

РУКОВОДСТВО К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ ПО БИОСТАТИСТИКЕ

Учебное пособие
для студентов
медицинских вузов

**РУКОВОДСТВО К ПРАКТИЧЕСКИМ
ЗАНЯТИЯМ ПО БИОСТАТИСТИКЕ
(учебное пособие)**



Алматы, 2020

ББК 51.1я73
УДК 614311(075.8)
P85

Рецензенты:

доктор PhD Сейдуанова Л.Б
к.м.н. Ашуева Н.И.

Руководство к практическим занятиям по биостатистике: учебное пособие / д.м.н., профессор Джайнакбаев Н.Т., д.м.н., доцент Орақбай Л. Ж., доктор PhD Адилханова А. Н., к.м.н. Ахметова Р.Л., к.м.н. Рахметова К.У., Алманиязова К.Н., Жунусова С.К., Сальменова Т.Б. – Алматы. -2020. – 94 с.

ISBN 978-601-08-0080-9

Изложение учебного материала пособия необходимо для освоения определенных знаний, навыков и умений бакалавров общественного здравоохранения. Приведены расчеты для учебных целей, для понимания сущности методов математической статистики и корректности их применения. На практике эти расчеты можно проводить с использованием специальных компьютерных программ.

Руководство предназначено для формирования новых подходов к преподаванию и изучению дисциплины «Общественное здоровье и здравоохранение», что необходимо для подготовки бакалавров общественного здравоохранения. Оно также может быть полезно исследователям.

Утверждено и рекомендовано к изданию на заседании Ученого совета НУО «КазРосМедУниверситет» протокол № 0/3 от «15» октября 2020 г.

ISBN 978-601-08-0080-9

© Джайнакбаев Н.Т., Орақбай Л. Ж. и др.

СОДЕРЖАНИЕ

	Предисловие	4
	Основные понятия и определения биостатистики	5
Тема 1	Введение. Основы биостатистики	6
Тема 2	Медицинская статистика. Основные разделы и задачи предмета.	9
Тема 3	Статистическая совокупность. Учетные признаки.	11
Тема 4	Статистические величины и их виды.	14
Тема 5	Средний уровень признака.	17
Тема 6	Нормальное распределение и его свойства.	19
Тема 7	Организация статистического исследования.	23
Тема 8	Планирование и организация статистических исследований.	27
Тема 9	Статистические гипотезы (нулевая и альтернативная гипотеза)	31
Тема 10	Параметрический метод оценки достоверности статистических гипотез.	33
Тема 11	Непараметрический метод оценки достоверности статистических гипотез.	37
Тема 12	Метод стандартизации	43
Тема 13	Ранговый метод и метод квадратов корреляционного анализа.	47
Тема 14	Регрессионный анализ.	50
Тема 15	Дисперсионный анализ в медицине и здравоохранении.	55
	Ситуационные задачи	60
	Тестовые задания	68
	Список источников	96

ПРЕДИСЛОВИЕ

XXI век – это век доказательной медицины. Это означает, что все большую роль в медицине играют методы точных наук, в первую очередь статистики. Современная концепция научно обоснованной медицины немыслима без биометрических исследований. Математико-статистическое описание данных медицинских исследований и оценка значимости различия производных величин, характеризующих эффективность профилактических, диагностических и лечебных мероприятий и процедур являются одним из основополагающих разделов доказательной медицины.

В настоящее время Казахстан реально находится в рыночной экономике, поэтому есть насущная необходимость экономить ресурсы, как человеческие, так и материальные. Одна из наиболее затратных статей бюджета страны – это расходы на здравоохранение. Однако возможности государства в целом и отдельно взятого гражданина не безграничны, и сегодня остро встает вопрос об эффективности используемых диагностических и лечебных инструментов. Требуется строгие доказательства обоснованности их применения, а также необходимо показать, какому проценту больных они помогут и в какой степени. Но эти данные невозможно получить без помощи статистических методов анализа. Это обусловлено естественным разнообразием биологических объектов, когда фактически невозможно встретить два одинаковых организма, а также влиянием психологических факторов на функциональное состояние, субъективизмом врача-исследователя в оценке таких состояний и т.д.

Надо отметить, что сегодня ни один серьезный медицинский журнал не примет статью без статистической обработки данных, без использования алгоритмов доказательной медицины. Однако, среди множества статистических методов врачу сложно выбрать наиболее адекватный, соответствующий конкретно им набранному материалу. Неправильный подбор методов приводит к неверным выводам, что в свою очередь чревато ошибкой диагностики и лечения.

Отдельно стоит вопрос интерпретации результатов статистической обработки. Наличие большого количества компьютерных программ по статистике позволяет проводить обработку данных с использованием различных многофакторных, многомерных математико-статистических методов – дисперсионного, регрессионного, кластерного и др. Но анализ результатов этих расчетов задача нетривиальная, требующая знаний как по статистике, так из области решаемых с ее помощью проблем.

Данное пособие направлено на выработку навыков применения статистических методов при решении различных медицинских задач.

Надеемся, что оно будет способствовать пониманию целей и задач биостатистики и явится своеобразным навигатором для тех, кто решил руководствоваться в своей деятельности принципом доказательности.

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ БИОСТАТИСТИКИ

Биостатистика - отрасль знаний о статистическом анализе групповых свойств и массовых явлений в биологии и медицине

Случайной величиной называется величина, которая в результате опыта может в определенных пределах принять то или иное значение, неизвестно заранее — какое именно

Генеральная совокупность (популяция)-множество всех обследуемых объектов, объединенных общими свойствами

Выборка- часть генеральной совокупности, по результатам анализа которой делается вывод обо всей генеральной совокупности

Независимые выборки- выборки, состоящие из разных объектов, при этом значение случайной величины в одной выборке не зависит от ее значений в другой выборке.

Зависимые выборки- выборки, состоящие из одних и тех же объектов, обследованных до и после воздействия.

Под термином «**распределение случайной величины**» понимается функция, показывающая вероятность (частоту встречаемости) всех возможных значений этой случайной величины

Доверительный интервал- интервал, в котором с некоторой вероятностью находится истинное (генеральное) значение числовой характеристики случайной величины

Уровень значимости α —максимально допустимая вероятность ошибки, которую может себе позволить исследователь, отвергая нулевую гипотезу (принимая альтернативную).

Уровень достоверности p – реальная вероятность ошибки в случае принятия альтернативной гипотезы

Термин **корреляция** означает взаимосвязь между двумя случайными величинами, когда изменение одной приводит к изменению в среднем другой величины.

Регрессия- метод анализа, позволяющий получить математическую модель (функцию) взаимосвязи зависимой величины с одной или несколькими независимыми величинами.

ТЕМА 1. ВВЕДЕНИЕ. ОСНОВЫ БИОСТАТИСТИКИ.

Цель: Формирование знаний по значению биostatистики

Задачи обучения: Изучение основ биostatистики

Основные вопросы темы:

1. Определение статистики
2. Разделы медицинской статистики
3. Задачи биostatистики
4. Методы биostatистики

Статистика — общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественными особенностями. Именно разнообразием качественных особенностей объясняется то, что для количественного описания явлений используется большое число самых разных статистических величин. Статистика устанавливает соответствие между идеальным миром и представлением о реальном мире. Статистика, изучающая вопросы, связанные с медициной, гигиеной и здравоохранением, называется *медицинской или санитарной*.

Разделы медицинской статистики:

- 1) статистика общественного здоровья;
- 2) статистика здравоохранения;
- 3) статистика научных исследований, или теоретическая медицинская статистика.

Если произвести единичное измерение какого-либо показателя, характеризующего состояние отдельного человеческого организма, это еще не статистика, но если измерить этот показатель у многих людей или у одного человека многократно за какой-то промежуток времени и представить этот показатель в обобщенном виде — это уже статистический показатель. Применяющиеся врачами цифровые данные, характеризующие нормальную функцию организма (нормальная частота пульса, артериального давления и т. д.), являются статистическими, так как они определены у большой группы здоровых людей. Следовательно, статистические данные — это результаты массовых исследований, характеризующие совокупность явлений, процессов, людей.

Статистика включает в себя общую теорию статистики, экономическую статистику и различные отраслевые статистики. Общая теория статистики рассматривает общие принципы и методы статистической науки. Экономическая статистика изучает с помощью статистических методов народное хозяйство в целом. Отраслевые статистики изучают различные отрасли народного хозяйства. Существует промышленная, торговая, сельскохозяйственная, транспортная, коммунальная, судебная, народного образования, демографическая, медицинская статистика и т. д.

В статистике свойство объектов или явлений, которое может быть наблюдаемо или измерено, называется *признаком*.

Статистика, изучающая вопросы, связанные с биологией, медициной, фармацией, гигиеной и здравоохранением, называется *биostatистикой*.

Роль биostatистики в практической и научной работе врача, медсестры, фармацевта велика.

Биostatистика применяет различные методы: сбор данных, их обобщение, анализ и подведение итогов, основанных на полученных наблюдениях.

Статистический анализ помогает добывать информацию из данных и оценивать качество этой информации.

Задачи биostatистики:

· количественное представление биологических фактов (измерение) — это выражение свойства отдельного биологического объекта в виде числа, варианты или значения переменной;

· обобщенное описание множества фактов (статистическое оценивание) – это расчет показателей и параметров, которые полноценно характеризуют свойства множества однотипных объектов или выборки;

· поиск закономерностей (проверка статистических гипотез) – это доказательство неслучайности отличий между сравниваемыми совокупностями, объектами, зависимости их характеристик от внешних или внутренних причин.

2. Первым этапом при проведении любого статистического исследования является сбор данных об анализируемом объекте или процессе в виде конкретных значений переменных.

Сбором статистических данных называется процесс получения информации об элементах исследуемой совокупности и их свойствах. Эти данные являются предметом статистической обработки и анализа.

Вторым этапом является анализ типов данных.

Типы данных: количественные, качественные и даты (рисунок 1).



Рисунок 1. Классификация типов данных

Основные типы данных делятся на количественные и качественные.

Количественные данные в свою очередь подразделяются на дискретные (прерывные) и непрерывные.

Дискретные данные – количественные данные, которые представлены только в виде целого числа, т.е. не могут иметь дробную часть. Например: количество детей.

Непрерывные данные – это данные, которые получают при измерении на непрерывной шкале, т.е. теоретически они могут иметь дробную часть.

Например: масса тела, рост, артериальное давление и т.д.

Непрерывные данные бывают интервальными и относительными.

Интервальные данные – вид непрерывных данных, которые измеряются в абсолютных величинах, имеющих физический смысл.

Относительные данные – вид непрерывных данных, отражающих долю изменения (увеличения или уменьшения) значения признака по отношению к исходному (или к какому-либо другому) значению этого признака. Эти данные являются безразмерными величинами или выражаются в процентах.

Качественные данные – подразделяются на номинальные и порядковые.

Номинальные данные – вид качественных данных, которые отражают условные коды неизмеряемых категорий (коды диагноза).

Порядковые данные - вид качественных данных, которые отражают условную степень выраженности какого-либо признака (стадии онкологических заболеваний, степени сердечной недостаточности).

Их основное отличие от дискретных количественных данных заключается в отсутствии пропорциональной шкалы для измерения выраженности признака.

Бинарные (дихотомические) данные - особо выделяемый вид качественных данных. Признак такого типа имеет лишь два возможных значения (пол, наличие или отсутствие какого-либо заболевания).

Особым видом данных являются *даты*. Поскольку в ряде случаев бывает необходимо произвести с ними некоторые арифметические действия (вычисление абсолютного периода времени между двумя событиями по датам этих событий).

Иногда выделяют также некоторые особые подтипы данных, являющиеся частными случаями вышеперечисленных типов: ранги, очки, визуальные аналоговые шкалы, цензурированные данные.

Перед тем как проводить углубленный статистический анализ, важно провести предварительный анализ данных. На этом этапе для сжатия и систематизации набора данных используют графические методы. Это позволяет оценить особенности набора данных и выявить аномалии, т.е. выбрать для дальнейшего анализа подходящие статистические методы.

Дискретные данные могут быть представлены в виде таблицы, столбиковой диаграммы, пиктограммы, круговой диаграммы, точечного рисунка.

Непрерывные данные могут быть представлены в виде группированной выборки, гистограммы, диаграммы «стебель с листьями» или «ящик с усами», кривой Лоренца и т.д.

Смешанные данные могут быть представлены в виде диаграммы рассеяния.

Контрольные вопросы:

1. Что является предметом изучения биологической статистики?
2. Каковы основные задачи, стоящие перед биологической статистикой?
3. Назовите основные области применения биологической статистики?
4. Назовите основные типы измерительных шкал?
5. Какими способами определяется надежность и достоверность в биостатистике?

ТЕМА 2. МЕДИЦИНСКАЯ СТАТИСТИКА. ОСНОВНЫЕ РАЗДЕЛЫ И ЗАДАЧИ ПРЕДМЕТА.

Цель: Ознакомить студентов с медицинской статистикой

Задачи обучения: Изучение медицинской статистики, его разделов и задач.

Основные вопросы темы:

1. Определение медицинской статистики
2. Задачи медицинской статистики
3. Разделы медицинской статистики

Статистика и статистические методы в оценке состояния здоровья населения и обусловленности здоровья входят в отраслевую статистику и широко используются врачами в практической и научной работе.

Одной из отраслей статистики является *медицинская статистика*, которая изучает количественную сторону массовых явлений и процессов в медицине. Медицинскую статистику обычно не принято дифференцировать на медицинскую и санитарную. Однако, по нашему мнению, более правильным было бы рассматривать санитарную статистику как один из разделов медицинской статистики. Кроме того, следует выделять статистический метод, применяемый в различного рода медицинских исследованиях (клинических, экспериментальных, гигиенических и т. д.)- Санитарная статистика, являясь разделом медицинской статистики, в свою очередь состоит из статистики здоровья и статистики здравоохранения. Статистика здоровья изучает здоровье общества в целом и отдельных его групп и устанавливает зависимость здоровья от различных факторов социальной среды. Статистика здравоохранения анализирует данные о сети медицинских и санитарных учреждений, их деятельности и кадрах, оценивает эффективность различных мероприятий по профилактике и лечению болезней.

Задачами медицинской статистики являются:

- 1) изучение состояния здоровья населения, анализ количественных показателей общественного здоровья.
- 2) выявление связей между показателями здоровья и различными факторами, оценка их влияния на здоровье населения.
- 3) изучение организации и деятельности здравоохранения.
- 4) оценка эффективности деятельности лечебно-профилактических учреждений.
- 5) оценка эффективности (медицинской, социальной, экономической) лечебных, профилактических, противоэпидемических мероприятий.
- 6) использование статистических методов при оценке и проведения клинических и экспериментальных медицинских и биологических исследований.

Медицинская статистика является методом социальной диагностики, поскольку она позволяет дать оценку состояния здоровья населения страны, региона и на этой основе разработать меры, направленные на улучшение общественного здоровья. Принципом статистики является применение ее для изучения не только отдельных но и массовых явлений, с целью выявления разных закономерностей. Эти закономерности проявляются, как правило, в массе наблюдений, то есть при изучении статистической совокупности.

Разделы медицинской статистики:

Статистика, связанная со здоровьем населения. Этот раздел занимается изучением состояния здоровья всего населения в целом. Также он затрагивает его отдельные группы. Сведения, как правило, собираются посредством проведения статистических анализов, а кроме того, сбора данных о составе и общей численности населения. При этом строго учитывается воспроизводство населения, а также его естественное движение наряду с физическим развитием, распространенностью различных болезней, продолжительностью жизни и так далее. Оценка показателей здоровья населения проводят, сопоставляя данные с общепринятым оценочным уровнем. В частности в динамике наблюдаются сведения, которые

получаются по различным регионам.

Это медицинская статистика в области здравоохранения. Раздел занимается решением вопросов по сбору, обработке и анализу информации обо всех учреждениях здравоохранения. При этом учитываются такие параметры, как размещение данных организаций, их оснащение, деятельность и кадры. В отношении кадров собирают информацию об общей численности врачей, а кроме того, о среднем и младшем медицинском персонале, а также о распределении лиц, согласно их непосредственным специальностям и длительности стажа работы. В отношении кадров также собираются сведения о прохождении переподготовки и так далее. В рамках проведения анализа деятельности лечебно-профилактических организаций выполняется сопоставление полученной информации с нормативными уровнями.

Клиническая статистика в медицине является разделом, который следит за использованием методик в рамках обработки экспериментальных, а кроме того, лабораторных результатов исследований. Такая статистика дает возможность с количественной точки зрения оценивать достоверность полученных результатов проведенного исследования. Благодаря ей удастся решать и ряд других задач, к примеру, речь идет об определении объемов необходимого количества наблюдений во время выборочных исследований, формировании экспериментальной и контрольной группы, об изучении наличия корреляционных и регрессионных взаимосвязей, а также об устранении качественной неоднородности групп и так далее.

Контрольные вопросы:

1. Что изучает медицинская статистика?
2. Что понимают под медицинской статистикой?
3. Что является предметом медицинской статистики?
4. Назовите задачи статистики.
5. Опишите разделы медицинской статистики
6. Где применяется медицинская статистика?

ТЕМА 3. СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ. УЧЕТНЫЕ ПРИЗНАКИ.

Цель: Сформировать у студентов знания о видах статистической совокупности

Задачи обучения: Изучение видов статистической совокупности

Основные вопросы темы:

1. Статистическая совокупность
2. Виды статистической совокупности
3. Генеральная совокупность
4. Выборочная совокупности

Статистическая совокупность — группа относительно однородных элементов (единиц наблюдения) в конкретных условиях времени и пространства. В зависимости от охвата единиц наблюдения (в связи с целью исследования) статистическая совокупность может быть генеральной и выборочной.

Статистическая совокупность формируется в зависимости от поставленной цели исследования. Например:

Цель исследования: изучить длительность пребывания больных в терапевтическом отделении стационара ГБ города Алматы за истекший год, факторы на нее влияющие.

В СООТВЕТСТВИИ С ЦЕЛЬЮ

Статистическая совокупность (объект наблюдения): группа больных, лечившаяся в терапевтическом отделении стационара ГБ г. Алматы в истекшем году.

Единица статистического наблюдения- это каждый первичный элемент (отдельное явление), составляющий статистическую совокупность. Число единиц наблюдения в статистической совокупности определяет объем исследования и обозначается буквой "n".

В нашем примере.

Единица наблюдения: больной, лечившийся в терапевтическом отделении стационара ГБ города Алматы в истекшем году.

Каждая единица наблюдения является носителем признака сходства и признака различия (учетных признаков). Признаки сходства служат основанием для объединения единиц наблюдения в статистическую совокупность. Признаки различия (собственно учетные признаки) - это те признаки, по которым единица наблюдения отличаются друг от друга. Всесторонний анализ этих признаков, а также взаимосвязь между ними позволяют изучить всю совокупность в целом, ее закономерности. В нашем примере:

Признаки сходства: прогноз, время госпитализации, место госпитализации.

Признаки различия: пол, возраст, количество проведенных койко-дней, исход лечения и т.д.

Статистика располагает определенными формулами для определения необходимого числа наблюдений в выборочной совокупности или уже готовыми таблицами.

Основные способы отбора единиц наблюдения для формирования выборочной совокупности: механический, случайный типологический и др.

По времени различают два вида наблюдения:

Единовременная - учет материала проводится на определенную дату, час. Например, перепись населения, учет больничных коек или числа врачебных должностей на 31 декабря каждого года.

Текущее - регистрация случаев по мере их возникновения за определенное время: за год, месяц, годы. Например, динамика заболеваемости населения района за 5 лет.

По способу наблюдения различают 3 вида:

1. Непосредственное собирание материала (медосмотр).
2. Взятие данных из документов (выкопировка).
3. Анкетный метод.

Статистическая совокупность в отличие от отдельных единиц наблюдения

(индивидуумов) имеет особые, только ей присущие свойства, к которым относятся:

1. Характер распределения изучаемого явления.
2. Средний уровень изучаемого явления, который дает его обобщающую характеристику.
3. Разнообразие (колеблемость, изменчивость) единиц наблюдения, составляющих совокупность.
4. Взаимосвязь между изучаемыми признаками.
5. Репрезентативность признаков выборочной совокупности по отношению к генеральной.

Учетные признаки различают:

1. По характеру - атрибутивные и количественные.
2. По роли в совокупности - факторные и результативные.

Атрибутивные качественные учетные признаки выражены словесно, имеют описательный характер. Например: пол, национальность, диагноз, метод лечения и др.

Количественные – учетные признаки выраженные числом. Например: АД, СОЭ, возраст, доза препарата, число дней лечения и др.

Факторные - признаки под влиянием которых изменяются другие, зависящие от них результативные признаки.

Результативные - признаки зависящие от факторных. Например: своевременная госпитализация и рациональное лечение (факторные признаки) снижают длительность пребывания больного в стационаре (результативный признак) каждая статистическая совокупность может рассматриваться как генеральная или как выборочная.

Генеральная совокупность состоит из всех единиц наблюдения, которые могут быть к ней отнесены в зависимости от цели исследования. Обычно ее рассматривают как приближающуюся к бесконечности. Например, если бы можно было бы изучить всех больных холециститом на всем земном шаре, то такая группа больных составила бы генеральную совокупность. В практике генеральную совокупность часто рассматривают в пределах конкретных границ (например, население юрода, дети школы, больные в отделении и др.).

Выборочная совокупность - это часть генеральной совокупности отобранная особыми методами. На основе анализа выборочной совокупности можно получить достаточно полное представление о закономерностях, присущих всей генеральной совокупности. Особенностью выборочной совокупности является ее репрезентативность т.е. представительность (типичность) всех признаков, составляющих ее по отношению к признакам генеральной совокупности.

Пути формирования статистической совокупности (собираение материала - статистическое наблюдение) следующие:

1. Сплошное наблюдение - регистрация всех случаев изучаемого явления (например, всеобщая перепись населения).
2. Выборочная регистрация - части случаев изучаемого явления.

Выборочную совокупность следует отбирать из генеральной на основе определенных правил. Для обеспечения репрезентативности выборочной совокупности, к ней предъявляют два основных требования.

▪ Выборочная совокупность должна быть максимально типичной, т.е. похожей на генеральную.

- Должна быть достаточной по объему (т.е. числу наблюдений).

Изучение того или иного явления с применением статистического метода требует от исследователя прежде всего умелого подхода к выбору объекта исследования, так называемой статистической совокупности.

Контрольные вопросы:

1. Статистическая совокупность
2. Единица статистического наблюдения
3. Признаки сходства и различия
4. Виды по способу наблюдения
5. Учетные признаки
6. Атрибутивные и количественные признаки
7. Результативные
8. Генеральная выборочная совокупности

ТЕМА 4. СТАТИСТИЧЕСКИЕ ВЕЛИЧИНЫ И ИХ ВИДЫ.

Цель: Ознакомление студентов с статистическими величинами и их видами.

Задачи обучения: Изучение видов статистической совокупности

Основные вопросы темы:

1. Абсолютные величины, виды
2. Относительные величины, виды
3. Сущность и значение средних величин

Абсолютные и относительные статистические величины.

Результаты статистического наблюдения регистрируются в виде первичных абсолютных величин. Абсолютная величина отражает уровень развития явления. В статистике все абсолютные величины являются именованными, измеряются в конкретных единицах. И в отличие от математического понятия абсолютные величины могут быть как положительными, так и отрицательными.

Абсолютные величины делятся на:

- 1) *Индивидуальные* – характеризуют размер признака отдельных единиц совокупности.
- 2) *Суммарные*. Характеризуют итоговое значение признака по определённой части совокупности. Они разделяются на:
 - а) *моментные* - показывают фактическое наличие на определённый момент или дату.
 - б) *интервальные* - итоговый накопленный результат за период в целом. В отличие от моментных, они допускают их последующее суммирование.

Абсолютная величина не даёт представления об изучаемом явлении, не показывает его структуру, соотношение между отдельными частями и развития во времени. Эти функции выполняют относительные показатели. *Относительная величина* – это обобщающий показатель, который даёт числовую меру соотношения двух сопоставляемых абсолютных величин. Основное условие правильного расчёта относительной величины – это сопоставимость сравниваемых показателей и наличие реальных связей между изучаемыми явлениями. Таким образом, по способу получения относительные показатели всегда величины производные, определяемые в форме коэффициентов, промилле и т.п.

Виды относительных величин:

- 1) экстенсивные коэффициенты;
- 2) интенсивные коэффициенты;
- 3) коэффициенты соотношения;
- 4) коэффициенты наглядности.

Экстенсивные коэффициенты характеризуют отношение части к целому, то есть определяют долю (удельный вес), процент части в целом, принятом за 100%. Используются для характеристики структуры статистической совокупности. Например: удельный вес (доля) заболеваний гриппом среди всех заболеваний в процентах; доля производственных травм среди всех травм у рабочих (отношение числа производственных травм к общему числу травм, умноженное на 100%).

Интенсивные коэффициенты отражают частоту (уровень распространенности) явления в своей среде. На практике их применяют для оценки здоровья населения, медико-

демографических процессов. Например: число случаев заболеваний с временной утратой трудоспособности на 100 работающих; число заболевших гипертонической болезнью на 100 жителей; число родившихся на 1000 человек (определяется как отношение числа родившихся за год к средней численности населения административной территории, умноженное на 1000). Интенсивные коэффициенты бывают общие и специальные. Общие: показатель рождаемости, общий показатель заболеваемости и др.; специальные (характеризуются более узким основанием): число женщин детородного возраста (плодовитость), число женщин, заболевших гипертонической болезнью, и др.

Коэффициенты соотношения характеризуют отношение двух самостоятельных совокупностей. Используются для характеристики обеспеченности (уровня и качества) медицинской помощью: число коек на 10000 человек; число врачей на 10000 жителей; число прививок на 1000 жителей (отношение числа лиц, охваченных прививками, к численности населения административной территории, умноженное на 1000).

Коэффициент наглядности определяет, на сколько процентов или во сколько раз произошло увеличение или уменьшение по сравнению с величиной, принятой за 100%. Используется для характеристики динамики явления. Например, число врачей в 1995 г. по сравнению с числом врачей в 1994 г., принятым за 100% (отношение числа специалистов в данном году к числу специалистов в предыдущем году, умноженное на 100%).

Сущность и значение средних величин.

В результате группировки единиц совокупности по величине варьирующего признака получают ряды распределения - первичную характеристику массовой статистической совокупности. Чтобы охарактеризовать такую совокупность в целом, часто пользуются средней величиной.

Средняя величина - обобщающая характеристика изучаемого признака в исследуемой совокупности. Она отражает его типичный уровень в расчете на единицу совокупности в конкретных условиях места и времени. Метод средних величин заключается в замене индивидуальных значений варьирующего признака единиц наблюдения. Средние величины применяются для оценки достигнутого уровня изучаемого показателя, при анализе и планировании производственно-хозяйственной деятельности предприятий (объединений), фирм, банков и других хозяйственных единиц; средние используются при выявлении взаимосвязей явлений, при прогнозировании, а также расчете нормативов.

Средняя величина всегда именованная, имеет ту же размерность (единицу измерения), что и признак у отдельных единиц совокупности. Основным условием научного использования средней величины является качественная однородность совокупности, по которой исчислена средняя. Поэтому очень важное правило - вычислять средние величины лишь по однородной совокупности единиц. Только при выполнении этого условия средняя как обобщающая характеристика отражает общее, типичное, закономерное, присущее всем единицам исследуемой совокупности. Прежде чем вычислять средние величины, необходимо произвести группировку единиц исследуемой совокупности, выделив качественно однородные группы.

Средняя, рассчитанная по совокупности в целом, называется общей средней, средние, исчисленные для каждой группы, - групповыми средними. Общая средняя отражает общие черты изучаемого явления, групповая средняя дает характеристику размера явления, складывающуюся в конкретных условиях данной группы.

Показатели вариации и способы их расчета.

При изучении явлений и процессов общественной жизни статистика встречается с разнообразной вариацией (изменчивостью) признаков, характеризующих отдельные единицы совокупности.

Величины признаков изменяются под действием различных факторов. Очевидно, что чем разнообразнее условия, влияющие на размер данного признака, тем больше его вариация. Например, размер заработной платы рабочих зависит от нескольких факторов: специальности, разряда, стажа работы, образования, состояния здоровья и т.д. Чем больше различия между

значениями факторов, тем больше вариация в уровне заработной платы. При характеристике колеблемости признака используют систему абсолютных и относительных показателей.

Абсолютные показатели вариации: Размах вариации $R = x_{\max} - x_{\min}$; . Среднее линейное отклонение. Дисперсия. Среднеквадратическое отклонение. Абсолютные показатели, кроме дисперсии, измеряются в тех же единицах, что и сам признак.

Относительные показатели вариации: Коэффициент осцилляции. Относительное линейное отклонение. Коэффициент вариации. Относительные показатели чаще всего выражаются в процентах. Размах колебаний, или размах вариации, представляет собой разность между максимальным и минимальным значениями признака в изучаемой совокупности:

$$R = x_{\max} - x_{\min}$$

Безусловным достоинством этого показателя является простота расчета. Однако размах вариации зависит от величины только крайних значений признака, поэтому область его применения ограничена достаточно однородными совокупностями. В частности, на практике он находит применение в предупредительном контроле качества продукции.

Точнее характеризует вариацию признака показатель, основанный на учете колеблемости всех значений признака. Поскольку средняя арифметическая является обобщающей характеристикой свойств совокупности, большинство показателей вариации основано на рассмотрении отклонений значений признака отдельных единиц совокупности от этой величины. К таким показателям относятся среднее линейное отклонение, дисперсия и среднее квадратическое отклонение, представляющие собой среднюю арифметическую из отклонений индивидуальных значений признака от средней арифметической. Среднелинейное отклонение рассчитывается из отклонений в первой степени, дисперсия и среднее квадратическое - из отклонений во второй степени.

Так как алгебраическая сумма отклонений индивидуальных значений признака от средней арифметической (согласно нулевому свойству) всегда равна нулю, то для расчета среднего линейного отклонения используется арифметическая сумма отклонений, т.е. суммируются абсолютные значения индивидуальных отклонений значений признака независимо от знака. Дисперсия имеет большое значение в статистическом анализе. Однако её применение как меры вариации в ряде случаев бывает не совсем удобным, потому что размерность дисперсии равна квадрату размерности изучаемого признака.

В таких случаях для измерения вариации признака вычисляют среднее квадратическое отклонение. Среднее квадратическое отклонение (представляет собой корень квадратный из дисперсии). Дисперсия и среднее квадратическое отклонение недостаточно полно характеризуют колеблемость признака, т.к. показывают абсолютный размер отклонений, что затрудняет сравнение изменчивости различных признаков. Для характеристики колеблемости явлений среднее квадратическое отклонение сопоставляется с его средней величиной и выражают в процентах. Такой показатель называется коэффициентом вариации.

Коэффициент вариации представляет собой отношение среднего квадратического отклонения к средней арифметической. Выражая коэффициент вариации в процентах, различные абсолютные среднеквадратические отклонения приводят к одному основанию и дают возможность сравнивать, оценивать колеблемость величин различных признаков. При помощи коэффициента вариации возможно, например, сравнение размера колеблемости производительности труда рабочих, занятых производством различных видов продукции, размера колеблемости урожая различных сельскохозяйственных культур и т.д.

Чем меньше коэффициент вариации, тем меньше колеблемость признака, и наоборот.

Относительное линейное отклонение определяется как отношение среднего линейного отклонения к средней арифметической в процентах.

Отношение размаха вариации к средней арифметической в процентах называется коэффициентом осцилляции. Самым распространенным относительным показателем колеблемости признака является

коэффициент вариации. Он более точно, чем абсолютный, характеризует различие колеблемости признаков. По величине коэффициента вариации можно судить о степени вариации признаков совокупности. Чем больше его величина, тем больше разброс значений вокруг средней, тем менее однородна совокупность по своему составу и тем менее представительна средняя.

Коэффициент вариации важен в тех случаях, когда нужно сравнивать средние квадратические отклонения, выраженные в разных единицах измерения.

Контрольные вопросы:

1. Статистические величины.
2. Абсолютные величины
3. Относительных величин
4. Показатели вариации и способы их расчета

ТЕМА 5.СРЕДНИЙ УРОВЕНЬ ПРИЗНАКА.

Цель: Ознакомление студентов с средними величинами, их характеристикой.

Задачи обучения: Изучение средних величин

Основные вопросы темы:

Сущность и задачи средних величин

Средняя величина – это обобщающая количественная характеристика совокупности однотипных явлений по одному варьирующему признаку.

Она отражает объективный уровень, достигнутый в процессе развития явления к определенному моменту или периоду.

Средняя представляет значение определенного признака в совокупности одним числом и элиминирует индивидуальные различия значений отдельных величин совокупности.

Необходимость сочетается со случайностью, поэтому средние величины связаны с Законом больших чисел. Суть этой связи в том, что при осреднении случайные отклонения индивидуальных величин от средней погашаются, а в средней отчетливо выявляется основная тенденция развития.

Важнейшая особенность средней величины – в том, что она относится к единице изучаемой совокупности и через характеристику единицы характеризует всю совокупность в целом.

Основные свойства средней величины:

1. Она обладает устойчивостью, что позволяет выявлять закономерности развития явлений. Средняя облегчает сравнение двух совокупностей, обладающих различной численностью.

2. Она помогает характеризовать развитие уровня явления во времени.

3. Она помогает выявить и охарактеризовать связь между явлениями.

Средние позволяют исключить влияние индивидуальных значений признака, т.е. они являются абстрактными величинами. Поэтому средние должны употребляться на основе сгруппированных данных.

Расчет средней. К расчету средней предъявляются два основных требования:

1. Среднюю нужно рассчитывать так, чтобы она погасила то, что мешает выявлению характерных черт и закономерностей в развитии явления, а не затушевывала развитие.

2. Средняя может быть вычислена только для однородной совокупности. Средняя, вычисленная для неоднородной совокупности, называется огульной.

Одинаковые по форме и технике вычисления средние в одних случаях могут быть огульными, а в других – общими в зависимости от того, с какой целью они интерпретируются.

Говоря о методологии исчисления средних, не надо забывать, что средняя всегда дает обобщенную характеристику лишь по одному признаку. Каждая же единица совокупности имеет много признаков. Поэтому необходимо рассчитывать систему средних, чтобы охарактеризовать явление со всех сторон.

Расчет средних величин производится по правилам, которые разрабатываются математической статистикой. Задача ОТС – дать смысловую, преимущественно экономическую интерпретацию результатам расчетов, произведенных по формулам.

Признак, по которому производится осреднение, называется осредняемым признаком. Величина осредняемого признака у каждой единицы совокупности называется ее индивидуальным значением.

Значение признака, которое встречается у групп единиц или у отдельных единиц и не повторяется, называется вариантом признака.

Контрольные вопросы:

1. Сущность и задачи средних величин
2. Средняя величина
3. Основные свойства средней величины:
4. Расчет средней
5. Требования к расчету

ТЕМА 6. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И ЕГО СВОЙСТВА.

Цель: Ознакомление студентов с нормальным распределением и его свойствами.

Задачи обучения: Изучение нормальным распределением и его свойствам.

Основные вопросы темы:

Закон распределения

Нормальный закон распределения (часто называемый законом Гаусса) играет исключительно важную роль в теории вероятностей и занимает среди других законов распределения особое положение. Это – наиболее часто встречающийся на практике закон распределения. Главная особенность, выделяющая нормальный закон среди других законов, состоит в том, что он является предельным законом, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях.

Можно доказать, что сумма достаточно большого числа независимых (или слабо зависимых) случайных величин, подчиненных каким угодно законам распределения (при соблюдении некоторых весьма нежестких ограничений), приближенно подчиняется нормальному закону, и это выполняется тем точнее, чем большее количество случайных величин суммируется. Большинство встречающихся на практике случайных величин, таких, например, как ошибки измерений, ошибки стрельбы и т.д., могут быть представлены как суммы весьма большого числа сравнительно малых слагаемых – элементарных ошибок, каждая из которых вызвана действием отдельной причины, не зависящей от остальных. Каким бы законам распределения ни были подчинены отдельные элементарные ошибки, особенности этих распределений в сумме большого числа слагаемых нивелируются, и сумма оказывается подчиненной закону, близкому к нормальному. Основное ограничение, налагаемое на суммируемые ошибки, состоит в том, чтобы они все равномерно играли в общей сумме относительно малую роль. Если это условие не выполняется и, например, одна из случайных ошибок окажется по своему влиянию на сумму резко преобладающей над всеми другими, то закон распределения этой преобладающей ошибки наложит свое влияние на сумму и определит в основных чертах её закон распределения.

Теоремы, устанавливающие нормальный закон как предельный для суммы независимых равномерно малых случайных слагаемых, будут подробнее рассмотрены в главе 13.

Нормальный закон распределения характеризуется плотностью вероятности вида:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (1)$$

Кривая распределения по нормальному закону имеет симметричный холмообразный

вид (рис. 1). Максимальная ордината кривой, равная $\frac{1}{\sigma\sqrt{2\pi}}$, соответствует точке $x = m$; по мере удаления от точки m плотность распределения падает, и при $x \rightarrow \pm\infty$ кривая асимптотически приближается к оси абсцисс.

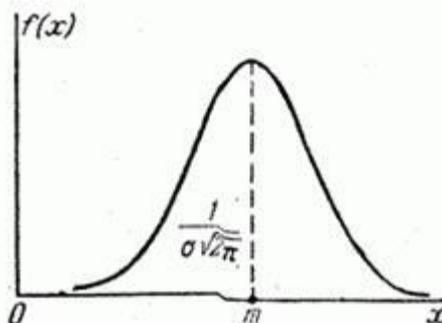


Рисунок 1.

Выясним смысл численных параметров m и σ , входящих в выражение нормального закона (6.1.1); докажем, что величина m есть не что иное, как математическое ожидание, а величина σ - среднее квадратическое отклонение величины X . Для этого вычислим основные числовые характеристики величины X - математическое ожидание и дисперсию.

$$M[X] = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Применяя замену переменной

$$\frac{x-m}{\sigma\sqrt{2}} = t,$$

имеем:

$$M[X] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}t + m)e^{-t^2} dt = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} te^{-t^2} dt + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt. \quad (1)$$

Нетрудно убедиться, что первый из двух интервалов в формуле (1) равен нулю; второй представляет собой известный интеграл Эйлера-Пуассона:

$$\int_{-\infty}^{\infty} e^{-t^2} dt = 2 \int_0^{\infty} e^{-t^2} dt = \sqrt{\pi}. \quad (2)$$

Следовательно,

$$M[X] = m,$$

т.е. параметр m представляет собой математическое ожидание величины X . Этот параметр, особенно в задачах стрельбы, часто называют центром рассеивания (сокращенно - ц. р.).

Вычислим дисперсию величины X :

$$D[X] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Применив снова замену переменной

$$\frac{x-m}{\sigma\sqrt{2}} = t,$$

имеем:

$$D[X] = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt.$$

Интегрируя по частям, получим:

$$D[X] = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t \cdot 2te^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \left\{ -te^{-t^2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2} dt \right\}.$$

Первое слагаемое в фигурных скобках равно нулю (так как e^{-t^2} при $t \rightarrow \infty$ убывает быстрее, чем возрастает любая степень t), второе слагаемое по формуле (2) равно $\sqrt{\pi}$, откуда

$$D[X] = \sigma^2.$$

Следовательно, параметр σ в формуле (6.1.1) есть не что иное, как среднее квадратическое отклонение величины X .

Выясним смысл параметров m и σ нормального распределения. Непосредственно из формулы (6.1.1) видно, что центром симметрии распределения является центр рассеивания m . Это ясно из того, что при изменении знака разности $(x-m)$ на обратный выражение (6.1.1) не меняется. Если изменять центр рассеивания m , кривая распределения будет смещаться вдоль оси абсцисс, не изменяя своей формы (рис. 3). Центр рассеивания характеризует положение распределения на оси абсцисс.

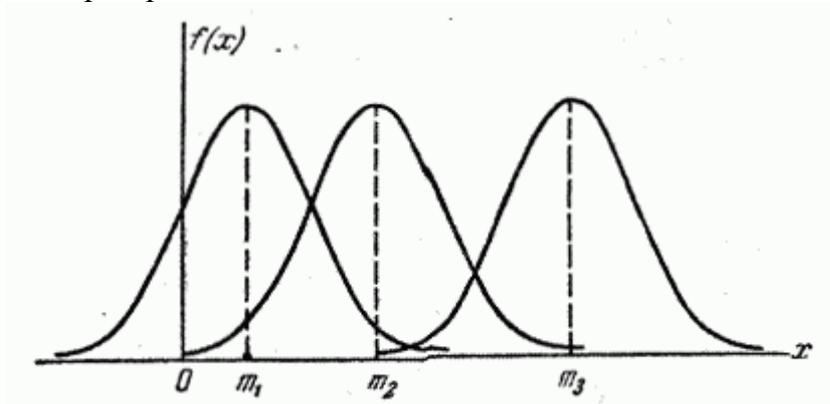


Рис. 3

Размерность центра рассеивания – та же, что размерность случайной величины X .

Параметр σ характеризует не положение, а самую форму кривой распределения. Это есть характеристика рассеивания. Наибольшая ордината кривой распределения обратно пропорциональна σ ; при увеличении σ максимальная ордината уменьшается. Так как площадь кривой распределения всегда должна оставаться равной единице, то при увеличении σ кривая распределения становится более плоской, растягиваясь вдоль оси абсцисс; напротив, при уменьшении σ кривая распределения вытягивается вверх, одновременно сжимаясь с боков, и становится более иглообразной. На рис. 6.1.3 показаны три нормальные кривые (I, II, III) при $m = 0$; из них кривая I соответствует самому большому, а кривая III – самому малому значению σ . Изменение параметра σ равносильно изменению масштаба кривой распределения – увеличению масштаба по одной оси и такому же уменьшению по другой.

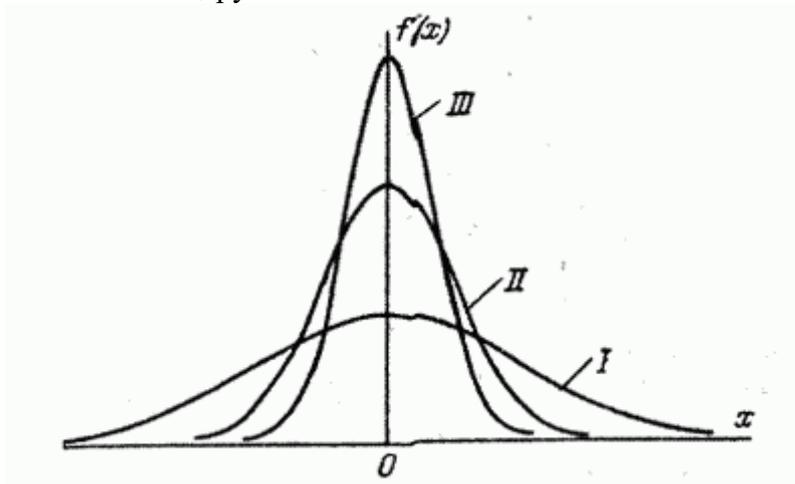


Рис. 4.

Размерность параметра σ , естественно, совпадает с размерностью случайной величины X .

В некоторых курсах теории вероятностей в качестве характеристики рассеивания для нормального закона вместо среднего квадратического отклонения применяется так называемая мера точности. Мерой точности называется величина, обратно пропорциональная среднему квадратическому отклонению σ :

$$h = \frac{1}{\sigma\sqrt{2}}.$$

Размерность меры точности обратная размерности случайной величины.

Термин «мера точности» заимствован из теории ошибок измерений: чем точнее измерение, тем больше мера точности. Пользуясь мерой точности h , можно записать нормальный закон в виде:

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2(x-m)^2}$$

Контрольные вопросы:

1. Нормальный закон распределения
2. Закон Гаусса
3. Случайные величины
4. Теории вероятностей

Цель: Ознакомление студентов с организацией статистического исследования.

Задачи обучения: Обучить организации статистического исследования

Основные вопросы темы:

1. Этапы исследований
2. Полные определения этапов

Независимо от вида исследования (социально-гигиеническое, клинико-статистическое и др.) его целесообразно проводить в определенной последовательности в соответствии с выполнением следующих *этапов*:

I — составление программы и плана исследования;

II — сбор материала;

III — статистическая обработка собранного материала;

IV — анализ полученных данных и формулировка выводов.

1.1. План исследования включает в себя целый ряд элементов. Во-первых, определяется *цель* исследования (или целевая установка), которая может быть структурирована на ряд подцелей (задач), число которых обычно составляет 3–6.

Далее следует знакомство с литературой, которое позволяет:

- получить представление об изучаемой проблеме;
- выбрать адекватную методику исследования;
- сформулировать рабочую гипотезу.

Рабочая гипотеза — это обоснованное предположение о результатах исследования. В случае несовпадения полученных результатов с рабочей гипотезой исследователь должен убедиться в: а) отсутствии ошибки в расчетах; б) адекватности выбранной методики исследования поставленным цели и задачам.

Очень важным элементом является выбор единицы наблюдения. *Единица наблюдения (счета)* — это первичный элемент статистической совокупности, наделенный всеми признаками, подлежащими изучению и регистрации.

В зависимости от степени охвата единиц наблюдения выделяют *два метода* исследования: сплошное (изучают все единицы наблюдения) и несплошное (выборочное).

Теоретическим обоснованием выборочного метода является закон больших чисел, сформулированный Бернулли: «При неограниченном увеличении числа однородных независимых опытов с практической достоверностью можно утверждать, что наблюдаемая частота случайного события будет сколь угодно мало отличаться от вероятности появления события в отдельном опыте».

Раскрывая смысл данного определения, следует выделить следующие основополагающие понятия:

Случайная величина — это величина, которая при реализации комплекса условий может принимать различные значения.

Статистическая вероятность, выражая численную меру объективной возможности появления того или иного события (при реализации определенного комплекса условий), является отношением числа опытов, в которых появилось событие, к общему числу опытов.

Соответствие, устанавливаемое между всеми возможными численными значениями случайной величины и вероятностями их появления, называется *законом распределения* (который описывает случайную величину с вероятностной точки зрения). Основными типами распределения являются: альтернативный, нормальный (симметричный, асимметричный — правосторонний, левосторонний, бимодальный) и др.

Для обеспечения качественной репрезентативности выборки необходимо избегать систематических ошибок, т.е. рандомизировать выборку. Процесс рандомизации аналогичен подбрасыванию монеты, обеспечивающему равные шансы каждой единице наблюдения попасть в ту или иную группу. Рандомизация уравнивает вероятность воздействия как учитываемых факторов, так и тех, о существовании которых мы не подозреваем (что является одним из условий доказательной медицины).

Возможны разные варианты формирования выборки:

Случайная выборка формируется путем отбора единиц наблюдения наугад: например, по первой букве фамилии, алфавита, по жребию и т.д.

Механическая выборка формируется с помощью механического (арифметического) подхода к отбору единиц наблюдения.

Типическая (типологическая) выборка — это выборка, при формировании которой генеральная совокупность предварительно разбивается на типы с последующим отбором единиц наблюдения из каждой типической группы, при этом число единиц наблюдения можно отобрать пропорционально численности типической группы (пропорциональный типологический отбор) и непропорционально, т.е. отбирая разное число наблюдений из каждой группы (непропорциональный типологический отбор).

Серийная выборка (гнездовой отбор) формируется с помощью отбора не отдельных единиц наблюдения, а целых групп, серий, или гнезд, в состав которых входят организованные определенным образом единицы наблюдения.

Метод многоступенчатого отбора. По количеству этапов различают одноступенчатый, двухступенчатый, трехступенчатый отбор и т.п.

Метод направленного отбора. Использование принципов направленного отбора позволяет выявить влияние неизвестных факторов при устранении влияния известных.

Когортный метод. Под когортой в демографии понимают совокупность людей, переживших одно и то же демографическое явление в течение одного и того же года. Иными словами, статистическую совокупность при этом методе составляют относительно однородные группы лиц, объединенные наступлением определенного демографического признака в один и тот же интервал времени.

Метод «копи-пара», или способ уравнивания групп (метод парных сочетаний). В основе его лежит подбор для каждой единицы наблюдения исследуемой группы «копи-пары» по одному или нескольким признакам..

Монографическое исследование— тщательное, глубокое изучение одного человека, одного учреждения, одного села и т.д. Монографическое исследование иногда проводят перед основным с целью разработки программы, изучения различных организационных вопросов. Оно нередко проводится по одной и той же единице наблюдения с определенными временными интервалами. При этом создаются оптимальные условия для изучения динамики факториальных и результативных признаков.

Метод основного массива охватывает большую часть единиц, изучаемого объекта наблюдения. Этот метод иногда называют несовершенным сплошным.

Объем исследования— это количество включенных в него единиц наблюдения (как отмечалось выше, определяется по специальным формулам).

Объект исследования— конкретная статистическая совокупность, подлежащая изучению (например, пациенты хирургического профиля, лечившиеся в поликлинике, больнице и т.д.).

План исследования должен включать в себя в такие вопросы, как:

- планируемые сроки работы;
- подбор, подготовка, обучение исполнителей;
- необходимые (и имеющиеся в наличии) материальные и финансовые ресурсы;
- источники информации, первичные учебные документы;
- тип работы (отчет, статья, диссертация, дипломная работа и т.д.).

1.2. Программа исследования - это перечень вопросов, подлежащих изучению. Вопросы программы по существу составляют признаки изучаемой совокупности. Выделение признаков единиц наблюдения с последующей их группировкой называется типологизацией. Признаки делятся на количественные (вариационные) и качественные (атрибутивные, типологические). К количественным признакам относятся такие, варианты которых отличаются друг от друга определенной величиной (возраст, рост и т.д.). Качественные признаки позволяют отличать единицу наблюдения по содержанию (социальной

принадлежности, характеру труда и пр.). Кроме того, признаки могут делиться на секторные (наличие в анамнезе факторов генетического, социального, профессионального риска) и результативные (наличие или отсутствие заболевания, характер его течения, исход).

Составляя программу исследования, во-первых, руководствуются его целью (при этом важно как включить в программу все необходимые вопросы, так и не перегрузить ее ненужными, излишними вопросами); во-вторых, он должен четко представлять, откуда (из каких источников) будет получена информация и насколько реально при этом получение ответов на те или иные вопросы.

2. Вторым этапом исследования является сбор материала, т.е. регистрации всех запланированных признаков для каждой единицы наблюдения. При этом на каждую единицу наблюдения заводится свой регистрационный документ (анкета, первичные учетные медицинские документы, специально разработанная карта исследования и т.д.). Сбор данных может осуществляться разными методами: анкетирование, интервьюирование, выкопировка сведений из первичной медицинской документации и др.

3. Третий этап обработки полученной информации начинается с проверки собранного материала, которая бывает количественной (соответствие количества единиц наблюдения и признаков запланированному) и логической (выявление логического несоответствия между признаками).

Далее необходимо провести *группировку*, т.е. разбить каждый признак на качественно однородные группы. При составлении группировки руководствуются:

- а) целью исследования;
- б) фактически собранным материалом;
- в) объемом исследования

Когда материал сгруппирован, его шифруют, т.е. придают каждой группе определенный символ (шифр) для облегчения последующей обработки.

Заканчивается третий этап составлением статистических таблиц. При этом необходимо провести сопоставление признаков, т.е. определить взаимосвязь и взаимозависимость между признаками, а затем выбрать те из них, которые представляют интерес для исследования.

Статистические таблицы в зависимости от количества содержащихся в них признаков разделяют на простые и сложные (а последние, в свою очередь, на групповые и комбинационные).

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

В *групповой* таблице подлежащее характеризуется одним или несколькими сказуемыми (преимущественно двумя), но признаки, характеризующие подлежащее, не связаны между собой.

В *комбинационной* таблице признаки, характеризующие подлежащее взаимосвязаны (таблица обычно содержит информацию о трех и более признаках).

Табличное подлежащее — это основной признак изучаемого явления оно обычно располагается слева по горизонтальным строкам таблицы.

Сказуемое - признаки, характеризующие подлежащее, — располагаются обычно сверху (столбцы или вертикальные графы таблицы).

При составлении таблиц должны соблюдаться определенные требования:

- таблица должна иметь четкое, краткое заглавие, отражающее суть таблицы;
- таблица должна иметь единую последовательную порядковую нумерацию;
- оформление таблицы заканчивается итогами по графам и строкам;
- в таблице не должно быть пустых клеток (если нет признака, ставится прочерк).

Наряду с табличной формой, статистическая информация может быть представлена графически; в виде диаграмм, картограмм, картодиаграмм.

Диаграмма - это графическое изображение статистических величин с помощью различных геометрических фигур и знаков.

Картограмма — географическая карта или ее схема, па которой различной краской или

штриховкой изображена степень распространения какого-либо явления на различных участках территории.

Картодиаграммой называется такое географическое изображение, когда на географическую карту или ее схему статистические данные наносятся в виде столбиковых, секторных, фигурных и других диаграмм.

Диаграммы чаще используются в медико-социальных исследованиях, в то время как картограммы и картодиаграммы — в медико-географических.

Диаграммы можно классифицировать различным образом. По назначению принято различать диаграммы сравнения, структурные и динамические диаграммы. Выделяют также линейные, плоскостные и объемные графические изображения.

При построении графических изображений следует соблюдать некоторые правила:

- каждая диаграмма должна иметь четкое, ясное, краткое название, отражающее ее содержание и порядковый номер;
- все элементы диаграммы (фигуры, знаки, окраска, штриховка) должны быть пояснены на самой диаграмме или в условных обозначениях (легенде);
- изображаемые графические величины должны иметь цифровые обозначения на самой диаграмме или в прилагаемой таблице;
- данные на диаграмме должны размещаться от большего к меньшему слева направо, снизу вверх и по часовой стрелке (но элемент «прочие» всегда располагается последним).

4. На четвертом этапе проводится анализ полученных данных, невозможный без расчета ряда статистических показателей.

Для проведения любого самостоятельного исследования необходимо изучить литературу, посвященную данной теме, отдельным ее аспектам, что позволит избежать многих ошибок и получить достоверные результаты. После ознакомления с литературой, надо сформулировать тему настоящего исследования. Для конкретизации предстоящего исследования, определения его характера, т. е. будет ли оно натуральным, проведенным в естественных, обычных, неизменных условиях, или, напротив, экспериментальным, когда условия его осуществления будут заранее определены.

Большинство исследований, в том числе и комплексные социально-гигиенические и клиничко-социальные, являются натуральными. Небольшая часть работ, в которых изучаются преимущественно вопросы внедрения новых организационных технологий в учреждения здравоохранения, являются экспериментальными. Так, например, в течение ряда лет в учреждениях здравоохранения проводился экономический эксперимент по расширению прав руководителей лечебно-профилактических учреждений в применении различных форм работы медицинского персонала, во внедрении различных методов стимулирования оплаты труда медицинского персонала, в использовании бюджетных и внебюджетных средств и т. д. Клинические исследования обычно проводят в условиях, определяемых самим исследователем, поэтому носят экспериментальный характер. В этих исследованиях, как правило, анализируется преимущество определенного нового метода диагностики, лечения, профилактики, реабилитации в сравнении с существующими.

Контрольные вопросы:

1. организацией статистического исследования.
2. составление программы и плана исследования;
3. сбор материала;
4. статистическая обработка собранного материала;
5. анализ полученных данных и формулировка выводов.
6. варианта формирования выборки

ТЕМА 8. ПЛАНИРОВАНИЕ И ОРГАНИЗАЦИЯ СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЙ.

Цель: Ознакомление студентов с планированием и организацией статистических исследований.

Задачи обучения: Изучить планирование и организация статистических исследований

Основные вопросы темы:

1. Планирование
2. Проведение исследования
3. Основные понятия и принципы теории планирования эксперимента

Организация и планирование экспериментальных исследований является важной составной частью исследовательской деятельности. В их основе лежит математическая теория планирования эксперимента, которая разрабатывает приемы и методы оптимальной организации эксперимента.

Планирование эксперимента как наука.

В прошлом эксперимент рассматривался как средство изучения явлений природы и биологических систем, в частности. Предполагалось, что проведение любого опыта должно удовлетворять требованиям чистоты. Это обозначает, что при проведении измерений, во-первых, требуется исключить все посторонние возмущения и воздействия на объект исследования и, во-вторых, предполагается полная воспроизводимость результатов измерений. Другими словами, объект исследовался как изолированное явление без учета его взаимодействия со средой.

В настоящее время в медико-биологических исследованиях все больший интерес проявляется исследованию сложных систем, содержащих много элементов, с большим количеством связей между ними и взаимодействующих с внешней средой, так называемых "плохоорганизованных" систем.

Проведение современных исследований имеет ряд особенностей:

1. Приходится выполнять большой объем измерений, как правило дорогостоящих, длительных, требующих высокой точности и т.п. В этих условиях эффективное планирование работ будет иметь первостепенно значение, поскольку никакая статистическо-математическая обработка данных плохо спланированного эксперимента не позволяет улучшать ситуацию.

2. В процессе работы широко используются ЭВМ, что увеличивает возможности эксперимента за счет применения разнообразных методов представления и анализа данных, автоматизации исследований, проведения модельных и численных экспериментов. Если раньше организацией планированием работ занимался сам экспериментатор, а математик привлекался только для обработки и представления результатов проведенных экспериментов, то в настоящее время работа математика должна начинаться уже на стадии подготовки эксперимента. Причем основное внимание требуется уделять нахождению оптимальных методов измерения данных и построению оптимальных процедур их математико-статистического анализа с учетом целей исследования, природы изучаемых явлений и процессов.

3. Все это говорит о том, что оптимизация исследований становится важным аспектом научной деятельности, в котором эксперимент сам по себе является предметом исследования, а теория планирования эксперимента рассматривается, как научное направление, являющееся методологической основой современных научных исследований.

4. Для того, чтобы эксперимент мог быть объектом изучения, он должен обладать рядом некоторых общих свойств вне зависимости от сферы предметной области исследования. К таким общим чертам эксперимента относятся:

- возможность исключать или оценивать влияние параметров, непринятых по тем или иным причинам к рассмотрению;
- необходимость определять точность измерительных приборов и полученных данных;
- необходимость уменьшать до разумных пределов число переменных в эксперименте;
- проверка приемлемости (правильности) полученных результатов и их точности;

- выбор и обоснование способа обработки экспериментальных данных и формы представления результатов;

- анализ полученных результатов и интерпретация их в терминах той предметной области, в которой эксперимент проводился.

Перечислим типовые задачи решаемые в процессе планирования любого экспериментального исследования.

- Получение предварительных сведений об изучаемом процессе или явлении. Под этим понимается работа с литературой, анализ результатов мнений экспертов, проведение уточняющих и отсеивающих экспериментов с целью очертить круг параметров подлежащих изучению.

- Получение формульных зависимостей. На основе имеющихся данных выдвигаются предположения о характере поведения и связях между параметрами. Строятся модели изучаемого объекта на вербальном или математическом уровне.

- Проверка гипотез о свойствах объекта. Собственно проведение эксперимента и обработка полученных результатов, а также интерпретация полученных результатов.

- Оптимизация свойств изучаемого объекта. На основе полученных результатов проводятся исследования по определению оптимальных значений характеристик изучаемого объекта, даются рекомендации по направлениям дальнейшего исследования.

Основные понятия и принципы теории планирования эксперимента.

Объект исследования - носитель некоторых неизвестных и подлежащих изучению свойств и качеств.

Предполагается, что всегда имеется некоторая априорная информация об объекте исследования (примерный перечень переменных, характеризующих объект). Даже не зная внутренней структуры и взаимосвязей между отдельными его элементами, объект исследования и может быть представлен в виде «черного ящика» (рис.1). Характеризующие его параметры могут быть объединены в группы:

X - входные контролируемые и управляемые параметры;

Z - входные контролируемые, но неуправляемые параметры;

E - неконтролируемые и неуправляемые параметры;

Y - выходные параметры (характеристика свойств объекта).



Параметры типа X, Y называются факторами, а Z, E - шумами или помехами.

Примерами помех могут служить ошибки измерительных приборов или методов анализа, неконтролируемых изменений характеристик используемых веществ в различных партиях, изменение свойств приборов в результате старения или износа, квалификация

Рисунок 2 Характеристики объекта исследования.

экспериментатора и т.п., то есть все те характеристики, которые могут не сказаться на качестве эксперимента, но будут существенны при сравнении различных серий экспериментов. Величина Y называется откликом системы, а ее зависимость от входных воздействий функцией отклика (геометрическое представление Y - поверхностью отклика).

Объектами исследования в теории планирования эксперимента являются:

- Реальные физические и биологические объекты (схемы, устройства, лабораторные животные, системы организма и т.п.).

- Физические модели реальных объектов (модель перцептрона, имитационные и аналоговые модели и т.п.).

- Математические модели реальных объектов (системы алгебраических, дифференциальных уравнений, статистические модели, случайные функции и т.п.).

• Источником помех при исследовании последних двух типов объектов являются округления и неточности численных расчетов и несоответствия модельных предположений их прототипам. Особенности статистических моделей является то, что они не описывают точное поведение объекта исследования в конкретном опыте, а характеризуют некоторые усредненные свойства, проявляющиеся при многократном воспроизведении опыта в идентичных условиях.

Эксперимент - система операций, воздействий и наблюдений, направленных на получение информации об объекте исследования. Различают активный и пассивный эксперименты. В активном эксперименте присутствуют факторы группы X и экспериментатор в соответствии с целями и задачами эксперимента может их целенаправленно изменять. Это наиболее распространенный тип эксперимента. В пассивном эксперименте присутствуют факторы группы Z и экспериментатор фиксирует функцию отклика им соответствующую (например, влияние количества солнечных дней на урожайность).

Опыт - воспроизведение исследуемого явления в определенных условиях проведения эксперимента, при возможности регистрации его результатов.

План эксперимента - совокупность данных, определяющих число, условия и порядок реализации опытов.

Планирование эксперимента - выбор плана эксперимента, удовлетворяющего заданным требованиям. Т.е. это вся совокупность действий от получения априорной информации до создания работоспособной модели, определения оптимальных условий или реализации какой либо другой цели исследования с требуемой точностью. Отличительной особенностью планирования эксперимента является то, что исследуемые факторы изменяются не последовательно, а одновременно по заданным правилам, гарантирующим свойства оптимальности исследования.

Область действия - область возможных значений фактора X при экспериментировании.

Точка плана - упорядоченная совокупность численных значений факторов, соответствующая условиям проведения опыта. **Область планирования** - область, в которой находятся точки плана, удовлетворяющие заданному плану эксперимента. По диапазону изменения факторов различают планирование на $\omega_{ар}$ когда $X_{t \min} < X_f < X_{t \max}$ и планирование на кубе, когда $0 < X_i < X$.

Спектр плана - совокупность всех точек плана.

Матрица плана - стандартная форма записи условий проведения эксперимента в виде прямоугольной таблицы, строки которой соответствуют опытам, а столбцы уровням факторов.

Матрица спектра плана – матрица, составленная из всех строк матрицы плана, отличающихся уровнем хотя бы одного фактора.

Модель ситуации и предварительный анализ данных

В основе использования любого из методов теории планирования эксперимента лежат определенные предпосылки о свойствах изучаемых объектов. Это могут быть предположения о виде законов распределения изучаемых характеристик, их свойствах (независимость, однородность, область действия и т.п.), характере связей между ними. Вся совокупность таких предпосылок носит название модель ситуации. Если модель ситуации неизвестна или не соответствует действительности, то спланировать оптимальный эксперимент, который, как правило, базируется нам вполне конкретном способе статистической обработки, практически невозможно.

Модель ситуации может быть различной в зависимости от целей исследования, модели объекта исследования, выбранного критерия оптимальности. Например, ситуации типа: оценка состояния - прогноз состояния системы, предельный режим - оптимальный режим функционирования, интерполяция - экстраполяция значений параметров, требуют различных планов эксперимента для изучения одной и той же исследуемой системы. Таким образом на решение о выборе плана эксперимента и способе обработки полученных результатов существенно влияет результат анализа модели ситуации.

Обоснование выбора модели ситуации является фундаментом для проведения дальнейших исследований. Оно предполагает первоначальный анализ результатов представления имеющихся данных с целью выявления количественных особенностей поведения изучаемых параметров и выдвижения предварительных гипотез о поведении изучаемого объекта.

Первоначальный анализ данных должен быть по возможности простым. Он ориентирован на задачи компактного описания и сравнения Данных, выявления общих тенденций и подчеркивания особенностей в поведении объекта исследования.

Специфика медико-биологических данных и классические задачи статистического анализа.

Огромное количество программ, пакетов и систем созданных для статистической обработки данных с использованием ЭВМ, предоставляет пользователю такое разнообразие всевозможных подходов к планированию экспериментальных исследований и анализу их результатов, что одной из главных проблем становится выбор методов, позволяющих наиболее адекватно провести оценки измеряемых характеристик и выявить существующие в изучаемой системе закономерности. Ясное понимание возможностей, границ и особенностей применения существующих приемов обработки данных имеет важнейшее значение для правильной постановки эксперимента и последующей интерпретации его результатов. Прежде всего это касается медико-биологических исследований, где представлен чрезвычайно широкий спектр решаемых задач, как с точки зрения сложности и многоплановости изучения биологических систем и организма в целом, так и с точки зрения особенностей медицинской измерительной аппаратуры и приемов исследования.

Данные медицинского или биологического эксперимента являются достаточно специфичными. Можно выделить ряд особенностей, обусловленных как спецификой организации самого исследования, так и сложностью объекта изучения, которые непосредственно могут сказываться на величине получаемых оценок и определяют выбор метода статистического анализа.

Контрольные вопросы:

1. Планирование эксперимента как наука.
2. Основные понятия и принципы теории планирования эксперимента
3. Объектами исследования в теории планирования эксперимента являются
4. Модель ситуации и предварительный анализ данных
5. Специфика медико-биологических данных и классические задачи статистического анализа.

ТЕМА 9. СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ (НУЛЕВАЯ И АЛЬТЕРНАТИВНАЯ ГИПОТЕЗА)

Цель: Ознакомление студентов со статистическими гипотезами.

Задачи обучения: изучение статистических гипотез

Основные вопросы темы:

1. Сравнительные оценки генеральных параметров

2. Гиптезы и распределения

Выборочные характеристики являются оценками генеральных параметров, которые, как правило, остаются неизвестными. Там же описаны точечные и интервальные способы оценки неизвестных параметров по значениям выборочных характеристик.

Ниже будут обсуждаться сравнительные оценки генеральных параметров по разности, наблюдаемой между сравниваемыми выборками. Это важно, так как ни одно исследование не обходится без сравнений. Сравнить приходится данные опыта с контролем, урожайность одной культуры с урожайностью другой, продуктивность одной группы животных с продуктивностью другой и т. д.

О преимуществе той или иной из сравниваемых групп судят обычно по разности между средними долями и другими выборочными показателями - величинами случайными, сопровождаемыми ошибками репрезентативности. Вопрос о достоверности выборочной разности с ее ошибкой приходится решать исходя из той или иной гипотезы, т. е. предположения или допущения относительно параметров сравниваемых групп, которое выражено в терминах вероятности и может быть проверено по выборочным характеристикам. Часто необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но имеются основания предположить, что он имеет определённый вид (назовём его A), выдвигают гипотезу: генеральная совокупность распределена по закону A . Таким образом, в этой гипотезе речь идёт о виде предполагаемого распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр равен определённому значению, выдвигают гипотезу: $=$. Таким образом, в этой гипотезе речь идёт о предполагаемой величине параметра одного известного распределения. Возможны и другие гипотезы: о равенстве параметров двух или нескольких распределений о независимости выборок и многие другие.

Статистической называют гипотезу о виде неизвестного распределения, или о параметрах известных распределений. Например, статистическими являются гипотезы:

- 1) генеральная совокупность распределена по закону Пуассона;
- 2) дисперсии двух нормальных совокупностей равны между собой

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй – о параметрах двух известных распределений. Гипотеза «на Марсе есть жизнь» не является статистической, поскольку в ней не идёт речь ни о виде, ни о параметрах распределения. Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая гипотеза. По этой причине эти гипотезы целесообразно различать.

Нулевой (основной) называют выдвинутую гипотезу. Конкурирующей (альтернативной) называют гипотезу, которая противоречит нулевой. Например, если нулевая гипотеза состоит в предположении, что математическое ожидание нормального распределения равно 10, то конкурирующая гипотеза, в частности, может состоять в предположении, что Коротко это записывают так: $=10$;

Различают гипотезы, которые содержат только одно и более одного предположений.

Простой называют гипотезу, содержащую только одно предположение. Например, если μ - параметр показательного распределения, то гипотеза $\mu = 5$ – простая. Гипотеза $\mu > 3$ (математическое ожидание нормального распределения равно 3 (известно) – простая.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза $\mu > 5$ состоит из бесчисленного множества простых вида $\mu = k$, где k - любое число, большее 5. Гипотеза $\mu = 3$ (математическое ожидание нормального распределения равно 3 (неизвестно) – сложная.

Контрольные вопросы:

1. Выборочные характеристики
2. сравнительные оценки генеральных параметров
3. достоверности выборочной разности
4. генеральная совокупность распределена по закону Пуассона;
5. дисперсии двух нормальных совокупностей равны между собой

ТЕМА 10. ПАРАМЕТРИЧЕСКИЙ МЕТОД ОЦЕНКИ ДОСТОВЕРНОСТИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Цель: Закрепить у студентов знания о параметрических методах оценки достоверности статистических гипотез.

Задачи обучения: Изучение параметрических методов оценки достоверности статистических гипотез.

Основные вопросы темы:

Параметрические методы оценки достоверности статистических гипотез

Малые выборки. Приведенные выше формулы и построенные на основании их оценочные таблицы применимы только при наличии относительно большого (практически не менее 30) числа наблюдений. При меньшем количестве наблюдений (так называемая малая выборка), что довольно часто имеет место в клинических и экспериментальных работах, для оценки достоверности результатов прибегают к специальным таблицам. (Напомним, что среднеквадратическое отклонение при малых выборках вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum (v - \bar{x})^2}{n - 1}}.$$

Это объясняется тем, что при небольших выборках распределение выборочных средних систематически отклоняется от кривой нормального распределения. Большие отклонения от генеральной средней при малых выборках являются более вероятными, чем при больших выборках.

Английский ученый В.Госсет (Стьюдент) исследовал распределение t для малых выборок и установил формулу плотности этого распределения.

Для определения пределов колеблемости полученной по данным малой выборки средней величины и оценки достоверности различий, сравниваемых средних (относительных величин) используют специальную таблицу критерия t (Стьюдента).

В графах 2.3 и 4 таблицы помещены величины **доверительного коэффициента** (t), показывающие во сколько раз разность сравниваемых величин при данном малом числе наблюдений должна превышать свою среднюю ошибку для того, чтобы эта разность могла быть признана достоверной с данным уровнем вероятности, а результаты статистического исследования – достаточно надежными. Числа графы 2-й исчислены для вероятности правильного заключения равной 0,95 (95%) и вероятности ошибки – 0,05 (5%), числа графы 3-й – с соответствующими вероятностями 0,99 (99%) и 0,01(1%); числа графы 4-й соответственно – 99,9 и 0,1%. Практически достаточно пользоваться числами графы 3 и даже 2 и только в случае необходимости, особенно большой точности, прибегать к числам графы 4.

Значение коэффициента t (Стьюдента) зависит не только от вероятности (p), но и от объема выборки (при $n' = n-1$). Из таблицы видно, что чем меньше выборка, тем больше значение t .

Обращаться к таблице следует по графе 1, в которой указано число степеней свободы $n' = n-1$, т.е. числу проведенных наблюдений, уменьшенному на единицу. Так, например, если после 8 испытаний действия спинномозговой анестезии на уровень кровяного давления установлено, что средняя величина снижения кровяного давления составляла 5,75 мм при средней ошибке 0,65, то из таблицы t видно, что при $n' = 8-1 = 7$; $t = 2,36$ (графа 2 приложения 11). Это значит, что с вероятностью ошибки не более чем 5% можно утверждать, что размеры снижения кровяного давления при спинномозговой анестезии находятся в пределах $5,75 \pm 2,36 \times 0,65$, т.е. в пределах $5,75 \pm 1,53$ или 4,22 – 8,26 мм; с вероятностью ошибки не более чем 1% можно утверждать, что размеры снижения

кровенного давления в результате спинномозговой анестезии составляют $5,75 \pm 3,50 \times 0,65$ (графа 3 приложения 11) или $3,48 - 8,02$ мм.

Если оценивается достоверность разности коэффициентов или средних, т.е.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{m_1^2 + m_2^2}}, \text{ то } n_1 + n_2 - 2.$$

Среднее падение АД при спинномозговой анестезии $\bar{x}_1 = 5,75 \text{ мм}$, а при эфирном наркозе $\bar{x}_2 = 3,75 \text{ мм}$; случайна ли разность $\frac{\bar{x}_1 - \bar{x}_2}{\frac{\sigma}{\sqrt{n}}}$ или действительно эфирный наркоз вызывает меньшее падение АД, чем спинномозговая анестезия?

Таблица 31. Падение артериального давления в зависимости от вида обезбоживания

Вид обезбоживания	Падение артериального давления в мм во время опыта
Спинномозговая анестезия (v ₁)	6,5 7 4 8 3 8 5
Эфирный наркоз (v ₂)	2 3 4 2 7 5 4 3

Вычисления \bar{x} , σ и m_x для каждого ряда можно произвести обычным путем, но для упрощения расчетов можно использовать следующую формулу, удобную для применения при малых числах наблюдений:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(\sum v_1^2 - n_1 \bar{x}_1^2 + \sum v_2^2 - n_2 \bar{x}_2^2)(n_1 + n_2)}{(n_1 + n_2 - 2)n_1 \cdot n_2}}}$$

Упрощение расчетов при использовании этой формулы достигается тем, что вместо вычисления σ и m ограничиваются определением $\sum v^2$ для каждого ряда чисел, что значительно облегчает вычислительную работу (v – отдельные наблюдения, варианты). В данном примере:

$$\begin{aligned} \sum v_1^2 &= 6^2 + 5^2 + 7^2 + 4^2 + 8^2 + 5^2 = 288, \text{ а} \\ \sum v_2^2 &= 2^2 + 3^2 + 4^2 + 2^2 + 7^2 + 5^2 + 4^2 + 3^2 = 132. \end{aligned}$$

\bar{x}_1 и \bar{x}_2 как указано было выше, равняются соответственно 5,75 и 3,75; их квадраты $\bar{x}_1^2 = 5,75^2 = 33,06$, $\bar{x}_2^2 = 3,75^2 = 14,06$; $n_1 = 8$ и $n_2 = 8$.

Подставив все эти числа в приведенную выше формулу, получим:

$$t = \frac{5,75 - 3,75}{\sqrt{\frac{(288 - 8 \cdot 33,06 + 132 - 8 \cdot 14,06)(8 + 8)}{(8 + 8 - 2) \cdot 8 \cdot 8}}} = \frac{2}{\sqrt{\frac{704}{896}}} = 2,25.$$

Оценивая t по данным приложения 2, видим, что при $n' = 8 + 8 - 2 = 14$ в графе 2 этой таблицы стоит величина 2.14. Следовательно, для достоверности утверждения неслучайности различия величин \bar{x}_1 и \bar{x}_2 , с вероятностью ошибки не более чем 0,05 (не более чем 5%) достаточно, чтобы t было не менее чем 2.14. В данном примере $t = 2,25$. Значит, действие двух приведенных видов обезбоживания на снижение кровяного давления действительно различно и это различие может считаться статистически доказанным.

Оценка достоверности интенсивных коэффициентов заболеваемости при наличии повторных заболеваний.

$$m = \sqrt{\frac{pq}{n}}$$

Формула средней ошибки показателя пригодна для оценки показателей только в случаях так называемого альтернативного варьирования, т.е. тогда, когда возможны только два исхода (умер или не умер, заболел данной болезнью или не заболел, привить против данного заболевания или не привит и т.п.).

По этой формуле можно исчислять средние ошибки коэффициентов смертности, летальности, а также заболеваемости теми болезнями, которыми, как правило, можно заболеть только один раз (хронические болезни – коронаросклероз, злокачественные опухоли, острозаразные заболевания, дающие длительный иммунитет, и т.п.) в течение жизни или хотя бы только один раз за период наблюдения (обычно год).

Определить среднюю ошибку по указанной выше формуле для коэффициентов общей заболеваемости (т.е. заболеваемости всеми болезнями, вместе взятыми) или заболеваемости с временной утратой трудоспособности неправильно.

Практически в течение года человек может болеть несколько раз различными болезнями или даже одной и той же болезнью, длящейся относительно недолго и не дающей стойкого иммунитета (например, грипп, острый катар верхних дыхательных путей, ангина, пневмония и др.). Случаев временной нетрудоспособности в связи с заболеванием также может быть несколько за год у одного и того же работающего не только в связи с заболеваниями некоторыми острыми болезнями, но и по поводу обострений хронических болезней.

В таких случаях средние ошибки показателей заболеваемости следует рассчитывать по формуле средней ошибки средних величин, т.е.

$$m = \frac{\sigma}{\sqrt{n}}$$

, строя вариационные ряды, где вариантами являются числа заболеваний или случаев временной нетрудоспособности в связи с заболеванием в течение года (0;1;2;3;4 и т.д.), а частотами – числа болевших данное число раз.

Однако такие расчеты, правильные теоретически, трудно осуществимы на практике, так как требуют кропотливой работы по распределению наблюдаемой группы населения на не болевших ни разу за год, болевших один раз, два раза и т.д.

Трудность этой работы зачастую заставляет вовсе отказываться от расчета средних ошибок коэффициентов заболеваемости, а следовательно, и от статистической оценки достоверности их разности.

В подобной ситуации допустимо приближенное вычисление средней ошибки показателей, предложенное В.А.Мозгляковой. Исходя из предположения, что распределение по числу заболеваний во многих случаях близко к так называемому распределению Пуассона, при котором наибольшие частоты соответствуют не средним, а наименьшим вариантам, В.А.Мозглякова предложила в целях упрощения пользоваться расчетом средних квадратических отклонений и средних ошибок уровней заболеваемости по формулам, пригодны для распределения Пуассона, а именно:

$$\sigma^2 = \bar{x}, \quad m = \sqrt{\frac{\bar{x}}{n}}$$

Хорошее соответствие фактического распределения кратности заболеваний теоретическому распределению Пуассона имеет место при числе наблюдений 100-150 и средней величине коэффициента заболеваемости 1,0 на 1 человека. Если коэффициент больше 1.5, то рекомендуемым расчетом не следует пользоваться.

Описанные методы оценки достоверности результатов статистического исследования с помощью критерия t (критерий Стьюдента) в основном пригодны для так называемого нормального распределения, т.е. такого, при котором крайние значения (самые малые и

самые большие) вариант встречаются редко, а наиболее часты варианты. Близкие по своей величине к средней арифметической ряда; или для состояния (да-нет, жив-умер и т.п.). Методы оценки достоверности различия параметров таких вариационных рядов называются параметрическими.

Однако характер распределения медико-биологических явлений нередко отличается от нормального. Проводя новые исследования, врач-экспериментатор часто не знает, какому закону варьирования будут следовать результаты, полученные в нескольких опытах, а относительно небольшое число проведенных наблюдений не позволяет ему определить форму распределения. В этих случаях оценку достоверности следует производить с применением так называемых непараметрических критериев.

Контрольные вопросы:

1. Малые выборки.
2. Доверительного коэффициента
3. Достоверность разности коэффициентов или средних
4. Оценка достоверности интенсивных коэффициентов заболеваемости при наличии повторных заболеваний
5. Формула средней ошибки показателя
6. Описанные методы оценки достоверности

ТЕМА 11. НЕПАРАМЕТРИЧЕСКИЙ МЕТОД ОЦЕНКИ ДОСТОВЕРНОСТИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ.

Цель: Закрепить у студентов знания о непараметрических методах оценки достоверности статистических гипотез.

Задачи обучения: Изучение непараметрических методов оценки достоверности статистических гипотез

Основные вопросы темы:

Непараметрические методы оценки достоверности статистических гипотез

Исследователь не должен руководствоваться предположением, которое нельзя проверить. При использовании непараметрических критериев риск ошибок в выводах минимален.

Рассмотрены в предыдущих разделах статистические параметры (средняя арифметическая, среднее квадратичное отклонение, коэффициент вариации, средняя ошибка), которые используют для анализа вариационных рядов, являются его параметрами и требуют представления выходных данных в количественном виде. Однако при проведении медицинских исследований достаточно часто придется использовать методы статистического анализа данных, представленных в полуколичественном, полукачественном и качественном виде. Совокупность статистических методов, которые позволяют оценить их результаты как в количественном (числовом), так и в полуколичественном и качественном виде объединяют в группу непараметрических критериев оценивания. Использование непараметрических критериев не нуждается в расчете параметров вариационного ряда. Здесь имеет значение порядок расположения вариантов в совокупности. Статистическое оценивание наблюдений с помощью непараметрических критериев, как правило, проще, чем оценивание параметрическими методами и не требует громоздких расчетов

подавляющее большинство параметрических статистических методик предусматривают наличие нормального распределения варианты в исследуемой совокупности. Но на практике встречаются не только нормальные, но и другие виды разделы признаков. При наличии таких ситуаций использование параметрических критериев повышает вероятность ошибок. Практическое применение непараметрических критериев, не связанное с определенной формой распределения исследуемых признаков, делает целесообразным их самостоятельное использование или в комплексе с параметрическими.

Невзирая на определенную простоту методик, надежность непараметрических критериев достаточно высока. Они могут быть использованы для оценивания достоверности медико-биологических результатов одной совокупности, разницы, двух и больше выборочных совокупностей.

Критерий знаков и критерий Вилкоксона используют для оценки достоверности разницы двумя взаимосвязанными совокупностями.

Критерий знаков позволяет включать в анализ до 100 пар наблюдений и базируется на подсчете числа однонаправленных результатов при их парном сравнении.

В таблице 19 приведена динамика скорости оседания эритроцитов (СОЭ) за 10-тидневный период лечения.

Основные этапы расчета по критериям знаков

1. Определение направленности разницы в сравнительных группах результатов. Динамика при этом обозначается соответствующими знаками: +, —, =. Из дальнейшего расчета исключают результаты без динамики (=).

2. Подсчет числа наблюдений с позитивными и негативными результатами. Из 10 приведенных изменения оказались у 9 больных.

Таблица 1. Динамика скорости оседания эритроцитов|эритроцита| (СОЭ)

Больные (№ п/п)	СОЭ		Направление разницы
	1-й день	2-й день	
1	13	23	+
2	22	15	-
3	16	18	+
4	20	14	-
5	19	11	-
6	25	13	-
7	23	12	-
8	20	13	-
9	17	18	+
10	18	18	=

3. Подсчет| числа знаков, которые реже встречаются. Снижение СОЭ (-) обнаружено у 6 больных, а прирост (+) зарегистрирован в трех случаях.

4. Сравнение меньшего числа знаков (критерий Z) с табличными критическими значениями для соответствующего числа наблюдений. Для $n = 9$ определен критерий $Z = 3$ выше предельного табличного ($Z_{0,05} = 2$).

Следовательно, нельзя сделать **вывод** о существенности динамики скорости оседания эритроцитов вероятности погрешности больше 5 % ($p > 0,05$).

T-критерий Вилкоксона предусматривает возможность попарного сравнения от 6 до 25 пар наблюдений.

Критерий Вилкоксона целесообразно использовать в тех случаях, когда оказываются неоднозначные количественные изменения исследуемого параметра (снижение и повышение). При этом учитывают не только направленность разницы, но и ее величину.

Методика анализа по T-критерию| Вилкоксона приведена в таблице 20.

1. Определяют разницу|разность| в парах|паре| наблюдения между конечным|концевым| и начальным|первоначальным| уровнями артериального давления.

2. Ранжирование полученных результатов за величиной разницы|разности| между показателями без учета направленности изменений|смены|. Результаты без динамики исключают из|с| последующего оценивания. Если два результата| имеют одинаковые абсолютные значения изменений|смены|, их ранги определяют как полгруппы|построчных| номеров.

3. Подсчет|вычисление| суммы однозначных рангов (позитивных|положительных| и негативных).

4. Оценивание за меньшей суммой рангов путем сравнения определенного T-критерия| с табличным значением при соответствующем числе пар|пары| наблюдений.

Таблица 2 Уровень артериального давления у больных гипертонической болезнью до и после лечения (мм рт. ст.).

Больные	Уровень АД		Разница	Ранг разницы	Меньшая сумма рангов
	ДО ЛЕЧЕНИЯ	ПОСЛЕ ЛЕЧЕНИЯ			
В.	210	175	-35	6,5	4
Д.	180	180	0	-	1
К.	185	140	-45	9	
Р.	160	185	+25	4	
Н.	175	145	-30	5	
П.	190	150	-40	8	
А.	155	160	+5	1	
С.	180	160	-20	3	
Ю.	200	155	-45	9	
Т.	170	155	-15	2	
					T=5

Критерий Вилкоксона T=5 не превышает табличного значения для данного числа наблюдений - n=9, T_{0,05} = 6.

Следовательно, можно сделать вывод о существенности (статистическую значимость) динамики артериального давления у больных после лечения.

Методика расчета критерия Колмогорова – Смирнова (λ^2) осуществляется в несколько этапов и рассмотрена в таблице 21.

1. Числовые значения двух вариационных рядов объединяют в один вариационный ряд, варианты которого упорядочивают в порядке роста.
2. Определяют частоты вариантов для обеих групп наблюдений.
 1. Определяют накопленные частоты для обеих групп.
 3. Определяют накопленные частицы|долю|, для чего накопленные частоты делятся|делящийся| на число наблюдений для каждой группы.
 4. Рассчитывается разница|разность| накопленных частиц|доли| групп X и Y без учета знаков.
 2. Определяют максимальную разницу|разность| — D = 0,51.
 3. Определяют критерий λ^2

$$\lambda^2 = \frac{D^2 * n_x * n_y}{n_x + n_y}$$

8. Сравняют полученный результат с предельным значением критерия Колмогорова-Смирнова|.

Если λ^2 больше предельного значения, разница между сравниваемыми группами является существенной.

Для данного задания $\lambda^2 = 1,10$. Сравнивая полученный результат с предельным значением $\lambda^2_{0,05} = 1,84$ и $\lambda^2_{0,01} = 2,65$, делаем вывод о несущественности разницы между сравниваемыми группами.

Таблица 3. Изменение радиоактивности крови подопытных животных, получавших (X) и не получавших (Y) лечение (в условных единицах).

Варианты по ряду X и Y	Частоты вариант по группам		Накопленные частоты по группам		Накопленные части по группам		Разница
	x	y	x	Sy	x	y	
24	2	3	2	0	0.22	0	0.22
26	3	0	5	0	0.56	0	0.56
28	1	2	6	2	0.67	0.25	0.42
30	1	1	7	3	0.78	0.38	0.40
32	1	0	8	3	0.89	0.38	0.51
34	1	1	9	4	1	0.50	0.50
36	0	1	9	5	1	0.63	0.37
38	0	1	9	6	1	0.75	0.25
40	0	1	9	8	1	1	0
nx=9 ny=8							

Попарное сравнение показателей не позволяет получить обобщающую оценку. В другом случае необходимо провести сравнение совокупности не только по обобщающим показателям, но и по характеру распределения признаков в исследуемых группах.

При проведении статистического анализа иногда необходимо оценить достоверность разницы большей от двух количеству показателей клинико-статистических групп. В указанных ситуациях наиболее целесообразным является использование критерия соответствия — χ^2 (критерий Пирсона), который рассчитывается по формуле:

$$\chi^2 = \sum \frac{(p - p_1)^2}{p_1}$$

где p - истинные частоты

p₁ - теоретические частоты

В обобщенном виде практическое значение критерия соответствия χ^2 заключается в:

- оценивании достоверности разницы|разности| между несколькими сравниваемыми группами при нескольких возможных результатах с разной|различной| степенью вероятности (например, три или четыре группы больных с разными|различными| методами лечения и их последствиями — разной|различной| частотой|осложнений|усложнения);

- определении наличия связи между двумя возбудителями (зависимость результатов лечения от возраста|века| больных, тяжести заболевания, связь между тяжестью патологии новорожденных и состоянием|станом| их физического развития);

- оценивании идентичности распределения|деления| частот в двух и больше совокупности (аналогичность распределения|деления| больных по уровню клинических параметров при разных|различных| степенях тяжести патологии). На основе «нулевой» гипотезы определяют «ожидаемые» результаты и сравнивают их с фактическими данными. Если разницы|разности| нет, можно сделать|совершить| вывод, что «нулевая» гипотеза подтвердилась. При наличии отличий фактических данных от теоретического раздела | определяют существенность разницы|разности| между сравниваемыми| группами.

Оценивание результатов (X^2) проводится за специальной таблицей. Существенной считается разница в том случае, когда величина рассчитанного коэффициента превышает табличное значение при достоверности не ниже 95% (вероятность погрешности менее 5% - $p < 0,05$).

Методику расчета коэффициента соответствия — X^2 рассмотрим на примере оценивание влияния методов лечения на их результаты.

1. Приведем фактические результаты по трём методам| лечения (таблица 4).

Таблица 4. Результаты лечения больных по отдельным методикам

Методика лечения	Всего больных	Результаты лечения – Р (фактические данные)		
		хорошо	удовлетворительно	неудовлетворительно
I	50	36	11	3
II	80	48	17	15
III	70	25	25	20
Всего	200 (100%)	109	53	38

2. Рассчитываем «ожидаемые» результаты согласно «нулевой» гипотезе, основой|основанием| которой|какой| является доказательство, что разницы|разности| между результатами лечения за отдельными методиками нет. В этом случае за основу|основание| берем общее деление|разделение| больных, вылеченных всеми методами. Числовая характеристика «нулевой» гипотезы составляет: хорошие результаты в целом имели 54,5 %, удовлетворительные — 26,5 % | и неудовлетворительные — 19 % больных. В соответствии с|соответственно| указанным делением|разделением| определяют «ожидаемые» данные результатов лечения за отдельными методиками (значение определяем в целых числах) — таблица. 5.

Таблица 5.

Методика лечения	Всего больных	Результаты лечения – P1 (ожидаемые данные)		
		хорошо	удовлетворительно	неудовлетворительно
I	50	27	13	10
II	80	44	21	15
III	70	38	19	13

Всего	200 %	109 (54,5%)	53 (26,5%)	38 (19%)
-------	-------	----------------	------------	----------

3. Сопоставляем фактические и теоретические данные (их разницу) с расчетом величины отклонения и учетом его направления (знака) — таблица. 6.

Таблица 6.

Методика лечения	(P - P1)		
	хорошо	удовлетворительно	неудовлетворительно
I	9 (36-27)	-2 (11-13)	-7 (3-10)
II	4 (48-44)	-4 (17-21)	0 (15-15)
III	-13 (25-38)	6 (25-19)	7 (20-13)
Всего	0	0	0

4. Рассчитываем квадрат отклонения теоретических данных от фактических и средний квадрат отклонения на одну «ожидаемую» группу. Данный этап расчета имеет такой вид в связи с тем, что на основе фактических отклонений невозможно определить его суммарную величину, поскольку она равняется нулю. При возведении отклонений в квадрат определяем их параметры для каждой группы $(p - p_1)^2$. Учитывая разное число больных в исследуемых группах величина отклонений может быть разной, потому квадрат их делимо на число соответствующих наблюдений каждой группы — $(p - p_1)^2 / p_1$. Проведя расчеты, определяем $(p - p_1)^2$ и $(p - p_1)^2 / p_1$. (таблица 7).

Таблица 7.

Методика лечения	$(P - P1)^2$			$(P - P1)^2 / P1$		
	хорошо	удовлетворительно	неудовлетворительно	хорошо	удовлетворительно	неудовлетворительно
I	81	4	49	3	0,3	4,9
II	16	16	0	0,4	0,8	0
III	169	36	49	4,4	1,9	3,8
$\Sigma=19,5$						

5. Определяем X^2 — итог результатов последнего этапа расчетов. В нашем случае $X^2 = 19,5$. Сравниваем его с табличным значением, учитывая число степеней свободы (n^1), которые определяют по формуле:

$$n^1 = (S - 1) (r - 1),$$

где S — число групп больных (для нашего примера |приклада| - три).

r — число результативных групп (три).

|Число степеней свободы $n^1 = (3 - 1) (3 - 1) = 4$.

Полученный результат превышает табличное значение χ^2 для $n^1 = 4$ по всем уровням достоверности. Следовательно, мы можем сделать вывод о существенности (достоверность) разницы между показателями при разных методах лечения — «нулевая» гипотеза не подтвердилась.

Контрольные вопросы:

1. В каких случаях целесообразно использовать параметрические критерии?
2. Какие преимущества имеют непараметрические критерии?
3. Какие непараметрические критерии используют для оценивания достоверности разности |разнорезультатов исследования во взаимосвязанных совокупностях?
4. Какие непараметрические критерии используют для оценки достоверности разности|разность| результатов исследования в независимой совокупности?

ТЕМА 12.МЕТОД СТАНДАРТИЗАЦИИ

Цель: Ознакомление студентов с методом стандартизации

Задачи обучения: Изучение методов стандартизации

Основные вопросы темы:

1. Применение методов стандартизации
2. Методы стандартизации
3. Этапы расчета стандартизованных показателей

При изучении общественного здоровья и здравоохранения в научных или практических целях исследователю нередко приходится доказывать влияние факторных признаков на результативные при сравнении двух или более совокупностей. С этой целью применяется целый ряд статистических приемов.

При сравнении двух неоднородных совокупностей по какому-либо признаку (составу) применяются методы стандартизации (прямой, обратный, косвенный).

Этот метод применяется при наличии полных сведений как о составе сравниваемых совокупностей, так и о распределении в них явления.

Прямой метод стандартизации

1. **Условия применения метода стандартизации.** Метод применяется при сравнении интенсивных показателей в совокупностях, отличающихся по составу (например, по возрасту, полу, профессиям и т.д.).

2. **Сущность метода стандартизации.** Он позволяет устранить (элиминировать) возможное влияние различий в составе совокупностей по какому-либо признаку на величину сравниваемых интенсивных показателей. С этой целью составы совокупностей по данному признаку уравниваются, что в дальнейшем позволяет рассчитать стандартизованные показатели.

3. **Стандартизованные показатели** — это условные, гипотетические величины, они не отражают истинных размеров явлений. Стандартизованные показатели свидетельствуют о том, каковы были бы значения сравниваемых интенсивных показателей, если бы были исключены различия в составах совокупностей.

4. **Назначение метода стандартизации.** Метод стандартизации применяется для выявления влияния фактора неоднородности составов совокупностей по какому-либо признаку на различия сравниваемых интенсивных показателей.

Этапы расчета стандартизованных показателей

• **I этап.** Расчет общих и частных интенсивных показателей:

- общих — по совокупностям в целом;
- частных — по признаку различия (полу, возрасту, стажу работы и т.д.).

• **II этап.** Определение стандарта, т.е. выбор одинакового численного состава среды по данному признаку (по возрасту, полу и т.д.) для сравниваемых совокупностей. Как правило, за стандарт принимается сумма или полусумма численностей составов соответствующих групп. В то же время стандартом может стать состав любой из сравниваемых совокупностей, а также состав по аналогичному признаку какой-либо другой совокупности. Например, при сравнении летальности в конкретной больнице по двум отделениям скорой помощи за стандарт может быть выбран состав больных любой другой больницы скорой помощи. Таким образом, так или иначе уравниваются условия среды, что дает возможность провести расчеты новых чисел явления, называемых "ожидаемыми величинами".

• **III этап.** Вычисление ожидаемых абсолютных величин в группах стандарта на основе групповых интенсивных показателей, рассчитанных на I этапе. Итоговые числа по сравниваемым совокупностям являются суммой ожидаемых величин в группах.

• **IV этап.** Вычисление стандартизованных показателей для сравниваемых совокупностей.

• **V этап.** Сопоставление соотношений стандартизованных и интенсивных показателей, формулировка вывода.

Задача – эталон

Задание. Используя метод стандартизации при сравнении уровней детальности в больницах А и Б, сделайте соответствующие выводы.

Возраст больных (в годах)	Больница А		Больница Б	
	число выбывших больных	из них умерло	число выбывших больных	из них умерло
До 40	600	12	1400	42
От 40 до 59	200	8	200	10
От 60 и старше	1200	60	400	24
Всего:	2000	80	2000	76

Этапы расчета стандартизованных показателей

I этапэап. Сначала определяют общие показатели летальности в больницах А и Б.
 Больница А: $80 \times 100 / 2000 = 4$ на 100 выбывших больных;
 Больница Б: $76 \times 100 / 2000 = 3,8$ на 100 выбывших больных.

Затем находят показатели летальности в зависимости от возраста больных. Например, в больнице А у больных в возрасте до 40 лет летальность составляет $12 \times 100 / 600 = 2\%$, а в больнице Б, соответственно, $42 \times 100 / 1400 = 3\%$.

Аналогично проводят расчеты и в других возрастных группах (см. сводную таблицу — I этап).

II этап. За стандарт принимают сумму выбывших больных по каждой возрастной группе в обеих больницах.

Возраст больных (в годах)	Число больных в больницах А и Б	Стандарт
До 40	600 + 1400	2000
От 40 до 59	200 + 200	400
От 60 и старше	1200 + 400	1600
Всего:	2000 + 2000	4000

III этап. Определяют ожидаемое число умерших в стандарте по каждой возрастной группе в больницах А и Б, с учетом соответствующих показателей летальности:

Возраст до 40 лет:	Возраст от 40 до 59:	Возраст 60 лет и старше:
Больница А. 100 — 2 2000 - X $X = 2 \times 2000 / 100 = 40$ Больница Б. 100 — 3 2000-X $X = 3 \times 2000 / 100 = 60$	Больница А. 100 — 4 400-X $X = 4 \times 400 / 100 = 16$ Больница Б. 100 — 5 400-X $X = 5 \times 400 / 100 = 20$	Больница А. 100 — 5 1600 - X $X = 5 \times 1600 / 100 = 80$ Больница Б. 100 — 6 1600 - X $X = 6 \times 1600 / 100 = 96$

Находят сумму ожидаемых чисел умерших в стандарте в больнице А ($40 + 16 + 80 = 136$) и больнице Б ($60 + 20 + 96 = 176$).

IV этап. Определяют общие стандартизованные показатели травматизма в больницах А и Б.
 Больница А. $136 \times 100 / 4000 = 3,4$ на 100 выбывших больных;
 Больница Б. $176 \times 100 / 4000 = 4,4$ на 100 выбывших больных.
 Результаты поэтапного расчета стандартизованных показателей летальности оформляют в виде таблицы:

Возраст больных (в годах)	Больница А		Больница Б		I этап		II этап стандарт (сумма составов больных обеих больниц)	III этап		
	выбыло больных	из них умерло	выбыло больных	из них умерло	летальность на 100 выбывших больных			ожидаемое число умерших в стандарте	б-ца А	б-ца Б
					б-ца А	б-ца Б				
До 40 лет	600	12	1400	42	2	3	2000	40	60	
От 40 до 59	200	8	200	10	4	5	400	16	20	
60 и старше	1200	60	400	24	5	6	1600	80	96	
Всего:	2000	80	2000	76	4,5	3,8	4000	136	176	
IV этап. Определение стандартизованных показателей							100	3,4	4,4	

V этап. Сопоставление соотношения интенсивных и стандартных показателей летальности в больницах А и Б.

Показатели:	Больница А	Больница Б	Соотношение А и Б
Интенсивные	4,0	3,8	А>Б
Стандартные	3,4	4,4	А<Б

Выводы

1. Уровень летальности в больнице А выше, чем в больнице Б.
2. Однако если бы возрастной состав выживших больных в этих больницах был одинаков, то летальность была бы выше в больнице Б.
3. Следовательно, на различия в уровнях летальности (в частности, на "завышение" ее в больнице А и "занижение" в больнице Б) оказала влияние неоднородность возрастного состава больных, а именно, преобладание в больнице А пожилых пациентов (60 лет и более) с относительно высоким показателем летальности, и наоборот, в больнице Б — больных в возрасте до 40 лет, имеющих низкие показатели летальности.

Контрольные вопросы:

1. Методы стандартизации
2. Прямой метод стандартизации
3. Условия применения метода стандартизации
4. Сущность метода стандартизации.
5. Стандартизованные показатели
6. Назначение метода стандартизации.
7. Этапы расчета стандартизованных показателей

ТЕМА 13.РАНГОВЫЙ МЕТОД И МЕТОД КВАДРАТОВ КОРРЕЛЯЦИОННОГО АНАЛИЗА.

Цель: Ознакомление студентов с ранговым методом и методом квадратов корреляционного анализа.

Задачи обучения: Изучение рангового метода корреляции.

Основные вопросы темы:

1. Корреляционная связь
2. Коэффициент ранговой корреляции Спирмена
3. Назначение рангового коэффициента корреляции

Ранговая корреляция – метод корреляционного анализа, отражающий отношения переменных, упорядоченных по возрастанию их значения. Наиболее часто ранговая корреляция применяется для анализа связи между признаками, измеряемыми в порядковых шкалах (см. шкалы измерительные), а также как один из методов определения корреляции качественных признаков. Достоинством коэффициентов ранговой корреляции является возможность их использования независимо от характера распределения коррелирующих признаков.

В практике наиболее часто применяются такие ранговые меры связи, как коэффициенты ранговой корреляции Спирмена и Кендалла. Первым этапом расчета коэффициентов ранговой корреляции является ранжирование рядов переменных (табл. 1). Процедура ранжирования начинается с расположения переменных по возрастанию их значений. Разным значениям присваиваются ранги, обозначаемые натуральными числами. Если встречается несколько равных по значению переменных, им присваивается усредненный ранг.

Таблица 1 -Ранжирование распределения показателей теста (n = 18)

Тестовая оценка	Порядковый номер	Ранг
20	1	1
17	2	2
16	3]	3,5
16	4]	
15	5	5
14	6]	7,5
14	7]	
14	8]	
14	9]	
12	10	10
10	11	11
9	12	12
7	13]	14,5
7	14]	
7	15]	
7	16]	
5	17	17
3	18	18

В таблице приведены данные для расчета коэффициентов ранговой корреляции. Во второй графе представлены ранжированные показатели по первому из сравниваемых распределений (оценка IQ, в третьей графе – соответствующие им данные теста зрительной памяти).

Коэффициент корреляции рангов Спирмена (r_s) определяется из уравнения:

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n},$$

где d_i – разности между рангами каждой переменной из пар значений X и Y;
 n – число сопоставляемых пар.

Используя данные таблицы 2, получаем:

$$r_s = 1 - \frac{6 \cdot 77,5}{10(100 - 1)} = 0,53.$$

Коэффициент корреляции рангов Кендалла τ определяется следующей формулой:

$$\tau = \frac{P - Q}{\frac{1}{2}n(n-1)},$$

где P и Q рассчитываются по таблице 12.

Так, в восьмой графе подсчитывается, начиная с первого объекта X, сколько раз его ранг по Y меньше, чем ранг объектов, расположенных ниже.

Соответственно, в девятой графе (S2) фиксируется, сколько раз ранг Y больше, чем ранги, стоящие ниже его в столбце X. Подставляя эти данные в формулу, получаем:

$$\tau = \frac{27 - 11}{\frac{10}{2}(10 - 1)} = 0,36.$$

При сопоставлении приведенных коэффициентов оказывается, что коэффициент τ более информативен, чем r_s , и рассчитывается проще. Поэтому на практике при расчете ранговой корреляции отдают предпочтение коэффициенту τ (табл. 2).

Таблица 2

Номер Испытуемого	IQ-оценка (X)	Зрительная память (X)	Ранг X	Ранг Y	d_i	d_i^2	P	Q
1	1,20	15	1	4	-3	9	5	2
2	1,00	15	2,5	4	-1,5	2,25	5	2
3	1,00	18	2,5	1	1,5	2,25	7	0
4	0,91	15	4,5	4	0,5	0,25	5	1
5	0,91	13	4,5	9	-4,5	20,25	0	3
6	0,90	13	6	9	-3	9	0	3
7	0,88	17	7	2	5	25	3	0
8	0,86	14	8	6,5	1,5	2,25	1	0
9	0,76	14	9	6,5	2,5	6,25	1	0
10	0,75	13	10	9	1	1	0	0
—	—	—	—	—	—	$\sum d_i^2 = 77,5$	$S_1 = 27$	$S_2 = 11$

Распределение IQ-оценок и показателей теста зрительной памяти

Метод наименьших квадратов (МНК, англ. OrdinaryLeastSquares, OLS) — математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных. Он может использоваться для «решения» переопределенных систем уравнений (когда количество уравнений превышает количество неизвестных), для поиска решения в случае обычных (не переопределенных) нелинейных систем уравнений, для аппроксимации точечных значений некоторой функции. МНК является одним из базовых методов регрессионного анализа для оценки неизвестных параметров регрессионных моделей по выборочным данным.

Контрольные вопросы:

1. Ранговая корреляция
2. Коэффициенты ранговой корреляции Спирмена и Кендалла.
3. Ранжирование распределения показателей теста
4. Коэффициент корреляции рангов Кендалла
5. Распределение IQ-оценок и показателей теста зрительной памяти

Цель: Ознакомление студентов с регрессионным анализом.

Задачи обучения: Формирование представления о регрессии

Основные вопросы темы:

1. Определение регрессионного анализа
2. Основная задача регрессионного анализа
3. Алгоритмы пошагового регрессионного анализа
4. Интервальная оценка для уравнения регрессии

Регрессионный анализ — это статистический метод исследования зависимости случайной величины y от переменных (аргументов) x_j ($j = 1, 2, \dots, k$), рассматриваемых в регрессионном анализе как неслучайные величины независимо от истинного закона распределения x_j .

Обычно предполагается, что случайная величина y имеет нормальный закон распределения с условным математическим ожиданием $\bar{y} = \varphi(x_1, \dots, x_k)$, являющимся функцией от аргументов x_j и с постоянной, не зависящей от аргументов дисперсией σ^2 .

Для проведения регрессионного анализа из $(k + 1)$ -мерной генеральной совокупности $(y, x_1, x_2, \dots, x_j, \dots, x_k)$ берется выборка объемом n , и каждое i -е наблюдение (объект) характеризуется значениями переменных $(y_i, x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{ik})$, где x_{ij} — значение j -й переменной для i -го наблюдения ($i = 1, 2, \dots, n$), y_i — значение результативного признака для i -го наблюдения.

Наиболее часто используемая множественная линейная модель регрессионного анализа имеет вид

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik} + \epsilon_i \quad (53.8)$$

где β_j — параметры регрессионной модели;

ϵ_j — случайные ошибки наблюдения, не зависящие друг от друга, имеют нулевую среднюю и дисперсию σ^2 .

Отметим, что модель (53.8) справедлива для всех $i = 1, 2, \dots, n$, линейна относительно неизвестных параметров $\beta_0, \beta_1, \dots, \beta_j, \dots, \beta_k$ и аргументов.

Как следует из (53.8), коэффициент регрессии β_j показывает, на какую величину в среднем изменится результативный признак y , если переменную x_j увеличить на единицу измерения, т.е. является нормативным коэффициентом.

В матричной форме регрессионная модель имеет вид

$$Y = X\beta + \epsilon, \quad (53.9)$$

где Y — случайный вектор-столбец размерности $n \times 1$ наблюдаемых значений результативного признака (y_1, y_2, \dots, y_n) ; X — матрица размерности $n \times (k + 1)$ наблюдаемых значений аргументов, элемент матрицы x_{ij} рассматривается как неслучайная величина ($i = 1, 2, \dots, n; j = 0, 1, \dots, k; x_{i0} = 1$); β — вектор-столбец размерности $(k + 1) \times 1$ неизвестных, подлежащих оценке параметров модели (коэффициентов регрессии); ϵ — случайный вектор-столбец размерности $n \times 1$ ошибок наблюдений (остатков). Компоненты вектора ϵ_i не зависят друг от друга, имеют нормальный закон распределения с нулевым математическим ожиданием ($M\epsilon_i = 0$) и неизвестной постоянной σ^2 ($D\epsilon_i = \sigma^2$).

На практике рекомендуется, чтобы значение n превышало k не менее чем в три раза.

В модели (53.9)

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{i1} & \dots & x_{ik} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{pmatrix}.$$

В первом столбце матрицы X указываются единицы при наличии свободного члена в модели (53.8). Здесь предполагается, что существует переменная x_0 , которая во всех наблюдениях принимает значения, равные единице.

Основная задача регрессионного анализа заключается в нахождении по выборке объемом n оценки неизвестных коэффициентов регрессии $\beta_0, \beta_1, \dots, \beta_k$ модели (53.8) или вектора β в (53.9).

Так как в регрессионном анализе x_j рассматриваются как неслучайные величины, а $M\varepsilon_i = 0$, то согласно (53.8) уравнение регрессии имеет вид

$$\tilde{y}_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik}, \quad (53.10)$$

для всех $i = 1, 2, \dots, n$, или в матричной форме:

$$\tilde{Y} = X\beta, \quad (53.11)$$

где \tilde{Y} — вектор-столбец с элементами $\tilde{y}_1, \dots, \tilde{y}_i, \dots, \tilde{y}_n$.

Для оценки вектора-столбца β наиболее часто используют метод наименьших квадратов, согласно которому в качестве оценки принимают вектор-столбец b , который минимизирует сумму квадратов отклонений наблюдаемых значений y_i от модельных значений \tilde{y}_i , т.е. квадратичную форму:

$$Q = (Y - X\beta)^T (Y - X\beta) = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \rightarrow \min_{\beta_0, \beta_1, \dots, \beta_k},$$

где символом «Т» обозначена транспонированная матрица.

Наблюдаемые и модельные значения результативного признака y показаны на рис. 53.1.

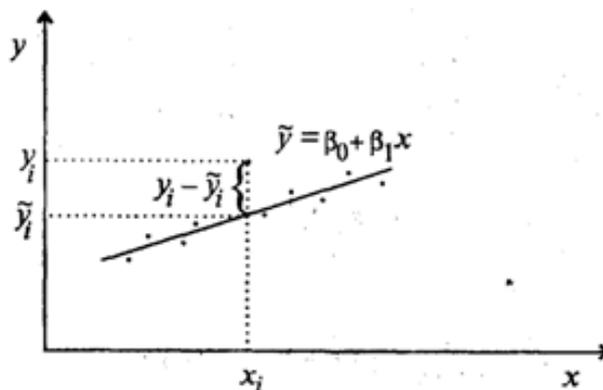


Рис. 53.1. Наблюдаемые и модельные значения результивного признака у

Дифференцируя, с учетом (53.11) и (53.10), квадратичную форму Q по $\beta_0, \beta_1, \dots, \beta_k$ и приравнивая частные производные к нулю, получим систему нормальных уравнений

$$\frac{\partial Q}{\partial \beta_j} = 0 \text{ для всех } j = 0, 1, \dots, k,$$

решая которую получим вектор-столбец оценок b , где $b = (b_0, b_1, \dots, b_k)^T$. Согласно методу наименьших квадратов, вектор-столбец оценок коэффициентов регрессии получается по формуле

$$b = (X^T X)^{-1} X^T Y, \quad (53.12)$$

где $b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_j \\ \vdots \\ b_k \end{pmatrix};$

X^T — транспонированная матрица X ;
 $(X^T X)^{-1}$ — матрица, обратная матрице $X^T X$.

Зная вектор-столбец b оценок коэффициентов регрессии, найдем оценку \hat{y}_i уравнения регрессии

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_k x_{ik}, \quad (53.13)$$

или в матричном виде:

$$\hat{y} = Xb,$$

где $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$.

Оценка ковариационной матрицы вектора коэффициентов регрессии b определяется выражением

$$S(b) = \hat{s}^2 (X^T X)^{-1}, \quad (53.14)$$

где

$$\hat{s}^2 = \frac{1}{n-k-1} (Y - Xb)^T (Y - Xb). \quad (53.15)$$

Учитывая, что на главной диагонали ковариационной матрицы находятся дисперсии коэффициентов регрессии, имеем

$$\hat{s}_{b_{(j-1)}} = \hat{s}^2 \left[(X^T X)^{-1} \right]_{jj}, \quad j = 1, 2, \dots, k+1. \quad (53.16)$$

Значимость уравнения регрессии, т.е. гипотеза $H_0: \beta = 0$ ($\beta_0 = \beta_1 = \beta_k = 0$), проверяется по F -критерию, наблюдаемое значение которого определяется по формуле

$$F_{\text{набл}} = \frac{Q_R / (k+1)}{Q_{\text{ост}} / (n-k-1)}, \quad (53.17)$$

где $Q_R = (Xb)^T (Xb)$,

$$Q_{\text{ост}} = (Y - Xb)^T (Y - Xb) = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

По таблице F -распределения для заданных α , $\nu_1 = k+1$, $\nu_2 = n-k-1$ находят $F_{\text{кр}}$.

Гипотеза H_0 отклоняется с вероятностью α , если $F_{\text{набл}} > F_{\text{кр}}$. Из этого следует, что уравнение является значимым, т.е. хотя бы один из коэффициентов регрессии отличен от нуля.

Для проверки значимости отдельных коэффициентов регрессии, т.е. гипотезы $H_0: \beta_j = 0$, где $j = 1, 2, \dots, k$, используют t -критерий и вычисляют $t_{\text{набл}}(b_j) = b_j / \hat{s}_{b_j}$. По таблице t -распределения для заданного α и $\nu = n-k-1$ находят $t_{\text{кр}}$.

Гипотеза H_0 отвергается с вероятностью α , если $t_{\text{набл}} > t_{\text{кр}}$. Из этого следует, что соответствующий коэффициент регрессии β_j значим, т.е. $\beta_j \neq 0$. В противном случае коэффициент регрессии незначим и соответствующая переменная в модель не включается. Тогда реализуется алгоритм пошагового регрессионного анализа, состоящий в том, что исключается одна из незначительных переменных, которой соответствует минимальное по абсолютной величине значение $t_{\text{набл}}$. После этого вновь проводят регрессионный анализ с числом факторов, уменьшенным на единицу. Алгоритм заканчивается получением уравнения регрессии со значимыми коэффициентами.

Существуют и другие алгоритмы пошагового регрессионного анализа, например с последовательным включением факторов.

Наряду с точечными оценками b_j генеральных коэффициентов регрессии β_j регрессионный анализ позволяет получать и интервальные оценки последних с доверительной вероятностью γ .

Интервальная оценка с доверительной вероятностью γ для параметра β_j имеет вид

$$b_j - t_{\alpha} \hat{s}_{b_j} \leq \beta_j \leq b_j + t_{\alpha} \hat{s}_{b_j} \quad (53.19)$$

где t_{α} находят по таблице t -распределения при вероятности $\alpha = 1 - \gamma$ и числе степеней свободы $\nu = n - k - 1$.

Интервальная оценка для уравнения регрессии \tilde{y} в точке, определяемой вектором-столбцом начальных условий $X^0 = (1, x^1, x^2, \dots, x^k)^T$ записывается в виде

$$\tilde{y} \in \left[(X^0)^T b \pm t_{\alpha} \hat{s} \sqrt{(X^0)^T (X^T X)^{-1} X^0} \right]. \quad (53.20)$$

Интервал предсказания \tilde{y}_{n+1} с доверительной вероятностью γ определяется как

$$\tilde{y} \in \left[(X^0)^T b \pm t_\alpha \hat{s} \sqrt{(X^0)^T (X^T X)^{-1} X^0 + 1} \right], \quad (53.21)$$

где t_α определяется по таблице t -распределения при $\alpha = 1 - \gamma$ и числе степеней свободы $\nu = n - k - 1$.

По мере удаления вектора начальных условий x^0 от вектора средних \bar{x} ширина доверительного интервала при заданном значении γ будет увеличиваться (рис. 53.2), где $\bar{x} = (1, \bar{x}_1, \dots, \bar{x}_k)$.

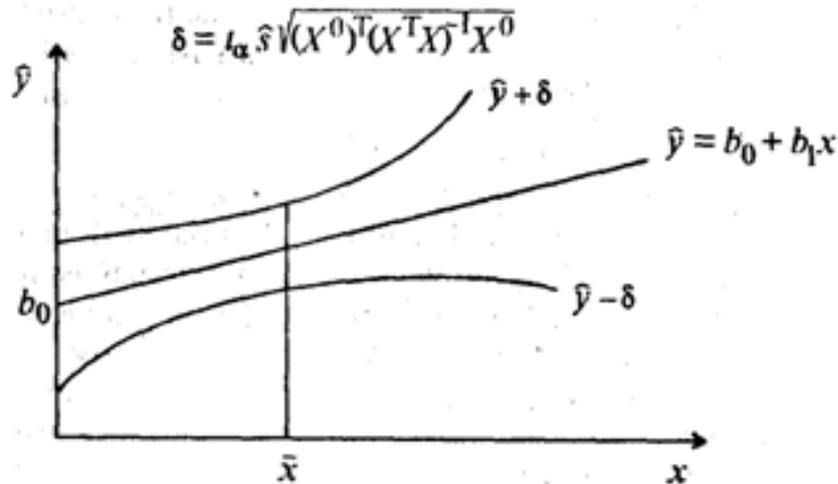


Рис. 53.2. Точечная \hat{y} и интервальная $[\hat{y} - \delta < \tilde{y} < \hat{y} + \delta]$ оценки уравнения регрессии $\tilde{y} = \beta_0 + \beta_1 x$.

Контрольные вопросы:

1. Регрессионный анализ
2. Значение результативного признака
3. Линейная модель регрессионного анализа
4. Основная задача регрессионного анализа
5. Наблюдаемые и модельные значения результативного признака
6. Алгоритмы пошагового регрессионного анализа

ТЕМА 15. ДИСПЕРСИОННЫЙ АНАЛИЗ В МЕДИЦИНЕ И ЗДРАВООХРАНЕНИИ.

Цель: Ознакомление студентов с дисперсионным анализом в медицине и здравоохранении.

Задачи обучения: Закрепление у студентов знания о принципах применения дисперсионного анализа

Основные вопросы темы:

1. Дисперсионный анализ
2. Методы дисперсионного анализа
3. Условия применения дисперсионного анализа

Дисперсионным анализом называют группу статистических методов, разработанных английским математиком и генетиком Р. Фишером в 20-х годах XX-го века для ряда экспериментальных задач биологии и сельского хозяйства.

Постановка задачи. Пусть даны генеральные совокупности X_1, X_2, \dots, X_k , где:

- все «*k*» генеральных совокупностей распределены нормально;
- дисперсии всех генеральных совокупностей одинаковы.

При этих условиях и заданном уровне значимости «*p*» требуется проверить нулевую

гипотезу о равенстве выборочных средних, т.е. $H_0: \bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_k$.

Каждая из генеральных совокупностей подвержена влиянию одного или нескольких *факторов*, которые могут изменять их средние значения.

Фактором называется показатель, который оказывает влияние на конечный результат.

Конкретную реализацию фактора называют *уровнем фактора*.

Значение измеряемого признака называют *откликом* на фактор.

Например, некоторое количество больных гипертонией разбиты случайным образом на «*k*» групп, каждой из которых назначен прием определенного лекарства. В результате контролируется среднее значение показателя изменения артериального давления.

В данном примере:

- значения показателя в «*i*»-ой группе, состоящей из «*n_i*» больных – это «*i*»-я выборка объема «*n_i*»;

- лекарство - это *фактор*, влияющий на величину контролируемого показателя;

- показатель изменения артериального давления - это *отклик* на воздействие фактора.

Предполагается, что по группам принимаемые лекарства различаются либо видом, либо дозой, либо еще каким-либо образом. Тогда воздействующий фактор подразделяется на некоторые составляющие, называемые *уровнями фактора*.

Если фактор оказывает воздействие на величину отклика, то нулевая гипотеза о равенстве средних $H_0: \bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_k$ отвергается.

В зависимости от количества изучаемых факторов дисперсионный анализ делится на *однофакторный* и *многофакторный*.

В примере с изменением артериального давления можно исследовать:

- фактор времени года (уровни: зима, весна, лето, осень);
- фактор места эксперимента (уровни: лечение в стационаре или дома);
- фактор режима (уровни: постельный, обычный или регулярные пешие прогулки на свежем воздухе) и т.п.

Выборочные данные для однофакторного дисперсионного анализа оформляют в виде таблицы (таблица 1).

Таблица 1.

Номер испытания	Уровень фактора А				
	A ₁	A ₂	...	A _k	
		x ₁₁	x ₁₂	...	x _{1k}
		x ₂₁	x ₂₂	...	x _{2k}

...
r	x_{r1}	x_{r2}	...	x_{rk}
Групповая средняя	$\bar{x}_{гр1}$	$\bar{x}_{гр2}$...	$\bar{x}_{грk}$

2. Основная цель дисперсионного анализа состоит в разбиении выборочной дисперсии на две компоненты:

- первая – это *факторная дисперсия*, она соответствует влиянию фактора на изменчивость средних значений;
- вторая – это *остаточная дисперсия*, она обусловлена случайными причинами и не влияет на изменчивость средних значений.

Для численной оценки влияния исследуемого фактора используют сравнение этих компонент с помощью F -критерия Фишера.

Факторная дисперсия ($S_{факт}^2$) – это дисперсия, которая соответствует влиянию фактора на изменение средних значений выборки:

$$S_{факт}^2 = \frac{SS_{факт}}{k-1} = \frac{r \sum_{j=1}^k (\bar{x}_{грj} - \bar{x})^2}{k-1}, \quad (4.1)$$

где $SS_{факт}$ - факторная сумма квадратов отклонений, k - количество уровней фактора, r - количество значений в каждой группе, \bar{x} - общая средняя, $\bar{x}_{гр}$ - групповая средняя.

Остаточная дисперсия ($S_{ост}^2$) – это дисперсия, возникающая по случайным причинам и не влияющая на изменение средних значений выборки:

$$S_{ост}^2 = \frac{SS_{ост}}{k(r-1)} = \frac{\sum_{j=1}^r (x_{j1} - \bar{x}_{гр1})^2 + \sum_{j=1}^r (x_{j2} - \bar{x}_{гр2})^2 + \dots + \sum_{j=1}^r (x_{jk} - \bar{x}_{грk})^2}{k(r-1)}, \quad (4.2)$$

где $SS_{ост}$ - остаточная сумма квадратов отклонений.

Общая дисперсия ($S_{общ}^2$) – это сумма факторной и остаточной дисперсий:

$$S_{общ}^2 = \frac{SS_{общ}}{r-1} = \frac{\sum_{j=1}^k \sum_{i=1}^r (x_{ji} - \bar{x})^2}{r-1}, \quad (4.3)$$

где $SS_{общ} = SS_{факт} + SS_{ост}$.

3. Схема применения однофакторного дисперсионного анализа:

1) $H_0: \bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_k$.

$H_1: \bar{x}_1 \neq \bar{x}_2 \neq \dots \neq \bar{x}_k$.

2) $p=0,05$ - уровень значимости.

$$3) F_{\text{факт}} = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}} \quad (4.4)$$

$$4) F_{\text{крит}}(p, f_1, f_2),$$

где $f_1=k-1$, $f_2=k(r-1)$ – число степеней свободы (табличные значения), k - количество уровней фактора, r - количество значений в каждой группе.

5) Если $F_{\text{факт}} \leq F_{\text{крит}}$, то « H_0 » принимается.

Если $F_{\text{факт}} > F_{\text{крит}}$, то « H_0 » отвергается.

Пример 4.1. Среди взрослого населения определенной возрастной категории фиксировалось число заболеваний дыхательных путей за два года. Цель исследования – статистическое доказательство влияния курения на заболеваемость дыхательных путей.

Случайным образом были отобраны 3 группы по 4 человека каждая, из них: 1 группа - некурящие, 2 группа - стаж курильщика - до 5 лет, 3 группа - стаж курильщика более 5 лет.

Таким образом, исследуемый фактор « A » - курение, уровни фактора, A_1 , A_2 , A_3 - стаж курильщика. Отклик на фактор курения - число заболеваний дыхательных путей.

Были получены 12 значений числа заболеваний – x_{ij} , где j - номер уровня фактора ($j=1, 2, 3$), i - номер элемента в соответствующей выборке (группе), $i=1, 2, 3, 4$:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix} = \begin{bmatrix} 1 & 3 & 3 \\ 0 & 2 & 4 \\ 1 & 2 & 5 \\ 2 & 1 & 3 \end{bmatrix}$$

Предполагаем, что $\{x_{ij}\}$ - выборка из нормальной генеральной совокупности.

Все данные необходимо занести в таблицу:

Номер испытания	Уровень фактора « A »		
	A_1	A_2	A_3
Групповая средняя	4/4=1	8/4=2	15/4=3,75

Решение.

$$1) H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3$$

$$H_1: \bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3$$

2) $p=0,05$ - уровень значимости.

3) Вычисляется:

3.1) Общая средняя:

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} = \frac{1 + 2 + 3,75}{3} = 2,25$$

3.2) Факторная сумма квадратов отклонений:

$$SS_{\text{факт}} = r \sum_{j=1}^k (\bar{x}_{\text{групп}} - \bar{x})^2 = 4 \cdot [(1 - 2,25)^2 + (2 - 2,25)^2 + (3,75 - 2,25)^2] = 15,5$$

3.3) Остаточная сумма квадратов отклонений:

$$\begin{aligned} SS_{\text{ост}} &= \sum_{i=1}^4 (x_{i1} - \bar{x}_{\text{групп1}})^2 + \sum_{i=1}^4 (x_{i2} - \bar{x}_{\text{групп2}})^2 + \sum_{i=1}^4 (x_{i3} - \bar{x}_{\text{групп3}})^2 = \\ &= [(1-1)^2 + (0-1)^2 + (1-1)^2 + (2-1)^2] + [(3-2)^2 + (2-2)^2 + (2-2)^2 + (1-2)^2] + \\ &+ [(3-3,75)^2 + (4-3,75)^2 + (5-3,75)^2 + (3-3,75)^2] = 6,75. \end{aligned}$$

3.4) Факторная дисперсия

$$S_{\text{факт}}^2 = \frac{SS_{\text{факт}}}{k-1} = \frac{15,5}{3-1} = 7,75$$

3.5) Остаточная дисперсия

$$S_{\text{ост}}^2 = \frac{SS_{\text{ост}}}{k(r-1)} = \frac{6,75}{3(4-1)} = 0,75$$

$$3.6) F_{\text{расч}} = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{7,75}{0,75} = 10,3$$

$$4) F_{\text{табл}}(0,05; 2; 9) = 4,26.$$

5) $F_{\text{расч}} > F_{\text{табл}}$, то « H_0 » отвергается, значит фактор курения значимо влияет на заболеваемость дыхательных путей.

Контрольные вопросы:

1. Дисперсионный анализ
2. Постановка задачи.
3. Фактор
4. дисперсионный анализ делится на однофакторный и многофакторный
5. Основная цель дисперсионного анализа
6. Факторная дисперсия
7. Остаточная дисперсия
8. Общая дисперсия
9. Схема применения однофакторного дисперсионного анализа

СИТУАЦИОННЫЕ ЗАДАЧИ

Задача № 1

При изучении заболеваемости населения двух районов города гепатитом В были получены следующие показатели, в районе А. -3,5%, в районе Б. -1,8%. Для суждения о влиянии уровня вакцинации на показатель заболеваемости врач счел необходимым использовать метод стандартизации. Какой этап метода стандартизации позволит врачу поставить два района в равные условия по охвату вакцинацией?

Можно ли на этом этапе сделать окончательный вывод о различиях в показателях заболеваемости населения в двух районах и влияющим на эти различия факторе?

Задача № 2

Графически изобразить распространенность наркомании в РК с 2013 по 2017 г. (на 100 000 населения)

Наименование	годы
--------------	------

показателей	2013	2014	2015	2016	2017
Число наркоманов в %	10,1	16,9	20,4	32,3	109,6

Задача № 3

Какой способ оценки достоверности результатов исследования необходимо применить для переноса результатов исследования на генеральную совокупность, если известно, что при изучении организации приема больных в одной из районных поликлиник города Н. среднее время на 1 обращение в регистратуру составил 4 мин, $m=\pm 1,5$ мин. Выборочно были изучены 1600 обращений пациентов в данную поликлинику.

1. Примените этот способ.
2. Сделайте соответствующий вывод.

Задача № 4

В двух цехах (№1 и №2) были изучены уровни травматизма. В первом цехе уровень травматизма выше, чем во втором.

Можно ли по представленным условиям в задаче сравнить показатели травматизма в цехах?

Нужны ли дополнительные сведения для вывода, если состав работающих в цехах различается по стажу работы (в годах)?

Обоснуйте свой ответ.

Задача № 5

При изучении заболеваемости населения двух районов города гепатитом В были получены следующие показатели, в районе А. – 3,5%, в районе Б. -1,8%. Для суждения о влиянии уровня вакцинации на показатель заболеваемости врач счел необходимым использовать метод стандартизации.

Какой этап метода стандартизации позволит врачу поставить два района в равные условия по охвату вакцинацией?

Можно ли на этом этапе сделать окончательный вывод о различиях в показателях заболеваемости населения в двух районах и влияющем на эти различия факторе?

Задача № 6

При оценке частоты ревматизма среди больных, отличающихся по регулярности диспансерного наблюдения, в двух отделениях поликлиники были получены следующие результаты.

Показатели	Отделение	Отделение
Интенсивные	1,2%	5,0%
Стандартизованные	4,0%	2,5%

Сделайте соответствующие выводы:

- 1.
- 2.

3.

Задача № 7

Средний вес новорожденных, родившихся у матерей с пороками сердца в роддоме №2 города А., составили 2,8кг, $m=\pm 0,3$ кг. С помощью какого способа оценки достоверности результатов исследования можно узнать аналогичный результат в генеральной совокупности?

Какая дополнительная информация необходимо для применения выбранной вами методики оценки достоверности?

Задача № 8

Вычислить показатели экстенсивности и интенсивности

Вариант 1

Заболеваемость кишечными инфекциями в Республики Казахстан за 2017г. (рассчитать на 100 000 населения)

Области	Население	Количество случаев
Акмолинская	894 200	3333
Алматинская	1 636 400	2 826
Южно-Казахстанская	1 999 100	9 289
Другие	11 236 500	38 007
По Казахстану	15 766 200	53 455

Задача № 9

Первичная заболеваемость в Республике Казахстан за 2017 г. по классам болезней (рассчитать на 100 000 населения)

Численность населения Республики Казахстан – 15 766 200

Заболевания	Количество случаев
Болезни органов дыхания	2 702 327
Болезни кожи и подкожной клетчатки	636 166,2
Травмы и отравления	629 041,4
Прочие	3 045 872,1
Всего	7 013 436,7

Задача №10

Рассчитать среднюю арифметическую взвешенную, сигму (меру колеблемости вариационного ряда), ошибку средней арифметической и доверительные границы.

Вариант 1.

Рост мальчиков 7 лет (см.)	
V (см)	P(число детей)
109	8
112	13
115	30
118	40
121	30
124	13
127	9
130	6

Задача № 11

Дать расчет показателей обеспеченности врачами и койками (на 10000тыс-населения 2017 г.)

Области	Население	Число врачей	Число коек
Всего по Казахстану	15 671 800	52 030	133 210
Павлодарская	875 700	2 688	8 362
г.Алматы	1 062 400	9 497	13 562

Задача № 12

Определить абсолютный рост или убывание. Вычислить показатели наглядности, динамики и темпов прироста.

Вариант 1.

Рождаемость в Казахстане (число родившихся на 1000 человек населения)

2013 г. – 19,9

2014 г. – 18,7

2015 г. – 16,7

2016 г. – 15,4

2017 г. – 14,7

Задача № 13

Решение ситуационный задачи:

Изучалась успеваемость студентов мед. института – неработающих и совмещающих учебу с работой. Получены следующие данные:

Студенты	Средний балл М	М
Неработающие	4,1	+0,09
Совмещающие работу	3,85	+0,05

с учебой		
----------	--	--

Определить достоверность разности показателей в сравниваемых группах.

Задача № 14

Дать расчет показателей обеспеченности врачами и койками (на 10 000 населения, 2017г.)

Области	Численность населения	Число врачей	Число больничных коек
Всего по Казахстану	15 671 800	52 030	133 210
Восточно – Казахстану	1 635 300	5 216	13 736
Алматы	1 062 400	9 497	13 566

Задача №15

Вычислить интенсивные и стандартизованные показатели.

Дать анализ (за стандарт принять среднее распределение по двум совокупностям)

Вариант 1

Заболеваемость кариесом учащихся средней школы №25 г. Алматы и школы-интерната №1 (показатели на 100 учащихся)

Возраст	Средняя школа №25			Школа-интернат №1		
	Число учащ-ся	Число заболев	Показ-ль интенсив	Число учащ-ся	Число заболев	Показ-ль интенсив
7-10	290	237		190	142	
10-14	270	245		150	129	
15-18	240	238		80	70	
Всего	800	720		420	341	

Задача № 16

Заболеваемость острыми респираторными заболеваниями дошкольников г. Алматы (по выборочным данным детской поликлиники №3)

Возраст	Организованные дети			Неорганизованные дети		
	Число детей	Число ОРЗ	Показ. инт. на 100дет	Число детей	Число ОРЗ	Показ. инт. на 100дет
1-3	100	230		120	190	
4-5	180	370		217	270	
6-7	170	300		130	101	

го	Все	450	900		467	561	
----	-----	-----	-----	--	-----	-----	--

Задача № 17

Решение ситуационный задачи:

Рост и окружность головы у 7 новорожденных детей

Рост ребенка	48	47	50	51	52	49	53
Окружность головы	36	37	38	38	39	39	40

1. Вычислить коэффициент корреляции рангов
2. Определить характер и силу связи между соответствующими признаками
3. Определить достоверность коэффициента корреляции.

Задача № 18

Решение ситуационный задачи:

Вариант 3.

Младенческая смертность (число умерших в возрасте до 1 года на 1000 родившихся живыми)

2013 г. – 26,2
 2014 г. – 28,0
 2015 г. – 27,3
 2016 г. – 25,3
 2017 г. – 24,2

Задача № 19

Возраст кормящей матери и количество сцеженного грудного молока при разовом кормлении

Возраст матери	16	18	24	27	32	35	37	39
Количество грудного молока	100	120	110	115	110	90	90	85

1. Вычислить коэффициент корреляции рангов
2. Определить характер и силу связи между признаками
3. Определить достоверность коэффициента корреляции.

Задача № 20

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

Заболевания	Пол		Возраст				всего
	муж	жен	до 20 лет	21-22 года	23-24 года	25 лет и более	
Гастрит							
Язвенная болезнь							
Язвенная болезнь 12-перстной кишки							
Прочие							
Итого							

Задача № 21

Обоснование выбора метода. Для решения задачи выбран по методу (Спирмена), так как каждый из признаков (жесткость воды и количество кальция) имеет числовое выражение; нет открытых вариантов.

Жесткость воды	Количество кальция в воде (в мг/л)
4	28
8	56
11	77
27	191
34	241
37	262

Задача №22

В результате анализа стационарной помощи больным острым панкреатитом в больнице А. были получены следующие данные:

Длительность стационарного лечения, койко-дней	Число пациентов, чел
14	2
15	6

16	12
18	10
21	5
Всего	35

Соответствует ли представленный вариационный ряд закону нормального распределения? Рассчитайте показатели вариационного ряда: среднюю арифметическую величину, моду, медиану, среднее квадратическое отклонение, коэффициент вариации, среднюю ошибку срендей арифметической.

Задача №23

В результате исследования, посвященного изучению состояния сердечно-сосудистой системы у спортсменов-лыжников, были получены следующие данные:

Частота сердечных сокращений, мин-1	Число исследуемых, чел
52	3
54	5
56	16
58	10
60	6
Всего	40

Соответствует ли представленный вариационный ряд закону нормального распределения? Рассчитайте показатели вариационного ряда: среднюю арифметическую величину, моду, медиану, среднее квадратическое отклонение, коэффициент вариации, среднюю ошибку срендей арифметической.

Задача №24

В исследования, посвященном изучению свойств нового антибиотика, обладающего высокой липотропностью, приняли участие пациенты отделения гнойной хирургии, страдающие ожирением. Наблюдалось следующее распределение исследуемых по массе:

Масса тела, кг	Число пациентов, чел
90	1
100	4
120	8
130	6
140	2
Всего	21

Соответствует ли представленный вариационный ряд закону нормального распределения? Рассчитайте показатели вариационного ряда: среднюю арифметическую величину, моду, медиану, среднее квадратическое отклонение, коэффициент вариации, среднюю ошибку срендей арифметической.

Задача №25

Главным врачом женской консультации было проведено исследование, посвященное влиянию курения во время беременности на здоровье плода. Согласно полученным данным, в

группе женщин, выкуривающих не менее 5 сигарет в день в течение первого триместра беременности, у 12 из 200 при прохождении планового УЗИ плода на 20-й неделе беременности был диагностирован порок развития плода. В контрольной группе из 400 некурящих женщин порок развития плода был диагностирован в 6 случаях.

Сделайте вывод о влиянии курения на риск формирования порока развития у плода.

Задача №26

Заведующим гепатологического отделения одной из инфекционных больниц был проведен анализ эффективности противовирусной терапии больных хроническим гепатитом В. Согласно полученным данным, из 450 больных, получавших противовирусные препараты, в течение 5 лет после постановки диагноза от развития печеночной недостаточности умерли 18 человек. Из 500 пациентов, не получавших специфической противовирусной терапии, за тот же период от развития печеночной недостаточности умерли 42 человека.

Сделайте вывод об эффективности противовирусной терапии хронического гепатита В.

Задача №27

Учебной частью одной из кафедр медицинского университета было проведено исследование успеваемости студентов в зависимости от посещаемости лекций. Для студентов, посетивших менее половины лекционного курса ($n=36$), средняя оценка на экзамене составила $3,2 \sigma=0,2$. Для студентов, посетивших более 90% лекций по предмету ($n=150$), средняя оценка на экзамене составила $4,5 \sigma=0,5$.

Сделайте вывод о достоверности различий успеваемости студентов в зависимости от посещаемости лекций по предмету.

ТЕСТОВЫЕ ЗАДАНИЯ

1. Точное определение статистики как науки:

- Наука о материальных явлениях, имеющих определенные, количественные характеристики.
- Общественная наука, использующая для точности информации разнообразные статистические методы обработки данных.
- Общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественными особенностями.
- Общественная наука, изучающая данные соответствующей научной дисциплины для раскрытия ее материальной сущности.

2. Основные разделы статистики:

- Статистика заболеваемости, демографическая статистика.
- Статистика здоровья населения, статистика здравоохранения.
- Статистика деятельности лечебно-профилактических организаций, демографическая статистика.
- Статистика здравоохранения, статистика рождаемости.

3. К области медицинской статистики не относится:

- Предупреждение и лечение заболеваний.
- Изучение здоровья населения в целом и его отдельных групп.

- в) Выявление и установление связей заболеваемости и смертности населения с различными факторами окружающей среды.
- г) Оценка эффективности воздействия новых медицинских препаратов на биологические организмы и социальный коллектив.

4. Объект статистического исследования:

- а) Единица наблюдения.
- б) Атрибутивные признаки.
- в) Количественные признаки.
- г) Статистическая совокупность.

5. Генеральная совокупность – это...

- а) Совокупность, состоящая из большого числа единиц наблюдения.
- б) Совокупность, отображенная по основному, наиболее важному признаку.
- в) Совокупность, состоящая из всех единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования.
- г) Совокупность, объем, который позволяет выявить общие закономерности изучаемого явления.

6. Предметом анализа единиц наблюдения в статистической совокупности служит

- а) Факторные признаки.
- б) Учетные признаки.
- в) Атрибутивные признаки.
- г) Результативные признаки.

7. Выборочная совокупность – это...

- а) Совокупность, сформированная с использованием выборочного метода.
- б) Разновидность генеральной совокупности, сформированная с учетом размера изучаемая явления.
- в) Совокупность, отобранная по какому-то признаку из нескольких совокупностей.
- г) Часть генеральной совокупности, отобранная специальным методом и служащая для характеристики генеральной совокупности.

8. Для основания объединения единиц наблюдения в статистической совокупности служит

- а) Признаки сходства.
- б) Атрибутивные признаки.
- в) Результативные признаки.
- г) Факторные признаки.

9. Первичный элемент статистической совокупности:

- а) Признак сходства.
- б) Учетный признак
- в) Единица наблюдения.
- г) Статистическая величина.

10. Учетные признаки по характеру классифицируются:

- а) Существенные и несущественные.
- б) Атрибутивные и количественные
- в) Факторные и количественные
- г) Результативные и факторные.

11. Учетные признаки по взаимосвязи классифицируются:

- а) Атрибутивные и факторные.
- б) Факторные и количественные

12. Репрезентативность – это...

- а) Соответствие признаков, отобранных для характеристики объекта исследования, цели и задачи исследования.
- б) Качественная и количественная представительность отобранной спец. методом части объекта стат. исследования, предназначенная для характеристики всего объекта в целом.
- в) Достаточный объем выборочной совокупности.
- г) Качественная однородность генеральной совокупности

13. Способы формирования статистической совокупности:

- а) По охвату, по времени наблюдения, по способу.
- б) По времени наблюдения.
- в) По полноте охвата.
- г) По типам группировок, по способам.

14. Существующие виды наблюдений в зависимости от полноты охвата:

- а) Сплошное, выборочное.
- б) Полное, неполное.
- в) Выборочное, полное.
- г) Генеральное, случайный отбор.

15. Существующие виды наблюдений в зависимости от времени:

- а) Ретроспективное, современное.
- б) Постоянное, прерывистое.
- в) Текущее, постоянное.
- г) Текущее, единовременное.

16. Существующие виды наблюдений в зависимости от способа:

- а) Выкопировка данных, механический отбор.
- б) Непосредственная регистрация, выкопировка данных, анамнестический.
- в) Анкетирование, опрос, непосредственная регистрация
- г) Сплошное, текущее, единовременное.

17. Репрезентативность выборки зависит:

- а) От цели и задач исследования.
- б) От однородности статистического материала.
- в) От численности и способа формирования выборочной совокупности.
- г) От объема выборочной совокупности.

**18. Характеристика переписи населения, как статистического исследования, с учетом:
1-охвата, 2- времени, 3- способа наблюдения:**

- а) 1-выборочное, 2-текущее, 3- непосредственная регистрации.
- б) 1-сплошное, 2- единовременное, 3-анамнестический.
- в) 1-сплошное, 2-текущее, 3-анкетирование
- г) 1- случайный отбор, 2-единовременное, 3- опрос.

19. Статистическое исследование начинается

- а) С постановки цели и задач исследования.
- б) С составлением программы исследования.

- в) С составления плана исследования.
- г) С определения единиц и наблюдения.

20. В статистики основная цель сбора достаточного числа наблюдений:

- а) Исключение случайных отклонений.
- б) Выявление основных закономерностей.
- в) Позволяет дать качественную характеристику статистической совокупности.
- г) Повышение достоверности статистических данных.

21. Выбор единиц наблюдения определяется:

- а) Программой анализа.
- б) Целью и задачами исследования.
- в) Программой разработки.
- г) Видом статистического исследования.

22. При выборе учетного документа необходимо руководствоваться.....

- а) Программой разработки.
- б) Программой анализа.
- в) Программой сбора материала.
- г) Объемом исследования

23. Программа сбора материала представляет собой....

- а) Первичный статистический документ с перечнем учетных признаков, подлежащих регистрации.
- б) Группировка учетных признаков.
- в) Инструкция по методике сбора материала.
- г) Сроки работы по отдельным этапам и операциям.

24. Программа разработки материала представляет собой....

- а) Сроки работы по отдельным этапам и операциям.
- б) Определения объекта наблюдения.
- в) Определения объекта наблюдения.
- г) Составления макетов статистических таблиц.

25. Однородность статистического материала достигается.....

- а) Ясностью целей и задач исследования.
- б) Достаточным числом наблюдений.
- в) Правильным отбором учетных признаков.
- г) Группировкой материала по основным, определяющим данное явление, признаком.

26. Группировка признаков проводится с целью....

- а) Выявить закономерности в изучаемом явлении.
- б) Выделить качественно однородные группы для установления определенных закономерностей в изучаемом явлении.
- в) Облегчить сбор и разработку материала.
- г) Облегчить составление макетов статистических таблиц.

27. В зависимости от характера учитываемых признаков различают виды группировок:

- а) Атрибутивные и количественные.
- б) Количественные и качественные.
- в) Вариационные и атрибутивные.
- г) Типологические и вариационные.

28. План статистического исследования предусматривает:

- а) Определения объекта исследования и способов формирования объекта исследования.
- б) Определения объекта исследования.
- в) Способы формирования объекта исследования.
- г) Способы анализа статистического материала.

29. Шифровка материала – это...

- а) Проставление определенных шифров.
- б) Условное обозначение зарегистрированных признаков.
- в) Введение сочетание цифр и знаков.
- г) Проставление порядковых номеров при разработке материала.

30. Типологическая группировка – это...

- а) Группировка по признакам, имеющих числовое выражение.
- б) Группировка по факторным и результативным признакам.
- в) Группировка по признакам, которые могут быть выражены словесно, описательно.
- г) Группировка по признакам единицы наблюдения.

31. Вариационная группировка – это...

- а) Группировка по признакам, имеющим числовое выражение.
- б) Группировка по признакам, которые могут быть выражены словесно.
- в) Группировка по факторным и результативным признакам.
- г) Группировка по атрибутивным и количественным признакам.

32. В статистическом исследовании используются виды таблиц....

- а) Групповые, комбинационные.
- б) Числовые, описательные.
- в) Одно - двух - трехпольные.
- г) Простые, групповые, комбинационные.

33. Простая таблица представляет собой

- а) Итоговую сводку данных по одному признаку.
- б) Объединение признаков из различных группировок.
- в) Общие сведения об изучаемых признаках
- г) Сводку по факторному признаку.

34. Групповую таблицу характеризует....

- а) Сводные данные с группировкой по двум и более признакам.
- б) Распределение по двум количественным признакам.
- в) Сводку по двум о более факторным признакам.
- г) Сводку по двум и более качественным признакам.

35. Комбинационная таблица представляет собой

- а) Распределение по качественному количественному признаку.
- б) Распределение по двум и более признакам.

- в) Сводные данные с сочетанием трех и более взаимосвязанных признаков.
- г) Распределение по факторному и результативному признаку.

36. В содержание 1-го этапа статистического исследования не входит...

- а) Определение объекта исследования.
- б) Контроль документов.
- в) Выбор единицы наблюдения.
- г) Составление плана исследования.

37. В содержание 2-го этапа статистического исследования не входит...

- а) Выкопировка данных.
- б) Статистический анализ материала.
- в) Непосредственная регистрация.
- г) Заполнение регистрационных бланков.

38. В содержание 3-го этапа статистического исследования не входит...

- а) Контроль собранного материала.
- б) Шифровка и раскладка карт по группам.
- в) Составление программ разработки материала.
- г) Подсчет и заполнения карт по группам

39. В содержание 4-го этапа статистического исследования не входит...

- а) Литературное оформление данных.
- б) Обобщение результатов.
- в) Составление макетов статистических таблиц.
- г) Качественная оценка выявленных закономерностей.

40. В статистическом исследовании после уяснения цели и задач ведущим является....

- а) Сбор материала.
- б) Разработка и сводка данных.
- в) Анализ полученных результатов.
- г) Составление плана и программы исследования.

41. В статистике используются виды величин....

- а) Достоверные, недостоверные, относительные.
- б) абсолютные, относительные, средние.
- в) Средние, абсолютные, большие.
- г) Генеральные, выборочные, средние.

42. Абсолютная величина – это...

- а) Величина, несущая информацию о размере явления.
- б) Величина, дающая качественную характеристику.
- в) Величина, используемая для обобщающей характеристики совокупностей.
- г) Величина, используемая для сравнения и сопоставления совокупностей.

43. Недостаток абсолютных величин заключается в ...

- а) Трудны для запоминания.
- б) Не дают обобщающей характеристики статистической совокупностей.
- в) Используется только для обобщающей характеристики изучаемого явления.
- г) Несут недостоверные сведения об изучаемом явлении.

44. Для характеристики частоты явлений следует употреблять показатели....

- а) Соотношения.
- б) Экстенсивные.
- в) Интенсивные.
- г) Наглядности.

45. Интенсивные показатели используются для.....

- а) Наглядно показать различия сравниваемых групп.
- б) Дать характеристику ряда, состоящего из однородных сопоставляемых величин.
- в) Показать долю части в целом.
- г) Судить о частоте явления.

46. Показатель наглядности можно вычислить на основании данных....

- а) На основании абсолютных величин.
- б) На основании любых величин (кроме интенсивных показателей).
- в) На основании относительных величин.
- г) На основании средних величин.

47. Вычисление какого показателя приведено ниже: 15400 случаев зарегистрированных заболеваний\1000\27320 жителей – 700%

- а) Экстенсивного
- б) Интенсивного
- в) Наглядности
- г) Соотношения

48. «В течение года зарегистрировано 11055 заболеваний, в том числе среди мужчин-44666, женщин-6589» из данных можно рассчитать показатель...

- а) Экстенсивные
- б) Наглядности
- в) Интенсивные
- г) Соотношения

49. «Численность населения города 8870000, число врачей стоматологов- 278, число зубных врачей- 520» из данных можно рассчитать показатель...

- а) Интенсивные
- б) Экстенсивные
- в) Соотношения
- г) Наглядности

50. «Число прошедших проф. осмотр -4200 человек, из них кариес зубов выявлен у 1570»из данных можно рассчитать показатель...

- а) Соотношения
- б) Экстенсивности
- в) Наглядности
- г) Интенсивные

51. Структуру изучаемого явления характеризует показатель....

- а) Динамического ряда
- б) Интенсивный
- в) Наглядности
- г) Экстенсивный

52. Методика вычисления экстенсивного показателя

- а) Отношение числа, выражающего величину данного явления, к величине всей совокупности
- б) Отношение части явления к целому явлению
- в) Отношение ряда чисел к одному из них, принимаемому за 100%
- г) Отношение абсолютного уровня последующего числа к предыдущему в %

53. Методика вычисления показателя наглядности

- а) Отношение числа, выражающего величину данного явления к величине всей совокупности
- б) Отношение части явления к целому явлению
- в) Отношение ряда сравниваемых однородных величин к одной из них, принятой за 100%
- г) Отношение абсолютного уровня последующего числа к предыдущему в %.

54. Определение показателя соотношения

- а) Изменение явления во времени.
- б) Распределение целого на части.
- в) Отношение между двумя самостоятельными совокупностями.
- г) Сопоставление ряда однородных величин, имеющих разный характер.

55. Определение интенсивного показателя

- а) Характеристика явления в среде, непосредственно с ним не связанной.
- б) Изменение явления во времени.
- в) Распределение целого на части.
- г) Частота явления в среде, не посредственно с ним связанной

56. Определение экстенсивного показателя

- а) Изменение явления по времени
- б) Распределение целого на части
- в) Частота явления в среде, непосредственно с ним связанной
- г) Характеристика явления в среде, непосредственно с ним не связанный

57. В статистике относительная величина обозначается буквой

- а) М
- б) m
- в) Р
- г) р

58. Экстенсивные показатели можно представить видом диаграммы....

- а) Секторная
- б) Столбиковая
- в) Линейная
- г) Радиальная

59. Динамику явления за х-замкнутый цикл времени можно изобразить видом диаграммы....

- а) Секторная
- б) Радиальная
- в) Столбиковая
- г) Любая

60. Для графического изображения могут быть использованы виды величин...

- а) Интенсивные и экстенсивные показатели
- б) Экстенсивные показатели соотношения
- в) Абсолютные и интенсивные показатели
- г) Любые величины

61. Наиболее правильное определение вариационного ряда

- а) Это ряд числовых значений определен. признака, отличающихся друг от друга по всей величине и нормально расположенных в ранговом порядке.
- б) Два ряда величин, изменяющихся в убывающем или возрастающем порядке
- в) Ряд числовых значений какого-то количественного порядка
- г) Статистический ряд, характеризующийся распределением чисел убывающем или возрастающем порядке

62. Определение варианты (V)

- а) Число, из которого состоит вариационный ряд
- б) Число простого вариационного ряда
- в) Число сгруппированного вариационного ряда.
- г) Числовое значение изучаемого признака

63. Определение частоты

- а) Частота, с которой встречаются каждая варианта
- б) Частота, показывающая число вариант в вариационном ряду
- в) Частота простого вариационного ряда
- г) Частота, показывающая разнообразие признака

64. Средняя величина – это...

- а) Величина, характеризующая типичный уровень признаков совокупности
- б) Величина, отражающая общие свойства статистической совокупности
- в) Обобщенная характеристика признака в статистической совокупности
- г) Размер признака в расчете на единицу однородной совокупности

65. В статистике средняя величина обозначается буквой...

- а) В
- б) М
- в) А
- г) Д

66. К средним величинам не относится

- а) Мода
- б) Медиана
- в) Средняя арифметическая
- г) Средняя геометрическая

67. Наиболее правильное определение средней арифметической простой

- а) Средняя величина, полученная путем простого суммирования вариант в деления суммы на общее число всех вариант.
- б) Средняя величина, получаемого из рядов, частоты которых равны единице или равны между собой.
- в) Средняя величина, получаемая как частное от деления суммы вариант на сумму частот.
- г) Средняя величина, получаемое как частное от деления суммы вариант на сумму их частот, равных единице.

68. Наиболее правильное определение средней арифметической взвешенной

- а) Наиболее часто употребляемая и встречающаяся в санитарной статистике и разновидность средней арифметической величины.
- б) средняя величина, получаемая из вариационных рядов, варианты которые встречаются с различной частотой
- в) Средняя величина, получаемая путем деления суммы вариант на общее число наблюдений
- г) Среднее из вариант, повторяющихся различное число раз

69. Наиболее правильное определение моды

- а) Мода-средняя величина, которая в простом ряду может определяться приближенно в сгруппированном по формуле.
- б) Мода-варианта, принимаемая за среднюю величину, используемая для оценки структуры изучаемого ряда.
- в) Мода-варианта, встречающаяся с наибольшей частотой
- г) Мода-средняя величина, обозначающая наиболее типичное значение изучаемого признака.

70. Наиболее правильное определение медианы

- а) Медиана называется варианта, которая находится в середине вариационного ряда.
- б) Средняя величина, обозначающая срединно-расположенную варианту в любых рядах.
- в) Медиана-это варианта, расположенная в середине вариационного и делящая его на 2 равные части.
- г) Средняя величина, обозначающая срединно-расположенную варианту в сгруппированном вариационном ряду.

71. Наиболее правильное определение оценки достоверности

- а) Полученные результаты характеризуют количественную репрезентативность совокупности.
- б) Определить вероятность прогноза результатов изучение данной совокупности
- в) Полученные результаты в изучаемой совокупности можно распределить на всю генеральную совокупность
- г) Установить вероятность б/ошиб. прогноза, с которым результаты, получены при изучении выбор. совокупности можно перенести на генерал. совокупность

72. Мерой достоверности величин является...

- а) Средняя ошибка
- б) Медиана
- в) Критерий достоверности
- г) Амплитуда ряда

73. К способу определения достоверности не относится...

- а) Определение.
- б) Вычисление ошибки средней величины
- в) Вычисление ошибки относительной величины
- г) Определение доверительных границ

74. Среднее квадратное отклонение (Q) характеризует ...

- а) Позволяет судить о частоте изучаемого явления
- б) Дает представление о разнообразии исследуемого признака

- в) Позволяет рассчитать амплитуду ряда
- г) Позволяет судить о правильности вариационного ряда

75. В статистике средняя ошибка обозначается буквой.....

- а) P
- б) T
- в) m
- г) M

76. Вероятность безошибочного прогноза обозначается буквой....

- а) p
- б) t
- в) m
- г) q

77. Уровень вероятности безошибочного прогноза зависит.....

- а) От величины доверительного интервала.
- б) От средних и относительных величин.
- в) От величины ошибок средних величин.
- г) От величины ошибок относительных величин

78. Если полученная при выборочном исследовании разность между сравнительными или относительными показателями недостоверна, следует:

- а) Использовать другие методы оценки достоверности
- б) увеличить число наблюдений
- в) Использовать другие виды величин
- г) Считать результаты окончательными

79. В медицинской статистике наиболее часто применяется метод стандартизации...

- а) Метод экспертных ошибок
- б) Косвенный метод
- в) Обратный метод.
- г) Прямой метод.

80. Метод стандартизации в медицинской статистике применяется с целью ...

- а) Для сравнения величин в пространстве и во времени
- б) Для динамического наблюдения
- в) Для сравнения качественно не однородных по составу групп
- г) Для изучения структуры явления

81. Первый этап стандартизации...

- а) Выбор и расчет данных
- б) Вычисление общих и по интенсивных групповых показателей.
- в) Расчет стандартизованных показателей
- г) Вычисление "ожидаемого" числа в соответствии со стандартом

82. Третий этап стандартизации...

- а) Вычисление общих и по групповых показателей
- б) Расчет стандартизованных показателей.
- в) Сравнение групп
- г) Вычисление "ожидаемого" величин в соответствии со стандартном

83. Выбор и расчет стандарта производится на этапе метода стандартизации...

- а) Первый этап
- б) Второй этап
- в) Третий этап
- г) Четвертый этап

84. Для отображения результатов исследований методом стандартизации используется вид таблицы...

- а) Столбиковая диаграмма, простая таблица
- б) Простая таблица
- в) Групповая таблица
- г) Комбинационная таблица

85. Сравнение групп по показателем производится на этапе метода стандартизации.

- а) Пятом
- б) Третьем
- в) Первом
- г) Четвертом

86. Расчет стандартизованных показателей производится на этапе метода стандартизации.

- а) Пятом
- б) Третьем
- в) Первом
- г) Четвертом

87. При расчетах методом стандартизации чаще всего используются показатели...

- а) Интенсивный показатель
- б) Экстенсивный показатель
- в) Показатель наглядности
- г) Показатель соотношения

88. Под медицинской статистикой понимают отрасль статистики, включающую:

- а) статистические методы по изучению здоровья населения
- б) совокупность статистических методов, необходимых для анализа деятельности ЛПУ
- в) совокупность статистических методов по изучению здоровья населения и факторов, влияющих на него, а также вопросов, связанных с медициной и здравоохранением
- г) статистические методы по изучению и совершенствованию управления в учреждениях здравоохранения

89. Предметом изучения медицинской статистики является информация:

- а) о здоровье населения
- б) о экологии
- в) о кадровой политике
- г) о санитарно-эпидемиологическом окружении

90. Статистический метод в медицине и здравоохранении применяется для:

- а) изучения психологической атмосферы
- б) изучения состояния и деятельности социальных органов
- в) планирования научных исследований

г) изучения общественного здоровья и факторов, его определяющих

91. Основными методами формирования выборочной совокупности являются:

- а) научный
- б) механический
- в) пилотажный
- г) первичный

92. Программа статистического исследования включает:

- а) составление программы сбора материала
- б) анализ социальных факторов
- в) определение объекта исследования
- г) определение исполнителей исследования

93. Результаты статистического исследования анализируются на основании:

- а) статистических (регистрационных) учетных документов
- б) амбулаторных карт
- в) статистических таблиц
- г) историй болезни

94. К экстенсивным показателям относятся:

- а) показатели рождаемости
- б) распределение числа врачей по специальностям
- в) показатели младенческой смертности
- г) обеспеченность населения врачами

95. К интенсивным показателям не относятся:

- а) показатель смертности
- б) показатели младенческой смертности
- в) показатель заболеваемости
- г) структура заболеваний по нозологическим формам

96. Вариационный ряд - это:

- а) ряд чисел, отражающих частоту (повторяемость) цифровых значений изучаемого признака
- б) ряд цифровых значений различных признаков
- в) ряд числовых измерений признака, расположенных в ранговом порядке и характеризующихся определенной частотой
- г) ряд любых чисел

97. Средняя арифметическая - это:

- а) варианта с наибольшей частотой
- б) разность между наибольшей и наименьшей величиной
- в) обобщающая величина, характеризующая размер варьирующего признака совокупности
- г) варианта, находящаяся в середине ряда

98. Средняя ошибка средней арифметической величины (ошибка репрезентативности) - это:

- а) средняя разность между средней арифметической и вариантами ряда
- б) величина, на которую полученная средняя величина выборочной совокупности отличается от среднего результата генеральной совокупности
- в) величина, на которую в среднем отличается каждая варианта от средней арифметической

г) средняя разность между числами

99. Корреляционная связь может устанавливаться между признаками....

- а) ростом и массой тела у детей
- б) частотой пульса и дыхания
- в) уровнем систолического и диастолического давления
- г) частотой случаев хронических заболеваний и острых заболеваний

100. Метод стандартизации применяется:

- а) для определения характера и силы связи между двумя признаками
- б) для определения очередности процессов
- в) для определения достоверности различия двух сравниваемых показателей
- г) для сравнения интенсивных показателей в неоднородных по составу совокупностях

101. Сущность метода стандартизации состоит в:

- а) установлении соответствия между сравниваемыми группами и эталоном (стандартом)
- б) устранении влияния различий в составе сравниваемых групп на величину обобщающих показателей
- в) установлении достоверности различий двух сравниваемых групп по какому-либо показателю
- г) установление достоверности и соответствия

102. При сравнении интенсивных показателей, полученных на однородных по своему составу совокупностях, необходимо применять:

- а) метод корреляции
- б) метод стандартизации
- в) оценку достоверности разности полученных показателей
- г) метод проб и ошибок

103. Динамический ряд - это:

- а) ряд числовых измерений определенного признака, отличающихся друг от друга по своей величине, расположенных в ранговом порядке.
- б) ряд числовых значений
- в) ряд величин, характеризующих результаты исследований в разных регионах
- г) ряд, состоящий из однородных сопоставимых величин, характеризующих изменения какого-либо явления во времени.

104. Способ преобразования (выравнивания) динамического ряда:

- а) укрупнение интервалов
- б) вычисление средней
- в) вычисление коэффициента вариации
- г) вычисление моды

105. Основными показателями динамического ряда не является:

- а) темп роста
- б) абсолютный прирост
- в) темп прироста
- г) сигмальное отклонение

106. Макеты статистических таблиц создаются на этапе статистического исследования:

- а) при составлении плана и программы исследований
- б) на этапе сбора материала

- в) на этапе статистической обработки материалов
- г) при проведении анализа результатов

107. Регистрация рождаемости и смертности относится к виду статистического наблюдения и методу статистического исследования....

- а) единовременное
- б) сплошной
- в) ежедневное
- г) выборочный

108. Наиболее объективную информацию о состоянии здоровья населения позволяют получить способов наблюдения:

- а) опрос
- б) анкетирование
- в) выкопировка данных из медицинской документации
- г) тестирование

109. К единовременному наблюдению относится:

- а) регистрация рождений
- б) перепись населения
- в) регистрация браков
- г) регистрация заболеваний

110. К текущему наблюдению не относится:

- а) регистрация случаев смерти
- б) регистрация случаев обращения в поликлинику
- в) учет родившихся
- г) перепись населения

111. Для экспертной оценки качества и эффективности медицинской помощи в женской консультации отобрана каждая десятая "Индивидуальная карта беременной и родильницы". Выборка является:

- а) случайной
- б) селективной
- в) когортной
- г) социальной

112. Единица наблюдения определяется в зависимости от:

- а) программы исследования
- б) плана исследования
- в) цели и задач исследования
- г) предмета исследования

113. Из перечисленных видов статистических таблиц наиболее информативной является:

- а) простая
- б) групповая
- в) комбинационная.
- г) сложная

114. Типологические группировки могут включать следующий признак:

- а) АД

- б) рост
- в) массу тела
- г) диагноз

115. Вариационные группировки могут включать следующие признаки:

- а) рост
- б) диагноз
- в) пол
- г) профессию

116. Экстенсивные показатели применяются для определения:

- а) частоты явления в совокупности (среде)
- б) удельного веса части в целом (внутри одной совокупности)
- в) соотношения несвязанных между собой совокупностей
- г) частоты соотношений

117. Экстенсивные показатели могут быть представлены следующим видом диаграммы:

- а) линейными
- б) секторными
- в) столбиковыми
- г) картограммами

118. Размер ошибки средней арифметической величины зависит от:

- а) типа вариационного ряда
- б) числа случаев
- в) способа расчета средней
- г) разнообразия изучаемого признака

119. Статистический метод позволяющий оценивать достоверность результатов, полученных при выборочных исследованиях:

- а) корреляция
- б) стандартизация
- в) определение доверительных границ
- г) оценка результатов

120. Для установления силы и характера связи между признаками нужно найти:

- а) коэффициент корреляции
- б) среднеквадратическое отклонение
- в) критерий достоверности
- г) стандартизованные показатели

121. Точное определение статистики как науки:

- А) Наука о материальных явлениях, имеющих определенные, количественные характеристики.
- Б) Общественная наука, использующая для точности информации разнообразные статистические методы обработки данных.
- В) Общественная наука, изучающая количественных сторону массовых общественных явлений в неразрывной связи с их качественными особенностями.
- Г) Общественная наука, изучающая данные соответствующей научной дисциплины для раскрытия ее материальной сущности.

122. Основные разделы статистики:

- А) Статистика заболеваемости, демографическая статистика.
- Б) Статистика здоровья населения, статистика здравоохранения.
- В) Статистика деятельности лечебно-профилактических организаций, демографическая статистика.
- Г) Статистика здравоохранения, статистика рождаемости.

123. К области медицинской статистики не относится:

- А) Предупреждение и лечение заболеваний.
- Б) Изучение здоровья населения в целом и его отдельных групп.
- В) Выявление и установление связей заболеваемости и смертности населения с различными факторами окружающей среды.
- Г) Оценка эффективности воздействия новых медицинских препаратов на биологические организмы и социальный коллектив.

124. Объект статистического исследования:

- А) Единица наблюдения
- Б) Атрибутивные признаки
- В) Количественные признаки
- Г) Статистическая совокупность.

125. Генеральная совокупность – это...

- А) Совокупность, состоящая из большого числа единиц наблюдения.
- Б) Совокупность, отображенная по основному, наиболее важному признаку.
- В) Совокупность, состоящая из всех единиц наблюдения, которые могут быть к ней отнесены в соответствии с целью исследования.
- Г) Совокупность, объём, который позволяет выявить общие закономерности изучаемого явления.

126. Предметом анализа единиц наблюдения в статистической совокупности служит....

- А) Факторные признаки
- Б) Учетные признаки
- В) Атрибутивные признаки
- Г) Результативные признаки

127. Выборочная совокупность – это....

- А) Совокупность, сформированная с использованием выборочного метода.
- Б) Разновидность генеральной совокупности, сформированная с учетом размера изучаемого явления.
- В) Совокупность, отобранная по какому-то признаку из нескольких совокупностей.
- Г) Часть генеральной совокупности, отобранная специальным методом и служащая для характеристики генеральной совокупности.

128. Для основания объединения единиц наблюдения в статистической совокупности служит:

- А) признаки сходства
- Б) атрибутивные признаки
- В) результативные признаки
- Г) факторные признаки

129. Первичный элемент статистической совокупности:

- А) признак сходства

- Б) учетный признак
- В) единица наблюдения
- Г) статистическая величина

130. Учетные признаки по характеру классифицируются:

- А) Существенные и несущественные
- Б) Атрибутивные и количественные
- В) Факторные и количественные
- Г) Результативные и факторные

131. Учетные признаки по взаимосвязи классифицируются:

- А) Атрибутивные и факторные
- Б) Факторные и количественные
- В) Существенные и несущественные
- Г) Факторные и результативные

132. Репрезентативность – это ..

- А) соответствие признаков, отобранных для характеристики объекта исследования, цели и задачи исследования.
- Б) Качественная и количественная представительность отобранной спец. Методом части объекта стат. Исследования, предназначенная для характеристики всего объекта в целом.
- В) Достаточный объем выборочной совокупности.
- Г) Качественная однородность генеральной совокупности.

133. Способы формирования статистической совокупности:

- А) по охвату, по времени наблюдения, по способу.
- Б) по времени наблюдения
- В) по полноте охвата
- Г) по типам группировок, по способам.

134. Существующие виды наблюдения в зависимости от полноты охвата:

- А) сплошное, выборочное
- Б) полное и неполное
- В) выборочное и полное
- Г) Генеральное, случайный отбор.

135. Существующий виды наблюдений в зависимости от времени:

- А) ретроспективное, современное
- Б) постоянное, прерывистое
- В) текущее, постоянное
- Г) текущее, единовременное.

136. Существующие виды наблюдений в зависимости от способа:

- А) Выкопировка данных, механический отбор
- Б) непосредственная регистрация, выкопировка данных, анамнестический.
- В) анкетирование, опрос, непосредственная регистрация.
- Г) сплошное, текущее, единовременное.

137. Репрезентативность выборки зависит:

- А) от цели и задач исследования
- Б) от однородности статистического материала
- В) от численности и способа формирования выборочной совокупности

Г) от объёма выборочной совокупности

138. Характеристика переписи населения, как статистического исследования, с учетом 1-охвата, 2-времени, 3-способа наблюдения:

- А) 1-выборочное, 2-текущее, 3-непосредственная регистрации
- Б) 1-сплошное, 2-единовременное, 3-анамнестический
- В) 1-сплошное, 2-текущее, 3-анкетирование
- Г) 1-случайный отбор, 2-единовременное, 3-опрос.

139. Статистическое исследование начинается:

- А) с постановки цели и задач исследования
- Б) С составления программ исследования
- В) с составления плана исследования
- Г) с определения единиц и наблюдения

140. В статистике основная цель сбора достаточного числа наблюдений:

- А) исключение случайных отклонений
- Б) выявление основных закономерностей
- В) Позволяет дать качественную характеристику статистической совокупности.
- Г) Повышение достоверности статистических данных

141. Выбор единиц наблюдения определяется:

- А) программой анализа
- Б) Целью и задачами исследования
- В) Программой разработки
- Г) Видом статистического исследования

142. При выборе учетного документа необходимо руководствоваться:

- А) Программой разработки
- Б) Программой анализа
- В) Программой сбора материала
- Г) Объёмом исследования

143. Программа сбора материала представляет собой:

- А) первичный статистический документ с перечнем учетных признаков, подлежащих регистрации.
- Б) Группировка учетных признаков
- В) Инструкция по методике сбора материала.
- Г) Сроки работы по отдельным этапам и операциям.

144. Программа разработки материала представляет собой...

- А) Сроки работы по отдельным этапам и операциям
- Б) Определение объекта наблюдения
- В) Определение объекта наблюдения
- Г) Составление макетов статистических таблиц.

145. Однородность статистического материала достигается:

- А) ясностью целей и задач исследования
- Б) достаточным числом наблюдения
- В) правильным отбором учетных признаков
- Г) группировкой материала по основным, определяющий данное явление, признаком.

146. Группировка признаков проводится с целью....

- А) Выявить закономерности в изучаемом явлении
- Б) Выделить качественно однородные группы для установления определённых закономерностей в изучаемом явлении.
- В) Облегчить сбор и разработку материала
- Г) Облегчить составление макетов статистических таблиц.

147. В зависимости от характера учитываемых признаков различают виды группировок:

- А) атрибутивные и количественные
- Б) количественные и качественные
- В) вариационные и атрибутивные
- Г) типологические и вариационные

148. План статистического исследования предусматривает:

- А) Определения объекта исследования и способов формирования объекта исследования
- Б) Определение объекта исследования
- В) Способы формирования объекта исследования
- Г) Способы анализа статистического материала

149. Шифровка материала – это..

- А) Проставление определённых шифров
- Б) Условные обозначения зарегистрированных признаков
- В) введение сочетания цифр и знаков.
- Г) проставление порядковых номеров при разработке материала

150. Типологическая группировка – это...

- А) группировка по признакам, имеющих числовое выражение
- Б) группировка по факторным и результативным признакам
- В) группировка по признакам, которые могут быть выражены словесно, описательно.
- Г) группировка по признакам единицы наблюдения

151. Вариационная группировка – это..

- А) группировка по признакам, имеющим числовое выражение
- Б) группировка по признакам, которые могут быть выражены словесно.
- В) группировка по факторным и результативным признакам.
- Г) группировка по атрибутивным и количественным признакам.

152. В статистическом исследовании используются виды таблиц..

- А) групповые и комбинационные
- Б) Числовые, описательные
- В) одно-двух-трехпольные
- Г) простые, групповые, комбинационные.

153. Простая таблица представляет собой ..

- А) итоговую сводку данных по одному признаку.
- Б) Объединение признаков из различных группировок
- В) Общие сведения об изучаемых признаках
- Г) сводку по факторному признаку

154. Групповую таблицу характеризует:

- А) Свободные данные с группировкой по двум и более признакам

- Б) Распределение по двум количественным признакам
- В) Сводку по двум и более факторным признакам
- Г) сводку по двум и более качественным признакам.

155. Комбинационная таблица представляет собой...

- А) распределение по качественному количественному признаку
- Б) распределение по двум и более признакам
- В) свободные данные в сочетании трех и более взаимосвязанных признаков.
- Г) Распределение по факторному и результативному признаку.

156. В содержании 1-го этапа статистического исследования не входит:

- А) определение объекта исследования
- Б) контроль документов
- В) выбор единицы наблюдения
- Г) составление плана исследования

157. В содержании 2-го этапа статистического исследования не входит:

- А) Выкопировка данных
- Б) Статистический анализ материала
- В) Непосредственная регистрация
- Г) Заполнение регистрационных данных.

158. В содержании 3-го этапа статистического исследования не входит:

- А) Контроль собранного материала
- Б) Шифровка и раскладка карт по группам
- В) составление программ разработки материала
- Г) подсчет и заполнения карт по группам

159. В содержании 4-го этапа статистического исследования не входит:

- А) Литературное оформление данных
- Б) Обобщение результатов
- В) Составление макетов статистических таблиц
- Г) Качественная оценка выявленных закономерностей

160. В статистическом исследовании после уяснения цели и задач ведущим является:

- А) сбор материала
- Б) Разработка и сводка данных
- В) Анализ полученных данных
- Г) Составление плана и программы исследования

161. В статистике используются виды величин:

- А) достоверные, недостоверные, относительные
- Б) абсолютные, относительные, средние
- В) средние, абсолютные, большие
- Г) генеральные, выборочные, средние

162. Абсолютная величина – это..

- А) величина, несущая информацию о размере явлений
- Б) величина, дающая качественную характеристику
- В) Величина, используемая для обобщающей характеристики совокупности
- Г) величина, используемая для сравнения и сопоставления совокупностей

163. Недостаток абсолютных величин заключается в...

- А) трудны для запоминания
- Б) не дают обобщающей характеристики статистической совокупностей
- В) используется только для обобщающей характеристики изучаемого явления
- Г) несут недостоверные сведения об изучаемом явлении

164. Для характеристики частоты явлений следует употреблять показатели:

- А) соотношения
- Б) экстенсивные
- В) интенсивные
- Г) наглядности

165. Интенсивные показатели используются для:

- А) Наглядно показать различия сравниваемых групп
- Б) Дать характеристику ряда, состоящего из однородных сопоставляемых величин.
- В) Показать долю части в целом.
- Г) Судить о частоте явлений.

166. Показатель наглядности можно вычислить на основании данных:

- А) на основании абсолютных величин
- Б) На основании любых величин (кроме интенсивных показателей)
- В) На основании относительных величин
- Г) на основании средних величин

167. Вычисление какого показателя приведено ниже: 15400 случаев зарегистрированных заболеваний 1000/27320 жителей – 700%

- А) экстенсивного
- Б) Интенсивного
- В) наглядности
- Г) соотношения

168. В течении года зарегистрировано 11055 заболеваний, в том числе среди мужчин – 44666, женщин – 6589». Из данных можно рассчитать показатель

- А) экстенсивные
- Б) наглядность
- В) интенсивные
- Г) соотношения

169. Численность населения 8870000, число врачей-стоматологов – 278, число зубных врачей – 520». Из данных можно рассчитать показатель:

- А) интенсивные
- Б) экстенсивные
- В) соотношения
- Г) наглядности

170. Число прошедших проф. Осмотр – 4200, из них кариес зубов выявлен у 1570» Из данных можно рассчитать показатель:

- А) соотношения
- Б) экстенсивности
- В) наглядности
- Г) интенсивности

171. Структуру изучаемого явления характеризует показатель...

- А) динамического ряда
- Б) интенсивный
- В) наглядности
- Г) экстенсивные

172. Методика вычисления экстенсивного показателя

- А) отношение числа, выражающего величину данного явления к величине всей совокупности
- Б) отношение части явления к целому явлению
- В) отношение ряда чисел к одному из них, принимаемому за 100%
- Г) Отношение абсолютного уровня последующего числа к предыдущему в %

173. Методика вычисления показателя наглядности:

- А) отношение числа, выражающего величину данного явления к величине всей совокупности
- Б) отношение части явления к целому явлению
- В) отношение ряда чисел к одному из них, принимаемому за 100%
- Г) Отношение абсолютного уровня последующего числа к предыдущему в %

174. Определение показателя соотношения

- А) изменения явления во времени
- Б) распределение целого на части
- В) отношение между двумя самостоятельными совокупностями
- Г) сопоставление ряда однородных величин, имеющих разный характер

175. Определение интенсивного показателя

- А) характеристика явлений в среде, непосредственно с ним не связанной
- Б) изменения явлений во времени
- В) распределение целого на части
- Г) частота явлений в среде, непосредственно с ним связанный

176. В статистике относительная величина, обозначается буквой

- А) М
- Б) m
- В) Р
- Г) р

177. Экстенсивные показатели можно представить видом диаграммы:

- А) секторная
- Б) столбиковая
- В) линейная
- Г) радиальная

178. Динамику явления за х-замкнутый цикл времени можно изобразить видом диаграммы:

- А) секторная
- Б) радиальная
- В) столбиковая
- Г) любая

179. Для графического изображения могут быть использованы виды величин:

- А) интенсивные и экстенсивные показатели
- Б) экстенсивные показатели соотношения

- В) абсолютные и интенсивные показатели
- Г) любые величины

180. Наиболее правильное определение вариационного ряда:

- А) это ряд числовых определённого признака, отличающихся друг от друга по всей величине и нормально расположенных в ранговом порядке.
- Б) два ряда величин, изменяющихся в убывающем или возрастающем порядке
- В) ряд числовых значений какого-то количественного порядка
- Г) статистический ряд, характеризующийся распределением чисел в убывающем или возрастающем порядке.

181. Определение варианты (V) –

- А) число из которого состоит вариационный ряд
- Б) число простого вариационного ряда
- В) число сгруппированного вариационного ряда
- Г) числовое значение изучаемого признака.

182. Определение частоты

- А) Частота, с которой встречаются каждая варианта
- Б) Частота, показывающая число вариант в вариационном ряду.
- В) Частота простого вариационного ряда
- Г) Частота показывающая разнообразие признака

183. Средняя величина – это...

- А) Величина, характеризующая типичный признаков совокупности
- Б) Величина, отражающая общие свойства статистической совокупности
- В) Обобщенная характеристика признака в статистической совокупности
- Г) размер признака в расчете на единицу однородной совокупности

184. В статистике средняя величина обозначается буквой.

- А) В
- Б) М
- В) А
- Г) Д

185. К средним величинам не относится

- А) Мода
- Б) Медиана
- В) Средняя арифметическая
- Г) средняя геометрическая

186. Наиболее правильное определение средней арифметической простой

- А) средняя величина, полученная путем простого суммирования вариант в деления суммы на общее число всех вариант
- Б) средняя величина получаемого из рядов, частоты которых равны единице или равны между собой
- В) средняя величина получаемая как частное от деления суммы вариант на сумму частот
- Г) Средняя величина, получаемая как частное от деления суммы вариант на сумму их частот, равных единице

187. Наиболее правильное определение средней арифметической взвешенной

- А) наиболее часто употребляемое и встречающееся в санитарной статистике и разновидность

средней арифметической величины.

Б) средняя величина, получаемая из вариационных рядов, варианты которые встречаются с различной частотой.

В) Средняя величина, получаемая путем деления суммы вариантов на общее число наблюдений

Г) Среднее из вариант, повторяющихся различное число раз.

188. Наиболее правильное определение моды :

А) Мода – средняя величина, которая в простом ряду может определять приближенно в сгруппированном по формуле.

Б) Мода- варианта , принимаемая за среднюю величину, используемая для оценки структуры изучаемого ряда.

В) Мода – варианта, встречающаяся с наибольшей частотой.

Г) Мода – средняя величина, обозначающая наиболее типичное значение изучаемого признака

189. Наиболее правильное определение медианы :

А) Медиана называется варианта, которая находится в середине вариационного ряда.

Б) средняя величина, обозначающая срединно-расположенную варианту в любых рядах.

В) Медиана – это варианта, расположенная в середине вариационного и делящая его на 2 равные части.

Г) средняя величина, обозначающая срединно расположенную варианту в сгруппированном вариационном ряду.

190. Наиболее правильное определение оценки достоверности:

А) Полученные результаты характеризуют количественную репрезентативность совокупности

Б) определить вероятность прогноза результатов изучения данной совокупности

В) Полученные результаты в изучаемой совокупности можно распределить на всю генеральную совокупность.

Г) установить вероятность безошибочного прогноза, с которым результаты, получены при изучении выборочной совокупности можно перенести на генеральную совокупность.

191. Мерой достоверности величин является:

А) средняя ошибка

Б) медиана

В) критерий достоверности

Г) амплитуда ряда

192. К способу определения достоверности не относится...

А) определение

Б) вычисление ошибки средней величины

В) вычисление ошибки относительной величины

Г) определение доверительных границ

193. Среднее квадратное отклонение характеризует..

А) позволяет судить о частоте изучаемого явления

Б) дает представление о разнообразии исследуемого признака

В) позволяет рассчитать амплитуду ряда

Г) позволяет судить о правильности вариационного ряда.

194. В статистике средняя ошибка обозначается буквой

А) Р

- Б) T
- В) m
- Г) M

195. Вероятность безошибочного прогноза обозначается буквой

- А) p
- Б) t
- В) m
- Г) q

196. Уровень вероятности безошибочного прогноза зависит:

- А) от величины доверительного интервала
- Б) от средних и относительных величин
- В) от величины ошибок средних величин
- Г) от величины ошибок относительных величин

197. Если полученная при выборочном исследовании разность между сравнительными или относительными показателями недостоверна, следует...

- А) использовать другие методы оценки достоверности
- Б) увеличить число наблюдений
- В) использовать другие виды величин
- Г) считать результаты окончательными

198. В медицинской статистике наиболее часто применяется метод стандартизации...

- А) метод экспертных ошибок
- Б) косвенный метод
- В) обратный метод
- Г) прямой метод

199. Метод стандартизации в медицинской статистике применяется с целью...

- А) для сравнения величин в пространстве и во времени
- Б) для динамического наблюдения
- В) для сравнения качественно неоднородных по составу групп
- Г) для изучения структуры явления

200. Первый этап стандартизации...

- А) выбор и расчет данных
- Б) вычисление общих и по интенсивным групповых показателей
- В) расчет стандартизованных показателей
- Г) вычисление ожидаемого числа в соответствии со стандартом

201. Третий этап стандартизации...

- А) вычисление общих и по интенсивным групповых показателей
- Б) расчет стандартизованных показателей
- В) сравнение групп
- Г) вычисление ожидаемого числа в соответствии со стандартом

202. Выбор и расчет стандарта производится на этапе метода стандартизации...

- А) первый этап
- Б) второй этап
- В) третий этап

Г) четвертый этап

203.Для отображения результатов исследований методом стандартизации используется вид таблицы...

- А) столбиковая диаграмма, простая таблица
- Б) простая таблица
- В) групповая таблица
- Г) комбинационная таблица

204.Сравнение групп по показателем производится на этапе метода стандартизации.

- А) пятом
- Б) третьем
- В) первом
- Г) четвертом

205.При расчетах методом стандартизации чаще всего используются показатели...

- А) интенсивный
- Б) экстенсивный
- В) наглядности
- Г) соотношения

206.Под медицинской статистикой понимают отрасль статистики, включающую.

- А) статистические методы по изучению здоровья населения
- Б) совокупность статистических методов, необходимых для анализа деятельности ЛПУ
- В) совокупность статистических методов по изучению здоровья населения и факторов, влияющих на него, а также вопросов, связанных с медициной и здравоохранением
- Г) статистические методы по изучению и совершенствованию управления в учреждениях здравоохранения

207.Предметом изучения медицинской статистики является информация...

- А) о здоровье населения
- Б) о экологии
- В) о кадровой политике
- Г) о санитарно-эпидемиологическом окружении

208.Статистический метод в медицине и здравоохранение применяется для...

- А) изучения психологической атмосферы
- Б) изучения состояния и деятельности социальных органов
- В) планирования научных исследований
- Г) изучения общественного здоровья и факторов, его определяющих

209.Основными методами формирования выборочной совокупности являются...

- А) научный
- Б) механический
- В) пилотажный
- Г) первичный

210.Программа статистического исследования включает..

- А) составление программы сбора материала
- Б) анализ социальных факторов
- В) определение объекта исследования

Г) определение исполнителей исследования

211. Результаты статистического исследования анализируются на основании...

- А) статистических учетных документов
- Б) амбулаторных карт
- В) статистических таблиц
- Г) историй болезни

212. К экстенсивным показателям относятся

- А) показатели рождаемости
- Б) распределение числа врачей по специальностям
- В) показатели младенческой смертности
- Г) обеспеченность населения врачами

213. К интенсивным показателям не относятся

- А) показатель смертности
- Б) показатели младенческой смертности
- В) показатель заболеваемости
- Г) структура заболеваний по нозологическим формам

214. Вариационный ряд-это...

- А) ряд чисел, отражающих частоту цифровых значений изучаемого признака
- Б) ряд цифровых значений различных признаков
- В) ряд числовых измерений признака, расположенных в ранговом порядке и характеризующихся определенной частотой
- Г) ряд любых чисел

215. Средняя арифметическая-это..

- А) варианта с наибольшей частотой
- Б) разность между наибольшей и наименьшей величиной
- В) обобщающая величина, характеризующая размер варьирующего признака совокупности
- Г) варианта, находящаяся в середине ряда

216. Средняя ошибка средней арифметической величины-это..

- А) средняя разность между средней арифметической и вариантами ряда
- Б) величина, на которую полученная величина средняя величина выборочной совокупности отличается от среднего результата генеральной совокупности
- В) величина, на которую в среднем отличается каждая варианта от средней арифметической
- Г) средняя разность между числами

217. Корреляционная связь может устанавливаться между признаками...

- А) ростом и массой тела у детей
- Б) частотой пульса и дыхания
- В) уровнем систолического и диастолического давления
- Г) частотой случаев хронических заболеваний и острых заболеваний

218. Метод стандартизации применяется..

- А) для определения характера и силы связи между двумя признаками
- Б) для определения очередности процессов
- В) для определения достоверности различия двух сравниваемых показателей

Г) для сравнения интенсивных показателей в неоднородных по составу совокупностях

219. Сущность метода стандартизации состоит в...

- А) установлении соответствия между сравниваемыми группами и эталоном
- Б) устранении влияния различий в составе сравниваемых групп на величину обобщающих показателей
- В) установлении достоверности различий двух сравниваемых групп по какому-либо показателю
- Г) установление достоверности и соответствия

220. При сравнении интенсивных показателей, полученных на однородных по своему составу совокупностях, необходимо применять..

- А) метод корреляции
- Б) метод стандартизации
- В) оценку достоверности разности полученных показателей
- Г) метод проб и ошибок

221. Динамический ряд-это..

- А) ряд числовых измерений определенного признака, отличающихся друг от друга по своей величине, расположенных в ранговом порядке
- Б) ряд числовых значений
- В) ряд величин, характеризующих результаты исследований в разных регионах
- Г) ряд, состоящий из однородных сопоставимых величин, характеризующих изменения какого-либо явления во времени

222. Способ преобразования динамического ряда

- А) укрупнение интервалов
- Б) вычисление средней
- В) вычисление коэффициента вариации
- Г) вычисление моды

223. Основными показателями динамического ряда НЕ является...

- А) темп роста
- Б) абсолютный прирост
- В) темп прироста
- Г) сигмальное отклонение

224. Макеты статистических таблиц создаются на этапе статистического исследования

- А) при составлении плана и программы исследований
- Б) на этапе сбора материала
- В) на этапе статистической обработки материалов
- Г) при проведении анализа результатов

225. Регистрация рождаемости и смертности относится к виду статистического наблюдения и методу статистического исследования...

- А) единовременное
- Б) сплошной
- В) ежедневное
- Г) выборочный

226. Наиболее объективную информацию о состоянии здоровья населения позволяют

получить способов наблюдения

- А) опрос
- Б) анкетирование
- В) выкопировка данных из медицинской документации
- Г) тестирование

227.К единовременному наблюдению относится...

- А) регистрация рождений
- Б) перепись населения
- В) регистрация браков
- Г) регистрация заболеваний

228.К текущему наблюдению не относится

- А) регистрация случаев смерти
- Б) регистрация случаев обращения в поликлинику
- В) учет родившихся
- Г) перепись населения

229.Для экспертной оценки качества и эффективности медицинской помощи в женской консультации отобрана каждая десятая «индивидуальная карта беременной и родильницы». Выборка является

- А) случайной
- Б) селективной
- В) когортной
- Г) социальной

230.Единица наблюдения определяется в зависимости от...

- А) программы исследования
- Б) плана исследования
- В) цели и задач исследования
- Г) предмета исследования

231.Из перечисленных видов статистических таблиц наиболее информативной является...

- А) простая
- Б) групповая
- В) комбинационная
- Г) сложная

232.Типологические группировки могут включать следующие признаки

- А) АД
- Б) рост
- В) массу тела
- Г) диагноз

233.Вариационные группировки могут включать следующие признаки

- А) рост
- Б) диагноз
- В) пол
- Г) профессию

СПИСОК ЛИТЕРАТУРЫ

1. Р. К. Биғалиева. Әлеуметтік медицина және денсаулық сақтауды басқару. - Алматы, «Эверо», 2016.
2. Методы научных исследований в медицине и здравоохранении / Элизабет Де Пой, Лаура Н. Гитлин; пер. с англ. под ред. В. В. Власова. — М.: ГЭОТАР-Медиа, 2017. — 432 с.: ил.
3. Кучеренко В.З. Применение методов статистического анализа для изучения общественного здоровья и здравоохранения.- Москва 2017г.
4. Общественное здравоохранение: учебник / А. А. Аканов [и др.]. — М.: Литтерра, 2017. — 496 с.
5. Общественное здоровье и здравоохранение: руководство к практическим занятиям: учебное пособие / В. А. Медик, В. И. Лисицин, М. С. Токмачев. — 2-е изд., испр. и доп. — М.: ГЭОТАР-Медиа, 2018. — 464 с.
6. .В.А.Медик, В.И.Лисицын, М.С.Токмачев. Общественное здоровье и здравоохранение: руководство к практическим занятиям:-2-е издание.испр.и доп. __ ГЭОТАР—Медиа, 2018г.
7. Основы доказательной медицины / Т. Гринхальх; пер. с англ. под ред. И. Н. Денисова, К. И. Сайткулова, В. П. Леонова. — 4-е изд., перераб. и доп. — М. : ГЭОТ АР-Медиа, 2018. — 336 с
8. Применение методов статистического анализа для изучения общественного здоровья и здравоохранении: учебное пособие/ ред. Кучеренко В.З.-4-е изд., перераб. и доп.- М.:ГЭОТАР-Медиа, 2011.-256 с.
9. Герасимов А.Н. Медицинская статистика: учебное пособие/ Герасимов А.Н.- М.:МИА, 2007.-480 с.
10. Триша Гринхальх Основы доказательной медицины. Москва, 2019. – 4 – издание.

