***Тема №13. Организация медицинского исследования.***

**Основные этапы научного исследования.**

В работе практикующего врача часто встречается необходимость проведения научного исследования. Планирование и оценка результата – неотрывная часть любого исследовательского процесса.

Выделяют следующие этапы статистического исследования:

*1-й этап —* определение целей и задач, составление плана и программы исследования;

*2-й этап* — наблюдение, сводка и группировка полученных статистических материалов, вычисление первичных итогов;

*3-й этап —* углубленная математико-статистическая обработка данных;

*4-й этап —* анализ полученных результатов, выводы.

На предыдущих занятиях изучалось, как правильно описать и сравнить уже имеющиеся данные. Эта тема посвящена особенностям первых двух этапов научного исследования: постановке цели и задач, оценке необходимых размеров выборки, методикам сбора данных и т.д.

***1-й этап***

На 1-ом этапе статистического исследования определяют:

* что делать (цели и задачи исследования);
* как делать (программа исследования). Выбор конкретных методов сбора, обработки и анализа из множества методов статистики напрямую зависит от цели и задач исследования;
* кто, когда и за чей счет будет делать (план исследования).

**Цель исследования** обуславливается конечным результатом, на достижение которого направлено это исследование, т.е. это то, ради чего оно проводится. Цель большинства статистических исследований в медицине — раскрытие взаимосвязи и оценка влияния тех или иных факторов на здоровье человека.

**Задачи исследования** отражают частные вопросы, которые необходимо последовательно решить, чтобы достигнуть конечной цели исследования. При определении конкретных задач обязательно учитывается главный принцип любого статистического исследования — достоверность исходной информации и объективность отражения изучаемого явления, что гарантирует обоснованность (валидность) выводов.

**План исследования** систематизирует решение организационных вопросов. В том числе: выбор места и сроков наблюдения, источников финансирования, субъекта исследования (организации и лиц, осуществляющих основные работы), подбор и обучение кадров, подготовку необходимых аппаратных и программных средств, регистрационных бланков и т.п. Для детальной проработки организационных и методических вопросов планом предусматривается проведение пробных (пилотажных) исследований.

**Программа исследования** включает программу сбора и программу разработки материалов исследования, обусловливает выбор **объекта** и **единицы наблюдения**, а также **учетных признаков**, подлежащих регистрации в ходе исследования. Как обсуждалось на предыдущих занятиях, учетные признаки по виду могут быть **качественными, порядковыми** или **количественными**, а по роли **факторными** или **результативными**.

Выбор единицы наблюдения и учетных признаков определяет весь ход и результаты статистического исследования.

***2-й этап***

В первую очередь на этом этапе проводится сбор данных.

*В обобщенном виде программа сбора — это перечень учетных признаков наблюдения, которые позволяют достаточно полно характеризовать каждую единицу наблюдения и факторы изучаемых явлений.* Конкретное воплощение этого перечня — набор вопросов, содержащихся в формуляре статистического наблюдения. Таким формуляром может быть анкета опроса, регистрационная карта наблюдения и т.п. Правильно разработанный формуляр статистического наблюдения является ключом всего исследования.

В целом, **по способу наблюдения** программой сбора могут предусматриваться следующие **варианты получения исходных данных**: *непосредственное наблюдение (регистрация), выкопировка данных из отчетно-учетных документов, опрос.*

**Непосредственное наблюдение** предполагает непосредственную регистрацию единиц наблюдения и их характеристик, либо измерение параметров с помощью технических средств (измерение жизненной емкости легких, форсированного выдоха, параметров кардиограмм и т.п.).

**Выкопировка данных** из отчетно-учетной документации предполагает использование в виде источника информации различных документов (история болезни, история развития ребенка, больничный лист). Этот способ получения информации требует предварительной экспертной оценки наличия документации в полном объеме, правильности заполнения и полноты записей в документах.

**Опрос** обеспечивает получение информации со слов опрашиваемого (респондента) методом интервью или заочным путем (почтовые, телефонные, прессовые опросы). Регистрация такой информации производится на специальные опросные листы или анкеты. Для качественного проведения опроса рекомендуется привлекать специалистов по разработке опросных листов.

Для получения оптимальных результатов посредством любой из существующих методик проведения опросов необходимо придерживаться следующих основных требований к содержанию опросного листа (анкеты):

* простота и доступность изложения вопросов и вариантов ответов, учет контингента, среди которого будет проводиться опрос. Например: опросы сотрудников управленческого аппарата, врачей, инженерно-технических работников и опросы санитарок — не должны быть одинаковыми;
* удобство для последующего ввода и обработки на компьютере. Необходимо помнить, что формализованные, краткие ответы, используемые с этой целью, могут не всегда соответствовать представлениям респондентов. Поэтому рекомендуется предусматривать варианты уклончивых или неопределенных ответов;
* по возможности содержать контрольные вопросы;
* идентификационные данные (возраст, место жительства, телефон и т.п.) желательно относить в конец опроса;
* оптимальность объема опроса. С одной стороны, размер анкеты должен обеспечивать получение максимума информации. С другой — анкета не должна быть утомительной для заполнения (усталость респондента и опрашивающего резко повышает число неустранимых ошибок) и затруднять своей обширностью последующую обработку.

**Получить неискаженную информацию по анкете, содержащей более 40—50 вопросов, на практике невозможно.**

Чрезмерное увеличение объема любой исходной информации, в том числе получаемой и не путем опроса, неизбежно ведет к увеличению так называемого «информационного шума», т.е. к росту числа помех. Этот «шум» складывается из механических систематических ошибок регистрации и ошибок, связанных с вариабельностью исходной информации. Достигая известного предела, такой «шум» просто подавляет искомую исследователем информацию.

Не является панацеей от упомянутого «шума» и использование самых точных и совершенных приборов и аппаратуры. Это связано, прежде всего, с большой вариабельностью (изменчивостью, случайностью) медико-биологических процессов, которая представляет собой сложное явление, складывающееся из нескольких составляющих. Одной из самых существенных среди них по негативным последствиям, с точки зрения результативности статистической обработки, является **аналитическая вариабельность**. Она возникает из-за расхождения между результатами измерений в одной пробе. Например, результаты подсчета клеточных элементов крови одного и того же зафиксированного мазка, проведенного несколько раз, даже одним и тем же специалистом, будут, хоть и не намного, отличаться.

**Внутрииндивидуальная вариабельность** характеризует расхождение значений нескольких замеров какого-либо параметра у одного и того же человека. Например: масса тела у одного и того же здорового человека в течение суток изменяется.

**Межиндивидуальная, или внутригрупповая**, вариабельность характеризует расхождение значений замеров какого-либо параметра у разных людей, даже если они здоровы.

К этим видам вариаций добавляется еще одна — **вариабельность межгрупповая**, изучение которой, собственно говоря, и является целью любого исследования.

***Высокую вариабельность медико-биологических данных можно снижать,*** используя специальные, порой весьма громоздкие, методики регистрации измеряемых параметров (стандартизация условий регистрации, многократная регистрация замеров и т.д.), а также употребляя специальные приемы статистической обработки. Однако, результативность таких операций остается во многих случаях невысокой. В лучшем случае, удается уменьшить один из компонентов обшей вариабельности. Например, в эксперименте, в котором сравниваются два наблюдения одной и той же группы до и после воздействия изучаемого фактора, исключается часть внутригрупповой вариации, связанной с индивидуальными различиями субъектов, поскольку рассматриваются просто разности между двумя измерениями («до» и «после») для каждой единицы наблюдения. Именно это учитывается при вычислении критерия Вилкоксона для зависимых выборок.

***По времени наблюдение может быть текущим или единовременным.***

**Текущее (непрерывное) наблюдение** предусматривает регистрацию данных по мере их возникновения за какой-либо длительный промежуток времени. Данные при этом виде наблюдения накапливаются во времени. Например: данные о заболеваемости по обращаемости, инфекционной заболеваемости, смертности, регистрация актов гражданского состояния ЗАГСами за год и т.д. Отдельные результаты, полученные таким путем, можно суммировать (помесячную заболеваемость можно суммировать по кварталам, за год).

**Единовременное (прерывное) наблюдение** предусматривает регистрацию данных в один момент времени, или по состоянию на один момент времени, так называемый критический момент наблюдения. Таким образом, проводится сбор данных при переписях населения.

**На определенный момент времени (на конец года)** учитывается численность населения, численность медицинского персонала, количество учреждений медицинской помощи. Заболеваемость по данным профилактических осмотров (патологическая пораженность) также регистрируется на определенный момент времени. Особенность полученных таким путем данных — их нельзя просто суммировать. Например: численность населения района в 2006 году составила 20000 человек, в 2007 году — 18000 человек. За суммарный период 2007—2008 годов можно рассчитать только среднегодовую численность населения. Если единовременное наблюдение регулярно повторяется (ежегодный учет численности населения, предприятий и т.п.), то **такое наблюдение называется периодическим.**

***По охвату статистической совокупности исследование может быть сплошное или не сплошное.*** Эта методическая особенность сбора данных определяет весь дальнейший ход и методику статистического анализа.

При сплошном статистическом исследовании группа наблюдения формируется путем полного охвата всех единиц **генеральной совокупности**. На практике сплошное исследование проводится крайне редко, поскольку осуществить такое наблюдение организационно очень трудно или физически невозможно из-за больших размеров генеральной совокупности или из-за отсутствия определенных границ этой совокупности. В ряде ситуаций, даже если генеральная совокупность ограничена в своих размерах, исследование объекта приводит к его уничтожению (анализ качества промышленных партий вакцин, сывороток, медикаментов). В этой ситуации проведение сплошных исследований также невозможно. К тому же, сплошные исследования во много раз дороже не сплошных. Существенный недостаток сплошных исследований — большие затраты времени на сбор данных, что приводит к затягиванию исследований на продолжительное время. Кроме того, формирование в ходе таких наблюдений больших массивов данных вызывает затруднения при обработке собранных материалов. Такие наблюдения используют, как правило, только для решения общегосударственных задач переписи населения, сбора информации об инфекционной заболеваемости и некоторых других задач.

***К методам не сплошного наблюдения относятся монографический метод, метод основного массива и, собственно, выборочный метод.***

**Монографический метод** применяется для подробного описания объекта, имеющего какие-либо яркие особенности. Например: медико-социальное обследование национальностей Крайнего Севера или социально-гигиеническое описание промышленного центра. Выводы, которые получаются путем таких обследований, относятся либо только к конкретному объекту исследования, либо могут быть распространены на весьма ограниченную группу аналогичных объектов.

**Метод основного массива** предусматривает обследование контингентов, которые могут быть сосредоточены на конкретном объекте. Например: изучение госпитализированной заболеваемости в стационаре. Данные о структуре заболеваний, тяжести течения и их прогнозе, полученные в этом исследовании, могут иметь значение только для решения частных вопросов. Судить о распространенности патологии за пределами такого стационара по этим данным нельзя.

**Собственно выборочное исследование охватывает выборочную совокупность или просто выборку из генеральной совокупности.** Такое исследование имеет ряд весьма существенных преимуществ перед сплошным наблюдением. Во-первых, оно дает значительную экономию средств и требует существенно меньше времени, чем сплошное. Во-вторых, при выборочном исследовании может быть достигнута большая глубина и детальность изучения вопроса. В-третьих, при меньшем числе наблюдений уменьшаются вероятности систематических ошибок наблюдения, возникающих, когда объектам приписывают искаженные данные.

**Виды выборок**

Процедура формирования выборки определяется методом, на основании которого отбираются единицы генеральной совокупности. При формировании выборки используются случайные (вероятностные) и неслучайные (детерминированные, невероятностные) методы.

Если все единицы генеральной совокупности имеют известный шанс (вероятность) быть включенными в выборку, то она называется **вероятностной** или **случайной**. Если этот шанс (вероятность) неизвестен, то выборка является **невероятностной**.

**Вероятностные выборки** предполагают, что в основе их формирования лежит принцип случайности, когда каждая единица генеральной совокупности имеет равные шансы попасть в выборку. Случайность при этом рассматривается не в обыденном смысле слова, а как некая система организации отбора, при которой именно случай определяет включение той или иной единицы плюральной совокупности в выборку.

**Простая случайная выборка** предполагает, что вероятность быть включенными в выборку известна и является одинаковой для всех единиц генеральной совокупности. При этом отбор производится независимо от признаков, которыми обладают отдельные единицы (то есть независимо от пола, возраста, образования, уровня доходов и т.д). Вероятность быть включенным в выборку определяется отношением се размера к объему генеральной совокупности.

**Систематическая выборка** также может использоваться только в том случае, когда исследователь располагает полным списком единиц генеральной совокупности. Техника формирования данной выборки предполагает определение выборочного шага («интервала скачка») и стартового номера. Первый определяется как отношение объема генеральной совокупности к размеру выборки, второй — случайным образом (лотерейным подбором, по таблице случайных чисел). Предположим, что проводится исследование успеваемости студентов. На первом курсе обучается 800 человек. Определено, что исследованием будет охвачено 100 студентов. Выборочный шаг равен 8 (800:100). Случайным образом определяется стартовый номер (к примеру, 3). Таким образом, в выборку попадут студенты-первокурсники под номерами 3, 11, 19, 27 и т.д.

Возможности использования систематической выборки во многом определяются наличием для различных видов генеральных совокупностей списков (например, спецификаций, телефонных справочников). Систематическая выборка была бы полностью адекватна случайной, если бы единицы генеральной совокупности попадали в нее случайным образом. Такому условию не удовлетворяет ни один реальный список. Поэтому на практике систематический отбор считают эквивалентным случайному, если порядок расположения единиц в списке никак не связан с исследуемыми признаками.

При достаточно простой процедуре формирования систематическая выборка даже при небольших своих размерах позволяет охватить сравнительно большие по объему генеральные совокупности. Данный метод является более экономичным и оперативным по сравнению с методом простого случайного отбора. В то же время, при использовании систематической выборки получаются менее репрезентативные результаты.

**Стратифицированная выборка** предполагает, что исследуемая совокупность подразделяется на страты, то есть на более мелкие совокупности по какому-либо признаку, который связан с исследуемой проблемой (например, взрослое население можно разбить на следующие страты: 18-25, 26-35, 36-45 лет и т.д.). Существенным моментом при стратификации является то, что каждая единица генеральной совокупности попадает только в одну страту. Из выделенных страт производится отбор единиц методами простой случайной или систематической выборки.

Если размер выборки для определенной страты пропорционален размеру страты по отношению ко всей совокупности, то выборка называется пропорционально стратифицированной. В случае непропорционально стратифицированной выборки необходимо использовать весовые коэффициенты, уравновешивающие размеры страт.

При применении данного метода обычно используются признаки, которые находятся в известной корреляции с проблемами, являющимися предметом исследования. Исходными принципами при выборе признаков для стратификации считаются: минимизация различий между единицами каждой страты и максимизация различий между стратами. Первый принцип обеспечивает однородность единиц каждой страты, то есть является гарантией ее гомогенности. Необходимость этого связана с предположением, что лица, попадающие в одну и ту же страту, будут проявлять достаточно сходное поведение. Соблюдение же второго принципа гарантирует достаточные различия в поведении респондентов разных страт. В противном случае, формирование стратифицированной выборки не имеет смысла.

Стратифицированная выборка является более точной по сравнению с простой случайной и предпочтительна, когда генеральная совокупность очень неоднородна по качественному составу. Однако, стратифицированная выборка может быть сформирована при наличии дополнительной информации обо всей генеральной совокупности или результатов предыдущих исследований.

*Приведем пример формирования стратифицированной выборки*. Предположим, что на определенном курорте в настоящее время отдыхает 2000 туристов, из которых: в возрасте 16-20 лет — 500 человек; 20-30 лет — 800; 30-40 лет — 400; свыше 40 лет — 300. Стратифицированная выборка может выглядеть следующим образом: *n = 200* человек, из которых: 50 человек в возрасте 16-20 лет; 80 — в возрасте 20-30 лет; 40 — в возрасте 30-40 лет; 30 — в возрасте свыше 40 лет.

**Кластерная выборка**, как и стратифицированная, предусматривает разделение генеральной совокупности на более мелкие. Основное различие между ними состоит в том, что при кластерной выборке производится отбор определенных частей (кластеров), из которых, в свою очередь, случайным образом осуществляется отбор лиц, попадающих в выборку. При стратификации в выборку включаются лица из каждой страты целевой совокупности. Тогда каждая страта является гомогенной по тому или иному признаку (например, полу, возрасту). В случае же кластерной выборки каждый кластер представляет собой уменьшенную модель генеральной совокупности. При этом единицы кластера должны быть максимально разнородными, а сами кластеры — как можно более однородными.

При исследовании здоровья населения страны кластеры могут быть сформированы по принципу территориальной привязки к определенному региону. Страна разбивается на четко определяемые части (кластеры), например, области. Считается, что выделенные кластеры являются идентичными и состояние здоровья населения отдельных областей характерно для страны в целом. Далее случайным образом выбирается одна из областей (один кластер), формируется выборка, проводится соответствующее исследование, а выводы относятся к совокупности населения всей страны.

Достоинство кластерной выборки состоит в том, что при ее формировании отсутствует необходимость (как в случае стратифицированной) в сведениях относительно распределения совокупности но отдельным признакам (полу, возрасту, образованию, специальному статусу и т.д.). Одним из самых существенных преимуществ кластерных выборок является и то, что при их использовании исследования отличаются относительно небольшими затратами вследствие компактности отдельных кластеров, в качестве которых могут рассматриваться не только территориально-географические образования.

На практике не всегда имеется в наличии полный список единиц генеральной совокупности, что затрудняет или делает в принципе невозможным использование вероятностной выборки. Поэтому при проведении исследований широко применяются **невероятностные (неслучайные)** выборки, отличающиеся простотой формирования, а также требующие меньших временных и финансовых затрат. В то же время, они не обладают такой точностью и высокой степенью представительности и достоверности, которая характерна для вероятностных (случайных) выборок. При формировании невероятностных выборок подбор единиц в определенной степени является преднамеренным, а не случайным. Поэтому, в отличие от вероятностных выборок, статистические выводы применительно ко всей генеральной совокупности в этом случае делать не вполне правомерно.

**Репрезентативность выборки**

Конечной целью изучения выборочной совокупности всегда является получение информации о генеральной совокупности. Для этого выборочное исследование должно удовлетворять определенным условиям. Одно из главных условий — **репрезентативность (представительность) выборки**. Как обсуждалось ранее, выделяют качественную и количественную репрезентативность.

***Случайность, гарантирующая качественную (структурную) репрезентативность статистических исследований, достигается выполнением ряда условий формирования выборочных групп (совокупностей):***

* Каждый член генеральной совокупности должен иметь равную вероятность попасть в выборку.
* Отбор единиц наблюдения из генеральной совокупности необходимо проводить независимо от изучаемого признака. Если отбор проводится целенаправленно, то и при этом необходимо соблюдать условия независимости распределения изучаемого признака.
* Отбор должен проводиться из однородных групп.

Соблюдение условий, гарантирующих максимальную близость выборочной и генеральной совокупностей, обеспечивается специальными способами отбора. В зависимости от способа формирования различают следующие выборки:

* Выборки, не требующие разделения генеральной совокупности на части (собственно, случайная повторная или бесповторная выборка).
* Выборки, требующие разбиения генеральной совокупности на части (механическая, типическая или типологическая выборки, когортная, парно-сопряженная выборки).

Собственно, случайная выборка формируется случайным отбором — наудачу. В основе случайного отбора лежит перемешивание. Например: выбор шара в спортлото после перемешивания всех шаров, выбор выигрышных номеров лотереи, случайный выбор карточек больных для исследования и т.п. Иногда используют случайные числа, получаемые из таблиц случайных чисел или с помощью генераторов случайных чисел. Согласно этим числам из заранее пронумерованного массива генеральной совокупности выбираются единицы наблюдения с номерами, соответствующими выпавшим случайным числам.

При составлении случайной выборки после того, как объект выбран, и все необходимые данные о нем зарегистрированы, можно поступать двояко: объект можно вернуть, или не вернуть в генеральную совокупность. В соответствии с этим **выборку называют повторной** (объект возвращается в генеральную совокупность) или **бесповторной** (объект не возвращается в генеральную совокупность). Поскольку в большинстве статистических исследований разница между повторной и бесповторной выборками практически отсутствует, то априорно принимается условие, что выборка повторная.

**Механическая выборка** — выборка, когда из обследуемой совокупности единицы наблюдения отбираются механически. Например: отбор каждого пятого или каждого десятого рабочего по карточкам отдела кадров предприятия или по амбулаторным картам поликлиники МСЧ.

**Типическая, типологическая** или **районированная** выборка предполагает разбивку генеральной совокупности на ряд качественно однородных групп. Например: при изучении заболеваемости студентов вуза для углубленного обследования на каждом курсе выбираются типичные по своему составу студенческие группы. Часто этот способ отбора комбинируется с другими способами. Например: территория города делится в зависимости от степени загрязнения на типичные районы, в этих районах путем случайного отбора формируются группы наблюдения.

**Когортный отбор** относится к целенаправленным отборам. При этом способе из генеральной совокупности отбираются лица (распределение на подгруппы при этом является неслучайным), объединенные моментом появления какого-либо признака или изучаемого воздействия, играющего существенную роль в исследовании (год рождения, начало болезни, прием препарата и т.п.).

**Исследование по типу случай-контроль** (СК) – тип эпидемиологического исследования, в котором распределение фактора риска сравнивается в группе пациентов с заболеванием и контрольной группе. Исследование (СК) относится к ретроспективным, поскольку исследователь, разделив пациентов на группы, по тому, есть или нет у них заболевание, выясняет у них информацию из прошлого.

Следует отдельно остановиться на использовании выборочного метода в санитарной статистике при изучении общей заболеваемости населения. Теоретические предпосылки выборочного метода были проверены в ходе специальных исследований. Так, В.С. Быховский и соавт. в 1928 году сделали параллельную обработку 132,8 тыс. карт с данными о заболеваниях сплошным методом и методом механического отбора каждой пятой карты. Анализ результатов этой обработки показал высокую репрезентативность данных выборочного исследования заболеваемости. Однако, вплоть до сегодняшнего дня, отсутствуют единые методические подходы проведения в широкой практике выборочных санитарно-статистических исследований.

**Оценка необходимой численности выборки**

Для того, чтобы выборочная совокупность была количественно репрезентативной по отношению к генеральной, необходимо первоначально оценить количество данных, которое требуется включить в выборочную совокупность.

**При неизвестной величине генеральной совокупности** величину выборки, гарантирующую репрезентативные результаты, если результат отражается показателем в виде **относительной величины (доли)**, определяют по формуле:

 , (1)

где р – величина показателя изучаемого признака, в %; *q = (100-p)*;

t – доверительный коэффициент, показывающий, какова вероятность того, что размеры показателя не будут выходить за границы предельной ошибки (обычно берется t = 2, что обеспечивает 95% вероятность безошибочного прогноза);

Δ — предельная ошибка показателя.

***Например:*** одним из показателей, характеризующих здоровье рабочих промышленных предприятий, является процент не болевших в течение года работников. Предположим, что для промышленной отрасли, к которой относится обследуемое предприятие, этот показатель равен 25%. Предельная ошибка, которую можно допустить, чтобы разброс значений показателя не превышал разумные границы, 5%. При этом показатель может принимать значения 25% ±5%, т.е. от 20% до 30%. Допуская t = 2, получаем



В том случае, **если показатель — средняя величина**, то число наблюдений можно установить по формуле:

 , (2)

где σ — среднее квадратическое отклонение, которое можно получить из предыдущих исследований, либо на основании пробных (пилотажных) исследований.

**При при условии известной численности генеральной совокупности** для определения необходимого размера случайной выборки в случае использования **относительных величин (доли)** применяется формула:

 (3)

**для средних величин** используется формула:

 , (4)

где N — численность генеральной совокупности.

Исходя из условий приведенного выше примера и принимая численность генеральной совокупности *N=500* рабочих, получаем:



Нетрудно заметить, что необходимая численность выборки при условии известной численности генеральной совокупности меньше, чем при неизвестной (соответственнo, 188 и 300 рабочих).

*В целом, число наблюдений, необходимое для получения репрезентативных данных, изменяется обратно пропорционально квадрату допустимой ошибки.*

**Основы группировки данных**

Программа разработки предусматривает реализуемые на втором этапе статистического исследования ***сводку*** и ***группировку*** статистических данных. Эти операции, осуществляемые на основе статистических таблиц, позволяют систематизировать полученные в ходе наблюдения данные, провести обработку и подсчет групповых итогов, расчеты простейших производных величин (статистических коэффициентов, средних величин). На этом же этапе, для повышения наглядности данных, предусматривается использование графических изображений. Иногда (в официальной статистике — почти всегда) на этом заканчивается весь процесс обработки собранных данных.

**Группировка — основа статистической разработки (систематизации) первичного материала.** Группировка, правильно спланированная на этапе подготовки исследования, позволяет облегчить регистрацию или повысить точность измерений на этапе сбора исходных данных без снижения результативности исследования в целом. Группировка собранной исходной информации определяет весь ход статистического анализа.

В ходе статистической разработки исследователю приходится сталкиваться со следующими вариантами группировок, каждый из которых имеет свои методические особенности:

* ***разделение анализируемой статистической совокупности на группы по тем или иным признакам.*** С такого рода группировкой приходится сталкиваться уже при подготовке программы сбора и в ходе реализации программы разработки любого исследования;
* ***объединение мелких однородных групп в более крупные.*** Этот вариант группировки применяется, как правило, уже в процессе статистической обработки данных, если выясняется несостоятельность мелких групп (малое число наблюдений, нечетко выраженный характер распределений и т.п.). Возможность такой группировки целесообразно предусмотреть уже на этапе подготовки программы сбора данных, т.е. обеспечить возможность укрупнения групп в соответствии с общепринятыми границами групп;
* ***комплексная группировка обеспечивает формирование комплексных оценок на основе многих учетных признаков, даже если они разнородны.*** Такая группировка часто делается на основе специально разрабатываемых алгоритмов или экспертных оценок (по аналогии — постановка диагноза на основе многих симптомов и результатов лабораторных обследований). Такая группировка представляет весьма сложную задачу и нередко сама по себе является самостоятельной целью исследования.

Выделяют следующие основные виды группировки в зависимости от конкретной цели статистического исследования (табл. 1).

*Таблица 1*

**Классификация статистических группировок**

|  |  |  |  |
| --- | --- | --- | --- |
| Аналитические | Структурные | Типологические | Специальные (балансовые, матричные и т.д.) |
| Характеризуют взаимосвязи между признаками, проявляют основные тенденции | Выявляют состав, структуру обследованных групп | Характеризуют основные группы (типы групп признаков) | Используются при составлении балансов предприятий и учреждений, отраслей промышленности и т.п. |

***Аналитическая группировка выявляет взаимосвязи между явлениями (признаками их характеризующими).*** При этом они подразделяются на факторные и результативные. Взаимосвязь проявляется в систематическом изменении результативного признака в связи с изменением факторного. Например: температура тела влияет на частоту пульса, в зависимости от величины роста изменяется вес и т.п.

***Структурная группировка выявляет состав, строение однородной в качественном отношении статистической совокупности.*** Например: состав больных по полу, возрасту, диагнозу и т.п. Сопоставление данных структурной группировки во времени дает представление о структурных сдвигах.

***Группировка типологическая — с ее помощью в статистической совокупности выделяются качественно однородные в существенном отношении группы.*** Например: группы больных с одинаковым диагнозом, с одинаковым исходом заболеваний и т.п.

При статистической разработке материала любого исследования необходимо учитывать существующие правила и стандарты определения группировочных признаков и границ групп (возрастно-половые группировки, группировки по категориям тяжести труда и т.п.). Несоблюдение правил формирования этих группировок ведет к потере ценности данных.

В национальной статистике выделяют следующие возрастно-половые группы (в границах точного возраста):

* + Дети до 3 лет. Эта группа находится под наблюдением детских консультаций и обслуживается детскими яслями. Из них часто выделяются дети в возрасте 1 и 2 года жизни;
  + Дошкольники — дети от 3 до 7 лет. Обслуживаются детскими садами;
  + Дети и подростки школьного возраста от 7 до 13 лет и 13—16 лет;
  + Подростки от 16 до 18 лет;
  + Трудоспособный контингент — мужчины 16—60 лет, женщины 16—55 лет;
  + Лица пенсионного возраста — мужчины 60 лет и старше, женщины 55 лет и старше;
  + Женщины репродуктивного возраста — обычно от 15 до 45 лет.

Иногда для группировки по возрасту используют одногодичные или (для взрослых) пятилетние, реже десятилетние интервалы. Группировка с пятилетним интервалом выглядит следующим образом: до 20 лет, 20—24, 25—29, 30—35 и т.д. С десятилетним интервалом: до 20 лет, 20—29, 30—39 и т.д. При изучении заболеваемости в связи с производственными факторами обычно используются аналогичные 5- и 10-летние интервалы группировок по стажу работы: 1—4 года, 5—9 лет, 10—14 лет и т.д. При этом, к группе 1—4 года относятся лица со стажем от 1 года до 4 лет 11 мес. 29 дней. Аналогично, к группе со стажем от 5 до 10 лет относятся лица, отработавшие от 5 лет до 9 лет 11 мес. и 29 дней.

В ряде случаев целесообразно расчленять отдельные крупные группы на более мелкие. Например: до 20 лет, 20—29, 30—39, 40—44, 45—49, 50—59, 60 и старше.

Группировка данных по своей сути представляет собой принцип классификации, т.е. установление принадлежности явлений и объектов к определенным классам. В государственной статистике для этого используются классификаторы — специальные справочники, инструкции и указатели в виде алфавитных и систематических словарей, дополняемых стандартным перечнем объектов и их групп. Использование классификаторов в официальной статистике является обязательным и имеет силу государственного стандарта. Основные классификаторы рассчитаны на длительное использование. Однако, с течением времени они пересматриваются, дополняются, в них вносятся необходимые коррективы. Так, с 1983 до 1999 года в СССР, а затем в России общепринятой считалась Международная статистическая классификация болезней, травм и причин смерти 9-го пересмотра. С 1999 года была введена классификация 10-го пересмотра (с 01.01.99 г. по Приказу Минздрава России № 3 от 12.01.98 г.).

В медико-биологических научных исследованиях использование государственных и международных классификаторов болезней и причин смерти не является строго обязательным. Однако, только их прямое применение или возможность путем вторичной группировки привести данные исследований в рамки, определяемые общепринятыми классификаторами, гарантирует сопоставимость статистических материалов о заболеваемости, причинах смерти разных категорий населения на различных территориях страны и за рубежом.