

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ УЗБЕКИСТАН

**УЧЕБНО-МЕТОДИЧЕСКИЙ КАБИНЕТ ПО ВЫСШЕМУ И СРЕДНЕМУ
МЕДИЦИНСКОМУ ОБРАЗОВАНИЮ**

ТАШКЕНТСКИЙ ПЕДИАТРИЧЕСКИЙ МЕДИЦИНСКИЙ ИНСТИТУТ

МЕДИЦИНСКАЯ СТАТИСТИКА
(методическое пособие для студентов)

ТАШКЕНТ - 2006

**МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ
УЗБЕКИСТАН**

**УЧЕБНО-МЕТОДИЧЕСКИЙ КАБИНЕТ ПО ВЫСШЕМУ И СРЕДНЕМУ
МЕДИЦИНСКОМУ ОБРАЗОВАНИЮ**

**ТАШКЕНТСКИЙ ПЕДИАТРИЧЕСКИЙ МЕДИЦИНСКИЙ
ИНСТИТУТ**

**«УТВЕРЖДАЮ»
Начальник Главного
управления кадров и
учебных заведений
Минздрава РУз**

Ш.Э.АТАХАНОВ
« _____ » _____ 2006 г.

МЕДИЦИНСКАЯ СТАТИСТИКА
(методическое пособие для студентов)

ТАШКЕНТ - 2006

Составители: зав.кафедрой общественного здоровья, организации и управления здравоохранением ТашПМИ, д.м.н., профессор Ш.Т.ИСКАНДАРОВА;
ассистенты кафедры: М.Т.КАРИЕВА, З.И.ОТАШЕХОВ, Г.А.ДЖАЛИЛОВА

Рецензенты:

1. доцент кафедры общественного здоровья, организации и управления здравоохранением Ташкентской Медицинской Академии, к.м.н., доцент КАРИМБАЕВ Ш.Д.
2. доцент кафедры общественного здоровья, организации и управления здравоохранением Ташкентской Медицинской Академии, к.м.н., доцент РУСТАМОВА Х.Е.
3. доцент кафедры общественного здоровья, организации и управления здравоохранением ТашПМИ, к.м.н. ФАЙЗИЕВА М.Ф.

СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ. ЭТАПЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

С помощью количественных характеристик, с учетом конкретных исторических условий статистика помогает выявить важнейшие закономерности различных процессов в экономической, социальной жизни общества, в его здоровье, а также в системе организации медицинской помощи населению.

Характерной особенностью статистики является применение ее для изучения массовых, а не единичных явлений. По единичным наблюдениям невозможно выявить, вскрыть общие, типичные особенности изучаемого процесса. На массе же наблюдений статистика, используя математические методы, устанавливает наиболее общие закономерности, характерные для исследуемых явлений. При этом она базируется в первую очередь на теории вероятностей и законе больших чисел.

Статистика – общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественными особенностями.

В медицинской статистике различают три основных раздела:

1. теоретические и методические основы медицинской статистики
2. статистику здоровья населения
3. статистику здравоохранения.

В задачи медицинской статистики входит:

- 1) выявление особенностей в состоянии здоровья населения и факторов, определяющих его;
- 2) изучение данных о сети, деятельности и кадрах лечебно-профилактических учреждений, а также данные о результатах лечебно-оздоровительных мероприятий, которые используют при поиске путей улучшения здоровья населения и дальнейшего совершенствования системы здравоохранения.

Кроме того, методы медицинской статистики применяют в экспериментальных, клинических, гигиенических и лабораторных исследованиях.

В медицине статистические приемы используют при:

1. клинико-гигиеническом нормировании факторов производственной среды
2. расчете доз лекарственных препаратов
3. определении стандартов физического развития
4. оценке эффективности примененных методов профилактики или иных заболеваний и т.д.

Статистический анализ также позволяет обосновать ту или иную тактику врача в предупреждении или лечении заболеваний.

Таким образом, каждый врач должен хорошо знать теоретические основы статистики, уметь правильно использовать статистические методы и оценивать информацию, накопленную в различных областях его деятельности.

СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ

Изучение тех или иных явлений с применением статистических методов требует от врача, прежде всего, умелого подхода к выбору объекта исследования (так называемой статической совокупности), единицы наблюдения и ее учитываемых признаков.

Статистическая совокупность – это группа, состоящая из большого числа относительно однородных элементов (единиц наблюдения), взятых вместе в известных границах времени и пространства.

Статистическая совокупность состоит из отдельных единичных наблюдений, однако, это не простая механическая их сумма, а специальным способом сформированная группа. Численность единиц наблюдения в совокупности определяет объем исследования и обозначается буквой «n».

В зависимости от конечной цели и задач исследования решается вопрос о первичном элементе статистической совокупности, который будет принят за единицу наблюдения.

Примерами статистической совокупности могут быть: население того или иного района, города, группа родившихся или группа умерших в данном году, группа больных тем или иным заболеванием и др.

Каждая единица наблюдения имеет много характеристик, однако учитываются только те из них, которые отвечают целям и решениям конкретных задач исследования. Эти признаки учитывают (регистрируют) и поэтому их называют учитываемыми. Вместе с тем каждый из этих признаков имеет свои градации. Например, возраст может иметь такие градации: до 20 лет, 20-24, 25-29 лет и т.д. Анализируя исходы лечения, выделяют больных с улучшением, без изменения, ухудшением процесса, а также лиц, у которых наступил смертельный исход.

Такие учитываемые признаки, как пол, возраст, место жительства, сроки заболевания и госпитализации, результаты клинических исследований, исходы лечения и другие, позволяют всестороннее изучить не только каждый элемент совокупности (единицу наблюдения), но и всю совокупность в целом.

По характеру учитываемые признаки делят на:

1. атрибутивные (описательные), выраженные словесно
2. количественные (выраженные числом).

К **атрибутивным** признакам относятся место жительства и др.

К **количественным** признакам относят: рост, массу тела, число дней лечения, количество холестерина в крови, количество белка в моче и т.д. Каждая величина количественного признака называется вариантной и обозначается буквой «V».

Врачом должно быть выявлено влияние отдельных признаков на изучаемое явление, поэтому различают факторные и результативные признаки.

Факторными называются такие учитываемые признаки, под влиянием которых изменяются другие, зависящие от них результативные признаки.

С изменением величины факторного признака происходит соответствующее возрастание или снижение числовых значений

результативного признака. Так, например, с увеличением возраста ребенка увеличивается его рост (возраст – факторный признак, рост – результативный признак). К факторным признакам следует отнести методы профилактики и лечения (или дозу лекарств), а также пол, возраст, профессию, образование, доход, т.е. признаки, которые могут прямо или косвенно повлиять на результативные признаки.

К результативным признакам можно отнести заболевание (диагноз), его исход (выздоровление, смерть, инвалидность), массу тела, рост уровень белка, холестерина, гемоглобина крови и др.

Каждая статистическая совокупность может рассматриваться как генеральная или как выборочная, от этого зависит интерпретация результатов исследования.

Генеральная совокупность состоит из всех единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования. Например, если бы можно было изучить всех больных ревматизмом, живущих на земном шаре, то такая группа больных составила бы генеральную совокупность.

При изучения здоровья населения генеральная совокупность, как правило, рассматривается в пределах конкретных границ, очерченных территориальным или производственным признаком, и поэтому включает в себя определенное число наблюдений.

Таким образом, в зависимости от цели исследования и его задач, изменяются границы генеральной совокупности, для этого используют основные признаки, ее ограничивающие.

В связи с невозможностью, а также часто и с нецелесообразностью или трудностью углубленного анализа всех единиц наблюдения, составляющих генеральную совокупность, исследование ограничивают только определенной частью единиц – выборочной совокупностью.

Выборочная совокупность – часть генеральной совокупности, отобранная специальным методом и предназначенная для характеристики генеральной совокупности. На основе анализа выборочной совокупности можно получить достаточно полное представление о закономерностях, присущих всей генеральной совокупности.

Выборочная совокупность должна быть репрезентативной, т.е. в отобранной части должны быть представлены все элементы и в том соотношении, как в генеральной совокупности. Иными словами, выборочная совокупность должна отражать свойства генеральной совокупности, т.е. правильно ее представлять.

Для обеспечения репрезентативности выборочной совокупности к ней предъявляют два основных требования:

1) качественная репрезентативность – она должна обладать основными характерными чертами генеральной совокупности; т.е. быть максимально на ее похожей. Собственно поэтому выборочную совокупность следует отбирать из генеральной на основе определенных правил, обеспечивающих объективность отбора составляющих ее единиц;

2) **количественная репрезентативность** – она должна быть достаточной по объему (числу наблюдений), чтобы более точно выразить особенности генеральной совокупности. Статистика располагает специальными формулами или же готовыми таблицами, по которым можно определить необходимое число наблюдений в выборочной совокупности.

Необходимо знать, что статистическая совокупность в отличие от отдельных единиц наблюдения (индивидуумов) имеет особые, только ей присущие свойства, к числу которых относят: характер распределения изучаемого явления; его средний уровень, который дает обобщающую характеристику изучаемому явлению; разнообразие (колеблемость, изменчивость) единиц наблюдения, составляющих совокупность, а также очень часто необходимо изучить взаимосвязь между изучаемыми признаками и репрезентативность признаков выборочной совокупности по отношению к генеральной.

С помощью специальных статистических методов выявляют эти свойства и получают обобщенные характеристики.

РАСПРЕДЕЛЕНИЕ ПРИЗНАКА В СТАТИСТИЧЕСКОЙ СОВОКУПНОСТИ

Элементы, из которых состоит совокупность, имеют различные по величине значения изучаемого признака и каждое из этих значений встречается в группе с неодинаковой частотой.

Характер распределения четко виден только на достаточно большой совокупности наблюдений. Изучая его, получают важную информацию о закономерностях, присущих тому или иному явлению, а также возможность правильно выбрать статистические критерии для обобщения.

Распределения, которые наблюдаются в медицинских, в том числе и в социально-гигиенических, исследованиях, довольно различны:

1. альтернативный,
2. нормальный (симметричный)
3. асимметричный (правосторонний, левосторонний)
4. двугорбый – бимодальный

В социально-гигиенических исследованиях очень популярным является **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.

Чаще других типов распределения встречается **нормальное (симметричное)** распределение. Обычно наблюдается нормальное распределение. 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. При этом наибольшее число случаев наблюдений приходится на середину ряда.

Встречаются явления, которые распределяются по типу асимметричного распределения. При асимметричном распределении наибольшее число случаев наблюдения скапливается не на уровне середины ряда, а сдвигается в сторону меньшего значения признака (правосторонняя асимметрия) или в сторону большего значения признака (левосторонняя асимметрия), или же скапливается по концам ряда (двугорбое бимодальное распределение).

Правосторонняя асимметрия характерна для распределения такого признака, как число детей в семье или кратность случаев семей, где 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 нетрудоспособности 1-2 (т.е. значительное число болеющих скапливается у наименьшей градации данного признака).

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, так как средняя величина роста у мальчиков и девочек различна.

ЭТАПЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Важнейшими элементами организации любого исследования, проводимого на достаточно большой совокупности, являются четыре его этапа.

Эти этапы осуществляют в строгой последовательности:

- 1) составление программы и плана исследования;
- 2) сбор материала;
- 3) разработка данных;
- 4) анализ, выводы, предложения, внедрение результатов исследования в практику.

Первый этап – составление программы и плана исследования – представляет важный раздел работы, который можно правильно осуществить только при наличии глубоких знаний по изучаемому вопросу и четкого представления о цели исследования. Программа и план подчиняются единой цели исследования: в программе указывают цели и задачи исследования, а план предусматривает организационные вопросы исследования.

Программа статистического исследования включает сведения, которые необходимо собрать, и указывает ведущие направления исследования. Она состоит из трех компонентов:

- 1) программы сбора материала;

- 2) программы разработки (табличной сводки);
- 3) программы анализа.

Программа сбора материала представляет собой карту (бланк) с перечнем вопросов, подлежащих обязательному изучению в процессе наблюдения. В заголовке каждой карты (бланка) указано, на кого она будет заполняться (формулировка единицы наблюдения), и перечислены ее признаки (учитываемые признаки).

Например, «Карта выбывшего из стационара». Также существуют официальные программы сбора материала (учетные формы), которые разработаны государственной статистической здравоохранения и служат для изучения состояния здоровья населения и оценки деятельности лечебно-профилактических и санитарно-эпидемиологических учреждений.

Однако в медицинской науке и практике здравоохранения для углубленного изучения ряда проблем научному работнику и врачу-практику требуется самому составлять специальную программу сбора – специальную карту. При составлении такой карты следует иметь в виду ряд важных правил. В карте все признаки должны соответствовать цели и задачам исследования.

Следует четко и кратко сформулировать эти вопросы. Не следует включать такие вопросы, на которые трудно ответить или получить точный ответ. Кроме того, рекомендуется на каждый вопрос предусмотреть возможные варианты ответов и дать так называемые «подсказки» с учетом определенных группировок, что облегчает заполнение карты. При этом каждой группе «вопрос-ответ» дается условный номер – шифр (код).

Пример вопросов, ответов и шифра

Вопрос (учитываемые признаки)	Ответ (подчеркнуть)	Шифр
1. Возраст	До 20 лет (включительно)	1
	20-39 лет	2
	40 лет и более	3
2. Пол	Мужчина	1
	Женщина	2
3. Профессия	Слесарь	1
	Токарь	2
	И т.д.	и т.д.

В зависимости от характера учитываемых признаков различают два вида группировок:

1. типологическую
2. вариационную.

Типологическая группировка – это группировка по атрибутивным признакам, которые могут быть выражены словесно, описательно (например, пол, виды заболеваний, профессия и др.).

Вариационная группировка – это группировка по признакам, имеющим числовое выражение (масса тела, рост, уровень артериального давления и др.).

Оба вида группировки могут быть представлены в различных сочетаниях и комбинациях.

Макеты таблиц строятся по определенным принципам, которые обеспечивают логичность и компактность каждой таблицы.

Таблицы должны иметь четкое и краткое заглавие. В таблице, как в грамматическом предложении, различают подлежащее (то, о чем говорится) и сказуемое (то, что объясняет подлежащее). **Статистическое подлежащее** – это основной признак изучаемого явления; оно располагается, как правило, по горизонтальным строкам таблицы. **Статистическое сказуемое** – признаки, характеризующие подлежащее, располагается в вертикальных графах таблицы. Оформление таблицы заканчивается итогами по графам и строкам.

Различают следующие виды таблиц:

1. простые
2. групповые
3. комбинационные (сложные).

Простой называется таблица, в которой представлена итоговая сводка данных лишь по одному признаку (табл.)

Таблица

Распределение больных гипертонической болезнью по стадиям заболевания
(макет)

Стадии гипертонической болезни	Число больных
1	
2	
3	
Итого...	

Простая таблица дает представление о перечне каких-то явлений по одному признаку, подвергнутому изучению.

Групповой называется таблица, в которой представлены данные с группировкой их по двум признакам, причем один из них находится в подлежащем таблицы, а другой – в сказуемом (табл.).

Таблица

Распределение больных гипертонической болезнью по полу, возрасту и
стадиям заболевания (макет)

Стадии гипертонической болезни	Пол		Возраст (в годах)					Всего
	Мужчины	Женщины	До 20	20-29	30-39	40-49	50 и старше	
1								
2								
3								
Итого...								

В данной групповой таблице попарно сочетаются признаки: стадия гипертонической болезни и пол, а затем стадия болезни и возраст. Групповая таблица может иметь и больше признаков в сказуемом: сроки госпитализации, сроки операции и др., однако с подлежащим они должны сочетаться только попарно: заболевания и сроки госпитализации, заболевания и сроки операции, заболевания и возраст, заболевания и пол и т.д.

Комбинационной называется таблица, в которой представлены сводные данные с неизменным сочетанием трех и более взаимосвязанных признаков (табл.)

Таблица

Распределение больных гипертонической болезнью по стадиям заболевания, возрасту и полу (макет)

Стадия гипертонической болезни	Возраст в годах								Всего	
	До 30		30-39		40-49		50 и старше			
	М	Ж	М	Ж	М	Ж	М	Ж	М	Ж
1										
2										
3										
Итого										

Подлежащее такой таблицы характеризуется сказуемым, в котором имеется комбинация из нескольких признаков.

В макетах таблиц еще до разработки собранного материала уже представлены основные направления исследования и анализа.

План исследования предусматривает организационные элементы работы.

В плане, прежде всего, следует указать объект исследования, т.е. совокупность явлений, предметов о которых должны быть собраны статистические сведения. **Объектами исследования** социальной гигиены могут быть все население области, города, района или отдельные его группы, больные лица, кадры учреждения здравоохранения и др.

Объект исследования (совокупность) должен быть четко определен в пространстве (территория), во времени (период), в объеме (число наблюдений (n)).

В плане обязательно следует указать, каким образом будет формироваться объект исследования (статистическая совокупность).

По полноте охвата различают сплошные и выборочные наблюдения.

По времени наблюдения – единовременные и текущие.

По способу получения информации могут быть непосредственные и анамнестические (опрос, составление анкет).

Наконец, необходимые сведения могут быть получены при **выкопировке** данных.

Сплошное наблюдение должно предусматривать регистрацию всех единичных случаев, составляющих генеральную совокупность. Сплошным методом собирают сведения о числе родившихся, умерших, обратившихся в поликлинику, инфекционных заболеваниях в целом по стране (или по другой территории), а также данные о численности больниц, врачей, среднего персонала и пр.

Организация выборочного наблюдения

Основные компоненты выборочного метода:

- 1) расчет объема выборочной совокупности (n);
- 2) способы специального отбора необходимой части единиц из генеральной совокупности (случайный, механический, типологический, серийный, парно-сопряженный, направленный, когортный);
- 3) оценка репрезентативности выборочных параметров.

К выборочному методу обращаются в тех случаях, когда необходимо провести углубленное исследование, соблюдая экономию сил, средств, времени.

Определение необходимого числа наблюдений (n) для выборочной совокупности осуществляется по формулам или специальным таблицам.

В плане необходимо предусмотреть способ отбора единиц для выборочной совокупности: случайный, механический, типологический, серийный.

Случайный отбор – это такой отбор, когда все единицы генеральной совокупности имеют равную возможность попасть в выборку. Наиболее распространенным способом такого отбора является **метод жеребьевки**, который заключается в том, что на каждую единицу наблюдения заготавливают небольшую карточку, в которой отмечают порядковый номер по списку. Все карточки должны иметь одинаковый вид; их перемешивают в закрытом ящике и затем в случайном порядке отбирают для данного наблюдения определенное количество карточек.

Механический отбор производится по какому-нибудь признаку (первая буква фамилии, номер истории болезни или по другому признаку), от которого зависят результаты исследования.

При механическом отборе единицы совокупности сначала располагаются в каком-либо порядке (по алфавиту или по номеру истории болезни). После этого производится механический отбор единиц через какой-нибудь интервал (каждую 4-ю или каждую 10-ю историю болезни). Интервал при механическом отборе предварительно рассчитывается. Для этого общее число единиц совокупности следует разделить на число единиц, которые надо отобрать. Например, из 3000 историй болезней нужно отобрать 300. В этом случае величина интервала для отбора историй болезни будет равна 10 ($3000 / 300 = 10$), что означает необходимость отбора одной через каждые 10 историй болезни.

Такой способ отбора является наиболее распространенным. Однако в некоторых случаях применение его становится ограниченным из-за трудоемкости.

Типологический отбор (типологическая, типичная выборка) позволяет производить выбор единиц наблюдения из типичных групп всей генеральной совокупности. Для этого сначала внутри генеральной совокупности все единицы группируются по какому-нибудь признаку в типичные группы (например, по возрасту). Из каждой такой группы производят отбор (случайным или механическим способом) необходимого числа единиц таким образом, чтобы соотношение размеров возрастных групп в выборочной совокупности сохранялось такое же, как и в генеральной совокупности.

Серийный отбор (серийная, гнездовая выборка) предусматривает выбор из генеральной совокупности не отдельных единиц, а выбор серий (гнезд). Для этой вся генеральная совокупность разбивается на относительно однородные серии (гнезда). Отбор серий осуществляется путем случайной или механической выборки. При этом отбор должен производиться так, чтобы для каждой серии генеральной совокупности была бы обеспечена одинаковая возможность быть отобранной в выборочную совокупность. В каждой отобранной серии обследуются все составляющие ее единицы наблюдения.

В некоторых случаях выборочная совокупность формируется в порядке комбинирования различных способов отбора.

Для того чтобы углубленно исследовать взаимосвязи менее изученных признаков при исключении признаков, влияние которых известно, формирование выборочной совокупности можно осуществлять по комплексу признаков, т.е. применять метод направленного отбора. Например, зависимость распространенности гипертонической болезни от возрастно-полового состава населения общеизвестна, поэтому при углубленном исследовании необходимо предусмотреть анализ совокупности лиц одного возраста и пола, т.е. она должна быть однородной по этим общеизвестным признакам.

В ряде случаев направленный отбор единиц наблюдения осуществляется не по одному, а одновременно по двум или нескольким признакам.

Одним из видов направленного отбора является **когортный метод**, который позволяет сделать своего рода срез в том месте изучаемого явления, в котором наиболее ярко проявляются те или иные закономерности.

Когорта – это статистическая совокупность, которая состоит из относительно однородных элементов, объединенных наступлением **определенного признака и прослеженного за один и тот же интервал времени**. Например, для определения численности детей, рожденных в семье (или количества 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, прослеживая детородную функцию за пять лет их супружеской жизни, получим когорту, состоящую из единиц наблюдения, однородных сразу по пяти признакам.

По времени наблюдения формирование совокупности осуществляется путем **текущего или единовременного наблюдения**.

При текущем наблюдении статистические данные собирают путем регистрации каждого отдельного случая по мере его возникновения за определенный промежуток времени. Таким путем изучают рождаемость, смертность, заболеваемость, нагрузку врачей и многие другие явления.

В порядке текущей регистрации изучаются те явления, которые зависят от времени наблюдения.

При единовременном наблюдении регистрацию данных приурочивают к какому-то моменту времени – сбор всех наблюдений осуществляется одновременно. Единовременное наблюдение отражает состояние изучаемого явления на определенный момент. Примером единовременного наблюдения являются перепись населения, медицинские осмотры населения, санитарные обследования объектов, взятие каких-то проб и т.д.

Сбор сведений различается также по способу наблюдений:

1. непосредственный,
2. выкопировка,
3. анкетный, опрос.

Способ непосредственного наблюдения предусматривает регистрацию сведений при непосредственном осмотре больного или здорового человека, а также при санитарно-гигиеническом обследовании какого-нибудь объекта. К непосредственному наблюдению относится запись данных проб, анализов и др.

Способ выкопировки – довольно распространенный метод в социально-гигиенических исследованиях. Например, нередко производится выкопировка сведений из историй болезни. Источником статистической информации, помимо историй болезни, могут быть и другие учетные медицинские документы (журналы, карточки, протоколы и др.), ведущиеся врачами в поликлинике, стационаре, диспансере, лаборатории. При выкопировке из этих документов берутся лишь необходимые данные, соответствующие цели исследования и предусмотренные его программой. Иногда выкопировку из документов проводят на так называемую фишку. Обычно это маленький листок плотной бумаги, куда выкопировываются два–три признака из какого-нибудь медицинского документа. Например, главному врачу срочно необходимы сведения о длительности нетрудоспособности больных, наблюдаемых разными участковыми врачами; в этом случае, выкопировываются из амбулаторных карт на фишку номер врачебного участка, диагноз, дни нетрудоспособности.

В социально-гигиенических исследованиях широко применяется **анамнестический способ** получения информации. Этот способ основан на регистрации сведений, полученных от больного или его близких родственников, их воспоминаний о событиях, которые ранее были в их жизни и могли явиться причиной заболевания.

Анамнестический метод может осуществляться двумя способами:

1. анкетным методом
2. методом опроса.

Анкетный метод предусматривает представление сведений о себе самом каждым из обследуемых лиц, которые заполняют специально

составленный бланк с вопросами – анкету. При этом способе возможны ошибки в связи с тем, что обследуемые могут не понять вопрос или сознательно исказить ответ.

Анкетный метод целесообразно применять в тех случаях, когда необходимо получить объективные сведения по глубоко личным, интимным сторонам жизни (взаимоотношения в семье, на работе, половой вопрос, данные об абортах и т.д.). Анкета в этих случаях должна быть, как правило, без указаний фамилии, т.е. анонимная.

Методом опроса сведения получают при беседе врача с обследуемым лицом по определенному кругу вопросов. Использование того или другого способа зависит от цели и содержания работы.

Второй этап статистического исследования – сбор материала. Этот этап исследования заключается в сборе материала, в осуществлении статистического наблюдения.

Под статистическим наблюдением понимается сам процесс регистрации, заполнения разработанных карт (программа сбора материала). Собираение материала (статистическое наблюдение) проводят строго по заранее составленной программе и плану исследования.

Третий этап статистического исследования – разработка материала. После того как материал собран, приступают к его разработке, которая складывается из нескольких последовательных операций (стадий):

- 1) контроль;
- 2) шифровка (кодирование);
- 3) раскладка карт по группам для подсчета;
- 4) составление сводки;
- 5) вычисление статистических показателей;
- 6) графическое изображение показателей.

Перед разработкой необходимо провести предварительный контроль за правильностью и полнотой сведений.

Шифровка (кодирование) проводится путем проставления условного знака, обозначающего каждый признак и его градацию. Шифровка проводится с целью облегчения дальнейших операций по разработке материала. Шифровать надо крупными яркими цифрами или буквами.

Шифровка может производиться одновременно с регистрацией данных на стадии сбора материала.

Четвертый этап – анализ, выводы, предложения, внедрение в практику. Он заключается в осмысливании полученных результатов, показателей, сопоставлении данных, выявлении закономерностей, в обобщении результатов.

Заканчивая исследование, необходимо представить кратко и четко сформулированные выводы и конкретные предложения по внедрению его результатов в практику здравоохранения.

Внедрение результатов в практику проводится путем написания статей, разработки методических рекомендаций и информационных писем, а также путем непосредственного внедрения нового метода лечения или диагностики, или же нового препарата в практическое здравоохранение.

ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

Статистическая совокупность характеризуется, как известно, качественными и количественными признаками.

Чтобы проанализировать ту или иную совокупность, необходимо сгруппировать отдельные варианты и представить группировку в виде таблицы или ряда. Наиболее проста группировка при качественной вариации (альтернативном распределении).

Например, распределение количества заболеваний по отдельным нозологическим формам болезней (диагнозам).

Название болезней	Количество
Гипертония	120
Инфаркт	10
Ангина	450
Грипп	150
Прочие болезни	1000
Итого:	1630

Или распределение детей 7 лет по уровню физического развития

Уровень физического развития	Количество детей
Выше нормы	15
Нормальное	85
Ниже нормы	20
Итого:	120

Как видно, в таблицах представлены суммы единиц наблюдения в абсолютных числах. Абсолютные числа имеют ограниченное познавательное значение, хотя сами по себе несут важную информацию о размере того иного явления (количество больных, умерших, родившихся). Ими можно пользоваться для характеристики редко встречающихся явлений (1 случай холеры, чумы, 5 случаев дифтерии). Основной недостаток абсолютных величин в том, что их нельзя применять для сравнения, то есть нельзя сравнивать размеры двух явлений или изучать изменения этих явлений во времени. Для этого необходимо все абсолютные числа, выражающие эти размеры, привести как бы к одному знаменателю.

Различают 4 вида относительных величин:

1. Интенсивный показатель.
2. Экстенсивный показатель.
3. Показатель соотношения.
4. Показатель наглядности.

Интенсивный показатель — или показатель частоты, распространенности, характеризует частоту явления во взаимосвязанной среде, которая сама продуцирует данное явление, т.е. показывает распространенность явления во

взаимосвязанной среде, которая сама продуцирует данное явление. Например, больные – среда, а число умерших – явление; студенты – среда, а заболевшие гриппом – явление.

Интенсивный показатель рассчитывается по следующей формуле:

Интенсивный показатель = явление х основание / величина среды, продуцирующая данное явление

где: **среда** – это все население или его определенная группа (мужчины, женщины, больные, работающие и т.д.);

основание – это единица с нулями (100, 1000, 10000, 100000 и т.д.). Основание применяется только для наглядности показателя, и чем меньше явление, тем больше основание. Если показатель рассчитывается на 100 населения, то он измеряется в процентах (%), если на 1000 населения – в промиллях (‰), на 10000 – в продецемиллях (‱).

явление – это процессы, протекающие среди населения (число умерших, родившихся, больных и т.д.)

Интенсивные показатели делятся на:

1. общие
2. специальные

Общие – характеризуют явление в целом.

Специальные – дают углубленный дифференцированный анализ.

Например: число мужчин в районе 30000, из них умерло в течение года 300. В этом случае интенсивный показатель или в данном случае показатель смертности среди мужчин составит:

$$\begin{array}{r} 30000 - 300 \\ 1000 - X, \\ \text{отсюда: } X = 300 \times 1000 / 30000 = 10 \text{ ‰ (промиллей)} \end{array}$$

То есть на каждые 1000 мужчин приходится 10 случаев смерти.

Экстенсивный показатель – это показатель распределения, который характеризует отношение части к целому и выражается, как правило, в процентах (реже в промиллях) к итогу.

Экстенсивный показатель = часть явления х 100 / целое явление

Экстенсивные показатели нужны для определения структуры статистической совокупности и сравнительной оценки, соотношения составляющих её частей (например, лейкоцитарная формула, структура заболеваемости, смертности по полу, возрасту и т.д.).

Сумма экстенсивных показателей всегда равна составит 100% (если показатели вычислялись на 100).

Например: при проведении медицинского осмотра детей в детском саду выявлено 30 случаев заболеваний, из них ОРЗ – 15, гепатит – 2, ветрянка – 5, другие заболевания – 8.

В этом случае экстенсивный показатель будет рассчитан следующим образом:

$$\begin{aligned} &\text{Для ОРЗ} \\ &30 - 100\% \\ &15 - X, \text{ т.е.} \\ &X = 15 \times 100 / 30 = 50\% \end{aligned}$$

$$\begin{aligned} &\text{Для гепатита} \\ &30 - 100\% \\ &2 - X, \text{ т.е.} \\ &X = 2 \times 100 / 30 = 6,7\% \end{aligned}$$

$$\begin{aligned} &\text{Для ветрянки} \\ &30 - 100\% \\ &5 - X, \text{ т.е.} \\ &X = 5 \times 100 / 30 = 16,7\% \end{aligned}$$

$$\begin{aligned} &\text{Другие заболевания} \\ &30 - 100\% \\ &8 - X, \text{ т.е.} \\ &X = 8 \times 100 / 30 = 26,6\% \end{aligned}$$

$$\text{Всего: } 50 + 6,7 + 16,7 + 26,6 = 100\%$$

Показатель соотношения – характеризует численное соотношение двух, не связанных между собой совокупностей, сопоставимых только логически, по их содержанию и соответственно показывает распространенность явления в среде, которая его не продуцирует. По методике вычисления этот коэффициент сходен с методикой вычисления интенсивного показателя, но различен по существу.

Показатель соотношения = явление х основание / среда, не продуцирующая данное явление

Например: обеспеченность населения врачами, койками, лечебно-профилактическими учреждениями и т.д.

Пример расчета: в городе с численностью 400000 населения работают 100 врачей, в этом случае обеспеченность населения врачами составит:

$$\begin{aligned} &400000 - 100 \\ &10000 - X, \\ &\text{т.е. } X = 100 \times 10000 / 400000 = 25,0\% \text{ (продецемиллей), или} \end{aligned}$$

говоря более простым языком, на каждые 10000 населения приходится 25 врачей.

Показатель наглядности. Показатели наглядности применяются с целью более наглядного и доступного сравнения абсолютных, относительных или средних величин. Чаще всего показатели наглядности используют при сравнении данных в динамике, чтобы в более наглядной форме представить закономерности изучаемых явлений во времени. Показатели наглядности показывают, на сколько процентов или во сколько раз произошло увеличение или уменьшение сравниваемых величин. При этом одна из сравниваемых величин приравнивается к 100, или 1 а остальные пересчитываются по отношению к этой величине.

Например:

Заболеваемость студентов ВУЗов гастритом (по данным осмотров)

Курс обучения	Заболеваемость в %	Показатель наглядности
1	6,0	100% или 1
2	12,0	200% или в 2 раза
3	15,0	250% или в 2,5 раза
4	10,0	166% или в 1,6 раза
5	18,0	300% или в 3 раза

Методика расчета:

Для студентов 2-го курса

6 – 100

12 – X, $X = 12 \times 100 / 6 = 200\%$,

т.е. на 2-м курсе заболеваемость студентов на 100% выше ($200 - 100 = 100$) по сравнению с заболеваемостью студентов на первом курсе.

Для студентов 3-го курса

6 – 100

15 – X, $X = 15 \times 100 / 6 = 250\%$,

т.е. на 3-м курсе заболеваемость студентов на 150% выше ($250 - 100 = 150$) по сравнению с заболеваемостью студентов на первом курсе.

Для студентов 4-курса

6 – 100

10 – X, $X = 10 \times 100 / 6 = 166\%$,

т.е. на 4-м курсе заболеваемость студентов на 66% выше ($166 - 100 = 66$) по сравнению с заболеваемостью студентов на первом курсе.

Для студентов 5-курса

6 – 100

18 – X, $X = 18 \times 100 / 6 = 300\%$,

т.е. на 5-м курсе заболеваемость студентов на 200% выше ($300 - 100 = 200$) по сравнению с заболеваемостью студентов на первом курсе.

ГРАФИЧЕСКИЕ ИЗОБРАЖЕНИЯ

Результаты статистического исследования обычно представляют в виде одного или нескольких рядов чисел, (абсолютных, относительных, средних), сведенных в статистические таблицы.

Для большей наглядности и лучшего усвоения эти результаты можно представить в виде различных графических изображений.

Графические изображения служат для наглядного представления статистических величин, позволяют глубже их проанализировать.

Различают следующие основные виды графических изображений:

1. Диаграммы (плоскостные, объемные).
2. Картограммы.
3. Картодиаграммы

При построении любого вида графических изображений следует соблюдать следующие правила:

- каждое графическое изображение должно иметь название (обычно ставится под изображением), в котором указывается его содержание, время и место;
- должно строится по определенному масштабу;
- для каждого графического изображения должны даваться пояснения (в виде условных обозначений) о примененной расцветке или штриховке;
- изображение должно строго соответствовать сущности изображаемых показателей.

ДИАГРАММЫ

Все диаграммы подразделяются на объемные и плоскостные. Практически любой вид диаграмм можно изобразить на плоскости (в одном измерении) или в объемном виде (в трех измерениях). Таким образом, объемные диаграммы отличаются от плоскостных только своим видом.

Кроме того, по характеру диаграммы делятся на:

1. линейные
2. столбиковые
3. ленточные
4. радиальные
5. секторные
6. внутрестолбиковые
7. внутреленточные
8. фигурные

Линейные диаграммы применяются для иллюстрации динамики изменения явления во времени.

Основой для её построения является прямоугольная система. На оси абсцисс «X» откладываются равные по масштабу промежутки времени, а по

оси ординат «У» - показатели явления (заболеваемость, численность населения и т.д.).

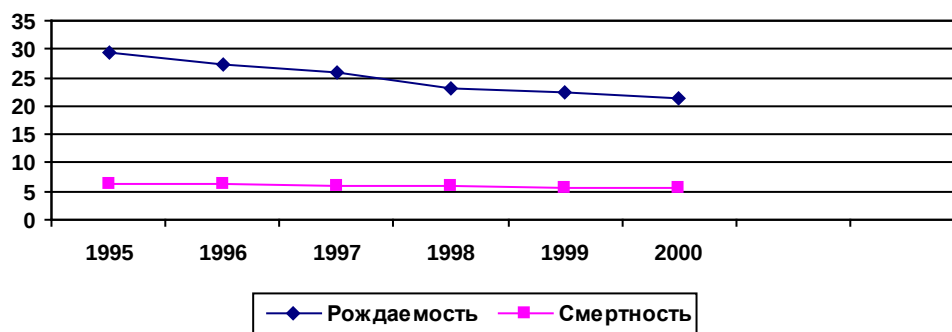
В тех случаях, когда на одной диаграмме изображают несколько явлений, линии наносят разного цвета или разной штриховки.

Типичным примером линейной диаграммы является температурная кривая, изменение уровней рождаемости, смертности и т.д.

По характеру линейные диаграммы бывают:

а) прямолинейные, б) восходящие, в) нисходящие, г) криволинейные

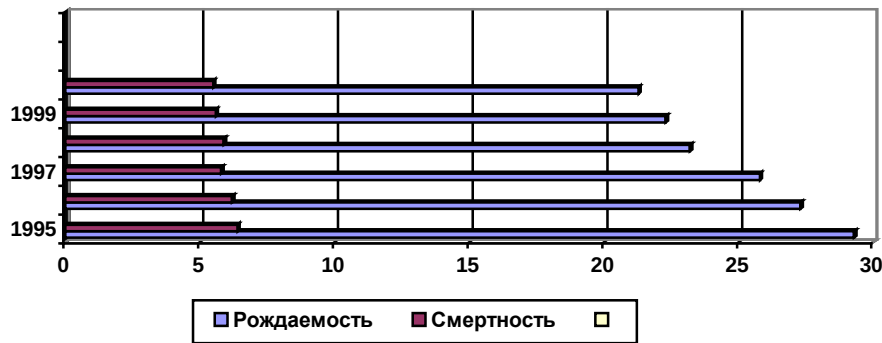
Например: изменение уровня рождаемости и смертности



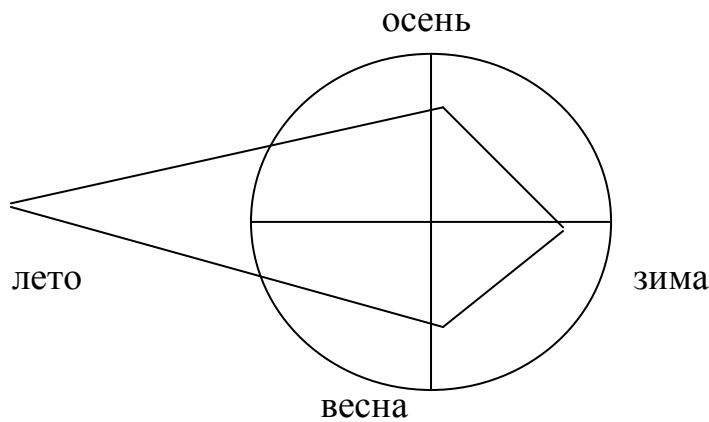
Столбиковые диаграммы применяются для иллюстрации однородных, но не взаимосвязанных между собой показателей (интенсивные, соотношения). Чаще применяются для изображения однородных показателей в двух и более сравниваемых совокупностях, но также могут отображать и динамику явления. Например: обеспеченность населения врачами и амбулаторно-поликлиническими учреждениями на 10000 населения можно изобразить следующим образом



Наряду со столбиковыми также применяются **ленточные диаграммы**. Основное их отличие от столбиковых диаграмм заключается в расположении. Так, если столбиковые диаграммы располагаются по вертикали, то ленточные располагаются по горизонтали. Условия же их использования такие же как и для столбиковых диаграмм.

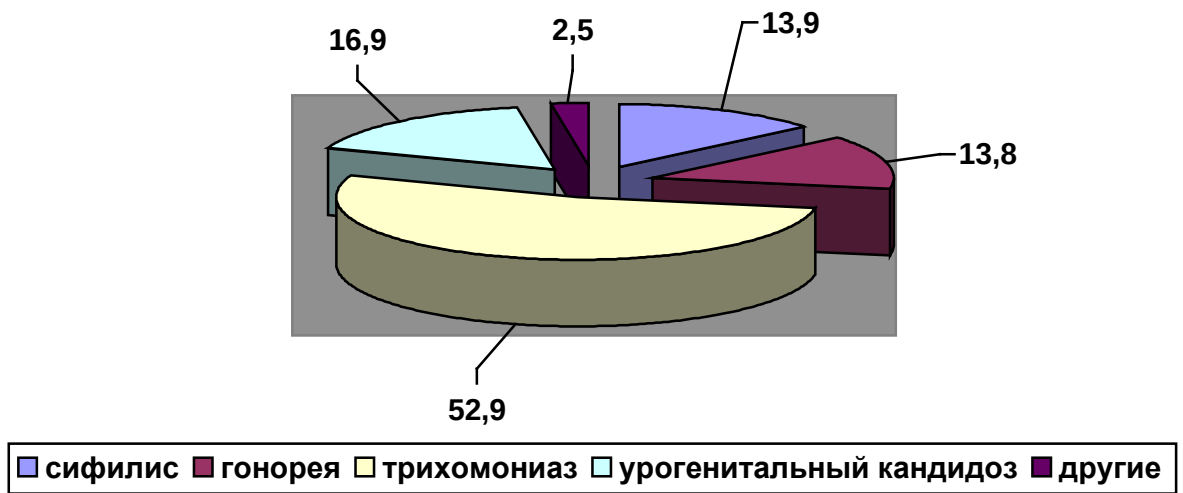


Радиальная диаграмма является разновидностью диаграммы построенной на полярных координатах. Её используют, когда надо изобразить изменение явления за замкнутый период времени (сутки, неделя, год). При построении радиальной диаграммы в качестве оси абсцисс используют окружность, разделенную на одинаковое число частей соответственно отрезкам времени того или иного цикла. Осью ординат служит радиус окружности или ее продолжение. Обычно за радиус окружности принято брать среднюю величину явления анализируемого цикла времени. Количество радиусов соответствует интервалам времени изучаемого цикла (12-радиусов – год, 7 неделя, 24 – часы). Например: изучение распространенности ОКЗ среди детей по сезонам года можно изобразить следующим образом



Для изображения структуры явления применяются **секторные диаграммы**, где в круге каждая её часть занимает соответствующий сектор. При этом радиус круга принимается равным 100%, а 1% явления равен 3,6 градусов.

Например: структура заболеваний, передающихся половым путем за 2000 год сифилис – 13,9%; гонорея – 13,8%; трихомониаз – 52,9%; урогенитальный кандидоз – 16,9%; другие – 2,5%

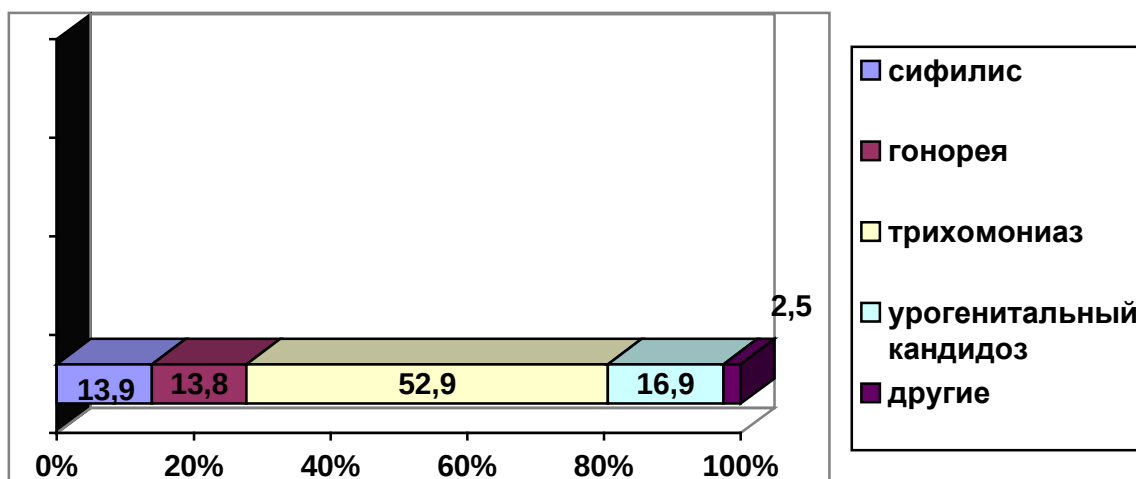


Для изображения структуры явления также применяются **внутристолбиковые и внутриленточные диаграммы**. При этом площадь столбика или ленты применяется за 100%, а каждому 1% явления соответствует 1 см.

Пример внутристолбиковой диаграммы: структура заболеваний, передающихся половым путем за 2000 год
сифилис – 13,9%; гонорея – 13,8%; трихомониаз – 52,9%; уrogenитальный кандидоз – 16,9%; другие – 2,5%



Пример внутриленточной диаграммы: структура заболеваний, передающихся половым путем за 2000 год
сифилис – 13,9%; гонорея – 13,8%; трихомониаз – 52,9%; уrogenитальный кандидоз – 16,9%; другие – 2,5%



Фигурные диаграммы применяются для изображения явления, как в сравниваемых совокупностях, так и в динамике. При этом изучаемое явление изображается в виде условных фигурок.

Например: изменение численности населения можно изобразить в виде человечков.

Картограмма представляет собой географическую карту (или схему карты) на которой отдельные территории заштрихованы или закрашиваются с различной интенсивностью в соответствии с уровнем показателей.

Для ее составления нужна географическая карта и для каждой группы показателей необходимо дать условную штриховку или цвет (степень окраски).

Картодиаграмма – представляет собой картограмму на которой наносятся также и диаграммы. Например: синоптическая карта, используемая для прогноза погоды; карта полезных ископаемых в географии и т.д.

Для изображения интенсивных показателей, показателей соотношения и наглядности применяются линейные, радиальные, столбиковые, ленточные и фигурных диаграммы, картограмма, картодиаграмма.

Экстенсивные показатели выражаются в виде секторных, внутрестолбиковых или внутриленточных диаграмм.

ДИНАМИЧЕСКИЕ РЯДЫ И ИХ АНАЛИЗ

Важной задачей медицины и здравоохранения является изучение здоровья населения, а также характера и объема деятельности лечебно-профилактических учреждений с учетом их изменения во времени. Для того чтобы проанализировать динамику того или иного процесса, необходимо уметь сопоставить динамические ряды разных типов, уметь их выравнять и анализировать.

Динамическим рядом называется ряд, состоящий из однородных сопоставимых величин, характеризующих изменение какого-либо явления за определенный промежуток времени.

Значения динамического ряда называют уровнями ряда. Уровни ряда могут быть представлены абсолютными, относительными или средними величинами.

Динамические ряды могут быть простыми и сложными. Простой динамический ряд состоит из абсолютных величин. Сложный динамический ряд состоит из относительных или средних величин.

В свою очередь простой динамический ряд может быть двух типов: моментный и интервальный.

Моментный ряд состоит из величин, характеризующих размеры явления на определенные даты – моменты уровни моментного ряда не подлежат дроблению.

Например: число населения на 1 января 2004 года.

Интервальный ряд – состоит из величин, характеризующих размеры явления за определенный интервал времени (сутки, неделя, декада, месяц, год).

Интервальный ряд в отличие от моментного ряда можно разделить на более мелкие периоды или укрупнить интервалы.

Например, в 1978 г. число родившихся было 1120, а за квартал в районе число родившихся составило примерно в 4 раза меньше – $1120:4 = 280$ детей.

Интервальные ряды могут состоять не только из чисел родившихся, но и из чисел умерших, заболевших и др., т.е. представляют изменение явления за те или иные промежутки времени.

Выбор интервала периода для интервального ряда (год, месяц, неделя, день, час и т.д.) в известной мере определяется степенью изменчивости явления (смертность, заболеваемость, рождаемость и т.д.). Чем медленнее изменяется явление во времени, тем крупнее могут быть периоды наблюдения.

Для анализа динамических рядов используют следующие показатели:

1. абсолютный прирост (или убыль),
2. темп прироста (убыли),
3. темп роста
4. значение 1 % прироста
5. показатель наглядности

Абсолютный прирост (убыль) – это разность уровней данного и предыдущего периодов.

Темп прироста (убыли) – процентное отношение абсолютного прироста к предыдущему уровню.

Темп роста (убыли) – процентное отношение последующего уровня к предыдущему уровню.

Значение 1% прироста – отношение абсолютного прироста к темпу роста.

Например, на 1 января соответствующего года число населения в городе составило (в тыс. человек):

2000 г. – 6956

2001 г. – 7151

2002 г. – 7409

2003 г. – 7530

2004 г. – 7745

Методика расчета и анализа динамического ряда:

Абсолютный прирост: для 2001 года составил: $7151 - 6956 = 195$

для 2002 года составил: $7409 - 7151 = 258$

для 2003 года составил: $7530 - 7409 = 121$

для 2004 года составил: $7745 - 7530 = 215$

Темп прироста: для 2001 года составил: $195 \times 100 / 6956 = 2,8\%$

для 2002 года составил: $258 \times 100 / 7151 = 3,6\%$

для 2003 года составил: $121 \times 100 / 7409 = 1,6\%$

для 2004 года составил: $215 \times 100 / 7530 = 2,9\%$

Темп роста: для 2001 года составил: $7151 \times 100 / 6956 = 102,8 \%$

для 2002 года составил: $7409 \times 100 / 7151 = 103,6\%$

для 2003 года составил: $7530 \times 100 / 7409 = 101,6\%$

для 2004 года составил: $7745 \times 100 / 7530 = 102,9\%$

Значение 1% прироста:

для 2001 года составил: $195 / 102,8 = 1,89$

для 2002 года составил: $258 / 103,6 = 2,49$

для 2003 года составил: $121 / 101,6 = 1,19$

для 2004 года составил: $215 / 102,9 = 2,08$

Динамический ряд не всегда состоит из уровней, последовательно изменяющихся в сторону снижения или увеличения. Нередко некоторые уровни в динамическом ряду представляют значительные колебания, что затрудняет возможность проследить основную закономерность, свойственную явлению в наблюдаемый период.

В этих случаях для выявления общей динамической тенденции рекомендуется произвести **выравнивание ряда**.

Существует несколько способов выравнивания динамического ряда:

1. укрупнение интервала
2. сглаживание ряда при помощи групповой

3. сглаживание ряда при помощи скользящей средней

Однако выравнивание уровней динамических рядов необходимо осуществлять только после глубокого и всестороннего анализа причин, обусловивших колебания этих уровней. Механическое выравнивание может искусственно сгладить уровни и завуалировать причинно-следственные связи.

Укрупнение интервала производят путем суммирования данных за ряд смежных периодов.

Например, помесечные числа заболеваний ангиной то увеличиваются, то уменьшаются. После укрупнения интервалов по кварталам года можно увидеть определенную закономерность, наибольшее число заболеваний приходится на летне-осенний период.

Вычисление групповой средней для каждого укрупненного периода производят путем суммирования смежных уровней соседних периодов, а затем полученную сумму делят на число слагаемых.

Выравнивание ряда путем вычисления групповой средней позволяет получать данные, иллюстрирующие довольно четкую тенденцию к постепенному изменению явления за определенный промежуток времени.

Вычисление скользящей средней позволяет каждый уровень заменить на среднюю величину из данного уровня и двух соседних с ним.

Ряд, выравненный при помощи скользящей средней, представляет последовательную тенденцию изменения явления.

То есть, скользящая средняя является простейшим способом выравнивания ряда. Этот метод дает возможность сгладить, устранить резкие колебания динамического ряда.

Например, на 1 января соответствующего года число населения в городе составило (в тыс. человек):

2000 г. – 7409

2001 г. – 7151

2002 г. – 6956

2003 г. – 7745

2004 г. – 7530

Как видно из приведенных данных они меняются волнообразно, поэтому прежде чем проводить анализ динамического ряда необходимо его выровнять.

Укрупнение интервала можно провести следующим образом:

Для 2000-2002 гг. – $7409 + 7151 + 6956 = 21516$

Для 2003-2004 гг. – $7745 + 7530 = 15275$

Вычисление групповой средней:

Для 2000-2002 гг. – $(7409 + 7151 + 6956) / 3 = 7172$

Для 2003-2004 гг. – $(7745 + 7530) / 2 = 7637,5$

Вычисление скользящей средней:

Для 2001 года: $(7409 + 7151 + 6956) / 3 = 7172$

Для 2002 года: $(7151 + 6956 + 7745) / 3 = 7284$

Для 2003 года: $(6956 + 7745 + 7530) / 3 = 7410,3$

СТАНДАРТИЗАЦИЯ

В практической деятельности врача нередко приходится сравнивать статистические показатели, вычисленные в качественно неоднородных по составу групп (по полу, возрасту, тяжести поступления больных в стационар). Необходимо иметь в виду, что на общие показатели, характеризующие уровень может оказать влияние неоднородность составов сравниваемых совокупностей. Так, например, чтобы сопоставить общие уровни летальности (число умерших на 100 больных) по двум больницам и сделать вывод о причинах различий этих показателей, необходимо, прежде всего, проанализировать, однороден ли по нозологическим формам состав больных, лечившихся в этих больницах. Безусловно, общий показатель летальности будет выше в той больнице, где в составе госпитализированных больных было больше лиц с тяжелыми хроническими заболеваниями. Наличие разного состава больных в этих больницах не позволяет делать правильные выводы о причинах различия показателей летальности.

Например, рождаемость, будет выше там, где среди населения больше лиц молодого возраста. А смертность будет выше там, где больше детей до 1 года и лиц пожилого возраста.

Устранить влияние неоднородности состава при сравнении показателей позволяет применение метода стандартизации.

Стандартизация – это метод расчета условных (стандартизованных) показателей, заменяющих общие интенсивные (или средние) величины в тех случаях, когда их сравнения затруднено из-за несопоставимости состава сравниваемых групп.

Стандартизованные показатели являются условными величинами и применяются в целях сравнения и отнюдь не служат для измерения фактического уровня какого-либо явления.

В статистике существует 3 метода стандартизации:

1. простой
2. косвенный
3. обратный

Выбор метода стандартизации зависит от имеющихся данных.

Прямой метод стандартизации является наиболее распространенным и применяется в том случае, когда исследователю известны, как структура сравниваемых совокупностей (по полу, возрасту, стажу, профессии и т.д.), так и количество родившихся, умерших, заболевших (по полу, возрасту, стажу, профессии) и т.д. Проще говоря когда имеются все данные как о числе наблюдений, так и структуре явления.

Косвенный метод применяется когда у исследователя отсутствуют сведения о возрастном, профессиональном составе больных, умерших и т.д., но имеются данные, т.е. имеются данные о числе наблюдений, но нет данных о структуре явления.

Обратный метод применяется когда у исследователя нет сведений о числе наблюдений, но имеются сведения о структуре явления.

Так как в социально-гигиенических и клинических исследованиях мы имеем все интересующие нас данные, как об объеме совокупности, так и структуре явления, то обычно применяется прямой метод стандартизации, предусматривающий возможность вычисления общих и погрупповых интенсивных, а также стандартизованных показателей.

Прямой метод стандартизации состоит из 5-ти этапов:

- 1 этап – расчет общих и специальных интенсивных показателей для двух сравниваемых совокупностей;
- 2 этап – выбор и расчет стандарта
- 3 этап – вычисление ожидаемых чисел относительно стандарта
- 4 этап – вычисление стандартизованных показателей,
- 5 этап – сравнение стандартизованных показателей с интенсивными, выводы.

Пример: На основе приведенных данных сделайте заключение о влиянии состава больных на показатель летальности в двух больницах.

Распределение больных и умерших по двум больницам

Отделение	Больница А		Больница Б	
	Число прошедших больных	Число умерших больных	Число прошедших больных	Число умерших больных
Терапевтическое	600	30	200	12
Хирургическое	300	6	700	21
Инфекционное	100	4	100	5
ВСЕГО	1000	40	1000	38

Так как нам необходимо сделать заключение о влиянии состава больных на показатель летальности, то необходимо рассчитать стандартизованные показатели. Нам известны сведения о числе больных (объем совокупности) и структуре явления (распределение больных по отделениям), поэтому мы можем воспользоваться прямым методом стандартизации.

Ход вычисления:

1 этап – вычисление интенсивных показателей. В данном случае мы рассчитываем показатель общей летальности и по отделениям в двух больницах.

Больница А

Для терапевтического отделения

$$600 - 30$$

$$100 - X$$

$$X = 30 \times 100 / 600 = 5\%$$

Больница Б

$$200 - 12$$

$$100 - X$$

$$X = 12 \times 100 / 200 = 6\%$$

Для хирургического отделения

$$300 - 6$$

$$100 - X$$

$$X = 6 \times 100 / 300 = 2\%$$

$$700 - 21$$

$$100 - X$$

$$X = 21 \times 100 / 700 = 3\%$$

Для инфекционного отделения

$$100 - 4$$

$$X = 4\%$$

$$100 - 5$$

$$X = 5\%$$

Общий показатель летальности

$$1000 - 40$$

$$100 - X$$

$$X = 4\%$$

$$1000 - 38$$

$$100 - X$$

$$X = 3,8\%$$

Показатели летальности в сравниваемых больницах (в %)

Отделение	Больница А	Больница Б
Терапевтическое	5,0	6,0
Хирургическое	2,0	3,0
инфекционное	4,0	5,0
Общий показатель	4,0	3,8

При сравнении показателя летальности в 2-х больницах по каждому отделению видно, что летальность в отделениях больницы Б выше, чем в больнице А, а общий показатель летальности для всей больницы А выше, чем в больнице Б.

Чем же объяснить такое различие?

Возможно, это зависит от неоднородности состава прошедших больных по отделениям. В больнице Б больше хирургических больных, у которых летальность наименьшая. В больнице А больше терапевтических больных, у которых летальность выше, по сравнению с хирургическими больными. Поэтому необходимо показать какова была бы летальность в обеих больницах, если бы число прошедших больных по отделениям было бы одинаковым. Для этого необходимо выбрать стандарт и рассчитать стандартизованные показатели относительно него.

2 этап – выбор и расчет стандарта.

За стандарт можно принять сумму или полусумму сравниваемых совокупностей, одну из сравниваемых совокупностей, или любую отвлеченную совокупность.

В нашем примере за стандарт берем полусумму прошедших больных по отделениям больниц А и Б

Расчет стандарта:

Всего по двум больницам полусумма больных составляет 1000 человек.

$$1. (1000 + 1000) / 2 = 1000$$

2. Терапевтическое отделение

$$(600 + 200) / 2 = 400$$

3. Хирургическое отделение

$$(300 + 700) / 2 = 500$$

4. Инфекционное отделение

$$(100 + 100) / 2 = 100$$

5. ВСЕГО

$$400 + 500 + 100 = 1000$$

Далее вычисляем ожидаемые числа относительно данного стандарта исходя из интенсивных показателей.

3 этап – вычисление стандартизованных показателей, т.е. ожидаемых показателей летальности по стандарту.

Терапевтическое отделение

В терапевтическом отделении больницы А летальность составила 5%, т.е. из 100 больных умерло 5. Сколько умерло бы, если число прошедших больных составляло 400 (стандарт)?

Составляем пропорцию: $100 - 5$

$$400 - X$$

$$X = 400 \times 5 / 100 = 20$$

Аналогично рассчитывается для терапевтического отделения больницы Б.

$$100 - 6$$

$$400 - X$$

$$X = 400 \times 6 / 100 = 24$$

Хирургическое отделение

В хирургическом отделении больницы А летальность составила 4%, т.е. из 100 больных умерло 2. Сколько умерло бы, если число прошедших больных составляло 500 (стандарт)?

Составляем пропорцию: $100 - 2$

$$500 - X$$

$$X = 500 \times 2 / 100 = 10$$

Аналогично рассчитывается для инфекционного отделения больницы Б.

$$100 - 3$$

$$500 - X$$

$$X = 500 \times 3 / 100 = 15$$

Инфекционное отделение

В инфекционном отделении больницы А летальность составила 4%, т.е. из 100 больных умерло 4. Сколько умерло бы, если число прошедших больных составляло 100 (стандарт)?

Составляем пропорцию: $100 - 4$, $X = 4$

Аналогично рассчитывается для инфекционного отделения больницы Б.

$$100 - 5, X = 5$$

Все результаты получаются в абсолютных числах.

Ожидаемые числа умерших в больнице А и Б (в абсолютных числах)

Отделения	Ожидаемые числа умерших в больнице А	Ожидаемые числа умерших в больнице Б
терапевтическое	20	24
хирургическое	10	15
инфекционное	4	5
ВСЕГО	$34 = (20+10+4)$	$44 = (24+15+5)$

4 этап – вычисление стандартизованных показателей

Согласно нашим расчетам в целом по больнице А умерло бы 34 человека из 1000 (стандарт), соответственно на 100 больных показатель летальности составит:

$$1000 - 34$$

$$100 - X$$

$$X = 34 \times 100 / 1000 = 3,4\%$$

В больнице Б умерло бы 44 на 1000, соответственно 100 показатель составит:

$$1000 - 44$$

$$X - 100$$

$$X = 44 \times 100 / 1000 = 4,4\%$$

5 этап – сравнение стандартизованных показателей с интенсивными и выводы

Для этого полученные данные необходимо представить в виде таблицы

Показатели	Больница А	Больница Б	Результаты сравнения
Стандартизованные	3,4%	4,4%	А < Б

Общие интенсивные	4%	3,8%	$A > B$
----------------------	----	------	---------

Вывод: сравнение стандартизованных показателей летальности в больницах А и Б, позволяет сделать заключение, что при одинаковом составе прошедших больных летальность была бы выше в больнице Б, чем в А.

При сравнении общих интенсивных показателей летальности результаты были противоположными (в $A > B$) в связи с тем, что на общие интенсивные показатели оказал влияние различный состав прошедших больных по отделениям. Более высокий показатель летальности в больнице А объясняется преобладанием в ней больных терапевтического профиля, имеющих самую высокую летальность, а более низкий показатель летальности в больнице Б обусловлен преобладанием в ней больных хирургического профиля, имеющих самую низкую летальность.

СРЕДНИЕ ВЕЛИЧИНЫ

В научных экспериментальных и клинических исследованиях для оценки общей ситуации, сравнения общих уровней в различных статистических группах находят широкое применение средние величины.

Средняя величина – это, величина, которая одним числом характеризует всю совокупность в целом. Средняя величина – есть обобщение качественно однородных, но количественно отличающихся друг от друга величин. Средняя величина как бы выражает то общее, что характерно для признака в данной совокупности, но отличается друг от друга по своей величине.

При использовании средних величин надо соблюдать два важнейших условия:

1. средняя величина должна быть рассчитана на основе качественно однородных статистических групп.
2. Для отражения характерного размера изучаемого признака средние величины должны быть вычислены на массовых материалах, т.е. в совокупности должно быть достаточно большое число наблюдений.

Группировка количественных признаков изучаемого явления производится путем составления, вариационных рядов.

Вариационный ряд – это ряд числовых значений изучаемого признака, отличающихся друг от друга по своей величине и расположенных в ранговом порядке (убывающем или возрастающем).

Характеристиками вариационного ряда являются:

1. Варианта (V) – числовое значение изучаемого признака,
2. частота (P) – повторяемость каждой варианты (частота встречаемости каждой варианты).
3. Общее число наблюдений (n), равное сумме частот – (P).

Вариационный ряд может быть простым, сгруппированным и интервальным.

В простом вариационном ряду каждая варианта встречается только один раз и составляется обычно при малом числе наблюдений <30

В сгруппированном вариационном ряду – варианты с одинаковой частотой встречаемости объединяются в группы, он составляется обычно при большом числе наблюдений >30 . Но также может составляться и при числе наблюдений меньше 30.

При большом числе наблюдений (более 100) для облегчения расчетов и экономии времени составляется интервальный вариационный ряд.

При составлении интервального вариационного ряда необходимо:

1.определить количество групп в ряду, которое будет зависеть от числа наблюдений, чем больше наблюдений, тем больше групп. Группы вариантов должны располагаться в определенном порядке (в восходящем или нисходящем). Количество групп определяется произвольно самим исследователем. Для удобства вычисления количество групп не должно превышать 10-12, при любом значении n .

2. Определить интервал между группами (i) используя формулу:

$$i = V_{\max} - V_{\min} / \text{число групп}$$

3.Определить границы и середину каждой группы.

4. Распределить изучаемую совокупность по группам.

Например: имеются данные о росте 56 студенток (n) 1 курса медицинского института 160, 158, 160, 163, 162, 170, 166, 165, 164, 170, 170, 168, 169, 175, 178 и т.д.

Определяем количество групп. У нас будет 7 групп, $V_{\max} = 178$, $V_{\min} = 158$, следовательно, интервал $i = 178 - 158 / 7 = 2,7$ или приравниваем к 3.

Составляем интервальный вариационный ряд:

Распределение студентов в зависимости от роста

Рост (в см)	Число студенток	Середина группы вариант
158-160	4	159
161-163	6	162
164-166	21	165
167-169	11	168
170-172	9	171
173-175	4	174
176-178	1	177

При составлении интервального ряда необходимо учесть, что интервалы в группах должны быть одинаковыми.

Таким образом, средняя величина является обобщающей характеристикой количественных признаков совокупности и рассчитывается из вариационных рядов.

В санитарной статистике различают следующие виды средних величин:

1. мода (M_o)
2. медиана (M_e)
3. средняя арифметическая (M).

Средняя величина имеет следующие свойства:

1. занимает среднее положение.
2. имеет абстрактный характер
3. сумма отклонений всех вариантов от условной средней величины равна нулю
4. средняя величина скрывает изменчивость признака и рассеянность вариационного ряда.

Мода – варианта, которая чаще других встречается в совокупности.

Медиана – варианта, занимающая в вариационном ряду срединное положение. Эта варианта делит вариационный ряд на две равные части. Для определения медианы необходимо найти середину ряда. При нечетном числе вариантов за медиану принимают варианту, имеющую срединное положение. При четном числе вариантов за медиану принимают среднюю величину двух центральных вариантов.

При сгруппированном вариационном ряде медиана вычисляется по специальной формуле.

Средняя арифметическая величина бывает простой, взвешенной и средней арифметической, рассчитанной по способу моментов.

Простая средняя арифметическая величина рассчитывается из простого вариационного ряда, в котором варианты встречаются только один раз и $n < 30$. Средняя арифметическая простая вычисляется по формуле:

$$M = \Sigma V / n, \text{ где}$$

Σ – знак суммы

V – варианты

n – число наблюдений

Например: измерен пульс у 9 человек: 65,60,61,75,70,76,62,68,63.

Составляем вариационный ряд:

V	P
60	1
61	1
62	1

63	1
65	1
68	1
70	1
75	1
76	1
	9

Так как каждая варианта встречается только один раз, то вычисляем простую среднюю арифметическую величину:

$$M = \Sigma V / n$$

В среднем получаем $M = (60+61+62+63+65+68+70+75+76) / 9 = 600 / 9 = 66,7$ ударов в минуту

Средняя арифметическая взвешенная вычисляется, когда варианты встречаются по несколько раз и число наблюдений меньше 30.

$$M = \Sigma V \times P / n, \text{ где}$$

Σ – знак суммы

V – варианты

P – частота встречаемости вариантов

n – число наблюдений

Например: измерен вес 120 девятилетних мальчиков: 24, 21, 28, 30, 32, 32, 35, 35, 30, 30, 24, 25, 36, 37, 38, 31, 29 и т.д.

Так как варианты встречаются по несколько раз, то целесообразно провести вычисление взвешенной средней арифметической величины путем составления сгруппированного вариационного ряда.

V	P	VP
21	1	21
22	1	22
23	2	46
24	3	72
25	4	100
26	8	208
27	14	378
28	26	728
29	28	696
30	15	450
31	9	279
32	4	128
33	2	66

34	1	34
35	2	70
36	1	36
37	2	74
38	1	38
Итого	124	3446

Для вычисления взвешенной средней арифметической величины применяется формула: $M = \Sigma V \times P / n$

Сначала мы перемножаем все варианты на соответствующую частоту наблюдения, находим сумму множителей, а затем дели на число наблюдений, в итоге получаем:

$M = 3446 / 124 = 27,7$ кг, т.е. средний вес девятилетних мальчиков, составляет 27,7 кг.

В тех случаях, когда варианты представлены большими числами (масса тела новорожденных, количество эритроцитов, лейкоцитов) или число наблюдений, выражено сотнями или тысячами, взвешенная средняя арифметическая вычисляется по способу моментов:

$$M = M_0 + \Sigma (d \times P) / n, \text{ где}$$

M_0 – условная средняя величина, обычно это мода

d – отклонение каждой варианты от условной средней

P – частота встречаемости вариантов

n – число наблюдений

Выражение $\Sigma (d \times P) / n$ называется моментом первой степени и показывает, на сколько отличается условная средняя величина от истинной средней.

Пример вычисления средней арифметической величины способом моментов:

Например: Измерен вес 120 девятилетних мальчиков: 24, 21, 28, 30, 32, 32, 35, 35, 30, 30, 24, 25, 36, 37, 38, 31, 29 и т.д.

Составляем вариационный ряд

V	P	d	dP	d ²	d ² P
21	1	-8	- 8	64	64
22	1	-7	- 7	49	49
23	2	-6	- 12	36	72
24	3	-5	- 15	25	75
25	4	-4	- 16	16	64
26	8	-3	- 24	9	72
27	14	-2	- 28	4	56
28	26	-1	- 26	1	26

29	28	0	0	0	0
30	15	+1	+15	1	15
31	9	+2	+18	4	36
32	4	+3	+12	9	36
33	2	+4	+8	16	32
34	1	+5	+5	25	25
35	2	+6	+12	36	72
36	1	+7	+7	49	49
37	2	+8	+16	64	128
38	1	+9	+9	81	81
Итого	$\Sigma = 124$		$\Sigma = - 34$		$\Sigma = 952$

Для расчета средней арифметической величины способом моментов мы вначале определяем условную среднюю величину, чаще всего это мода или величина, которая встречается в вариационном ряду наибольшее количество раз. В нашем примере это «29», так как это величина встречается 28 раз.

Затем определяем отклонение каждой варианты от моды. d – определяется по следующей формуле: $V - M_o$.

Таким образом, получаем, для 1-ой варианты это $21 - 29 = - 8$,

для 2-ой варианты это $22 - 29 = - 7$ и т.д.

Затем чисто арифметически находим произведение $d \times P$, d^2 и d^2P .

Подставляя полученные значения в формулу $M = M_o + \Sigma (d \times P) / n$, получаем: $M = 29 + (-34) / 124 = 29 - 0,27 = 28,73$

Внутреннюю структуру совокупности характеризует среднее квадратическое отклонение G (сигма), которая вычисляется по нижеследующим формулам:

1. для простой средней арифметической величины:

$$G = \pm \sqrt{\Sigma d^2 / n}$$

2. для взвешенной средней арифметической величины:

$$G = \pm \sqrt{\Sigma d^2 \times P / n}$$

3. для средней арифметической величины, рассчитанной по способу моментов:

$$G = \pm \sqrt{\Sigma d^2 \times P / n - (\Sigma d \times P / n)^2}$$

В статистике G – также называется моментом второй степени и используется для:

1. характеристики нормы и патологии. Для этого определяют предел или интервал между средней арифметической величиной и средним квадратическим отклонением.

Формула – средние данные (нормы) $M \pm 1G$

- данные выше средних $M + 2G$
- данные ниже средних $M - 2G$
- высокие данные $M + 3G$
- низкие данные $M - 3G$

Все значения, которые находятся за пределами $3G$, являются патологическими.

2. для определения концентрированности вариационного ряда. Для этого применяется следующее уравнение:

$M \pm 1G - 68,3\%$;

$M \pm 2G - 95,5\%$;

$M \pm 3G - 99,0\%$.

То есть, если при $M \pm 1G$ вокруг средней величины сконцентрировано 68,3% единиц наблюдения;

при $M \pm 2G - 95,5\%$ единиц наблюдения;

при $M \pm 3G - 99,0\%$ единиц наблюдения, то вариационный ряд является концентрированным. В противном случае составленный нами вариационный ряд будет рассеянным.

Рассчитаем концентрированность вариационного ряда для нашего последнего примера.

Для начала определим значение G . Так как средняя арифметическая величина рассчитана способом моментов, то G будет рассчитана по формуле:

$$G = \pm \sqrt{\frac{\sum d^2 \times P}{n} - \left(\frac{\sum d \times P}{n}\right)^2}$$

В нашем примере $G = \pm \sqrt{952 / 124 - (-34 / 124)^2} = \pm 2,8$

Всего число наблюдений (n) у нас = 124, $M = 28,73$ кг, $G = \pm 2,8$.

Находим интервал $M \pm 1 G$, т.е. $28,73 \pm 2,8$. У нас получается интервал от 25,93 ($28,73 - 2,8$) до 31,53 ($28,73 + 2,8$). При этом все число наблюдений равно 100%, а число наблюдений в данном интервале – X .

Отсюда составляем пропорцию: $124 - 100\%$

$104 - X$,

$$X = 104 \times 100 / 124 = 83,9\%.$$

Вывод: для того, чтобы вариационный ряд был концентрированным в интервале $M \pm 1 G$ должно располагаться не менее 68,3% единиц наблюдения, у нас получилось 83,9%. Значит, наш вариационный ряд является концентрированным.

ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Оценить достоверность результатов исследования это, значит, узнать, с какой вероятностью можно перенести результаты, полученные на выборочной совокупности, на всю генеральную совокупность.

Оценка достоверности необходима для того, чтобы по выборочной совокупности можно было судить о генеральной совокупности. Однако при этом надо знать несколько мы ошибаемся, взяв для изучения определенный объем (часть генеральной совокупности).

Для оценки достоверности результатов исследования необходимо:

1. вычислить ошибку репрезентативности (m)
2. определить доверительные границы
3. оценить достоверность одного показателя или разности показателей.

В отличие от ошибок методических, арифметических и логических ошибки репрезентативности не могут быть устранены, если не осуществлен переход на генеральное исследование. Например: если мы делим буханку хлеба на части, то, сколько бы мы не старались добиться точности все равно ошибка будет допущена, так как эта ошибка является независимой от исследователя. Однако ошибка может быть сведена к достаточно малой величине путем увеличения числа наблюдений (n) выборочного исследования.

Ошибка репрезентативности вычисляется по следующим формулам:

Ошибка репрезентативности относительной величины вычисляется по формуле:

$$m_p = \text{SQR } P \times q / n$$

где: P – относительная величина (показатель)

n – число наблюдений

q – величина, обратная показателю, вычисляемая следующим образом:

а) если показатель P выражен в процентах, то $q = 100 - P$

б) если показатель P выражен в промиллях (‰), то $q = 1000 - P$

в) если показатель P выражен в процепемиллях (‱), то $q = 10000 - P$ и

т.д.

Ошибка репрезентативности средней арифметической величины вычисляется по формуле:

$$m_m = G / \text{SQR } n$$

где: M – средняя арифметическая величина

G – среднее квадратическое отклонение

n – число наблюдений

При числе наблюдений меньше 30, в формулах расчета ошибок репрезентативности в знаменатель вводится поправочный коэффициент 1, поэтому формулы приобретают следующий вид:

$$m_p = \text{SQR } P \times q / n-1 \text{ или } m_m = G / \text{SQR } n-1$$

Исходя из вышесказанного каждая средняя или относительная величина, полученная на выборочной совокупности должна быть, представлена со своей ошибкой репрезентативности.

По величине ошибок определяют доверительные границы.

Доверительные границы – границы средних (или относительных) величин, выход за пределы которых вследствие случайных колебаний имеет незначительную вероятность. Доверительные границы выражаются следующими формулами:

$$P_{\text{ген}} = P_{\text{выб}} \pm t \times m_p \quad \text{или} \quad M_{\text{ген}} = M_{\text{выб}} \pm t \times m_m$$

Где: $M_{\text{ген}}$ – значение средней величины, $P_{\text{ген}}$ – значение относительной величины, полученных при изучении генеральной совокупности;

$M_{\text{выб}}$ и $P_{\text{выб}}$ – значение средней или относительной величины для выборочной совокупности;

m_p и m_m – ошибки репрезентативности выборочных показателей;

t – доверительный критерий Стьюдента (который устанавливает при планировании исследования).

Следовательно, определение доверительных границ есть определение крайних значений минимального и максимального значений для средней или относительной величины. Критерий достоверности, доверительный коэффициент (t) избирается исследователем исходя из необходимости получения результата с определенной степенью точности. Его величина зависит от числа наблюдений в данной выборочной совокупности, определяется по специальным таблицам. Данный критерий определяется по специальным таблицам. Согласно табличным данным, при $t = 1$ достоверность полученных данных составляет 68,3%, при $t = 2$ достоверность полученных данных составляет 95,5%, при $t = 3$ достоверность полученных данных составляет 99,0%,

Вероятность безошибочного прогноза – это вероятность, с которой можно утверждать, что в генеральной совокупности средняя или относительная величина будет находиться в определенных пределах.

Достоверность одного показателя определяется соответственно по формулам:

$$\text{Для относительной величины: } t = P / m_p \geq 3$$

$$\text{Для средней величины: } t = M / m_m \geq 3$$

Это означает, что если величина показателя больше своей ошибки в три и более раза, то вычисленный показатель достоверен и мы можем по нему судить о всей генеральной совокупности.

В медицине и биологии часто приходится сравнивать результаты исследований, полученные в разных группах. Например, сравнивают среднюю частоту пульса, давления, дыхания, длительность лечения в опытной и

контрольной группам. При их сравнении оценивается не только полученные данные, но и достоверность разности показателей.

Достоверность разности показателей, вычисленных при выборочном исследовании означает, что вывод о их различии может быть перенесен на соответствующие генеральные совокупности. Достоверность разности сравниваемых величин оценивается следующими формулами:

$$\text{Для относительных величин: } t = P_1 - P_2 / \text{SQR } m_1^2 + m_2^2 \geq 2$$

$$\text{Для средних величин: } t = M_1 - M_2 / \text{SQR } m_1^2 + m_2^2 \geq 2$$

Где: M_1 , M_2 , P_1 , P_2 – величины, полученные при выборочных исследованиях

m_1 и m_2 – их средние ошибки

t – критерий достоверности

Разность двух величин достоверна с вероятностью безошибочного прогноза на 95,5% при критерии достоверности $t = 2$. Для медико-биологических исследований эта степень вероятности вполне достаточна. Если же $t < 2$ степень вероятности безошибочного прогноза менее 95% при этом нельзя с достаточной уверенностью утверждать, что различия между этими величинами достоверны. В этом случае необходимо увеличить объем изучаемой совокупности. Если при увеличении числа наблюдений разность остается недостоверной, то можно сказать, что между сравниваемыми величинами нет различий.

Пример: у студентов медиков исследовали частоту пульса (в минуту) до и после сдачи экзамена. Частота пульса до экзамена составила $M_1 = 94,2$ удара в минуту, $m_1 = \pm 3,9$ удара в минуту; после экзамена $M_2 = 82,0$ удара в минуту, $m_2 = \pm 4,1$ удара в минуту.

По формуле определяем достоверность разности показателей (t).

$$t = M_1 - M_2 / \text{SQR } m_1^2 + m_2^2 \geq 2$$

$$t = 94,2 - 82 / \text{SQR } 3,9^2 + 4,1^2 = 12,2 / \text{SQR } 15,21 + 16,81 = 12,2 / 5,7 = 2,1$$

В данном случае разность между двумя средними величинами более 2, что позволяет с вероятностью безошибочного прогноза более чем на 95% утверждать, что эта разность достоверна и экзамен является фактором, действительно действующим на нервно-психическое состояние, что выражается в увеличении частоты пульса.

ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ПРИЗНАКАМИ. КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Корреляция – латинское слово, означающее взаимосвязь, взаимозависимость. Различают функциональную и корреляционную связь между признаками или явлениями.

При функциональной связи каждой отдельной величине соответствует определенная взаимосвязанная с ней другая величина, т.е. изменение одной величины строго ведет к изменению другой величины на определенное значение. Например: с изменением радиуса меняется площадь круга, или с увеличением температуры нагрева увеличивается объем тела и т.д.

При корреляционной связи одной величине соответствует несколько других, варьирующих значений взаимосвязанного признака или явления. Примером корреляционной связи может служить зависимость веса тела от роста. Если измерить вес группы детей одинакового роста, то величина веса у них будет различна, и варьировать в определенных пределах, так как на величину веса влияет не только рост человека, а также питание, состояние здоровья, нервно-психическое состояние и т.д.

Для измерения и оценки связи необходимо вычислить коэффициент корреляции, который дает представление о размере и характере связи между явлениями или признаками.

По характеру связь между признаками или явлениями может быть прямой и обратной.

При прямой связи – изменение одного признака ведет к изменению взаимосвязанного с ним другого признака в одном и том же направлении.

Например: с увеличением роста ребенка увеличивается его вес. Прямая связь обозначается знаком плюс (+).

При обратной связи – с изменением одного признака в одном направлении взаимосвязанный с ним другой признак изменяется в противоположенном направлении. Например, с увеличением эффективности проводимых профилактических прививок, число инфекционных заболеваний уменьшается; или с увеличением температуры воздуха, уменьшается число бронхитов. Коэффициент корреляции, характеризующий обратную связь, обозначается знаком (-).

Размер связи колеблется от ± 1 до 0.

По величине коэффициента корреляции различают полную, сильную, среднюю, слабую связь и отсутствие связи.

Размер связи представлен в таблице.

Размер связи	Прямая связь (+)	Обратная связь (-)
Полная (функциональная)	+ 1	- 1
Сильная	(+ 0,7) – (+ 1)	(- 1) – (- 0,7)
Средняя	(+ 0,3) – (+ 0,69)	(- 0,3) – (- 0,69)
Слабая	0 – (+ 0,29)	(- 0,29) – 0
Отсутствие связи	0	0

В статистике существует несколько методов определения коэффициента корреляции. В частности, к ним относятся:

1. метод рангов (метод Спирмена)
2. метод квадратов (метод Пирсона).

Метод Спирмена или метод ранговой корреляции – является наиболее простым методом определения связи, но менее точным. Этот способ применяется при:

1. небольшом числе наблюдений, т.е. не более 30 парных величин
2. когда нет необходимости в точных расчетах измерения размера связи, а нужны лишь ориентировочные данные,
3. когда один из признаков или оба выражены качественными характеристиками.

Коэффициент корреляции в данном случае определяется по формуле:

$$\rho = 1 - 6 \Sigma d^2 / n (n^2 - 1)$$

Где : ρ – коэффициент ранговой корреляции

n – число парных членов, признаков

d – разность рангов

Σ – знак суммы

1 и 6 – постоянные числа

Например: определите характер и размер связи между стажем работы ткачих и понижением слуха у них.

Стаж работы (X)	Понижение слуха в 100% (Y)	ранги		Разность рангов	
		X	Y	d (X-Y)	d ²
До 1 года	1,0	1	1	0	0
1 – 4 года	6,0	2	2	0	0
5 – 9 лет	10,0	3	4	- 1	1
10 – 14 лет	8,0	4	3	+1	1
15 – 19 лет	15,0	5	6	- 1	1
20 – 24 лет	12,0	6	5	+1	1
25 – 29 лет	15,0	7	6	+1	1
30 и более	20,0	8	7	0	0
Итого					5

Расположить значения в ранговом порядке – это значит определить их месторасположение в порядке убывания или возрастания.

$d = X - Y$ или разность между рангами

Коэффициент корреляции рассчитываем по формуле:

$$\rho = 1 - 6 \Sigma d^2 / n (n^2 - 1)$$

$$\rho = 1 - 6 \times 5 / 8 \times (64 - 1) = 1 - 30 / 504 = 1 - 0,06 = + 0,94$$

Вывод: Коэффициент корреляции + 0,94, указывает на прямую сильную связь, т.е. с увеличением стажа работы увеличивает показатель снижения слуха у ткачих.

Метод квадратов или метод Пирсона применяется в тех случаях, когда:

1. число парных вариантов не превышает 30
2. при необходимости определения более точного размера связи
3. если признаки имеют только количественное выражение

Коэффициент корреляции методом Пирсона вычисляется по формуле:

$$r = \Sigma d_x \times d_y / \text{SQR } \Sigma d_x^2 \times \Sigma d_y^2$$

где d_x и d_y – отклонение вариантов от своих средних величин
 r – коэффициент корреляции

Например: определить характер и размер связи между максимальным и минимальным давлением у лиц в возрасте от 25 до 45 лет.

Максимальное давление (X)	Минимальное давление (Y)	d_x	d_y	d_x^2	d_y^2	$d_x \times d_y$
126	76	-5	-18	25	324	+90
126	81	-5	-13	25	169	+65
129	81	-2	-13	4	169	+26
126	94	-5	+0	25	0	0
136	96	+5	+2	25	4	+10
136	101	+5	+7	25	49	+35
136	101	+5	+7	25	49	+35
130	102	-1	+8	1	64	-8
131	102	0	+8	0	64	0
135	104	+4	+10	16	100	+40
$\Sigma = 1311$	$\Sigma = 938$			$\Sigma = 171$	$\Sigma = 992$	$\Sigma = +293$

$$n = 10$$

Последовательность вычисления:

1. Находим простую среднюю арифметическую величину для вариационного ряда X и Y, т.е. M_x и M_y

M_x = сумма всех значений вариационного ряда X / число наблюдений
 $M_x = \Sigma V_x / n$, т.е. $M_x = 1311 / 10 = 131,1$; для облегчения дальнейших расчетов полученное значение можно округлить до 131;

M_y = сумма всех значений вариационного ряда Y / число наблюдений
 $M_y = \Sigma V_y / n$, т.е. $M_y = 938 / 10 = 93,8$; для облегчения дальнейших расчетов полученное значение можно округлить до 94;

2. Определяем отклонение – d по формуле:

$$d_x = V_x - M_x; \quad d_y = V_y - M_y$$

$$d = 126 - 131,1 = -5,1 \text{ и т.д.}$$

3. Возводим отклонение в квадрат

4. Находим сумму квадратов

5. Находим произведение отклонений

6. Вычисляем коэффициент корреляции

$$r = +293 / \text{SQR } 171 \times 992 = +0,71$$

7. Вывод коэффициент корреляции +0,71 указывает на прямую и сильную связь между уровнем максимального и минимального давления.

Для того, чтобы определить на сколько результат полученный, при выборочном исследовании отличается от результата, полученного при сплошном исследовании (генеральной совокупности) необходимо вычислить ошибку репрезентативности коэффициента корреляции (m_r)

$$m_r = \text{SQR } 1 - r^2 / n - 2$$

Для нашего примера ошибка будет равна: $m_r = \text{SQR } 1 - (0,71)^2 / 10 - 2 = 0,25$

Таким образом, коэффициент корреляции указывает на направление и силу взаимосвязи.

РЕГРЕССИЯ

Коэффициент корреляции указывает лишь на направление и силу связи между двумя переменными величинами, но не дает возможности судить о том, как количественно меняются величины одного по мере изменения другого признака. Ответ на этот вопрос дает метод расчета коэффициента регрессии.

Регрессия – это функция, позволяющая по величине одного коррелируемого (связанного) признака определить средние величины другого признака. С помощью регрессии можно определить, на какую величину изменится одна величина при изменении другой величины на единицу.

Коэффициент регрессии определяется по формуле:

$$R = r \times G_y / G_x;$$

где G – среднее квадратическое отклонение признаков,

r – коэффициент корреляции

На основе предыдущего примера из таблицы вычислим G

Как показано в таблице, нами составлен простой вариационный ряд, поэтому среднее квадратическое отклонение будет определяться по формуле:

$G = \text{SQR } E d^2 / n - 1$, т.к. число наблюдений меньше 30, то в знаменатель введена поправка 1.

$$G_x = 171 / 10 - 1 = \pm 4,4 \quad G_y = \text{SQR } 992 / 10 - 1 = \pm 10,5$$

Отсюда вытекает, что $R = 0,71 \times 10,5 / 4,4 = 1,7$

Следовательно, при изменении максимального артериального давления на 1 мм.рт.ст. минимальное давление увеличивается в среднем на 1,7 мм.рт.ст.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Боярский А.Я. Общая теория статистики. – М. – 1977. – 460 с.
2. Искандаров Т.И., Маматкулов Б. Санитария-статистик ва ижтимоий-гигиеник тадқиқот услублари. – Ташкент. – 1994. – 200 с.
3. Мерков А.М., Поляков Л.Е. Санитарная статистика. – Л. – 1974. – 432 с.
4. Серенко А.Ф., Ермакова В.В. Социальная гигиена и организация здравоохранения. – М. – 1984. – 562 с.
5. Шиган Е.Н. Статистические методы и вычислительная техника в социально-гигиенических исследованиях. – М. – 1977. – 250 с.