенных наступлением определенного признака и прослеженного за один и тот же интервал времени.

**Пример.** Для определения численности детей, рожденных в семье (или числа беременностей, родов, аборт) за определенный интервал супружеской жизни, исследуется когорта лиц, имеющих единый срок вступления в брак и единую продолжительность периода супружеской жизни. Взяв за единицу наблюдения молодых (до 30 лет) супругов, проживающих на территории одного города, вступивших в первый брак в течение определенного года, и прослеживая детородную функцию за пять лет их супружеской жизни, получим когорту, состоящую из единиц наблюдения, однородных сразу по пяти признакам.

К **несовершенным сплошным методам** исследования относится метод основного массива, позволяющий выбрать для изучения наиболее представленную часть генеральной совокупности. В соответствии с методом основного массива для исследования отбирается от 70% и более единиц изучаемого объекта. Недостаток метода заключается в том, что не включенная в исследование часть общей совокупности может отличаться от основного массива.

**Монографический метод** используется для детального описания отдельных единиц совокупности.

**Третий этап исследования — сводка и группировка материала.** Собранную информацию следует переносить с бумажных носителей в электронную базу данных. Доступен для большинства пользователей персональных компьютеров программный продукт Access — программа для создания баз данных.

Собранные данные проверяют, систематизируя по группам и подгруппам таким образом, чтобы можно было описать характер распределений внутри них, обнаружить взаимосвязи между признаками, тенденции развития явления во времени, рассчитать статистические показатели, характеризующие процессы и явления. Сводка — это еще и технические операции по распределению данных по группам, таблицам и подсчет итогов.

На этом этапе обрабатывают материалы исследования, характеризуют всю исследуемую совокупность, рассчитывают различные статистические показатели, собранные данные проверяют на наличие ошибок, составляют таблицы, динамические ряды, визуализируют материал.

В современных условиях все эти операции можно осуществить, используя программу Excel, которая позволяет составлять сводные таблицы, группировать данные, рассчитывать различные статистические показатели, строить

диаграммы, оценивать характер распределения признаков в статистической совокупности, проверять выдвинутые гипотезы.

Рекомендуется также использовать пакеты прикладных профессиональных статистических программ STATISTICA, IBM SPSS. Эти программные продукты дают возможность осуществить факторный, кластерный, дискриминантный анализы, прогнозирование и другие сложные виды статистических анализов. Более подробно о данных видах анализа можно прочитать в книге Дж.-О. Кима, Ч.У. Мьюллера, У.Р. Клекки и соавт. «Факторный, дискриминантный и кластерный анализ» (1989).

Перед тем как заносить цифровые данные в таблицы, полученный материал группируют с учетом признаков изучаемого явления. Группировка — это объединение единиц статистической совокупности в однородные группы в соответствии со значениями одного или нескольких признаков. Различают четыре вида группировок: типологическую и вариационную, аналитическую и комбинационную.

**Типологическая группировка** строится на основе качественных признаков (табл. 2.5).

**Таблица 2.5.** Типологическая группировка обратившихся в поликлинику по полу

Пол	Численность, %
Мужчины	
Женщины	
<b>Итого</b>	

В основу **вариационной группировки** положено числовое выражение признака (численность населения, уровень заболеваемости, рост, масса тела, параметры артериального и венозного давления и т.п., табл. 2.6).

**Таблица 2.6.** Вариационная группировка по числу обращений в поликлинику в течение года

Число обращений в течение года	Число обратившихся, %
1	
2	
3	
4 и более	
<b>Итого</b>	

**Аналитические группировки** позволяют установить, как факторный признак влияет на результативный (табл. 2.7).

**Таблица 2.7.** Аналитическая группировка: зависимость заболеваемости от курения

Статус	Число единиц наблюдения	Средний уровень заболеваемости на 1000 населения
Курящие		
Некурящие		
<b>Итого</b>		



**Комбинационные группировки** строятся по иерархической системе (табл. 2.8). Группы, выделенные по одному признаку, делятся на подгруппы по значениям других признаков. Построение комбинационной группировки требует больших размеров совокупности, иначе при образовании большого числа групп появляются малочисленные и (или) пустые интервалы.

**Таблица 2.8.** Группировка комбинационная: распределение работающих металлургических заводов в зависимости от семейного положения, пола и возраста

Пол	Состоит в браке			Не состоит в браке		
	Возраст, лет					
	До 20	21–40	40	До 20	21–40	40 более
Мужчины						
Женщины						
<b>Итого</b>						

Многомерные группировки предназначены для выделения групп, однородных по совокупности признаков. Эту задачу решают кластерным методом, в основе которого лежит идея разбиения исходного множества на непересекающиеся подмножества (кластеры, группы, таксоны). Элементы этих подмножеств либо подобны друг другу, либо наименее удалены друг от друга в  $n$ -мерном пространстве признаков.

При обработке информации названные виды группировок используют в различных сочетаниях.

**Завершающий (четвертый) этап исследования** — формулировка выводов и предложений по внедрению его результатов в практику здравоохранения. На этом этапе исследователь анализирует полученные данные, сравнивает их со стандартами и нормативами, с данными по другим учреждениям, территориям, оценивает показатели в динамике, делает выводы на основе тех задач, которые выдвинуты на первом этапе, литературно оформляет работу.

Использование результатов исследования в практической работе медицинских организаций, образовательном процессе в вузах подтверждается актами внедрения.

## 2.5. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ, ДОВЕРИТЕЛЬНАЯ ВЕРОЯТНОСТЬ, УРОВЕНЬ СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ

Выборка из генеральной совокупности позволяет получить точечную оценку интересующего параметра и вычислить стандартную ошибку для определения точности оценки. **Стандартная ошибка** используется для вычисления интервальной оценки — доверительного интервала для параметра популяции.

**Доверительный интервал** расширяет оценки в обе стороны некоторой величиной, кратной стандартной ошибке найденной величины. Доверительные границы интервала обычно отделяют запятой и ставят в скобки.

Величина доверительного интервала задается вероятностью безошибочного прогноза (доверительная вероятность, надежность). Величина доверительной вероятности может задаваться коэффициентом Стьюдента ( $t$ ).

При достаточном числе наблюдений значения коэффициента  $t$  и доверительной вероятности соотносятся следующим образом: если  $t = 1$ , то с доверительной вероятностью в 68,3% результаты выборочного исследования могут быть перенесены на генеральную совокупность; при  $t = 2$  вероятность перенесения результатов выборочного исследования на генеральную совокупность увеличивается до 95,5% и при  $t = 3$  — до 99,7%.

При интерпретации доверительных интервалов необходимо учитывать их ширину. Широкий доверительный интервал указывает на неточную оценку, узкий — на точную оценку. Ширина доверительного интервала зависит от размера стандартной ошибки, объема выборки и числовой изменчивости данных. Исследования с небольшим набором данных дают широкие доверительные интервалы.

Доверительный интервал принято определять для средней величины, медианы, относительной частоты встречаемости признака в совокупности и т.д. Доверительный интервал с определенной долей уверенности (95 или 99%) показывает, в каких пределах находится изучаемый показатель в генеральной совокупности. Например, 95% доверительный интервал представляет собой область, в которую попадает истинное значение изучаемого показателя в 95% случаев. Иными словами, можно с 95% надежностью сказать, что истинное значение изучаемого показателя в генеральной совокупности будет находиться в пределах 95% доверительного интервала.

Значению **доверительной вероятности** соответствует свой уровень статистической значимости ( $P$ ). **Уровень статистической значимости** выражает вероятность нулевой гипотезы (вероятность того, что в сравниваемых совокупностях отсутствуют различия в распределении частот). Чем выше уровень статистической значимости, тем меньше можно доверять утверждению о том, что различия существуют.

Для вероятности безошибочного прогноза 95% уровень статистической значимости  $P = 1 - 0,95 = 0,05$ . Для доверительной вероятности 99%  $P = 0,01$ .

Таким образом, статистическая значимость полученных на выборке данных представляет собой меру уверенности в их истинности. Уровень статистической значимости находится в убывающей зависимости от доверительной вероятности. Высокая статистическая значимость соответствует низкому уровню доверия к найденной по выборке величине. Именно уровень статистической значимости представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю генеральную совокупность. Выбор того или иного уровня значимости в большинстве случаев произволен. Для медицинских и биологических исследований допустима доверительная вероятность не менее 95%, соответственно уровень значимости составит не более 0,05. В случаях, когда необходима особая уверенность в полученных результатах, в качестве критического уровня статистической значимости принимают  $P = 0,01$  или  $P = 0,001$ .

Интерпретация уровней статистической значимости:

- $P \geq 0,1$  — данные согласуются с нулевой гипотезой;
- $P \geq 0,05$  — есть сомнения в истинности как нулевой, так и альтернативной гипотез;
- $P < 0,05$  — нулевая гипотеза может быть отвергнута;
- $P \leq 0,01$  — нулевая гипотеза может быть отвергнута, сильный довод;
- $P \leq 0,001$  — нулевая гипотеза не подтверждается, очень сильный довод.

Исследователю необходимо различать клиническую и статистическую значимость полученных результатов. Целесообразно употреблять термин «**клиническая важность**». Клинически важное заключение — заключение о том, что при тестируемом методе лечения есть последствия для здоровья пациента. Статистически значимое заключение основано на вероятности. Статистическая значимость отражает влияние случая на результат. Клиническая важность отражает биологическую ценность полученных в ходе исследования данных. Ошибочно делать вывод о «тенденции к значимости» при клинически важных, но статистически не значимых результатах. Результаты исследования не могут ни «стремиться к значимости», ни «приближаться к значимости». Вместо использования этих выражений следует указать отмеченную разность и 95% доверительный интервал для нее.

## 2.6. ОПИСАНИЕ КАЧЕСТВЕННЫХ ПРИЗНАКОВ

### Абсолютные и относительные величины

В медицине и здравоохранении нередко пользуются абсолютными величинами. Они несут важную информацию о численности и составе населения регионов, городов, районов, числе родившихся, заболевших, умерших, сменивших место жительства и т.д. Эти данные используют для организации медицинской помощи населению и решения других проблем.

Абсолютная величина должна быть указана, если данные получены на малых выборках (менее 20 единиц наблюдения), потому что в этом случае процентные значения оказываются значительно больше абсолютного числа единиц наблюдения.

Если число единиц наблюдения в исследовании от 20 до 100, проценты должны быть представлены в виде целых чисел. Например, 81% и 19%. При числе наблюдений более 100 относительная величина указывается с одним разрядом десятичной дроби.

В некоторых случаях абсолютные числа без их преобразования в относительные величины (показатели) имеют ограниченное познавательное значение.

**Пример.** Об эффективности противоэпидемических мероприятий по поводу инфекционных заболеваний можно судить на основании рассмотрения данных об их распространенности в двух городах без учета численности населения городов и определения соответствующих показателей.

Различают следующие виды относительных величин: интенсивные, экстенсивные, соотношения и наглядности.