

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Башкирский государственный медицинский университет»  
Министерства здравоохранения Российской Федерации

# СТАТИСТИЧЕСКИЕ МЕТОДЫ В МЕДИЦИНЕ И ЗДРАВООХРАНЕНИИ

Учебное пособие



Уфа  
2018

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»  
МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
(ФГБОУ ВО БГМУ Минздрава России)

СТАТИСТИЧЕСКИЕ МЕТОДЫ  
В МЕДИЦИНЕ  
И ЗДРАВООХРАНЕНИИ

УЧЕБНОЕ ПОСОБИЕ

Уфа  
2018

УДК 61:31 (075,8)  
ББК 51.1 (2 Рос), 02  
С 78

Рецензенты:

Профессор, д.м.н., зав. кафедрой организации здравоохранения,  
общественного здоровья и медицинского права ФГБОУ ВО  
«Саратовский государственный медицинский университет  
им. В.И. Разумовского» Минздрава России *М.В. Еругина*

Профессор, д.м.н., зав. кафедрой общественного здоровья и организации  
здравоохранения с курсом медицинской информатики ФГБОУ ВО  
«Самарский государственный медицинский университет»  
Минздрава России *А.А. Гильманов*

**Статистические методы в медицине и здравоохранении:** учеб.  
**С 78** пособие: / сост.: Н.Х. Шарафутдинова, Э.Ф. Киреева, И.Е. Николаева,  
М.Ю. Павлова, Р.М. Халфин, М.А. Шарафутдинов, М.В. Борисова, А.Б.  
Латыпов, А.Ш. Галикеева. – Уфа: ФГБОУ ВО БГМУ Минздрава России,  
2018. – 131 с.

Учебное пособие подготовлено в соответствии с требованиями ФГОС ВО и ООП по специальностям 31.05.01 «Лечебное дело», 31.05.02 «Педиатрия», 31.05.03 «Стоматология», 32.05.01 «Медико-профилактическое дело» для изучения дисциплины «Общественное здоровье и здравоохранение» на основании рабочих программ (2017г.) и действующего учебного плана (2017г.).

В нем изложены основы применения статистического анализа в медицине и здравоохранении. На наглядных примерах рассмотрены конкретные методики математической статистики в приложении к практике медико-биологических, клинических и социально-гигиенических исследований.

Пособие позволяет обучающимся медицинских вузов познакомиться с организацией статистического исследования и приобрести практические навыки статистической обработки данных и их анализа. Материалы пособия могут быть использованы обучающимися при подготовке к практическим занятиям и экзаменам.

Предназначено для обучающихся по специальностям 31.05.01 «Лечебное дело», 31.05.02 «Педиатрия», 31.05.03 «Стоматология», 32.05.01 «Медико-профилактическое дело».

Рекомендовано в печать Координационным научно-методическим советом и утверждено решением редакционно-издательского совета ФГБОУ ВО БГМУ Минздрава России.

**УДК 61:31 (075,8)**  
**ББК 51.1 (2 Рос), 02**

© ФГБОУ ВО БГМУ Минздрава России, 2018  
© Н.Х. Шарафутдинова, Э.Ф. Киреева, И.Е. Николаева,  
М.Ю. Павлова, Р.М. Халфин, М.А. Шарафутдинов,  
М.В. Борисова, А.Б. Латыпов, А.Ш. Галикеева, 2018

## ОГЛАВЛЕНИЕ

Список сокращений и условных обозначений.....	5
Введение.....	6
1. ДОКАЗАТЕЛЬНАЯ МЕДИЦИНА.....	8
1.1. Доказательная медицина. Критерии доказательности.....	8
1.2. Основные дизайны клинических исследований.....	10
1.3. Рандомизированное контролируемое испытание.....	12
2. ОРГАНИЗАЦИЯ И МЕТОДИКА СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ.....	14
2.1. Этапы статистического исследования.....	14
2.2. Требования к оформлению статистических таблиц.....	22
2.3. Графические изображения.....	25
3. ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ.....	32
4. СТАНДАРТИЗОВАННЫЕ КОЭФФИЦИЕНТЫ.....	36
4.1. Прямой метод стандартизации.....	38
4.2. Косвенный метод стандартизации.....	39
4.3. Обратный метод стандартизации.....	40
5. ДИНАМИЧЕСКИЕ РЯДЫ.....	43
6. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ.....	48
7. СРЕДНИЕ ВЕЛИЧИНЫ. ВАРИАЦИОННЫЙ РЯД. ПОКАЗАТЕЛИ ВАРИАЦИИ.....	53
7.1. Вариационный ряд, средние величины.....	53
7.2. Среднее квадратическое отклонение.....	57
8. ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ.....	62
8.1. Определение средней ошибки показателя равных или близких к 0 или 100 % .....	64
8.2. Доверительный интервал. Оценка достоверности средних и относительных величин.....	65
9. ОСНОВНЫЕ МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ КЛАССИФИКАЦИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ.....	67
10. ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ.....	72
10.1. Сравнение двух независимых выборок. Критерий Стьюдента ( $t$ ).....	72

10.2. Сравнение двух зависимых выборок. Критерий Стьюдента ( $t$ ) (разностный метод).....	74
11. НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ.....	77
11.1. Метод сравнения двух независимых выборок. Критерий Манна-Уитни ( $U$ ).....	77
11.2. Методы сравнения двух зависимых выборок. Критерий Вилкоксона ( $T$ ).....	81
11.3. Критерий согласия ( $\chi^2$ ).....	83
11.4. Точный метод Фишера.....	83
11.5. Степени свободы.....	85
12. Материалы для самоподготовки.....	87
12.1. Тестовые задания.....	87
12.2. Ситуационные задачи .....	92
12.3. Эталоны ответов для тестовых заданий .....	94
12.4. Ответы к ситуационным задачам.....	94
13. Рекомендуемая литература.....	100
14. Приложения.....	101
15. Словарь терминов.....	121

## СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

### *Сокращения*

ДМ – доказательная медицина

ДИ – доверительный интервал

РКИ – рандомизированное клиническое испытание

Табл. – таблица

ТМФ – точный метод Фишера

Рис. – рисунок

### *Условные обозначения*

М – средняя арифметическая величина

$S_v$  – коэффициент вариации

$\rho$  – коэффициент ранговой корреляции

n – число наблюдений;

r – коэффициент линейной корреляции

T – Критерий Вилкоксона

t – коэффициент Стьюдента

$\sigma$  – среднее квадратическое отклонение

U – критерий Манна-Уитни

## ВВЕДЕНИЕ

В учебном пособии определены базовые понятия медико-биологической статистики, представлены классификация и характеристика методов статистического анализа данных, рассматривается методика проведения анализа на конкретных примерах, таким образом, чтобы студент-медик, взяв это учебно-методическое пособие и определив цель исследования, мог самостоятельно получить необходимый результат.

Материал изложен на общих методологических подходах к организации и проведению научного исследования в медицине и здравоохранении. Результаты всех медицинских исследований, и аналитических, и описательных зависят от правильной организации научного исследования, надлежащего сбора, анализа и интерпретации соответствующих числовых данных. Состоятельность таких исследований и их результатов приобретает особую актуальность в связи с повсеместным распространением и принятием концепции «медицины, основанной на доказательствах» (evidence-based medicine, ЕВМ), которая предполагает, что к широкому применению в медицинской практике должны рекомендоваться те вмешательства, эффективность и безопасность которых подтверждена в высококачественных исследованиях, выполненных на основе единых методологических принципов.

Настоящее учебное пособие составлено в соответствии в соответствии с требованиями ФГОС ВО и ООП по специальностям 31.05.01 «Лечебное дело», 31.05.02 «Педиатрия», 31.05.03 «Стоматология», 32.05.01 «Медико-профилактическое дело», с учебным планом, рабочей программой дисциплины «Общественное здоровье и здравоохранение, экономика здравоохранения».

Процесс изучения вопросов, освещенных в данном учебном пособии, направлен на формирование следующих компетенций у обучающихся: по специальности 31.05.01 «Лечебное дело» и 31.05.02 «Педиатрия» ПК-4 – способность и готовность к применению социально-гигиенических методик сбора и медико-статистического анализа информации о показателях здоровья населения и здравоохранения; ПК-20 – го-

товность к анализу и публичному представлению медицинской информации на основе доказательной медицины; по специальности 31.05.03 «Стоматология» ПК-4 – способность и готовность к применению социально-гигиенических методик сбора и медико-статистического анализа информации о стоматологической заболеваемости, ПК-17 – готовность к анализу и публичному представлению медицинской информации на основе доказательной медицины; по специальности 32.05.01 «Медико-профилактическое дело» ОПК-7 – способность применять современные методики сбора и обработки информации, проводить статистический анализ и интерпретировать результаты, изучать, анализировать тенденции, прогнозировать развитие событий и состояние популяционного здоровья.



# 1. ДОКАЗАТЕЛЬНАЯ МЕДИЦИНА

Доказательная медицина (ДМ, англ. Evidence-based medicine — медицина, основанная на доказательствах) — подход к медицинской практике, при котором решения о применении профилактических, диагностических и лечебных мероприятий принимаются исходя из имеющихся доказательств их эффективности и безопасности, а такие доказательства подвергаются поиску, сравнению, обобщению и широкому распространению для использования в интересах пациентов (Evidence Based Medicine Working Group, 1993). Термин «доказательная медицина» был предложен группой канадских учёных из Университета Мак-Мастера в 1990 году (Т. Гринхальх, 2004).

В основе ДМ лежит проверка эффективности и безопасности методик диагностики, профилактики и лечения в клинических исследованиях. В большинстве стран стали общепризнанными некоторые правила проведения клинических исследований, изложенные в стандарте GCP (Good Clinical Practice, «Надлежащая клиническая практика»), а также правила производства лекарственных средств (стандарт GMP) и выполнения лабораторных исследований (стандарт GLP).

## 1.1. Доказательная медицина.

### Критерии доказательности

Важнейшим механизмом, позволяющим врачу ориентироваться в огромном количестве публикуемых медицинских данных, является рейтинговая система оценки научных исследований. Однако до настоящего времени отсутствует общепринятая шкала по оценке степени доказательности.

Категорий доказательности исследования может быть – от 3 до 7. Меньшей цифре соответствует меньшая доказательность. Также выделяют три категории доказательности исследования.

**Категория I** – крупные, рандомизированные, контролируемые исследования, данные мета-анализа, систематических обзоров.

**Категория II** – когортные исследования и исследования типа случай-контроль.

**Категория III** – неконтролируемые исследования и консенсусы специалистов.

Практические рекомендации по диагностике и лечению могут опираться как на данные исследований, так и на их экстраполяцию. С учетом этого обстоятельства рекомендации подразделяют на 3-5 уровней, которые принято обозначать латинскими буквами – А, В, С, D, E.

*А – доказательства убедительны:* есть веские доказательства в пользу применения данного метода;

*В – относительная убедительность доказательств:* есть достаточно доказательств в пользу того, чтобы рекомендовать данное предложение;

*С – достаточных доказательств нет:* имеющихся доказательств недостаточно для вынесения рекомендации, но рекомендации могут быть даны с учетом иных обстоятельств;

*D – достаточно отрицательных доказательств:* имеется достаточно доказательств, чтобы рекомендовать отказаться от применения данного метода в определенной ситуации;

*E – веские отрицательные доказательства:* имеются достаточно убедительные доказательства, чтобы исключить данный метод из рекомендаций.

*В странах Западной Европы и Канаде выделяют три уровня доказательности – А, В, С. Рекомендации уровня А отличаются наиболее высоким уровнем достоверности и базируются на результатах исследований, относимых к I категории доказательности. Достоверность рекомендаций уровня В также достаточно высока, при их формулировании используются материалы исследований II категории или экстраполяции исследований I категории доказательности. Рекомендации уровня С строятся на основании неконтролируемых исследований и консенсусов специалистов (III категории доказательности) или содержат экстраполяции рекомендаций I и II категории.*

*Американское Агентство по политике здравоохранения и научным исследованиям* предложило более простую шкалу с выделением следующих уровней доказательств в порядке убывания их достоверности:

I a – мета-анализ РКИ;

I b – как минимум одно РКИ;

II a – хорошо организованное контролируемое исследование без рандомизации;

II b – как минимум одно хорошо организованное квази-экспериментальное исследование;

III – не экспериментальные исследования;

IV – отчет экспертного комитета или мнение и/или клинический опыт авторитетных специалистов, где главную роль играет субъективный фактор, в частности личный опыт врача и мнение экспертов (Т. Гринхальх, 2004).

Группами экспертов разрабатываются научно-обоснованные клинические рекомендации по определенным проблемам. Результаты систематических обзоров используют фармацевтические компании в качестве аргументов для включения своих препаратов в национальный формуляр.

## **1.2. Основные дизайны клинических исследований**

Для каждой структуры исследования сегодня сформировались требования к его подготовке, специфические стандарты качества, обработки и представления данных. От структуры исследования зависит главное – доказательность его результатов. Все виды исследований расположены в порядке убывания достоверности их результатов.

**Систематический обзор** – исследование, в ходе которого проводится критический анализ и оценка (а не просто совместное рассмотрение) результатов других исследований для того, чтобы ответить на заранее сформулированный ясный клинический вопрос при помощи методов, позволяющих свести к минимуму возможность появления систематической ошибки. Для этого проводится систематический поиск всех опубликованных и неопубликованных РКИ по данному вопросу, проводится их качественная оценка и врачу предоставляется готовое решение: эффективно вмешательство или нет, либо имеющейся информации недостаточ-

но для использования вмешательства. Деятельностью по подготовке систематических обзоров занимается всемирное Кокрановское Сотрудничество врачей (названо так в честь английского эпидемиолога А. Кокрана, впервые сформулировавшего концепцию ДМ). Сотрудничество формирует также реферативную базу данных, включающую рефераты публикаций о контролируемых и рандомизированных клинических испытаниях, отвечающих современным стандартам качества их проведения.

**Мета-анализ** – обзор, в котором при помощи количественного метода оценки обобщены данные нескольких исследований, а итоговый результат представлен в виде одного средневзвешенного показателя.

**Рандомизированное клиническое испытание (РКИ)** – исследование диагностического теста, метода профилактики или лечения, в котором участники в случайном порядке (подобному подбрасыванию монеты, например, в конверты вкладываются директивы оперировать/лечить консервативно и при поступлении больного врач вскрывает конверт и следует директиве). Распределяются в основную и контрольную группы, после чего за ними осуществляется наблюдение для определения эффекта вмешательства.

**Слепое исследование** – когда пациент не знает, к какой группе он находится (например, к группе плацебо или основной).

**Двойное слепое исследование** – когда ни доктор, ни пациент не знают, к какой группе относится пациент. Примеры: лучше ли данный препарат по сравнению с плацебо или другим препаратом при данном заболевании; лучше ли лапароскопическая аппендэктомия по сравнению с открытым классическим способом.

**Когортное исследование (обсервационное) – проспективное** исследование факторов, которые могут послужить причиной развития того или иного заболевания. В ходе него из лиц без изучаемого клинического исхода формируются две группы, в одной из которых участники подвергаются воздействию вредного фактора, а в другой – нет. При последующем наблюдении сравнивают частоту развития клинического исхода. Используется также и в тех случаях, когда невозможно выполнить рандомизированное испытание. Примеры: приводит ли курение к развитию рака

легких; каковы физическое развитие и успехи в учебе у детей школьного возраста, родившихся недоношенными.

**Исследования «случай-контроль»** – исследование, при котором в основную группу включают пациентов с уже развившимся клиническим исходом, в контрольную – представителей популяции. Затем анализируются факторы (путем опроса, анализа историй болезней, амбулаторных карт), которые могли оказать возможное влияние на развитие заболевания. Такой дизайн – практически единственная возможность изучения редких заболеваний. Примеры: увеличивает ли наклонное положение тела во время сна риск внезапной смерти ребенка; есть ли причинная связь между высоковольтными линиями электропередач и возникновением лейкозов.

**Описание случаев** – это история болезни одного пациента. Чаще всего это описание ранее не встречавшейся побочной реакции. Описание серии случаев – описание нескольких пациентов со сходной патологией. Пример: в 1965 году на основании описания серии случаев рождения детей с аномалиями развития, из продажи был отозван препарат талидомид, впоследствии подтвердился его тератогенный эффект.

### **1.3. Рандомизированное контролируемое испытание**

Важнейший принцип научно-обоснованной медицинской практики связан с критическим анализом информации: “вес” каждого факта тем больше, чем строже научная методика исследования, в ходе которого факт получен. “Золотым стандартом” считаются рандомизированные контролируемые исследования (РКИ). Индивидуальный врачебный опыт и мнение экспертов или “авторитетов”, рассматриваются как не имеющие достаточной научной основы. РКИ предполагают наличие опытной и контрольной групп, пациентов распределяют по группам случайным образом (*рандомизация*), следя при этом, чтобы группы не различались по параметрам, влияющим на исход заболевания. Врач-исследователь, а тем более сам пациент не знают, получает ли больной плацебо или лекарство (*двойной слепой метод*). Все пациенты прослеживаются в течение определенного, часто весьма длительного отрезка времени (перспективное исследование), по истечении которого сравнивается частота наступления

клинически важных конечных точек (выздоровление, смерть, осложнения) в опытной и контрольной группах. Нередко для проведения подобных исследований привлекаются тысячи и десятки тысяч больных, в разных научных центрах и странах.

РКИ в современной медицинской науке является общепризнанным эталоном научного исследования для оценки клинической эффективности.

1. *Сравнение конечных результатов должно быть проведено в 2-х группах пациентов:*

1) Контрольная группа (лечение не проводится, лечение проводится стандартное, традиционное лечение, пациенты получают плацебо);

2) Группа активного лечения (проводится лечение, эффективность которого исследуется).

2. *Группы пациентов должны быть сопоставимы и однородны, по крайней мере, по следующим показателям:*

- клинические особенности заболевания и сопутствующая патология;

- возраст, пол, расовая принадлежность.

3. *Группы пациентов должны быть репрезентативны* (то есть, количество пациентов в каждой группе должно быть достаточным для получения статистически достоверных результатов).

4. *Распределение пациентов на группы должно происходить случайным образом* (рандомизация).

5. *При оценке результатов должен быть исключен субъективный фактор.* Для этого используют несколько типов исследований:

- простое слепое (о принадлежности к определенной группе не знает пациент);

- двойное слепое (о принадлежности к определенной группе не знают пациент и лечащий врач);

- тройное слепое (о принадлежности к определенной группе не знают пациент, лечащий врач и организатор испытания).

6. *Должны быть использованы объективные критерии конечных результатов:*

- смертность от данного заболевания;

- общая смертность;
- частота развития «больших» осложнений, в особенности опасных для жизни и инвалидизирующих;
- частота повторных госпитализаций;
- оценка качества жизни.

7. *Результаты должны быть практически значимы и информативны*, что может быть осуществлено только при достаточно длительном наблюдении за пациентом и низким числом отказов пациентов от продолжения участия в исследовании.

**Вопросы для самоподготовки:**

1. Что такое доказательная медицина?
2. Перечислите критерии доказательства
3. Дайте определение рандомизированного клинического исследования.
4. Что такое простое- и двойное слепое исследование?
5. Что такое систематический обзор?
6. Дайте определение мета-анализу.

## **2. ОРГАНИЗАЦИЯ И МЕТОДИКА СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ**

*Статистика* – наука, изучающая закономерности массовых явлений методом обобщающих показателей. В здравоохранении статистические методы исследования используют для:

- 1) изучения здоровья населения;
- 2) анализа, оценки и планирования медицинской помощи;
- 3) специальных научных исследований.

### **2.1. Этапы статистического исследования**

Статистическое исследование состоит из 4 основных этапов:

- I этап – составление плана и программы исследования.
- II этап – сбор материала (на специально разработанные первичные документы).
- III этап – обработка и сводка собранного материала.

- IV этап статистический анализ.

**I этап статистического исследования – составление плана и программы исследования.**

*План исследования* включает:

а) *определение цели и задач* исследования. Формулируется главная цель исследования и подцели (или задачи), последовательное решение которых обеспечивает достижение главной цели;

б) *знакомство с литературой* позволяет получить представление об изучаемой проблеме, выбрать адекватную методику исследования, выяснить, не проводились ли в недавнем времени аналогичная работа (в связи с чем может исчезнуть необходимость в проведении данного исследования). На основании знакомства с литературой составляется также рабочая гипотеза;

в) *составление рабочей гипотезы* (то есть предположения о результатах исследования). В случае несовпадения полученных результатов с рабочей гипотезой исследователь обязан проверить, во-первых, не допущена ли в работе счетная ошибка; во вторых, насколько адекватной поставленной цели была выбрана методика исследования;

г) *определение единицы наблюдения* – первоначального элемента, из которого складывается статистическая совокупность (или первичного носителя информационных признаков). Выбор единицы наблюдения определяется главным образом целью и задачами исследования;

д) *выбор метода исследования.*

**Генеральная совокупность** – это совокупность, в которой изучаются все единицы наблюдения, метод считается сплошным.

**Выборочная статистическая совокупность** – это совокупность, на которой в исследование берется часть генеральной совокупности. Выборочная совокупность должна быть **репрезентативной**, то есть отражать свойства генеральной совокупности.

Существуют специальные формулы для определения необходимого числа наблюдений, обеспечивающего количественную репрезентативность выборки.

Условием обоснованного расчета необходимого числа наблюдений является допущение возможной ошибки, то есть предположение макси-



мально допустимого отклонения результатов выборочного исследования от генеральных значений.

Формула для определения предельной ошибки показателя:

$$\Delta = t \sqrt{\frac{pq}{n}}, \text{ где}$$

$\Delta$  – предельная ошибка показателя;

$p$  – величина показателя (для изучаемого признака);

$q = (1 - p)$  или  $(100 - p)$  – в зависимости от того, в каких величинах выражается показатель;

$n$  – число наблюдений;

$t$  – коэффициент, показывающий, с какой вероятностью (надежностью) можно гарантировать достоверность полученного результата. Обычно  $t$  берется равным 2.

Исходя из формулы предельной ошибки показателя, можно вывести формулу числа наблюдений:

$$n = \frac{t^2 pq}{\Delta^2}$$

Для расчета необходимого числа наблюдений может использоваться и следующая формула:

$$n = \frac{Nt^2 pq}{N\Delta^2 + t^2 pq}, \text{ где}$$

$N$  – численность генеральной совокупности.

В случае изучения количественных признаков применяется формула предельной ошибки средней величины:

$$\Delta = \frac{\sigma}{\sqrt{n}} t \quad n = \frac{t^2 \sigma^2}{\Delta^2}, \text{ где}$$

$\sigma$  - среднее квадратное отклонение (его величина определяется из предыдущих исследований или при собственном, то есть пробном исследовании).

**Качественная репрезентативность** выборки может быть достигнута несколькими путями (важно, чтобы единица генеральной совокупности имела равную возможность попасть в выборочную совокупность):

а) типологическая выборка – предполагает разбивку изучаемого объекта на ряд однотипных качественных групп, из которых далее отби-

раются единицы для наблюдения (с учетом размера этих групп – пропорциональный отбор или равномерно-непропорциональный отбор);

б) случайный отбор – по жребию, начальной букве фамилии и т.д.;

в) механический – отбор каждого десятого и т.д. медицинского документа;

г) гнездовой – первоначальный выбор определенных групп (гнезд) с последующим их изучением сплошь или выборочно (механически или случайно);

д) направленный (в соответствии с заранее разработанной методикой).

Следующим вопросом, отражаемым в плане исследования, является его характер, который может быть единовременным и текущим.

**При единовременном наблюдении** исследователь получает информацию об объекте на момент времени (например, состояние здоровья какой-либо группы населения по данным профилактического осмотра); **при текущем наблюдении** – в течение определенного времени (месяца, года и т.д.).

**Объем исследования** – это количество включенных в него единиц наблюдения.

**Объект исследования** – конкретная статистическая совокупность, которую изучают.

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

- планируемые сроки работы;

- необходимые и имеющиеся в наличии материальные и финансовые ресурсы;

- тип работы (отчет, статья, диссертация, курсовые или дипломная работа и т.д.).

**Программа исследования** – это перечень вопросов, подлежащих изучению.

Составляя программу своего исследования, исследователь, руководствуется целью работы (при этом важно как включать в программу все необходимые вопросы и не перегружать ее излишними вопросами); во-вторых, он должен четко представлять, откуда, из каких источников, будет получена информация, чаще всего, из медицинской документации, от-

веты на все сформулированные вопросы.

Вопросы программы составляют **признаки изучаемой совокупности**, которые делятся на *качественные* (атрибутивные, типологические) и *количественные* (вариационные).

### **II этап статистического исследования – сбор материала.**

Включает регистрацию всех запланированных признаков для каждой единицы наблюдения. При этом на каждую единицу наблюдения заводится свой регистрационный документ (анкета, учетно-отчетная медицинская документация, специально разработанная карта исследования и т.д.). Сбор данных может осуществляться разными методами: анкетирование, интервьюирование, выкопировка сведений из первичной медицинской документации и т.д.

### **III этап – разработка и сводка данных.**

Включает следующие последовательно выполняемые исследователем действия:

**1. Контроль** – проверка собранного материала с целью отбора учетных документов, имеющих дефекты для их последующего исправления, дополнения или исключения из исследования. Например, в анкете не указан пол, возраст или нет ответов на другие поставленные вопросы.

В этом случае необходимы дополнительные данные, например из официальных документов (амбулаторных карт, историй болезни и т.п.). Если эти данные не могут быть получены из дополнительных учетных документов, привлеченных исследователем или от самого человека, то некачественные карты (анкеты) должны быть исключены из исследования.

**2. Группировка материала** – распределение собранного материала по атрибутивному и/или количественному признакам (типологическая или вариационная). Например, группировка студентов по курсам обучения: I курс, II курс, III курс, IV курс, V курс, VI курс.

**3. Шифровка** – применение условных обозначений выделяемых признаков. При ручной обработке материала шифры могут быть цифровые, буквенные; при машинной – только цифровые.

Например, Буквенная шифровка: Пол: муж. М  
жен. Ж

Цифровая шифровка:

Возрастная группировка	Шифр
до 20 лет	1
21-29	2
30-39	3
40-49	4

**4. Сводка материала** – занесение полученных после подсчета цифровых данных в таблицы.

**5. Вычисление статистических показателей и статистическая обработка материала.**

**6. Графическое изображение полученных показателей.**

**IV этап – статистический анализ результатов. внедрение полученных результатов в практику.**

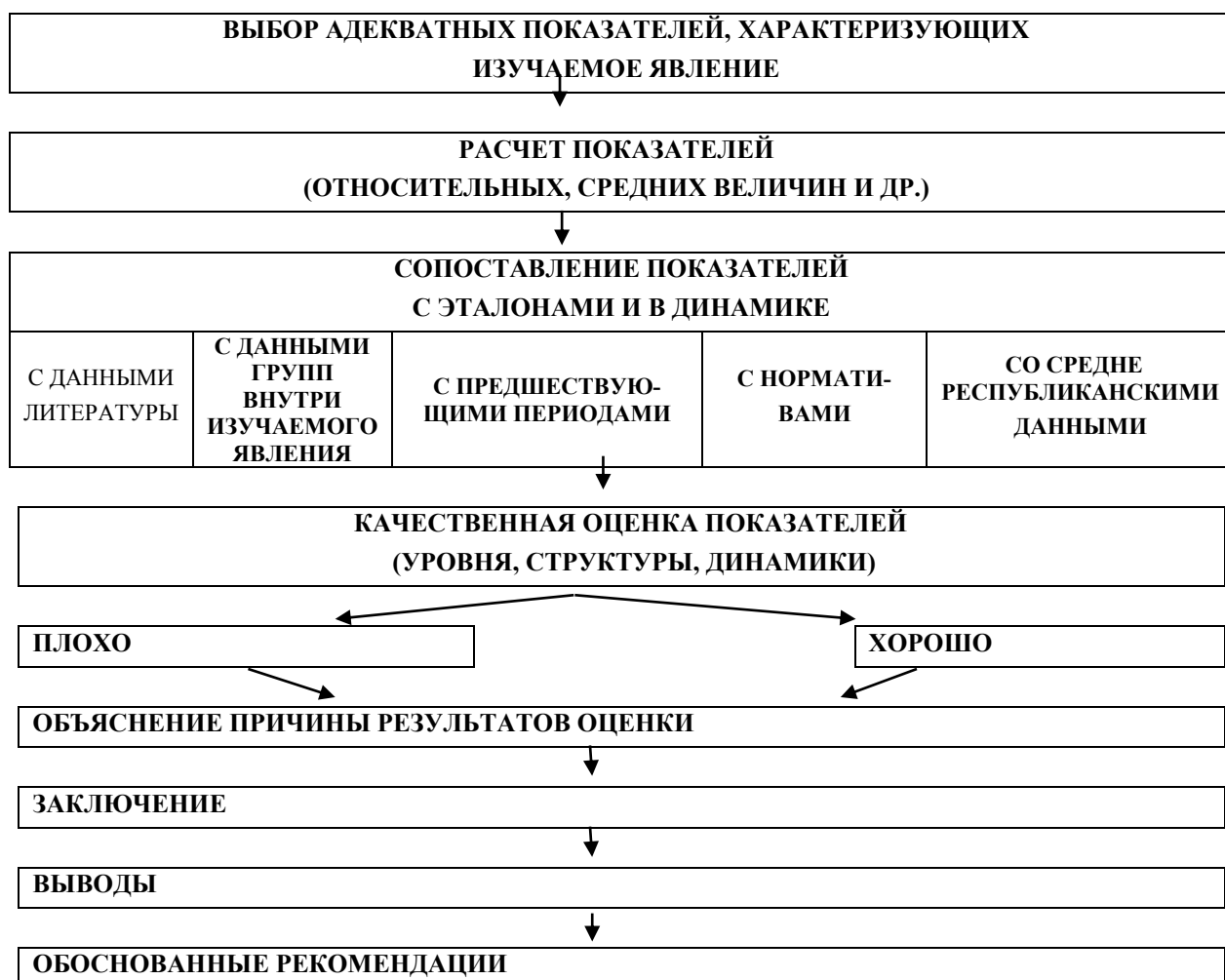


Схема 1. Последовательность анализа различных явлений

Пример составления плана и программы статистического исследования:

**Тема:** Анализ заболеваемости болезнями органов дыхания работающих на Хлопчатобумажном комбинате (ХБК) в городе Н за 2017 год.

**Цель исследования:** разработать комплексный план оздоровительных мероприятий по профилактике болезней органов дыхания у работающих на ХБК

**Задачи исследования:**

1. Составить адекватную методику исследования.
2. Изучить распространенность болезней органов дыхания в различных цехах ХБК.
3. Изучить и выявить факторы условий труда, влияющие на распространенность болезней органов дыхания у работающих на ХБК

**План исследования:**

1. Объект – работающие на ХБК города Н в 2017 г.
2. Пути формирования объекта.
  - 2.1. По охвату – наблюдение выборочное, метод отбора механический.
  - 2.2. По времени – текущее.
  - 2.3. По виду – наблюдение непосредственное (микrokлиматических параметров) и путем выкопировки (данных о заболеваемости).
3. Объем исследования – 100 работающих.
4. Сроки исследования – с 1 января 2017 года по 31 декабря 2017 года.
5. Место исследования – ХБК города Н.
6. Руководитель: зав. цеховым терапевтическим отделением поликлиники.
7. Исполнители: цеховой терапевт и помощник санитарного врача отделения гигиены труда ФГУЗ «Центр гигиены и эпидемиологии в РБ».
8. Материально-техническое обеспечение:
  - лабораторное оборудование и приборы для изучения микrokлиматических условий;
  - канцелярские товары;
  - без дополнительных ассигнований.

**Программа исследования:**

**1. Программа сбора**

1.1. Единица наблюдения – работающий на ХБК города Н в 2017 г.

1.2. Первичный учетный документ для изучения заболеваемости верхних дыхательных путей составляется в виде таблицы данных.

№ п/п	Учетные признаки	Градация признака	Шифр
1.	Пол	Мужской	1
		Женский	2
2.	Возраст	до 19	1
		20-29	2
		30-39	3
		40-49	4
		50-59	5
		60 и более	6
3.	Стаж работы	до 5	1
		6-10	2
		11-15	3
		16-20	4
		21-25	5
		26 и более	6
4.	Профессия	пряильщица	1
		ткачиха	2
		ленточница	3
		наладчик	4
		механик	5
		другие	6
5.	Скорость движения воздуха	ниже нормы	1
		в пределах нормы	2
		выше нормы	3
6.	Температура воздуха	ниже нормы	1
		в пределах нормы	2
		выше нормы	3
7.	Диагноз	Острый назофарингит	1
		Острый бронхит	2
		Хронический бронхит	3
		Пневмония	4
		ХОБЛ	5
		Др. _____	6
8.	Число обострений в год	не болел	0
		1-3 раза	1
		4-6 раз	2
		7 и более	3
9	Частота ОРВИ	не болел	0
		1-3 раза	1
		4-6 раз	2
		7 и более	3

## **2. Программа разработки**

### **2.1. Лист намеченных таблиц**

№ п/п		Простая таблица	Групповая таблица	Комбинационная
1.	Пол			+
2.	Возраст		+	
3.	Стаж работы			+
4.	Профессия		+	
5.	Скорость движения воздуха			+
6.	Температура воздуха		+	
7.	Диагноз	+	+	+
8.	Число случаев заболевания	+		

Оформить макеты таблиц: простой, групповой, комбинационной (см. раздел данной методички «Статистические таблицы»).

### **2.2. Требования к оформлению статистических таблиц**

1. Сверху-справа над названием таблицы пишется слово “Таблица” и дается нумерация арабскими цифрами без проставления знака “№”.

2. Каждая таблица должна иметь название, из которого становится ясным круг излагаемых вопросов, географические границы статистической совокупности, период времени за который или момент времени, на который относятся данные. Иногда в названии подчеркиваются основные выводы, которые могут быть сделаны из таблицы. Точку в конце заголовка не ставят. Переносы в заголовках не допускаются. В названии указываются единицы изменения, если они неодинаковые, то в верхних или боковых заголовках (заголовки граф называются верхними, а заголовки строк – боковыми) следует указать, в каких единицах приведены данные. Если каждая строчка имеет свою особую единицу измерения, то для их обозначения следует указать, в каких единицах приведены данные. Если каждая строчка имеет особую единицу измерения, то для их обозначения следует отводить специальную графу.

3. В таблице все слова пишутся полностью. Допускаются только общепринятые сокращения /например, мм, млрд, кг и т.д./.

4. Если таблица большая и не умещается на одном лице, ее переносят на следующую страницу, на которой заголовочную часть таблицы не пишут, а повторяют только нумерацию граф. В длинных таблицах полезно оставлять двойной промежуток после каждых пяти или десяти строк.

5. Желательно давать нумерацию граф. Первая графа предназначена для боковых заголовков и потому не входит в нумерацию.

Не следует строить громоздких таблиц с очень большим числом граф, иногда целесообразно некоторые графы объединить под рубрикой “прочие”. Однако она не должна охватывать более 10 % общего итога.

6. Рекомендуется пользоваться двойными горизонтальными и вертикальными линиями для отделения табличных клеток от заголовков и отделения некоторых частей таблицы друг от друга.

7. Большое значение в статистической таблице имеет итоги. Они бывают вертикальные и горизонтальные. Вертикальные итоги получаются в результате суммирования всех чисел графы, а горизонтальные – строки. Без итогов таблица не может считаться законченной.

Следует различать “итога” и “всего”. “Итого” является итогом для определенной части, а “Всего” – для всей совокупности, для всей таблицы в целом.

8. Если нет необходимости в том, чтобы приводить все элементы итога, следует ограничиться некоторыми из них, спустив остальные. В таких случаях делают приписку “в том числе” или “из них” и заголовок строки несколько отодвигают вправо.

9. При употреблении в клетках таблицы многозначных чисел следует каждые три цифры, отсчитывая их справа, отделять друг от друга /напр. 4 582 973/.

10. Отсутствие чисел в клетке может иметь ряд причин, и эти причины должны быть показаны в таблице и обозначены знаками:

(...) – отсутствие сведений о данном факте;

(-) – отсутствие самого явления;

(X) – вообще не подлежит заполнению.

11. При монтаже таблицы надо стремиться к тому, чтобы числа, которые особенно важно сравнивать друг с другом, были расположены ря-



дом (например, показатели за смежные периоды, показатели экспериментальной и контрольной групп, показатели плана и его выполнения).

12. Необходимо указывать источник данных, включенных в таблицу. При построении таблицы могут быть использованы *сноски*. Они могут относиться либо к отдельным строчкам, либо к графам, либо к клеткам. Сносками пользуются для того, чтобы указать на ограничительные обстоятельства, которые надо принять во внимание при чтении таблицы.

От сносок надо отличать *примечания*: они относятся обычно ко всей таблице и содержат в себе указания о методах, положенных в основу получения тех или иных показателей. Примечания к таблицам рекомендуется давать тут же после таблицы.

13. В каждой табличной клетке должно быть только одно число. Не рекомендуется представлять цифровой материал в виде числителя и знаменателя.

Различают *три вида таблиц*: простые, групповые и комбинационные.

В *простой* таблице приводится подсчет по основным делениям изучаемой группы или совокупности (табл. 1).

В *групповой* таблице различается статистическое подлежащее (основной изучаемый признак, например заболевания) и статистические сказуемые (признаки, характеризующие основной, например возраст и пол). В групповой таблице подлежащее сочетается каждый раз только с одним из признаков сказуемого (табл. 2).

При одновременном сочетании подлежащего таблицы с несколькими признаками сказуемого таблица носит название *комбинационной* (табл. 3).

Таблица 1

**Распределение болезней органов дыхания по нозологическим формам у работников ХБК в г. Н в 2017 году**

Нозологические формы	Число заболеваний	
	абс. ч.	% к итогу
Острая респираторная инфекция		
Острый бронхит		
Пневмония		
Прочие		
<b>Всего</b>		

Таблица 2

**Распределение больных по нозологическим формам  
болезней органов дыхания, полу и возрасту, работающих  
на ХБК г. в г. Н за 2017 г.**

Нозологические формы	Пол		Итого	Возраст, лет						Итого
	м	ж		до 20	20-29	30-39	40-49	50-59	60 и более	
Острая респираторная инфекция										
Острый бронхит										
Пневмония										
Прочие										
<b>Всего</b>										

Таблица 3

**Распределение больных по нозологическим формам болезней  
органов дыхания в зависимости от стажа работы  
и пола на ХБК в г. Н за 2017 г.**

Нозологические формы	Стаж работы (в годах)																		Итого
	до 5 лет			6-10			11-15			16-20			21-25			25 и более			
	м	ж	Итого	м	ж	Итого	м	ж	Итого	м	ж	Итого	м	ж	Итого	м	ж	Итого	
Острый назофарингит																			
Острый бронхит																			
Хронический бронхит																			
Пневмония																			
Прочие																			
<b>Всего</b>																			

### 2.3. Графические изображения

Графический метод является следующим шагом после табличного метода в понимании материала при статистическом анализе. Графики дают общее представление об основных тенденциях, взаимоотношениях между статистическими величинами, делают числа наглядными, доступными, понятными и интересными. При помощи графических изображений легче уяснить закономерности развития распределения, изменения явления. Поэтому очень важно правильно подобрать для характеристики изучаемого явления соответствующий график и правильно его построить.

Каждое графическое изображение так же, как и статистическая таблица, должно иметь название, в котором бы отражался круг изображаемых явлений, их географическая и временная характеристика. Название графиков дается под ним. Слово “диаграмма” в названии не пишется, пишется слово “Рис.” с проставлением соответствующего порядкового номера без знака “№”.

**Пример:** Рис. 1. Структура диспансерной группы больных по нозологическим формам в больнице № 1 г. Уфы за 2017 год. На стендах, предназначенных для выставки, графические изображения оформляются без масштабных шкал, названия их располагают над рисунком без слова “Рис”.

Шкала, на которую ориентируется размер диаграммы, должна начинаться с нуля (знак “0” можно не обозначать). При выборе верхнего показателя шкалы необходимо учитывать максимальную величину изображаемого показателя.

Разделительные черточки масштабной шкалы располагаются на строго определенных расстояниях друг от друга на осях абсцисс и ординат. На уровне этих черточек слева от оси ординат и вниз от оси абсцисс указываются масштабные знаки. Количество делений масштабной шкалы должно быть минимальным.

Геометрические знаки, краски, фигуры, штриховки должны быть пояснены. Условные обозначения располагаются ниже рисунка (иногда справа) в строчку. Размеры их должны быть небольшими.

*Различают следующие виды графических изображений:* диаграммы, картограммы и картодиаграммы.

**Диаграмма** – графическое изображение статистических данных при помощи линий или геометрических фигур, а также предметов, наглядно показывающее соотношение между сравниваемыми величинами.

В зависимости от формы и способов изображения **различают следующие виды диаграмм:**

1. **Линейные или координатные**, на которых линиями изображены динамические изменения величин изучаемого явления;

2. **Плоскостные**, дающие сравнительное изображение при помощи геометрических фигур, имеющих два измерения (прямоугольник, круг, квадрат и т.п.);

3. **Объемные** (пространственные, стереограммы), на которых данные изображены геометрическими телами, имеющими три измерения (шар, куб, призма и т.д.);

4. **Изобразительные** (фигурные), в которых для большей наглядности материал представлен не геометрическими фигурами, а фигурами-образами, в которых отображается содержательный смысл сравниваемых объектов.

**Линейная диаграмма** применяется для иллюстрации частоты явления, изменяющегося во времени. Линейная диаграмма употребляется для изображения динамики явления. На линейной диаграмме можно изображать абсолютные и средние величины, показатели интенсивности, соотношения и наглядности, иллюстрирует значение ряда величин, нанесенных в виде точек на систему координат и соединенных линиями (рис. 1).

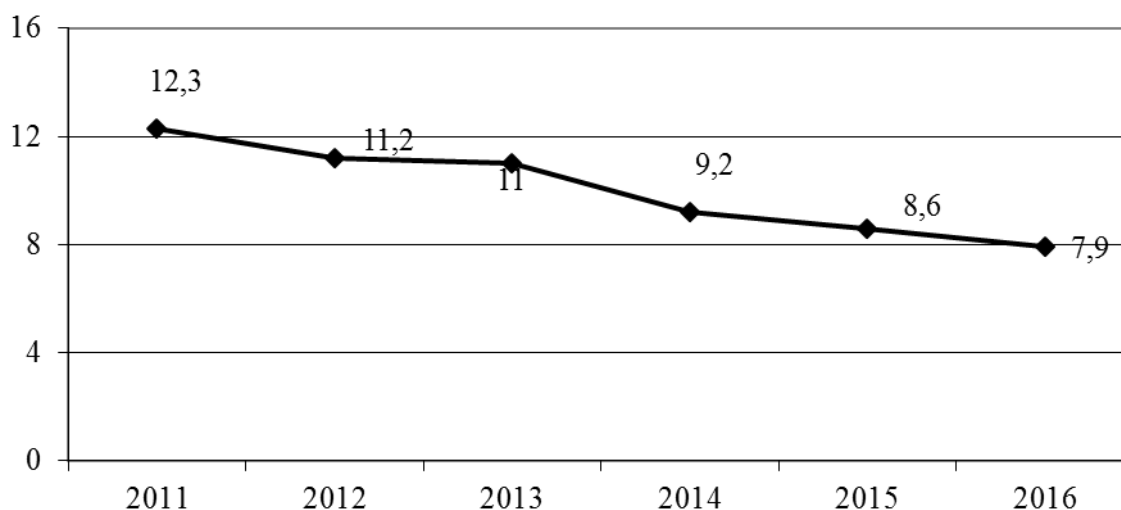


Рис. 1. Динамика младенческой смертности в г. Н за 2011-2016 гг. (%).

**Радиальная диаграмма** является частным видом линейной диаграммы, построенной на полярных координатах. Радиальная диаграмма используется при необходимости изобразить динамику явления за замкнутый цикл времени (сутки, неделя, месяц, год). При построении радиальной диаграммы в качестве оси абсцисс используется окружность, разделенная на одинаковое число частей, соответственно отрезкам времени

того или иного цикла. По технике построения радиальные диаграммы отличаются друг от друга в зависимости оттого, что взято в качестве пункта отсчета – центр круга (рис. 2) или окружность.

Осью ординат при этом служит радиус окружности или его продолжение. Обычно за радиус окружности принято брать среднюю величину явления анализируемого цикла времени. Количество радиусов соответствует интервалам времени изучаемого цикла.

**Плоскостные диаграммы** строятся по принципу сравнения площадей тех или иных геометрических фигур, соответствующих по своим размерам сравниваемым величинам признаков. Наибольшее применение для плоскостных диаграмм имеет прямоугольник и круг. Плоскостные диаграммы делятся на следующие ряды: столбиковые, пирамидальные, внутрисклонные, секторные.

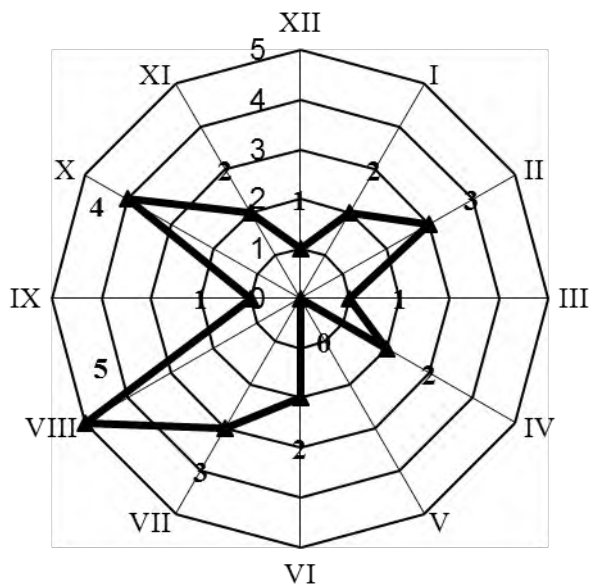


Рис. 2. Распределение числа умерших детей до 1 года по месяцам в г. Н. за 2016 г.

**Столбиковые диаграммы** представляют собой график, в котором различные величины представлены расположенными в высоту “столбиками” одинаковой ширины и разной высоты (рис. 3).

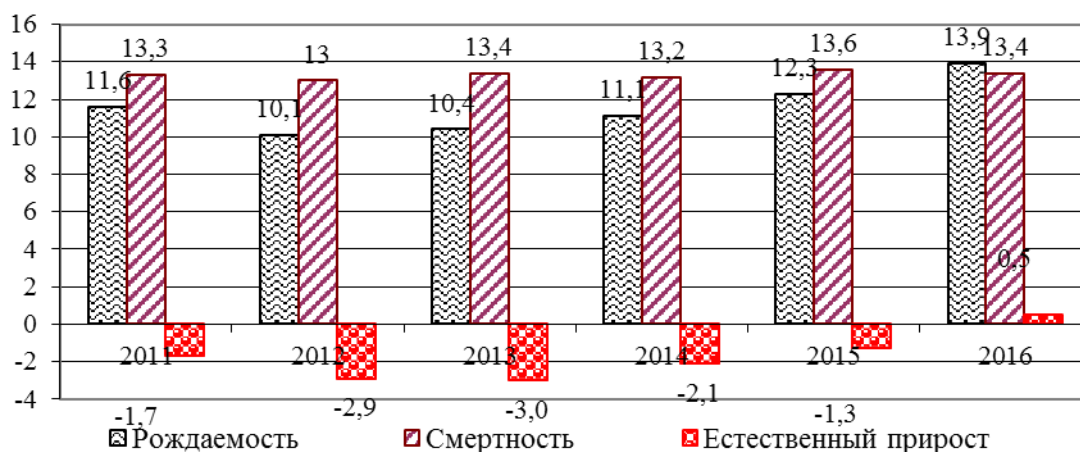


Рис.3. Динамика естественного движения населения в городе Н за 2011-2016 гг. (‰)

Столбиковые диаграммы применяются для изображения следующих статистических показателей: абсолютных и средних величин, показателей интенсивных наглядности и соотношения.

Построение столбиковой диаграммы требует одной масштабной шкалы, которая определяет высоту каждого столбика. Столбики должны быть даны одинаковом расстоянии друг от друга.

**Ленточная диаграмма** отличается от столбиковой тем, что прямоугольники, изображающие числа, располагаются не по вертикали, а по горизонтали. На ленточных диаграммах изображаются те же виды величин, что и на столбиковых.

**Внутристолбиковая диаграмма** предназначена для изображения структуры явления, т.е. экстенсивных показателей. Берется прямоугольник произвольной ширины и высоты, вся площадь которого принимается равной 100 % (рис. 4).

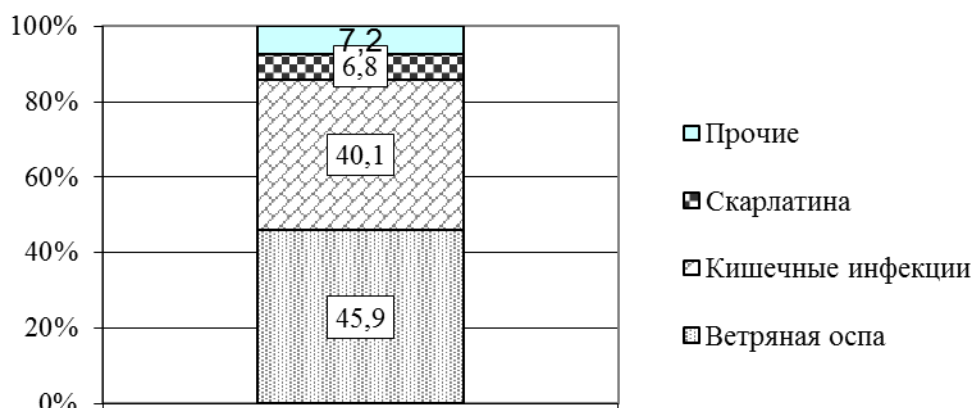


Рис. 4. Структура инфекционных заболеваний в районе К. по нозологическим формам за 2016 г.

**Секторная диаграмма** является наиболее распространенным графическим изображением распределения явления, для характеристики удельных весов отдельных частей целого, для выявления структуры сдвигов, т.е. для изображения экстенсивных показателей. Секторная диаграмма строится следующим образом. Сначала вычеркивается основной круг, площадь которого принимается равной ста процентам. Внутри его вычеркивается маленький круг, в которой вписывается итоговая цифра – 100%. Далее проценты структуры переводятся в градусы из расчета, 1 % равен 3,6°. Затем устанавливается исходная точка, лежащая на “12 часах” по окружности. Начиная от этой точки по часовой стрелке откладываются с помощью транспортира отрезки в градусах, соответствующие элементам структуры. Полученные точки соединяются прямыми линиями с центром круга, и образовавшиеся сектора будут соответствовать элементам структуры. Начинать откладывать от большего элемента к меньшим. Группа “прочие” откладывается в последнюю очередь, независимо от ее величины (рис. 5).

При сравнительном изображении явлений в динамике во втором изображении, которое располагается рядом справа, надо придерживаться порядка следования показателей первого изображения. Условные обозначения элементов структуры даются под рисунком или справа от него. Секторные диаграммы сохраняют свою выразительность при делении совокупностей на небольшое число частей – не более 4-5.

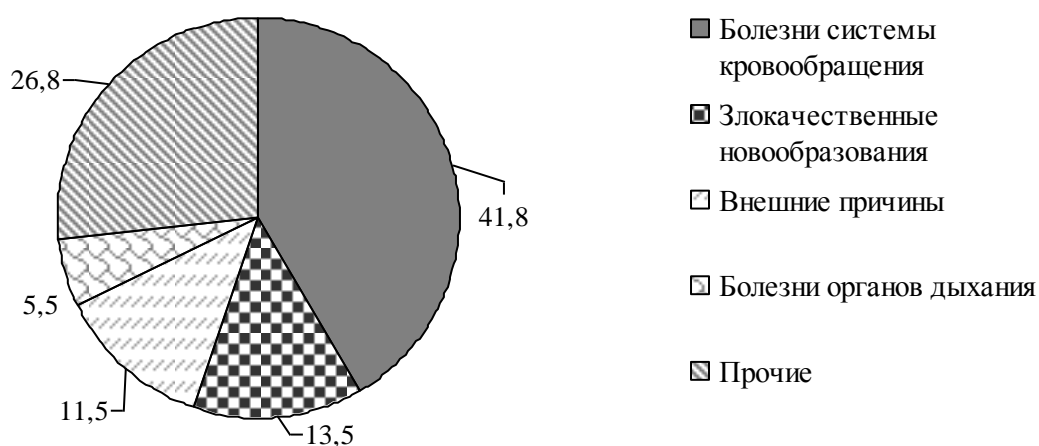


Рис. 5. Структура смертности населения города Н за 2016 г., %.

**Картограммой** называется особая геометрическая карта, на которой при помощи некоторых графических символов (штриховкой, цветом, точками и т.п.) представлены статистические данные в пространственной группировке (рис. 6).

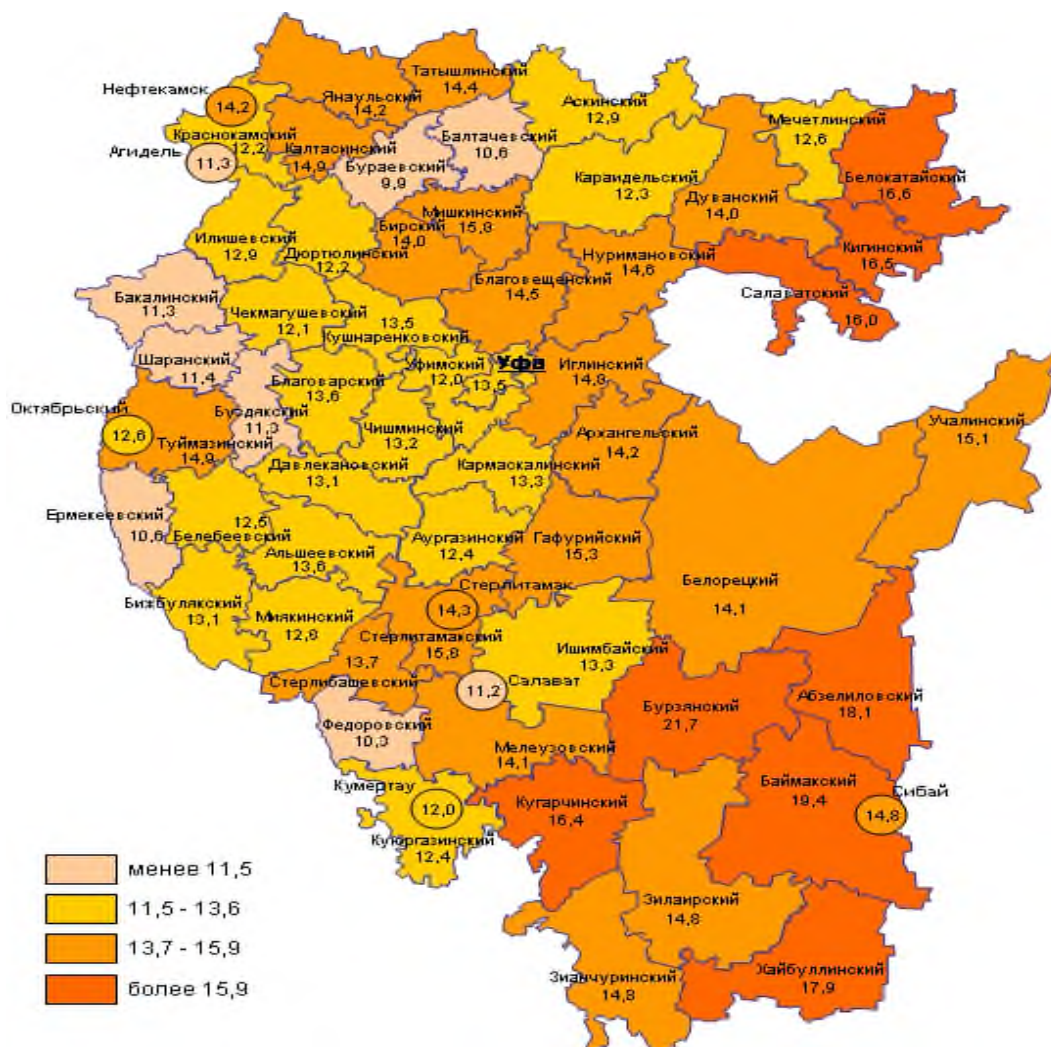


Рис.6. Общие коэффициенты рождаемости по муниципальным образованиям Республики Башкортостан в 2009 году (число родившихся на 1000 населения)

**Картодиаграмма** представляет собой сочетание схематической географической карты с диаграммой. Это сочетание заключается в том, что фигуры, используемые в диаграмме, разбросаны по карте в соответствии с той территорией, которые они представляют. Диаграммные фигуры картограммы статистически характеризуют данную территорию по одному или нескольким признакам.



### **Вопросы для самоподготовки:**

1. Отличие цели от задач исследования.
2. Основные элементы первого этапа.
3. Отличие плана от программы исследования.
4. Виды наблюдения по времени, по охвату, по способу сбора материала.
5. Методы отбора выборочной совокупности из генеральной.
6. Методы расчета необходимого объема наблюдения.
7. Основные элементы третьего этапа исследования.
8. Содержание статистического анализа (□□ этап)
9. Виды статистических таблиц.
10. Каковы правила оформления статистических таблиц?
11. Какова методика построения линейной диаграммы?
12. Какова методика построения секторной диаграммы?
13. Какова методика построения радиальной диаграммы?
14. Какова методика построения внутрестолбиковой диаграммы?

## **3. ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ**

В результате группировки и табличной сводки материалов наблюдения исследователь получает абсолютные величины. В ряде случаев этих абсолютных величин достаточно для характеристики размеров изучаемых явлений и процессов. Так, например, абсолютные величины применяются при учете ряда редких инфекционных заболеваний (малярия, дифтерия, трахома, СПИД и другие). Большое практическое значение для правильного планирования медицинской помощи населению имеют также абсолютные величины численности населения и его отдельных возрастных групп; численность медицинского персонала и лечебно-профилактических учреждений; количество больничных коек и т.д.

Однако при рассмотрении абсолютных величин чаще всего можно сделать только некоторые предварительные выводы, и для дальнейшего анализа возникает необходимость в преобразовании этих величин в производные величины: относительные и средние. Необходимость перевода

абсолютных величин в относительные можно пояснить простым примером.

В районе "А" выявлено в 1996 году 22 случая острых желудочно-кишечных заболеваний, а в районе "Б" за этот период обнаружено 24 подобных случаев заболеваний. Можно ли сказать, что в районе "Б" выше уровень заболеваемости этими болезнями? Конечно же, нет. Для того чтобы ответить на этот вопрос, надо знать численность населения в данных районах. Допустим в районе "А" проживает 20 тысяч, а в районе "Б" – 30 тысяч человек.

Относя число случаев желудочно-кишечных заболеваний в каждом районе к числу их жителей, получаем, в расчете на 10 000 человек, следующие величины:

$$\text{Район "А"} \quad \frac{22 \cdot 10000}{20000} = 11 \text{‰}$$

$$\text{Район "Б"} \quad \frac{24 \cdot 1000}{30000} = 8 \text{‰}$$

Вот теперь мы можем сделать заключение, что заболеваемость острыми желудочно-кишечными заболеваниями выше в районе "А".

Относительные величины (статистические коэффициенты), получаемые из соотношения двух сравниваемых чисел, для удобства сопоставления обычно умножаются на какое-либо круглое число (100, 1000, 10000, 100000 и т.д.), которое называется базой или основанием. В результате полученные коэффициенты приобретают форму "процентов" (%), "промилле" (‰), "продецимилле" (‰‰), "просантимилле" (‰‰‰) и т.д. Чем реже встречается изучаемое явление, тем больше числовое основание следует избрать с тем, чтобы не было коэффициентов меньше единицы, которыми неудобно пользоваться.

По своему содержанию статистические коэффициенты, чаще всего применяемые в медицинской статистике, разделяются на три вида: 1) коэффициенты экстенсивности (показатели распределения, структуры, доли, удельного веса); 2) коэффициенты интенсивности (показатели частоты, распространенности); 3) коэффициенты (показатели) соотношения; 4) коэффициенты наглядности.

**Экстенсивные коэффициенты** характеризуют распределение явления или среды на его составные части, его внутреннюю структуру или отношение частей к целому (удельный вес).

При вычислении экстенсивных коэффициентов мы имеем дело только с одной статистической совокупностью и ее составом. Экстенсивные коэффициенты обычно выражаются в процентах.

$$\frac{\text{Часть явления}}{\text{Явление в целом}} \cdot 100 \%$$

Например, в 1998 году число умерших от болезней системы кровообращения составило 715 случаев, в том числе 452 случая от ИБС. Если принять все случаи смерти за 100, а случаи смерти от ИБС за X (икс), то доля ИБС среди всех умерших составит:

$$\frac{452}{715} \cdot 100\% = 63,2\%$$

В качестве примеров экстенсивных коэффициентов, применяемых в медицине и здравоохранении, можно назвать структуру заболеваемости населения; распределение госпитализированных больных по отдельным нозологическим формам; лейкоцитарную формулу и т.д.

Необходимо помнить, что экстенсивными показателями следует пользоваться для характеристики состава совокупности (явление, среда) в данном месте в данное время. Для динамических сравнений эти показатели непригодны. Сравнение удельных весов позволяет судить лишь об их порядковом номере в структуре (заболеваемости, смертности и т.д.), но не дает возможности говорить о частоте, распространенности данного явления. Для этой цели всегда необходимо знать численность среды, в которой происходит явление, и вычислить интенсивные коэффициенты.

**Интенсивные коэффициенты** характеризуют частоту (интенсивность, уровень, распространенность) явления в среде, в которой оно происходит и с которой непосредственно органически связано, за определенный промежуток времени, чаще всего за год.

При вычислении интенсивных коэффициентов необходимо знание двух статистических совокупностей, одна из которых представляет среду, а вторая – явление. Среда продуцирует это явление.

В демографической и санитарной медицинской статистике в качестве среды часто рассматривается население и при расчете к нему относят

то или иное явление, например, число заболеваний за год, число рождений за год, число смертей за год и т.д. При вычислении показателя младенческой смертности средой является количество новорожденных и к нему относят число умерших детей в возрасте до 1 года.

Если вычисляется коэффициент летальности при туберкулезе, то средой будут все больные туберкулезом, а явлением – умершие от туберкулеза. Коэффициенты интенсивности рассчитываются на основании 100, 1000, 10000, 100000 и т.д. в зависимости от распространенности явления. Однако в практике здравоохранения существуют общепринятые положения. Так, летальность, частота осложнений выражаются в %, рождаемость, смертность, младенческая смертность всегда выражаются в промилле (‰), заболеваемость в ‰‰.

Техника вычисления интенсивных коэффициентов выглядит следующим образом:

$$\frac{\text{явление}}{\text{среда}} \times 1000$$

Например, в городе "Н" в 1996 г. жителей в возрасте 70-79 лет было 9845 человек; из этого числа в течение года умерло 784 человека. Для вычисления коэффициента смертности лиц в возрасте 70-79 лет необходимо составить и решить следующую пропорцию:

$$\begin{array}{l} 9845 - 784 \\ 1000 - x \\ x = \frac{784 \cdot 1000}{9845} = 79,6\% \circ \end{array}$$

**Коэффициент соотношения** (обеспеченности) характеризует численное соотношение двух, не связанных между собой совокупностей, сопоставляемых только логически, по их содержанию. К ним относятся такие показатели, как число врачей или число больничных коек на 1000 населения; количество различных лабораторных исследований или число переливаний крови на 100 больных и т.д.

По методике вычисления коэффициенты соотношения сходны с интенсивными коэффициентами, хотя различны с ними по существу.

**Коэффициент наглядности** характеризует динамику изучаемого явления и означает, на сколько процентов или во сколько раз увеличилось или уменьшилось явление.

Например, показатель рождаемости в 1987 г. составил 19,0 ‰, а в 1997г.- 10,7 ‰. Для вычисления коэффициента наглядности обычно базовый показатель берут за 100 %.

$$19,0 ‰ - 100 \%$$

$$10,7 ‰ - x$$

$$x = 56,3 \%$$

$$100 \% - 56,3 \% = 43,7 \%$$

Вывод: показатель рождаемости снизился на 43,7 %.

#### **Вопросы для самоподготовки**

1. Когда и где используются относительные величины?
2. Виды относительных величин?
3. Методика расчета интенсивных показателей.
4. Методика расчета экстенсивных показателей.
5. Методика расчета показателей соотношения.
6. Методика расчета показателей наглядности.
7. В каких единицах измеряется каждый из перечисленных показателей?

## **4. СТАНДАРТИЗОВАННЫЕ КОЭФФИЦИЕНТЫ. МЕТОД СТАНДАРТИЗАЦИИ**

Общие интенсивные коэффициенты (рождаемости, смертности, младенческой смертности, заболеваемости и т.д.) правильно отражают частоту явлений при их сопоставлении лишь в том случае, если состав сравниваемых совокупностей однороден. Если же они имеют неоднородный возрастно-половой или профессиональный состав, различие по тяжести болезни, по нозологическим формам или по другим признакам, то ориентируясь на общие показатели, сравнивая их, можно сделать неправильный вывод о тенденциях изучаемых явлений и истинных причинах разницы общих показателей сравниваемых совокупностей.

Например, больничная летальность на терапевтическом отделении №1 в отчетном году составила 3 %, а на терапевтическом отделении №2 в том же году – 6 %. Если оценивать деятельность этих отделений по общим показателям, то можно сделать вывод о неблагоприятности на 2 тера-

певтическом отделении. А если предположить, что состав лечившихся на этих отделениях различается по нозологическим формам или по тяжести заболеваний госпитализированных, то наиболее правильным способом анализа является сопоставление специальных коэффициентов, рассчитанных отдельно для каждой группы больных с одинаковыми нозологическими формами или тяжестью заболеваний, так называемых "повозрастных коэффициентов".

Зачастую, в сравниваемых совокупностях наблюдаются противоречивые данные. Кроме того, даже при наличии одинаковой тенденции во всех сравниваемых группах не всегда удобно пользоваться набором показателей, а предпочтительнее получить единую суммарную оценку. Во всех подобных случаях прибегают к методу стандартизации, то есть к устранению (элиминации) влияния состава (структуры) совокупностей на общий, итоговый показатель.

Следовательно, метод стандартизации применяется тогда, когда имеющиеся различия в составе сравниваемых совокупностей могут повлиять на размеры общих коэффициентов.

Для того чтобы устранить влияние неоднородности состава сравниваемых совокупностей на величину получаемых коэффициентов, их приводят к единому стандарту, то есть условно допускается, что состав сравниваемых совокупностей одинаков. В качестве стандарта можно принять состав какой-либо близкой по существу третьей совокупности, средний состав двух сравниваемых групп или, проще всего, состав одной из сравниваемых групп.

Стандартизованные коэффициенты показывают, каковы были бы общие интенсивные показатели (рождаемости, заболеваемости, смертности, летальности и т.д.), если бы на их величину не оказывала влияние неоднородность в составах сравниваемых групп. Стандартизованные коэффициенты являются условными величинами и применяются исключительно для анализа в целях сравнения.

**Существуют три метода стандартизации: прямой, косвенный и обратный (Керриджа).**

Рассмотрим применение этих трех методов стандартизации на примерах, взятых из статистики злокачественных новообразований. Как из-

вестно, с возрастом значительно повышаются коэффициенты смертности от злокачественных новообразований. Отсюда следует, что если в каком-либо городе будет относительно высока доля людей пожилых возрастов, а в другом – преобладать население среднего возраста, то даже при полном равенстве санитарных условий жизни и медицинской помощи в обоих сравниваемых городах неизбежно общий коэффициент смертности населения от злокачественных новообразований в первом городе будет выше, чем тот же коэффициент во втором городе.

Для того, что нивелировать влияние возраста на общий показатель смертности населения от злокачественных новообразований, необходимо применить стандартизацию. Только после этого можно будет сравнивать полученные коэффициенты, и сделать обоснованный вывод о большем или меньшем уровне смертности от злокачественных новообразований в целом в сравниваемых городах.

#### **4.1. Прямой метод стандартизации.**

В нашем примере его можно применять в том случае, когда известен возрастной состав населения и есть информация для расчета повозрастных коэффициентов смертности населения от злокачественных новообразований (числа умерших от злокачественных новообразований в каждой возрастной группе).

**Методика вычисления стандартизованных коэффициентов прямым методом:**

Метод складывается из четырех последовательных этапов (табл. 2а приложения).

*Первый этап.* Вычисление "повозрастных" коэффициентов смертности от злокачественных новообразований (отдельно для каждой возрастной группы).

*Второй этап.* Выбор стандарта осуществляется произвольно. В нашем примере за стандарт взят возрастной состав населения в городе "А".

*Третий этап.* Расчет "ожидаемых" числе. Мы определяем, сколько бы человек умерло от злокачественных новообразований в каждой возрастной группе населения города "Б" при имеющихся повозрастных пока-

зателях смертности от злокачественных новообразований в этом городе, но при возрастном составе города "А" (стандарт).

Например, в возрастной группе "до 30 лет":

6 – 100000

$$x_1 - 350000 \quad x_1 = \frac{6 \cdot 350000}{100000} = 21.0,$$

или в возрастной группе "40-49 лет":

140 – 100000

$$x_3 - 95000 \quad x_3 = \frac{140 \cdot 95000}{100000} = 133.0 \quad \text{и т.д.}$$

*Четвертый этап.* Расчет стандартизованных коэффициентов. Сумму "ожидаемых" чисел (1069.0) мы предполагаем получить из общей численности населения города "А" (700000). А сколько же умерших от злокачественных новообразований приходится на 100000 населения?

1069 – 700000

$$x_{ст} - 100000 \quad x_{ст} = \frac{1069 \cdot 100000}{700000} = 152.7\% \text{ ооо.}$$

Из наших результатов можно сделать следующий вывод: если бы возрастной состав населения "Б" был бы такой же, как в городе "А" (стандарт), то смертность населения от злокачественных новообразований в городе "Б" была бы существенно выше (152.7 ‰ против 120.2 ‰).

## 4.2. Косвенный метод стандартизации

Метод применяется, если специальные коэффициенты в сравниваемых группах известны, но мало достоверны. Это наблюдается, например, когда числа заболевших очень малы и, следовательно, вычисляемые коэффициенты будут существенно меняться в зависимости от прибавления одного или нескольких случаев заболеваний.

**Методика вычисления стандартизованных коэффициентов косвенным методом:**

Метод можно разбить на три этапа (табл. 2б приложения).

*Первый этап.* Состоит в выборе стандарта. Так как нам обычно неизвестны специальные коэффициенты сравниваемых групп (коллективов), то за стандарт берутся специальные коэффициенты какого-то хорошо изученного коллектива. В рассматриваемом примере таковыми могут



служить повозрастные показатели смертности от злокачественных новообразований в городе "С".

*Второй этап* включает вычисление "ожидаемых" чисел умерших от злокачественных новообразований. Допуская, что повозрастные коэффициенты смертности в обоих сравниваемых городах равны стандартным, определяем сколько бы умерло людей от злокачественных новообразований в каждой возрастной группе.

*На третьем этапе* вычисляются стандартизированные коэффициенты смертности населения от злокачественных новообразований. Для этого действительное число умерших относят к суммарному "ожидаемому" числу, и результат умножают на общий коэффициент смертности стандарта:

$$\frac{\text{Действительное число умерших}}{\text{Ожидаемое число умерших}} \times \text{Общий коэфф. смертности стандарта}$$

Вычисление стандартизированных коэффициентов смертности от злокачественных новообразований проводится следующим образом:

$$\text{Для города N: } \frac{754 \cdot 125}{866,8} = 108,7\% \text{ } \circ \circ \circ$$

$$\text{Для города M: } \frac{590 \cdot 125}{587,1} = 125,6\% \text{ } \circ \circ \circ$$

Следовательно, более низкий общий коэффициент смертности населения в городе М (118.0 ‰ против 130.0 ‰ в городе N) объясняется более благоприятной возрастной структурой населения в этом городе.

### **4.3. Обратный метод стандартизации (Керридж, 1958 г.)**

Обратный метод применяется при отсутствии данных о возрастном составе населения, когда имеются лишь сведения о возрастном составе больных или умерших, то есть данные обратные тем, что использовались при косвенном методе. Метод дает менее точные результаты. Они тем точнее, чем более дробные возрастные интервалы применяются при стандартизации. Важно также выбрать подходящий, близкий к сравниваемым контингентам, стандарт. Стандартом в этом случае служат возрастные коэффициенты смертности или заболеваемости.

Например, в городе N за последние 10 лет несколько увеличились коэффициенты смертности населения от злокачественных новообразований с 115,5 ‰ в 2006 г. до 119,0 ‰ в 2016 г. (табл. 2в приложения).

За это время численность населения возросла с 800000 до 900000 человек и, по-видимому, возрастной состав в сравниваемые годы был различен.

*Первый этап* состоит из выбора стандарта. Примем за стандарт по-возрастные коэффициенты смертности от злокачественных новообразований на 100000 населения в 2010 г., в год переписи, когда эти коэффициенты были определены с достаточной точностью.

*Второй этап* включает в себя вычисление "ожидаемой" численности населения города, при этом допускается, что по-возрастные коэффициенты смертности от злокачественных новообразований в 2006 и 2016 гг. были такими же, как и в 2010 г.

В графах 3 и 5 таблицы – "ожидаемая" численность населения по возрастным группам и суммарная в 2006 и 2016 гг. Для вычисления "ожидаемой" численности населения делим число умерших в каждой возрастной группе на соответствующие по-возрастные коэффициенты смертности от злокачественных новообразований принятого за стандарт населения, и результат умножаем на 100000.

Например, для того, чтобы в возрасте до 30 лет коэффициент смертности от злокачественных новообразований составлял 4.0 на 100000 при наличии 21 умершего в этом возрасте в 2006 г., численность населения данного возраста в этом году должна составлять:

$$\text{в 2006 г. } \frac{21 \times 100000}{4.0} = 525000 \text{ человек,}$$

$$\text{в 2016 г. } \frac{18 \cdot 100000}{4.0} = 450000 \text{ человек.}$$

Таким же образом определяем "ожидаемую" численность населения для всех остальных возрастных групп населения. В результате подсчета оказалось, что "ожидаемая" численность населения в 2006 году составляла 890548 человек, а в 2016 году – 840024 человек.

Расхождение "ожидаемых" и фактических чисел населения вызвано различием действительных и принятых за стандарт по-возрастных коэффициентов смертности населения от злокачественных новообразований.

На третьем этапе стандартизации для устранения указанного различия делим "ожидаемые" числа населения на фактические и умножаем на принятый за стандарт коэффициент смертности.

Для 2006 г. это составляет  $\frac{890548}{800000} \cdot 121.0 = 134.7 \text{ ‰}$ ,

а для 2016 г.  $\frac{840024}{900000} \cdot 121.0 = 112.9 \text{ ‰}$ .

Отсюда можно сделать вывод, что некоторый рост общих коэффициентов смертности населения города N от злокачественных новообразований был вызван только изменением возрастного состава населения. После применения стандартизации и элиминирования влияния изменений возрастного состава оказалось, что за истекшие 10 лет население города стало реже умирать от злокачественных новообразований.

Необходимо еще раз подчеркнуть, что выбор конкретного метода стандартизации зависит от того, насколько полный статистический материал имеется в наличии. Прямой метод дает более надежные результаты, но в случае невозможности его применения следует использовать косвенный или обратный метод стандартизации: они достаточно точны для практического применения. Стандартизация позволяет нам сделать правильный вывод о том, имеется ли действительно разница общих интенсивных коэффициентов в сравниваемых коллективах или эти различия зависят только от неодинаковой структуры сравниваемых совокупностей.

#### **Вопросы для самоподготовки:**

1. В каких случаях применяется стандартизация показателей?
2. В чем сущность прямого метода стандартизации?
3. Для чего применяются стандартизованные показатели?
4. Какие существуют варианты выбора стандарта?
5. Этапы расчета стандартизованных показателей.
6. Расчет стандартизованных показателей.
7. Стандартизованные показатели истинные или условные (абстрактные)?
8. В чем сущность обратного метода стандартизации?
9. В чем сущность косвенного метода стандартизации?

## 5. ДИНАМИЧЕСКИЕ РЯДЫ

При изучении изменений какого-либо явления во времени составляется динамический ряд.

*Динамическим рядом* называется совокупность однородных статистических величин, показывающих изменение какого-либо явления на протяжении определенного промежутка времени.

Величины, составляющие динамический ряд, называются *уровнями ряда*.

Уровни динамического ряда могут быть представлены:

- абсолютными величинами;
- относительными величинами (в том числе показателями интенсивными, экстенсивными, соотношения);
- средними величинами.

Динамические ряды бывают **двух видов**:

\* *Моментный* динамический ряд состоит из величин, характеризующих явление на какой-то определенный момент (дату). Например, каждый уровень может характеризовать численность населения, численность врачей и т.д. на конец какого-то года.

\* *Интервальный* динамический ряд состоит из величин, характеризующих явление за определенный промежуток времени (интервал). Например, каждый уровень такого ряда может характеризовать смертность, рождаемость, заболеваемость, среднегодовую занятость койки за какой-то год.

Примеры:

*Интервальный динамический ряд, состоящий из величин; полученных за интервальный период времени.*

Динамика рождаемости в РФ (на 1000 жителей):

2010 – 13,4	2013 – 9,4	2016 – 8,9
2011 – 12,1	2014 – 9,6	
2012 – 10,7	2015 – 9,8	

**Моментный динамический ряд, состоящий из абсолютных величин.**

Динамика среднегодовой численности коек в РБ:

2010 – 52870	2013 – 51386	2016 – 51162
2011 – 52375	2014 – 51594	2017 – 51202
2012 – 52105	2015 – 51360	

Динамический ряд можно подвергнуть преобразованиям, целью которых является выявление особенностей изучаемого процесса, а также достижение наглядности в характеристике того или иного явления.

Для определения тенденции изучаемого явления рассчитывают показатели динамического ряда.

**Показатели динамического ряда:**

- абсолютный прирост;
- показатель наглядности;
- показатель роста (снижения);
- темп прироста (снижения).

**Абсолютный прирост** представляет собой разность между последующим и предыдущим уровнем. Измеряется в тех же единицах, в которых представлены уровни ряда.

**Показатель наглядности** показывает отношение каждого уровня ряда к одному из них (чаще начальному) принятому за 100 %.

**Темп роста (снижения)** показывает отношение каждого последующего уровня к предыдущему, принятому за 100 %.

**Темп прироста (убыли)** показывает отношение абсолютного прироста (снижения) каждого последующего уровня к предыдущему уровню, принятому за 100 %.

Если показатель роста (убыли) показывает, сколько процентов от предыдущего уровня составляет последующий уровень, то темп прироста показывает, на сколько процентов увеличился (снизился) последующий уровень по сравнению с предыдущим. Поэтому, темп прироста можно рассчитать и по следующей формуле:

$$\text{темп прироста} = \text{показатель роста} - 100 \%$$

Динамический ряд и его показатели могут быть представлены в виде таблицы (табл. 9 приложения).

Выделяют *два метода расчет динамического ряда*:

- *Цепной* метод используется, когда сравнение проводится с предыдущим периодом.

- *Базисный* – когда сравнение проводится с базисным (первым в динамическом ряду) периодом.

### ***Методы выравнивания динамического ряда***

Иногда динамика изученного явления представлена не в виде непрерывно меняющегося в одном направлении явления, а скачкообразными изменениями.

В таких случаях используют различные методы выравнивания динамического ряда:

- укрупнение интервалов;
- расчет скользящей средней;
- метод наименьших квадратов.

***Укрупнение*** интервала можно производить за определенные промежутки времени (за квартал, за один, два, три года и т.д.).

***Пример выравнивания динамического ряда с помощью укрупнения интервалов*** (табл. 4)

Произведено *укрупнение интервала за два года* и рассчитана средняя длительность пребывания больного на койке для каждого интервала.

2010 – 2011  $(16,1 + 14,7) / 2 = 15,4$

2012 – 2013  $(13,4 + 11,6) / 2 = 12,5$

Показатели преобразованного динамического ряда рассчитываются по общепринятой методике.

Таблица 4

### **Средняя длительность пребывания больного на терапевтической койке в больнице Н. за 2010-2017 гг.**

Годы	Средняя длительность пребывания больного на терапевтической койке (в днях)	Укрупненный интервал (годы)	Средняя длительность пребывания больного на терапевтической койке (в днях)
2010	16,1		
		2010-2011	15,4
2011	14,7		
2012	13,4		
		2012-2013	12,5
2013	11,6		

Продолжение табл. 4

2014	11,7		
		2014-2015	11,4
2015	11,2		
2016	11,1		
		2016-2017	10,9
2017	10,7		

Влияние случайных колебаний на уровни динамического ряда можно устранить и с помощью скользящей средней. При ее расчете лучше использовать интервалы, включающие три хронологических периода.

**Пример выравнивания динамического ряда методом скользящей средней** (табл.5).

Таблица 5

**Динамика показателей рождаемости в г. Н за 2010-2016 гг.**

Годы	Показатели рождаемости	Скользящая средняя
2010	13,4	-
2011	12,1	12,0
2012	10,7	10,7
2013	9,4	9,9
2014	9,6	9,6
2015	9,8	9,4
2016	8,9	-

Для выравнивания динамического ряда произведено *вычисление скользящей средней с использованием интервала в три года.*

$$2011 \text{ г.} \quad (13,4 + 12,1 + 10,7) / 3 = 12,0$$

$$2012 \text{ г.} \quad (12,1 + 10,7 + 9,4) / 3 = 10,7$$

$$2013 \text{ г.} \quad (10,7 + 9,4 + 9,6) / 3 = 9,9$$

$$2014 \text{ г.} \quad (9,4 + 9,6 + 9,8) / 3 = 9,6$$

$$2015 \text{ г.} \quad (9,6 + 9,8 + 8,9) / 3 = 9,4$$

Однако этот метод исключает из анализа средние величины первого и последнего уровня.

**Метод наименьших квадратов** позволяет наиболее точно выравнивать тенденции изучаемого явления. Он позволяет рассчитать точки прохождения такой прямой линии, от которой имеющаяся эмпирическая находится на расстоянии наименьших квадратов от других возможных линий.

Динамический ряд в случае применения данного метода должен иметь менее 5 хронологических дат, количество их должно быть нечетным, а интервалы между ними – одинаковыми.

*Пример выравнивания динамического ряда методом наименьших квадратов* приведен в таблице 6.

Таблица 6

**Динамика младенческой смертности в г. Н  
(на 1000 родившихся живыми) за 2012-2016 гг.**

Хронологические даты, годы	Младенческая смертность, у	Порядковый номер хронологической даты от центральной, х	$x*y$	$x^2$	Выравненные уровни младенческой смертности
2012	18.0	-2	-36.0	4	19.0
2013	19.9	-1	-19.9	1	18.7
2014	18.6	0	0	0	18.4
2015	18.1	1	18.1	1	18.1
2016	17.4	2	34.8	4	17.8
	$\Sigma_y = 92.0$		$\Sigma_{xy} = -3$	$\Sigma x^2 = 10$	

Даты искомой прямой линии округляются по следующей формуле:

$$y_1 = \alpha_0 + \alpha_1 \times x, \text{ где}$$

$\alpha_0$  – это хронологическая средняя (значение центральной хронологической даты), которая вычисляется по формуле:

$$\alpha_0 = \frac{\Sigma_y}{S}, \text{ где}$$

S – сумма хронологических дат (периодов);

$\Sigma_y$  – сумма всех значений изучаемого явления.

$$\alpha_0 = \frac{92,0}{5} = 18,4$$

$\alpha_1$  – это коэффициент поправки искомого расстояния, который определяется по формуле

$$\alpha_1 = \frac{\Sigma_{xy}}{\Sigma_{x^2}}$$

x – порядковый номер (расстояние) хронологических дат от центральной, принятой за 0.



Сумма произведений  $x \times y$  определяется с учетом алгебраических знаков.

$$\alpha_1 = \frac{-3}{10} = -0.3$$

Зная величины  $\alpha_0$  и  $\alpha_1$  подставляем их в уравнение:

$y_1 = \alpha_0 + \alpha_1 \times x$  и, придавая последовательные значения чисел ряда  $x$ , получим выравненный динамический ряд младенческой смертности.

2012		$y_1 = 18,4 + (-0,3) \times (-2) = 19,0$
2013		$y_2 = 18,4 + (-0,3) \times (-1) = 18,7$
2014		$y_3 = 18,4$
2015		$y_4 = 18,4 + (-0,3) \times (1) = 18,1$
2016		$y_5 = 18,4 + (-0,3) \times (2) = 17,8$

### Вопросы для самоподготовки:

1. Что такое динамический ряд?
2. Какие виды динамических рядов вы знаете?
3. Каковы показатели динамического ряда?
4. Каковы методы расчета показателей динамического ряда?
5. В каких случаях проводится выравнивание рядов динамики?
6. Какие вы знаете методы выравнивания динамического ряда?
7. Как графически можно представить динамический ряд?

## 6. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

**Корреляция** – понятие, означающее взаимосвязь между признаками. Согласно диалектико-материалистическому учению о природе и обществе, все явления и процессы, совершающиеся в них, взаимосвязаны и взаимообусловлены. Задача каждой науки – вскрыть и изучить наиболее существенные связи между явлениями и процессами.

В социальной гигиене и организации здравоохранения, в различных разделах медицины и биологии часто приходится проводить статистический анализ связей всевозможных признаков в совокупности. Необходимо уметь изучать особенности этих связей, определять их размеры и характер, а также оценивать их достоверность.

Различают две формы проявления количественных связей между явлениями или процессами: функциональную и корреляционную.

Под **функциональной** понимают такую связь, при которой любому значению одного из признаков соответствует строго определенное значение другого (радиуса круга соответствует определенная площадь круга, скорость свободно падающего тела определяется величиной ускорения силы тяжести и времени падения). Функциональная связь характерна для физико-химических процессов.

В социально-гигиенических исследованиях, а также в клинической медицине и биологии зависимости между явлениями носят характер – характер **корреляционной** связи. При корреляционной связи значению каждой средней величины одного признака соответствует несколько значений другого взаимосвязанного с ним признака. Всем известно, что рост и масса тела человека связаны между собой. У группы лиц с одинаковым ростом наблюдаются различные колебания массы тела. Однако эти колебания массы тела варьируют в определенных размерах – вокруг своей средней величины.

Между уровнем температуры тела человека и числом сердечных сокращений существует также зависимость. При одинаковой температуре тела у различных людей наблюдаются индивидуальные колебания частоты сердечных сокращений, варьирующие вокруг своей средней.

Важно отметить, что корреляционная связь проявляется лишь в массе наблюдений, т.е. в совокупности. Используя методы корреляции, важно помнить о возможности измерять связь между различными признаками только лишь в *качественно однородной совокупности*. Нельзя, например, сопоставлять рост и массу тела людей, состоящих из лиц разного пола и возраста.

Корреляция может быть представлена в виде: Таблицы, графика или коэффициента корреляции.

Таблицы и графики дают лишь представление о наличии и направлении связи. Так, между температурой воздуха и числом случаев бронхита существует корреляционная связь. При этом с повышением температуры воздуха число бронхитов уменьшается. Об этой закономерности можно судить по данным, представленным в таблице и на графике. Однако измерить и оценить статистическую достоверность этой связи можно

лишь при помощи специального коэффициента корреляции ( $r_{xy}$ ) и его средней ошибки ( $m_r$ ).

**Коэффициент корреляции** ( $r_{xy}$ ) одним числом измеряет силу связи между изучаемыми явлениями и дает представление о ее направлении.

**По направлению связь может быть прямой и обратной:**

- при **прямой** связи с увеличением значений одного признака возрастает среднее значение другого признака. Например: с повышением температуры тела увеличивается частота пульса у большинства инфекционных больных; с увеличением роста ребенка увеличивается масса его тела. Коэффициент корреляции, характеризующий прямую связь, обозначается знаком плюс (+).

- при **обратной** связи: с увеличением одного признака убывает среднее значение другого признака. Например, чем ниже температура воздуха в осенний период, тем выше заболеваемость детей острым бронхитом, обозначается знаком минус (-). (табл. 7). По силе связи коэффициенты корреляции колеблются от единицы (полная связь) до нуля (отсутствующие связи).

Таблица 7

**Схема оценки силы и направления корреляционной связи по коэффициенту корреляции**

Сила связи	Прямая (+)	Обратная (-)
Полная	+ 1,0	- 1,0
Сильная	От + 0,70 до + 0,99	От - 0,70 до - 0,99
Средняя	От + 0,30 до + 0,69	От - 0,30 до - 0,69
Слабая	От + 0,10 до + 0,29	От - 0,10 до - 0,29
Связи нет	0	0

**Коэффициент линейной корреляции.**

При небольшом числе наблюдений ( $n \leq 30$ ), при не сгруппированных данных коэффициент линейной корреляции определяется по следующей формуле:

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

$x$  и  $y$  – смежные варианты сопоставляемых вариационных рядов;  $d_x$  и  $d_y$  – отклонение каждой переменной (варианты) от своей средней арифметической ( $M_x$  и  $M_y$ ).

Ошибка коэффициента линейной корреляции. Для того чтобы убедиться в том, что коэффициент корреляции, вычислительный по данным выборочного исследования, будет соответствовать размеру связи в генеральной совокупности, необходимо определить среднюю ошибку коэффициента корреляции ( $m_r$ ) и критерий  $t$ :

$$m_r = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}; \quad t = \frac{r_{xy}}{m_r},$$

$t$  – оценивается по таблице критерия  $t$ , где при  $n = n - 2$  ( $n$  – число парных вариантов) будет соответствовать вероятности наличия связи ( $p$ ).

Кроме того, для оценки достоверности коэффициента корреляции можно использовать специальную таблицу для малых выборок (табл.8). Преимущество этой таблицы заключается в том, что полученную величину коэффициента корреляции можно оценить без предварительных расчетов  $m_r$  и  $t$ , а путем сравнения  $r_{xy}$  со стандартным коэффициентом корреляции, рассчитанным и представленным в приложении 1 для различной степени вероятности и различного числа наблюдений. Методика пользования таблицей: в вертикальном первом ряду представлены значения  $n = (n - 2)$ , в верхней строке – желательная степень вероятности наличия связи ( $p$ ), при которой коэффициент корреляции можно считать достоверным. Цифры внутри таблицы – стандартные коэффициенты корреляции. Вычисленный  $r_{xy}$  должен быть  $\geq r_{\text{табл. (стандарт)}}$ .

### ***Коэффициент ранговой корреляции***

В некоторых случаях измерение направления и силы связи можно осуществлять с помощью так называемого коэффициента ранговой корреляции ( $\rho$ ) и его ошибки ( $m_\rho$ ).

*Коэффициент ранговой корреляции* для измерения взаимосвязи между парными признаками *применяют при следующих условиях:*

- 1) при небольшом числе наблюдений (не более 30 парных величин);
- 2) когда нет необходимости в точных расчетах уровня силы связи, а нужны лишь ориентировочные данные;
- 3) когда признаки имеют не только количественные, но и качественные (описательного характера) значения;
- 4) когда ряды распределения имеют открытые варианты (например,  $< 20$  или  $> 40$ ).

При расчете коэффициента ранговой корреляции ( $\rho$ ) не имеет значения характер связи: прямолинейная или криволинейная.

Формула расчета коэффициента ранговой корреляции:

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

где  $\rho$  – коэффициент ранговой корреляции;

$d$  – разность рангов,  $n$  – число пар.

Формула расчета ошибки коэффициента ранговой корреляции:

$$m_\rho = \sqrt{\frac{1 - \rho^2}{n - 2}}.$$

Оценка достоверности  $\rho$  осуществляется по тем же принципам, как и  $r_{xy}$  с помощью критерия  $t$  и числа степени свободы  $n$ :

$$t = \frac{\rho}{m_\rho}; \quad n = n - 2$$

Результаты сравниваются с табличными критериями  $t_{\text{табл.}}$  (приложение 1).

Таблица 8

**Стандартные коэффициенты корреляции, которые считаются достоверными (по Л.С. Каминскому)**

Число степеней свободы $n=n-2$	Уровень вероятности $p$ (в %)			Число степеней свободы $n=n-2$	Уровень вероятности $p$ (в %)		
	95,0	98,0	99,0		95,0	98,0	99,0
<b>1</b>	0,997	0,999	0,999	<b>12</b>	0,532	0,612	0,661
<b>2</b>	0,950	0,980	0,990	<b>13</b>	0,514	0,592	0,641
<b>3</b>	0,878	0,934	0,959	<b>14</b>	0,497	0,574	0,623
<b>4</b>	0,811	0,882	0,917	<b>15</b>	0,482	0,558	0,606
<b>5</b>	0,754	0,883	0,874	<b>16</b>	0,468	0,542	0,590
<b>6</b>	0,707	0,789	0,834	<b>17</b>	0,456	0,528	0,575
<b>7</b>	0,666	0,750	0,798	<b>18</b>	0,444	0,516	0,561
<b>8</b>	0,632	0,716	0,765	<b>19</b>	0,433	0,503	0,549
<b>9</b>	0,602	0,685	0,735	<b>20</b>	0,423	0,492	0,537
<b>10</b>	0,576	0,658	0,708	<b>25</b>	0,381	0,445	0,487
<b>11</b>	0,553	0,634	0,684	<b>30</b>	0,349	0,409	0,449

Преимущество этой таблицы заключается в том, что полученную величину коэффициента корреляции можно оценить без предварительных расчетов  $m_r$  и  $t$ , а путем сравнения  $r_{xy}$  со стандартным коэффициентом корреляции, рассчитанным и представленным в приложении 1 для раз-

личной степени вероятности и различного числа наблюдений. Методика пользования таблицей: в вертикальном первом ряду представлены значения  $n = (n - 2)$ , в верхней строке – желательная степень вероятности наличия связи ( $p$ ), при которой коэффициент корреляции можно считать достоверным. Цифры внутри таблицы – стандартные коэффициенты корреляции. Вычисленный  $r_{xy}$  должен быть  $\geq r_{\text{табл. (стандарт)}}$ .

#### **Вопросы для самоподготовки:**

1. Какие виды связи вы знаете?
2. Что такое корреляционный анализ,
3. Какие различают связи по направлению и силе?
4. Метод расчета коэффициента корреляции Пирсона.
5. Метод расчета коэффициента корреляции Спирмена.
6. Принцип выбора коэффициентов корреляции?

## **7. СРЕДНИЕ ВЕЛИЧИНЫ. ВАРИАЦИОННЫЙ РЯД. ПОКАЗАТЕЛИ ВАРИАЦИИ**

В медицинской статистике широко используются средние величины. Они применяются для характеристики здоровья населения: рождаемости, заболеваемости, инвалидности, смертности, в описании симптомов и течения различных болезней, физического развития отдельных контингентов, при обобщении результатов научных экспериментов. Средними величинами измеряют длительность течения болезни, сроки от начала заболевания до операции, сроки госпитализации или реабилитации, продолжительность жизни больных. При характеристике организации амбулаторно-поликлинической помощи населению используются такие понятия, как среднее число врачебных посещений на одного жителя в год, средняя численность населения на терапевтическом и педиатрическом участке и т.д.

### **7.1. Средние величины. Вариационный ряд.**

*Средняя* – это величина, которая одним числовым значением дает представление обо всей статистической совокупности. Какие требования

предъявляются средним величинам? Во-первых, средние величины следует вычислять только на качественно однородном материале. Так, например, при характеристике физического развития новорожденных в исследуемую группу должны быть отобраны младенцы одного пола. Во-вторых, при определении средних величин должно быть достаточное число наблюдений в выборочной совокупности.

Различают *несколько видов средних величин*: мода, медиана, средняя арифметическая, средняя геометрическая, средняя гармоническая.

*Модой* в вариационном ряду называется варианта, которая среди других встречается наиболее часто. Практическое значение моды заключается в том, что, не проводя порой достаточно сложных расчетов, ориентируясь на моду, можно знать примерное значение средней величины.

*Медианой* называется варианта, делящая вариационный ряд пополам. Практическое значение медианы заключается в том, что в симметричном вариационном ряду, котором от середины в обе стороны находится равное число вариантов, она по своему значению наиболее близка к средней величине.

*Средние арифметические величины* в зависимости от метода расчета делятся на среднюю арифметическую простую, среднюю арифметическую взвешенную, среднюю арифметическую способом моментов, среднюю арифметическую в сгруппированном или интервальном ряду.

Для расчета средней арифметической величины, прежде всего числовые значения (варианты) располагают в возрастающем или, напротив, в убывающем порядке, т.е. составляют *вариационный ряд*.

Различают следующие *виды вариационных рядов*:

1. *Простой*, когда каждая варианта встречается только один раз (пример 1).

2. *Взвешенный*, когда отдельные значения вариант начинают повторяться, нужно указать частоту встречаемости (Р) каждой варианты, когда варианты встречаются с различной частотой. Во взвешенном вариационном ряду среднюю арифметическую можно определить двумя методами: средняя арифметическая «взвешенная» и по способу моментов (пример 2).

3. **Сгруппированный**, когда варианты группируют в том случае, если большой объем наблюдений. Кроме того, различают:

1. **Непрерывный вариационный ряд**, который определяется в случае, когда числовые значения изучаемого признака могут выражаться дробными числами (рост, вес, масса тела, температура тела, содержание в крови и мочи их ингредиентов и т.д.).

2. **Дискретный вариационный ряд**, когда признаки выражены целыми числами (частота дыхания, пульс и т.д.).

Ниже на примерах рассмотрены методики вычисления средней арифметической в разных вариационных рядах.

**Пример 1.** Вычисление средней арифметической простой:

V <sub>см</sub>	P
116	1
118	1
119	1
121	1
122	1
125	1
127	1
128	1
130	1
	n=9

В простом вариационном ряду средняя арифметическая простая определяется по формуле  $M = \frac{\sum V}{n}$   $M = \frac{1106}{9} = 122,9$  см

**Пример 2.** Вычисление средней арифметической «взвешенной»:

V <sub>см</sub>	P	V×P
116	1	116
118	4	472
119	7	833
121	8	968
122	10	1220
125	16	2000
127	15	1905
128	10	1280
130	2	260
	n=73	$\sum VP = 9054$



Средняя арифметическая взвешенная определяется по формуле:

$$M = \frac{\sum VP}{n} \quad M = \frac{9054}{73} = 124,02 \text{ см}$$

Этот способ определения средней величины является неудобным ввиду необходимости проведения больших расчетов и применяется, в основном, при наличии счетной техники.

**Пример 3.** Вычисление средней арифметической способом моментов:

V см	P	a×(V-M <sub>1</sub> )	a×P
116	1	-9	- 9
118	4	-7	-28
119	7	-6	-42
121	8	-4	-32
122	10	-3	-30
<u>125</u>	16	0	0
127	15	2	30
128	10	3	30
130	2	5	10
	n =73		$\sum aP = -71$

В вариационном ряду выбирается варианта, которая наиболее часто встречается (**мода**), её принимают за условную среднюю величину (M<sub>1</sub>). В нашем примере 125. Находим отклонения всех других вариантов от условной средней, затем сумму произведений отклонений всех вариантов ( $\sum aP$ ) делим на общее число наблюдений ( $\frac{\sum aP}{n}$  момент первой степени).

Момент первой степени и является той величиной, которая показывает, насколько условная средняя варианта отличается от фактической или истинной средней. Напишем формулу:

$$M = M_1 + \frac{\sum aP}{n} \quad M = 125 + \frac{-71}{73} = 124,02 \text{ см}$$

При большом количестве наблюдений число встречающихся размеров вариант может быть очень большим; тогда рекомендуются варианты объединять в группы, причем каждая группа должна иметь равное число значений вариант (иметь равный интервал) 2, 3 ... и т.д.

**Пример 4.** Вычисление средней арифметической в сгруппированном вариационном ряду:

$V_1 - V_2$ (см)	P	$a_i$	$a_i \times P$
110-112	1	-4	- 4
113-115	5	-3	-15
116-118	11	-2	-22
119-121	12	-1	-12
122-124	17	0	0
125-127	14	1	14
128-130	9	2	18
131-133	4	3	12
	$n=73$		$\sum a_i P = -9$

**Условной средней ( $M_1$ )** в сгруппированном вариационном ряду является середина наиболее часто встречающейся группы (122-124), которая определяется в зависимости от изучаемого признака двумя способами:

1. **В непрерывном вариационном ряду**, условная средняя рассчитывается как полусумма первых значений смежных (соседних) групп.

2. **В дискретном вариационном ряду** – как полусумма начала и конца наиболее часто встречающейся группы, взятой за условную среднюю.

В данном примере ряд непрерывный (рост восьмилетних мальчиков), в данном примере условная средняя равняется:  $\frac{122 + 125}{2} = 123,5$  см.

Отклонения ( $a_i$ ) в сгруппированном вариационном ряду определяем как условные, выраженные в интервальных значениях (при определении отклонения пренебрегаем интервалом). В нашем примере интервал (разница между значениями групп)  $i = 3$  см.

Напишем формулу:  $M = M_1 + i \frac{\sum a_i P}{n}$        $M = 123,5 + 3 \frac{-9}{73} = 123,1$  см

## 7.2. Показатели вариации.

### Среднее квадратическое отклонение ( $\delta$ )

Средние величины, взятые сами по себе без оценки колеблемости вариационных рядов, из которых они вычислены, имеют ограниченное

значение. Сказанное поясним на **примере 5**: Рассмотрим два вариационных ряда.

V <sub>1</sub>	P <sub>1</sub>	V <sub>2</sub>	P <sub>2</sub>
7	1	4	1
9	1	8	1
10	1	10	1
11	1	12	1
13	1	16	1
	n=5		n=5

В первой группе разность или колеблемость вариант незначительна, чем во второй группе. Средняя арифметическая, равная в обоих случаях 10-ти, более типична для первого вариационного ряда. Приблизительно о варьировании (колеблемости) можно судить *по амплитуде* вариационного ряда – разности между максимальным и минимальным размерами вариант.

**Среднее квадратическое отклонение** более точно характеризует степень варьирования полученных данных среднее квадратическое отклонение, которое указывает на разброс данных вокруг средней величины одной выборки из генеральной совокупности, относится к показателям описательной статистики **и применяется для:**

1. *Оценки колеблемости вариационного ряда.* Правило трех сигм в симметричном вариационном ряду в пределах значения одной сигмы от величины средней арифметической, т.е.  $M \pm 1 \delta$  находится 68,3 % вариант от их общего числа. В пределах двух сигм ( $M \pm 2 \delta$ ) находится 95,5 % вариант, в интервале трех сигм ( $M \pm 3 \delta$ ) – 99,7 % вариант вариационного ряда. При нормальном распределении почти весь вариационный ряд укладывается в интервале ( $\pm 3 \delta$ ) от значения средней арифметической (M).

2. *Оценки физического развития.* Индивиды со значениями признака в пределах  $M \pm 1 \delta$  оцениваются как имеющие нормальное развитие, а этот интервал считают нормой. Индивиды со значением по признаку в пределах от  $+1 \delta$  до  $+2 \delta$  или от  $-1 \delta$  до  $-2 \delta$  оцениваются как имеющие развитие выше или ниже нормального, т.е. как субнорма. Если варианта находится в пределах от  $+2 \delta$  до  $+3 \delta$  или от  $-2 \delta$  до  $-3 \delta$ , то такой индивид расценивается как высокий или низкий (субаномалия).

3. *Оценки изменчивости нескольких вариационных рядов.* В тех случаях, когда сравниваются ряды, имеющие одну и ту же систему измерений, (например, рост или масса тела) можно сделать выводы непосредственно по величине среднего квадратического отклонения. При характеристике неоднородных рядов, когда значения одних представлены в метрах, других в килограммах, следует использовать коэффициент вариации:

$$C = \frac{\delta}{M} \times 100\%$$

*Критерии оценки коэффициента вариации.*

- низкий, если его величина не превышает 10,0%;
- средний, если его величина колеблется в пределах от 10,0 до 20,0%;
- высокий, – больше 20,0%.

4. *Применяется для оценки достоверности различий средних величин.*

Рассмотрим примеры вычисления среднего квадратического отклонения ( $\delta$ ).

**Пример 6.** Расчет среднего квадратического отклонения в простом вариационном ряду:

Vсм	P	d(V-M)	d <sup>2</sup>
116	1	-6,9	47,6
118	1	-4,9	24,0
119	1	-3,9	15,2
121	1	-1,9	3,6
122	1	-0,9	0,8
125	1	2,1	4,4
127	1	4,1	16,8
128	1	5,1	26,0
130	1	7,1	50,4
	n=9		$\sum d^2 = 188,8$

Последовательность расчета:

1. Находим отклонение (d) каждой варианты от истинной средней (V-M). Для данного вариационного ряда M = 122,9 (пример).
2. Отклонение возводим в квадрат (d<sup>2</sup>).

3. Находим сумму квадратов отклонений ( $\sum d^2$ ).

4. Сумму квадратов отклонений делим на число наблюдений и извлекаем корень квадратный.

Напишем формулу:  $\delta = \pm \sqrt{\frac{\sum d^2}{n}}$

При числе наблюдений  $n < 30$  формула следующая:

$$\delta = \pm \sqrt{\frac{\sum d^2}{n-1}} \quad \delta = \pm \sqrt{\frac{188,8}{8}} = \pm 4,8 \text{ см}$$

**Пример 7.** Расчет среднего квадратичного отклонения во взвешенном вариационном ряду (способ среднеарифметический):

Vсм	P	d	d <sup>2</sup>	d <sup>2</sup> P
116	1	-8	64	64
118	4	-6	36	144
119	7	-5	25	175
121	8	-3	9	72
122	10	-2	4	40
125	16	1	1	16
127	15	3	9	135
128	10	4	16	160
130	2	6	36	72
n=73				$\sum d^2P = 878$

1. Находим отклонения вариантов от истинной средней  $M=124,02$  (пример). Для упрощения расчетов возьмем  $M=124$  см.

2. Отклонения возводим в квадрат ( $d^2$ ).

3. Квадрат отклонений умножаем на частоту ( $d^2P$ ).

4. Находим сумму квадратов отклонений ( $\sum d^2P$ ).

5. Сумму квадратов отклонений делим на число наблюдений и извлекаем корень квадратный.

Напишем формулу:  $\delta = \pm \sqrt{\frac{\sum d^2P}{n}}$   $\delta = \pm \sqrt{\frac{878}{73}} = \pm 3,4 \text{ см}$

Если средняя арифметическая (M) рассчитывалась по способе моментов, то среднее квадратичное отклонение определяется по следующей методике.

**Пример 8.** Расчет среднего квадратического отклонения во взвешенном вариационном ряду моментов:

Vсм	P	a	aP	a <sup>2</sup>	a <sup>2</sup> P
116	1	-9	-9	81	81
118	4	-7	-28	49	196
119	7	-6	-42	36	252
121	8	-4	-32	16	128
122	10	-3	-30	9	90
<u>125</u>	16	0	0	0	0
127	15	2	30	4	60
128	10	3	30	9	90
130	2	5	10	25	50
	n=73		∑ aP = 71		∑ a <sup>2</sup> P = 947

Последовательность расчета:

1. Находим отклонения (a) вариант от условной средней (M<sub>1</sub>=125).
2. Отклонения умножаем на частоту встречаемости вариант (aP).
3. Находим сумму отклонений (∑ aP) и делим на число наблюдений ( $\frac{\sum aP}{n}$ ) – это момент первой степени.
4. Отклонения возводим в квадрат (a<sup>2</sup>).
5. Квадрат отклонений умножаем на частоту (a<sup>2</sup>P).
6. Находим сумму квадратов отклонений (∑ a<sup>2</sup>P) и делим на число наблюдений ( $\frac{\sum a^2P}{n}$ ) – момент второй степени.
7. Из момента второй степени вычитаем момент первой степени, возведенный в квадрат, извлекаем корень квадратный.

Напишем формулу и определим сигму:

$$\delta = \pm \sqrt{\frac{\sum a^2P}{n} - \left(\frac{\sum aP}{n}\right)^2} \quad \delta = \pm \sqrt{\frac{947}{73} - \left(\frac{71}{73}\right)^2} = \pm 3,4 \text{ см}$$

При определении средней арифметической величины в сгруппированном вариационном ряду отклонения (a) определяются в условных интервальных отклонениях (пример).

**Формула расчета среднего квадратического отклонения в сгруппированном вариационном ряду:**

$$\delta = \pm i \sqrt{\frac{\sum a^2P}{n} - \left(\frac{\sum aP}{n}\right)^2}, \text{ где } i \text{ – интервальное отклонение.}$$

**Упрощенный способ расчета среднее квадратичное отклонения:**

$$\delta = \pm \frac{V_{\max} - V_{\min}}{K}, \text{ где}$$

K – специальный коэффициент, величина которого определяется числом наблюдений по таблице С.И. Ермолаевой (табл.9).

Для нашего примера среднее квадратичное отклонение упрощенным методом:  $\delta = \pm \frac{130 - 116}{4,79} = \pm 2,9 \text{ см}$

Таблица 9

**Коэффициенты для вычисления среднего квадратичного отклонения по амплитуде (таблица С.И. Ермолаевой)**

	0	1	2	3	4	5	6	7	8	9
0	-	-	1,13	1,69	2,06	2,33	2,53	2,70	2,85	2,97
1	3,08	3,17	3,26	3,34	3,41	3,47	3,53	3,59	3,64	3,69
2	3,73	3,78	3,82	3,86	3,90	3,93	3,96	4,00	4,03	4,06
3	4,09	4,11	4,14	4,16	4,19	4,21	4,24	4,26	4,28	4,30
4	4,32	4,34	4,36	4,38	4,40	4,42	4,43	4,45	4,47	4,48
5	4,50	4,51	4,53	4,54	4,56	4,57	4,59	4,60	4,61	4,63
6	4,64	4,65	4,66	4,68	4,69	4,70	4,71	4,72	4,73	4,74
7	4,75	4,77	4,78	4,79	4,80	4,81	4,82	4,83	4,83	4,84
8	4,85	4,86	4,87	4,88	4,89	4,90	4,91	4,91	4,92	4,93
9	4,94	4,95	4,96	4,96	4,97	4,98	4,99	4,99	5,00	5,01
	100	200	300	400	500	600	700	800	900	1000
	5,02	5,49	5,76	5,94	6,07	6,18	6,28	6,35	6,42	6,48

## **8. ОЦЕНКА ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ**

В статистических исследованиях применяются 2 вида наблюдений – сплошное и выборочное. Самые надежные результаты можно получить при применении сплошного метода, т.е. при изучении генеральной совокупности.

Между тем изучение генеральной совокупности связано со значительной трудоемкостью. Поэтому в медико-биологических исследованиях, как правило, проводятся выборочные наблюдения. С тем, чтобы полученные при изучении выборочной совокупности данные можно было перенести на генеральную совокупность, необходимо провести оценку достоверности результатов статистического исследования. Выборочная совокупность может недостаточно полно представлять генеральную совокупность, поэтому выборочным наблюдениям всегда сопутствуют **ошибки репрезентативности** (средней ошибки ( $m$ )). *По размерам средней ошибки ( $m$ ) можно судить, насколько найденная выборочная средняя величина отличается от средней генеральной совокупности.* Малая ошибка указывает на близость этих показателей, большая ошибка такой уверенности не дает. На величину средней ошибки средней арифметической влияют следующие два обстоятельства.

Во-первых, однородность собранного материала: чем меньше разбросанность вариант вокруг своей средней, тем меньше ошибка репрезентативности.

Во-вторых, число наблюдений: средняя ошибка будет тем меньше, чем больше число наблюдений.

Средняя ошибка средней арифметической вычисляется по следующей формуле:

$$m_M = \pm \frac{\delta}{\sqrt{n}}$$

Средняя ошибка для относительных величин вычисляется по формуле:

$$m_p = \pm \sqrt{\frac{P \cdot q}{n}}, \text{ где}$$

$P$  – величина показателя в расчете на 100, 1000, 10 000 и т.д.;

$q$  – разность между основанием, на которое рассчитывается показатель, и его конкретным числовым значением (100 –  $P$ , 1000 –  $P$ , 10 000 –  $P$  и т.д.).

При  $n < 30$  в знаменателе  $n - 1$ .

Рассмотрим методику вычисления средних ошибок средних и относительных величин на примерах:

**Пример.** Средний рост восьмилетних мальчиков составил – 125,5 см, среднее квадратическое отклонение  $\delta = \pm 3,4$  см,  $n = 73$

$$m_M = \pm \frac{3,4}{\sqrt{73}} = \pm 0,4 \text{ см}$$



**Пример.** Численность детей в возрасте до года по данным детской поликлиники составила 450, из них ни разу не болели 100 детей. Необходимо определить "Индекс здоровья "(процент ни разу не болевших детей) и вычислить ошибку для данного показателя.

$$\text{Индекс здоровья} = \frac{100}{450} \times 100 = 22,2\%$$

$$m_p = \pm \sqrt{\frac{22,2 \cdot 77,8}{450}} = 1,9\%$$

### **8.1. Определение средней ошибки показателя равных или близких к 0 или 100 %**

При проведении наблюдений в небольших группах, возможно получение показателя равного 0 или 100 %. Вычисление средней ошибки показателя по известной формуле:

$$m = \pm \sqrt{\frac{P \cdot q}{n}}$$

не представляется возможным, т.к. в обоих случаях величина  $P = 0$ .

Вместе с тем, у показателя, равного 0 или 100%, тоже имеется средняя ошибка, но вычисляется по следующей формуле (А.М. Мерков, Л.Е. Поляков, Л.,1974):

$$m = \pm \frac{t^2 \times 100}{n + t^2}, \text{ где}$$

t-доверительный коэффициент Стьюдента;

n-число наблюдений.

Среднюю ошибку целесообразно рассчитывать по указанной формуле и в тех случаях, когда показатели незначительно отличаются от 0 или 100%. Посмотрим на примерах:

Допустим послеоперационная летальность в хирургическом отделении стационара составила 1,3 %, число оперированных – 560. Определим среднюю ошибку показателя послеоперационной летальности. При этом будем исходить из того, что в медико-биологических исследованиях достаточной является достоверность 95 %, которой соответствует значение критерия t Стьюдента, равное 2.

$$m = \frac{2^2 \times 1,3}{560 + 2^2} = \pm 0,009\%$$

Определим среднюю ошибку, когда показатели отличаются незначительно от 100 %. Допустим эффективность внутриматочной контрацепции у 130 обследованных составила 99 %.

$$m = \frac{2^2 \times 99}{130 + 2^2} = \pm 2,9 \%$$

## 8.2. Доверительный интервал.

### Оценка достоверности средних и относительных величин

*Доверительный интервал* (confidence interval) – статистический показатель, позволяющий оценить, в каких пределах может находиться истинное значение параметра в популяции; диапазон колебаний истинных значений. Величины, полученные в исследованиях на выборке больных, отличаются от истинных величин в популяции вследствие влияния случайности.

Это оценка интервала генерального среднего, которая содержит информацию об уровне значимости, а также о величине и направлении рассматриваемого эффекта лечения, по сравнению с контрольным результатом. Он представлен в виде интервала значений, который включает в себя оцениваемый генеральный параметр, с определенной и указанной вероятностью, называемой доверительной. Он всегда симметричен относительно среднего арифметического. Интервал оценивает вариабельность элементов выборки. Широкий доверительный интервал означает высокий уровень выборочной ошибки ( $m$ ).

Доверительный интервал сужается, если:

- возрастает объем выборки;
- уменьшается изменчивость данных;
- уровень доверия, необходимый для генерального среднего, уменьшается.

При оценке достоверности средних или относительных величин руководствуются следующим правилом: *средняя арифметическая или относительная величина при числе наблюдений в выборочной совокупности 30 и более должны превышать свою ошибку не менее чем в 2 раза.*

$$\frac{M}{m} > 2 \quad \text{или} \quad \frac{P}{m} > 2$$

В рассматриваемых примерах средняя арифметическая, характеризующая рост восьмилетних мальчиков и показатель «индекс здоровья» превышают свои ошибки соответственно  $\frac{125,5}{0,4} = 313 \cdot \frac{22,2}{1,9} = 11$  раз, что соответствует высокой степени их статистической достоверности с вероятностью более чем 99,7%. Высказанное положение вытекает из теории «вероятности», под которой понимается числовая мера объективной возможности появления случайного события.

**Вероятность** – число, которое находится между 0 и 1, или между 0% и 100%. Математиками определено, что той или иной вероятности, выраженной в процентах, соответствует определенное значение критерия  $t$  Стьюдента. Так, например, вероятности равной  $P = 68,3 \%$  соответствует  $t = 1,0$ , вероятности  $P = 95,5 \%$  –  $t = 2,0$  и вероятности  $P = 99,7 \%$  –  $t = 3,0$ .

***В медико-биологических исследованиях событие является статистически достоверным, если вероятность его появления соответствует значению критерия  $t$  Стьюдента, равное 2.***

Средняя ошибка позволяет не только оценить достоверность относительного показателя или средней величины, но и найти доверительные границы средней величины или относительного показателя в генеральной совокупности.

$$M_{\text{ген.}} = M_{\text{выб.}} \pm t \times m$$

$$P_{\text{ген.}} = P_{\text{выб.}} \pm t \times m$$

Как уже было сказано, величина средней ошибки указывает, насколько средняя величина и относительный показатель выборочной совокупности отличаются от соответствующих величин в генеральной совокупности. Величина  $t \times m$  является тем доверительным интервалом по отношению к средней или относительной величине, в котором с определенной степенью вероятности можно ожидать нахождения средней или относительной величины в генеральной совокупности.

**Пример.**

$$M_{\text{выб.}} = 125,5 \text{ см}; m = \pm 0,4 \text{ см.}$$

При 95 % вероятности  $t = 2$ , при 99,7 % –  $t = 3$ .

$$M_{\text{ген.}} = 125,5 \text{ см} \pm 2 \times 0,4 \text{ см} = 124,7 - 126,3 \text{ см}$$

$$M_{\text{ген.}} = 125,5 \text{ см} \pm 3 \times 0,4 \text{ см} = 124,3 - 126,7 \text{ см}$$

Таким образом, с вероятностью 95 % можно ожидать, что средняя будет находиться в пределах от 124,7 до 126,3 см и с вероятностью 99,7 % – в пределах от 124,3 до 126,7 см.

Понятно, что действительное значение средней в генеральной совокупности можно получить только при обследовании всех 8-летних мальчиков, но подобное исследование нецелесообразно, т.к. средняя арифметическая статистически достоверна ( $p > 99,7 \%$ ), а доверительный интервал для средней в генеральной совокупности является весьма незначительным –  $tm=3 \times 0,4$  т.е. всего по 1,2 см от средней выборочной совокупности в большую и меньшую сторону.

## **9. ОСНОВНЫЕ МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ. КЛАССИФИКАЦИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ**

Обычно перед исследователем встает ряд вопросов, для ответа на которые необходимо воспользоваться различными методами статистического анализа данных. В одном исследовании для решения его задач необходимо использовать как можно меньшее число различных методов. Анализ данных служит лишь для выдвижения или проверки научных гипотез.

*Методы статистического анализа данных принято классифицировать по нескольким принципам:*

### **1. По количеству анализируемых признаков:**

- *одномерные* (анализ каждого признака в отдельности);
- *двумерные* (одновременный анализ двух признаков, например, анализ связей изучаемых признаков – ассоциации или корреляции);
- *многофакторные* (анализ трех признаков и более одновременно, например, многофакторный дисперсионный анализ, многофакторный регрессионный анализ, дискриминантный анализ).

### **2. По статистическим принципам, лежащим в основе методов:**

- *параметрические*. Применяются главным образом для анализа нормально распределенных количественных признаков;

- *непараметрические*. Применяются в остальных случаях:

а) для анализа количественных признаков независимо от вида их распределения;

б) для анализа качественных признаков.

В целом, непараметрические методы (в случае использования их на малых выборках) являются менее мощными по сравнению с параметрическими, т.е. иногда не позволяют выявить статистические закономерности, которые могут быть выявлены с помощью параметрических методов. В то же время непараметрические методы более надежны в случаях, когда есть сомнения в том, что анализируемый признак имеет нормальное распределение. Для нормального распределения признаков параметрические и непараметрические методы дают близкие результаты.

Наиболее мощным является так называемые параметрические методы, в которых для получения окончательного результата необходимо предварительно определить значение ряда статистических параметров. Примером параметрических методов могут быть критерий Стьюдента ( $t$ ) и дисперсионный анализ.

Параметрические методы применимы лишь при нормальном распределении вариант в генеральной совокупности. Чем больше распределение отклоняется от нормального, тем значительно меньше параметрические методы теряют свою мощность, то есть возрастает вероятность не выявить существующие статистические закономерности. В этих условиях более простые непараметрические методы могут оказаться более мощными, чем параметрические.

Если параметрические методы оказались менее мощными, чем непараметрические и такие методы, как, например, Стьюдента –  $t$ , дисперсионный анализ и др., не выявили достоверности различий сравниваемых выборок, а использованные непараметрические (Вилкоксона-Манна-Уитни –  $U$ , Вилкоксона –  $T$ ,  $\chi^2$  – Фридмана,  $\chi^2$  – Макнимара, Кокрена –  $Q$ , и др.) их выявили, то можно сделать предположение о значительном отклонении вариант в данной совокупности от нормального распределения (Нечаев И.Е., 1970).

В статистике руководствуются правилом: начинать сравнение вариационных рядов с применения наиболее простых и наименее трудоемких методов.

Статистические пакеты позволяют начинать сравнение вариационных рядов с параметрических методов (критерий Стьюдента –  $t$ , дисперсионный анализ). Такая практика особенно выгодна в тех случаях, когда анализируются результаты многих исследований (физиологических, биохимических, гематологических, психофизиологических, социологических и др.), полученных у одних и тех же лиц в динамике. В этом случае вариационные ряды удобно сравнивать по единой программе. Однако напомним, что будет ошибкой сравнение их параметрическими методами, когда оно будет заканчиваться получением недостаточно достоверных данных. В этих случаях исследование следует продолжить с применением непараметрических критериев.

### **3. По зависимости или независимости сопоставляемых выборок:**

Классификация наиболее важных статистических методов, которая может быть использована при выборе теста для решения конкретной задачи, представлена в таблице 10.

Тесты для *независимых выборок*. Используются в случае, если при формировании выборок объекты исследования набирались в группы независимо друг от друга (например, были рандомизированы в две группы или лечились в разных больницах).

Тесты для *зависимых выборок*. Используются в двух случаях:

а) когда в динамике анализируются данные, полученные в одной группе больных (например, до и после лечения);

б) когда анализируются выборки, сформированные путем подбора пар по полу, возрасту, стадии заболевания и т.д.

**Рекомендуемые к использованию статистические критерии (методы)  
в зависимости от задачи статистического исследования и типа данных**

Задача	Методы	
	Параметрические (для количественных нормально распределенных признаков)	Непараметрические (для количественных признаков независимо от вида распределения, для качественных – порядковых или номинальных признаков)
Выполнение описательной статистики	Вычисление средних значений, средних квадратичных отклонений и т.д.	Вычисление медиан и интерквартильных интервалов, пропорций
Сравнение двух независимых групп по одному признаку	<i>t</i> -критерий Стьюдента для независимых выборок	Критерии Манна-Уитни, Колмогорова-Смирнова, Вальда-Вольфовица, $\chi^2$ , точный критерий Фишера
Сравнение двух зависимых групп по одному признаку	<i>t</i> -критерий Стьюдента для зависимых выборок	Критерий Вилкоксона, критерий знаков, критерий МакНемара
Сравнение трех независимых групп и более по одному признаку	ANOVA- однофакторный дисперсионный анализ	ANOVA по Краскелу-Уоллису, медианный критерий, критерий $\chi^2$
Сравнение трех зависимых групп и более по одному признаку	Критерий Кокрана	ANOVA по Фридмену, критерий Кокрана
Анализ взаимосвязи двух признаков	Корреляционный анализ по Пирсону	Критерий $\chi^2$ , корреляционный анализ по Спирмену, Кендаллу, гамма и др.
Одновременный анализ трех признаков и более	Регрессионный анализ Дискриминантный анализ Факторный анализ Кластерный анализ	Логистический регрессионный анализ, логлинейный анализ, анализ древовидных диаграмм, анализ конъюнкций и др.

В таблице 11 предложены примеры выбора адекватных методов статистической обработки данных в зависимости от задач медико-биологических исследований.

**Примеры выбора методов при статистической обработке  
данных медико-биологических исследований**

<i>Задача исследования</i>	<i>Методы обработки</i>
Оценить достоверность различий средней продолжительности одного случая заболевания рабочих в двух цехах.	Критерий Стьюдента ( $t$ ) для средних величин.
Определить существенность различий четырех замеров ( $n_1=n_2\leq 20$ ) роста детей (нормальное распределение признака).	Критерий Лорда ( $O$ ).
Оценить значимость различных уровней АД у студентов двух групп.	Критерий Вилкоксона-Манна-Уитни ( $U$ ).
Определить достоверность различий уровней паратгормонов у больных двух групп ( $n=20$ ), если в одной группе выявлено 6 наибольших значений.	Критерий Мостеллера.
Оценить достоверность различий СОЭ у больных двух групп (при различном уровне и качестве лечения).	Критерий Розенбаума ( $Q$ ).
Определить уровень трансаминазы в крови здоровых лиц и больных.	Критерий Уайта ( $K$ ).
Установить зависимость показателей реакции утомляемости организма (сенсомоторные реакции) в начале и в конце рабочей смены.	Максимум – критерий для разностей пар.
Оценить степень утомления рабочих, несущих вахту по нескольким сменам в течение суток.	Однофакторный дисперсионный анализ.
Оценить влияние условий труда на уровень ЗВУТ у рабочих нескольких предприятий.	Однофакторный дисперсионный анализ.
Определить надежность работы аппаратчиков с различным стажем работы, выполняющих труд нескольких степеней напряженности.	Двухфакторный дисперсионный анализ.
Определить влияние пола рабочих и профиля цеха на ЗВУТ рабочих.	Двухфакторный дисперсионный анализ.
Оценить влияние профессии, стажа работы и места жительства на уровень заболеваемости острым бронхитом.	Трехфакторный дисперсионный анализ.
Оценить поведенческие реакции животных в эксперименте в период наблюдения.	Критерий Фридмана ( $\chi^2$ ).
Определить существенность динамики показателей крови после эндогенной и экзогенной нагрузки.	Критерий Вилкоксона – Вилкоккс.
Сравнить и оценить достоверность различий уровней заболеваемости по данным двух поликлиник.	Критерий Стьюдента ( $t$ ) для относительных величин.
Определить влияние прививок на показатели заболеваемости населения.	Критерий согласия ( $\chi^2$ ) (алгебраический метод).
Выявить различия в показателях функционального состояния организма двух малых ( $n < 10$ ) групп лиц.	Точный метод Фишера (ТМФ).



<i>Задача исследования</i>	<i>Методы обработки</i>
Оценить достоверность токсического эффекта химического вещества в опыте на животных (контроль – опыт).	Критерий Ван – дер – Вадена(x).
Оценить силу действия препарата (сильное, слабое).	Критерий( $\chi^2$ ) Макнимара
Оценить влияние профессии на состояние здоровья рабочих (болели, не болели).	Дисперсионный анализ для альтернативных вариантов.
Оценить влияние времени госпитализации на исход лечения.	Критерий( $\chi^2$ ) по Р. Рунioni.
Оценить достоверность различий показателя физического развития (в баллах) в двух малых независимых выборках.	Критерий Ван – дер – Вардена ( $\chi$ ).
Установить достоверность различий посещения врача мужчинами и женщинами в условиях городской поликлиники.	Критерий итерации по Веберу.

## 10. ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ

Напомним, что параметрические методы применяются главным образом для анализа нормально распределенных количественных признаков.

### 10.1. Сравнение двух независимых выборок.

#### Критерий Стьюдента ( $t$ )

Метод Стьюдента применим для сравнения выборок, отдельные наблюдения в которых распределены нормально.

Для оценки различий двух сравниваемых выборок методом Стьюдента необходимо предварительно определить некоторые статистические параметры этих выборок. Поэтому метод называют **параметрическим**. Критерий Стьюдента ( $t$ ) может использоваться для оценки достоверности различия двух средних величин как в независимых (несвязанных), так и в зависимых (связанных) группах.

С целью проверки высказанных гипотез применяются различные статистические приемы, среди которых наиболее простым является следующий:

а) для средних величин  $t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$ ;

б) для относительных величин  $t = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}}$ , где

$M_1$  и  $P_1$  – более выраженные по своей величине средняя или относительная;  
 $M_2$  и  $P_2$  – средняя или относительная величина, которые по своей величине меньше в сравнении с  $M_1$  и  $P_1$ .

Если полученное значение критерия  $t$  Стьюдента окажется равным 2, что соответствует  $p = 95 \%$ , это является достаточным в медико-биологических исследованиях.

Критерий Стьюдента может иметь разные значения в зависимости от этого вероятность различия между показателями может составлять 95%, 99,7% и 99,9 %. Но может быть и так, что  $t > 2$ . В таком случае вероятность различий  $P > 95 \%$ , а практический вывод заключается в том, что нельзя утверждать о статически достоверном различии сравниваемых и относительных величин. Исследователю можно рекомендовать увеличить число наблюдений с тем, чтобы окончательно решить вопрос о влиянии изучаемого фактора на резульативный признак.

**Пример.** Проведем оценку достоверности различий показателей, характеризующих “индекс здоровья” детей двух районов:

$$P_1 = 28 \% \quad m_1 \pm 0,5 \% \quad n_1 = 250 \quad P_2 = 26 \% \quad m_2 = \pm 0,4 \% \quad n_2 = 310$$

$$t = \frac{28 - 26}{\sqrt{0,5^2 + 0,4^2}} = \frac{2}{\sqrt{0,41}} = 3,1$$

В связи с тем, что критерий достоверности оказался равным 3,1 ( $t=3,1$ ), который соответствует по таблице стандартных значений критерия  $t$  Стьюдента вероятности 99,7 %, можно утверждать о наличии статически достоверных различий между показателями “индекса здоровья” детей двух районов.

В медицинской литературе чаще вероятность достоверности выражается не критерием  $t$ , а уровнем значимости ( $p$ ), который является дополнением доверительной вероятности до 100 % (или до 1,0). Так вероятности 95 % (0,95) соответствует уровень значимости 0,05 ( $1,0 - 0,95 = 0,05$ ), вероятности 99 % – 0,01 ( $1 - 0,99 = 0,01$ ), вероятности 99,9 % – 0,001 ( $1 - 0,999 = 0,001$ ).

В противоположность доверительной вероятности уровень значимости указывает не на степень надежности полученного показателя или вывода, а на возможную вероятность его ошибочности. Как уже отмечалось выше, в математической статистике минимальным значением достоверности считается вероятность в 95 % (0,95) или же уровень значимости 0,05. Чем меньше уровень значимости, тем больше достоверность, т.е.  $0,001 < 0,01 < 0,05$ .

Оценить достоверность различий в уровнях значимости нужно также по таблице Стьюдента. Определим уровень значимости по найденному критерию  $t=3,1$  (см. пример). Берется сумма чисел наблюдений (в случае, если число наблюдений меньше 30, то вычитается 1, а если в обеих группах  $n < 30$ , то вычитается два). На нашем примере  $250 + 310 = 660$ . В таблице Стьюдента (приложение 1) эта цифра близка к  $\infty$ . Для критерия  $t = 3,1$  при данном числе наблюдений уровень значимости будет  $p < 0,01$ , т.к.  $t = 3,1 <$  чем  $3,29$ , но  $> 1,96$  (Найденный критерий  $t$  должен быть больше табличного значения).

Таким образом можно сделать вывод о наличии достоверной разницы между показателями «индекса здоровья» детей двух районов с уровнем значимости  $< 0,01$  ( $P < 0,01$ ), что соответствует вероятности достоверности 99%.

## **10.2. Сравнение двух зависимых выборок.**

### **Критерий Стьюдента (t) (разностный метод)**

Объектом сравнения являются данные повторного исследования тех или иных показателей у одних и тех же лиц, изменение которых ожидается в связи с воздействием изучаемого фактора. Таким образом, сравниваются два ряда сопряженных (попарно связанных) показателей, полученных в динамическом исследовании, например: до и во время действия фактора, до и после, во время и после действия и т.д. (Урбах В.Ю., 1963; Каминский Л.С., 1964).

Определение достоверности факториальных влияний при сравнении двух зависимых выборок проводится так называемым разностным методом. Он основан на определении среднего арифметического разностей между показателями. Критерий Стьюдента ( $t$ ) для сравнения зависимых

выборки является наиболее мощным, но лишь в случае нормального (или незначительно отличающегося от нормального) распределения разностей величин в выборках. Исследование одних и тех же лиц в динамике наиболее эффективно, чем исследование лиц основной и контрольной групп, так как при таком сравнении выборок меньше влияние на среднее квадратическое отклонение ( $\delta$ ) и ошибку репрезентативности ( $m$ ) оказывают исходные (фоновые) индивидуальные особенности каждого обследованного. Сравнение двух функционально связанных (зависимых, сопряженных) выборок производится по следующей схеме:

1. Определяется разность показателей каждой из функционально связанных пар. Из двух сравниваемых рядов данных образуется один ряд – разностей.

2. Последовательно определяются среднее арифметическое, среднее квадратическое отклонения и стандартная ошибка полученного ряда разностей.

3. Критерий  $t$  определяется по формуле

$$t_d = \frac{M_d}{m_d}, \text{ где}$$

$t_d$  – аргумент распределения критерия Стьюдента ( $t$ ) для разницы значений;

$M_d$  – среднее арифметическое для ряда разностей;

$m_d$  – стандартная ошибка ряда разностей.

4. По табл. 1 приложения определяется критическое значение и затем, на основании сравнения полученного значения  $t$  с табличным – достоверность различий сравниваемых выборок.

**Пример:** Сотрудники предприятия исследовались в начале и конце дня. Определялось время сенсомоторной реакции (табл. 12). Ожидалось ухудшение показателей в конце рабочего дня. Задача статистической обработки заключалась в установлении значимости ожидаемого ухудшения показателей.

Среднее квадратическое отклонение разностей ( $\sigma_d$ ) определяем по размаху (амплитуде) вариационного ряда и значению  $K$  (табл. 6 приложения):

$$\sigma_d = \frac{x_{\max} - x_{\min}}{K} = \frac{0,36}{3,08} = 0,117$$

Стандартная ошибка разностей:

$$m_d = \frac{\sigma_d}{\sqrt{n}} = \frac{0,117}{3,16} = 0,037$$

$$t_d = \frac{M_d}{m_d} = \frac{0,165}{0,037} = 4,5$$

Таблица 12

**Время сенсомоторной реакции у работников предприятия (n=10)**

<i>Время реакции, с</i>		<i>Разности</i>
<i>в начале дня</i>	<i>в конце дня</i>	
0,51	0,71	+ 0,20
0,70	0,71	+ 0,01
0,82	0,72	+ 0,10
0,65	0,85	+ 0,20
0,46	0,88	+ 0,42
0,61	0,91	+ 0,30
0,75	0,85	+ 0,10
0,81	0,91	+ 0,10
0,51	0,81	+ 0,30
0,66	0,78	+ 0,12

Сумма разностей  $\Sigma d=1,65$

Среднее арифметическое разностей  $M_d=0,165$ .

Обращаемся к табл. 4 приложения (граничные значения  $t$ ). Число степеней свободы  $K$  при сравнении зависимых выборок равно числу пар минус 1, а не  $n_1 + n_2$ , так как выборки зависимы; они "не свободны", а взаимосвязаны. Таким образом,  $K=10-1=9$ . Для  $K=9$  и пороге значимости  $P=0,05$  табличное значение для одностороннего критерия равно 1,833, для  $P = 0,01$  при  $K = 9$ ; следовательно, различия сравниваемых вариационных рядов достоверны. Действие фактора привело к значимому увеличению времени сложной сенсомоторной реакции ( $p < 0,01$ ). "Нулевая" гипотеза отвергается с вероятностью ошибочного прогноза менее 0,01.

## 11. НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ

Непараметрические методы статистической обработки полученных данных используются при распределении полученных данных отличным от нормального, при анализе качественных данных.

### 11.1. Метод сравнения двух независимых выборок.

#### Критерий Манна-Уитни ( $U$ )

Метод сравнения двух независимых выборок с применением критерия  $U$  не требует нормального распределения значений в выборках и равенства объема выборок (Закс Л., 1976). Сравняются две независимые выборки ( $A$  и  $B$ ) объемом  $n_1$  и  $n_2$  следующим образом:

1. Оба вариационных ряда объединяют и значения распределяют в порядке нарастания величины (упорядочивают значения общего ряда).

2. Каждому значению придают ранг и отмечают, к какому из рядов ( $A$  или  $B$ ) он относится. Определяют сумму рангов в первом ( $\Sigma R_1$ ) и втором ( $\Sigma R_2$ ) вариационных рядах.

3. Вычисляют  $U_1$  и  $U_2$  по формуле:

$$U_1 = n_1 n_2 - \frac{n_1(n_1 + 1)}{2} - \Sigma R_1;$$

$$U_2 = n_1 n_2 - \frac{n_2(n_2 + 1)}{2} - \Sigma R_2.$$

Поясним алгоритм определения  $U_1$  и  $U_2$  на п р и м е р е (по Заксу Л., 1976). В результате лечения двух групп больных ( $A$  и  $B$ ) были определены следующие величины СОЭ:

- выборка  $A$  (вариационный ряд): 7 9 13 14 16 17 18 19 ( $n_1=8$ );
- выборка  $B$  (вариационный ряд): 3 5 6 8 10 11 12 15 ( $n_2=8$ );
- объединяем ряда  $A$  и  $B$  и упорядочиваем их. Каждому значению придаем ранг и отмечаем, в какой группе выборок он относится;
- значения: 3 5 6 7 10 14 17 18 20 22 36 39 40 48 49 50;
- ранги: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16;
- выборка:  $B B B A B A B B B A A B A A A A$ ;

- определяем сумму рангов в выборках  $A$  и  $B$ .

В выборке  $A$ :  $\Sigma R_1 = 4 + 6 + 10 + 11 + 13 + 14 + 15 + 16 = 89$ .

В выборке  $B$   $\Sigma R_2 = 1 + 2 + 3 + 4 + 5 + 7 + 8 + 9 + 12 = 51$ ;

- вычисляем:

$$U_1 = 8 \times 8 = \frac{8(8+1)}{2} - 89 = 11;$$

$$U_2 = 8 \times 8 = \frac{8(8+1)}{2} - 47 = 53;$$

Из приведенных формул видно, что различия значений  $U_1$  и  $U_2$  зависят от  $n$  и различий сумм рангов для каждого ряда. Пользуясь табл. 3 приложения, сравниваем наименьшее значение  $U_1$  в данном примере с табличным. Для  $n_1 = n_2 = 8$  критическое значение  $U$  при пороге надежности  $p = 0,05$  (для двустороннего критерия) равно 13. Достоверность сравниваемых вариационных рядов при заданном пороге надежности устанавливается тогда, когда наименьшее значение  $U$  меньше табличного. В приведенном примере различия достоверны с вероятностью ошибочного прогноза ( $p$ ) меньше 5 % ( $p < 0,05$ ).

В тех случаях, когда в выборках, упорядоченных по величине значений, имеются одинаковые значения, им присваивается средний ранг. Если, например, в вариационном ряду имеются значения: 3 3 4 5 5 5 5 9 9 10 13 15 ( $n=14$ ), то им присваиваются соответствующие ранги: 1,5; 1,5; 3; 5,5; 5,5; 5,5; 5,5; 8,5; 8,5; 10; 12; 12; 12; 14. Так, в ряду оказалось 2 значения по 3, им присваиваются ранги 1 и 2, средний ранг для них будет равен

$$\frac{1+2}{2} = 1,5.$$

4 значения вариационного ряда имеют величину 5, они занимают в вариационном ряду места 4-7. Им присваиваются ранги 4, 5, 6, 7, а средний ранг будет равен

$$\frac{4+5+6+7}{4} = 5,5 \text{ и т.д.}$$

Достоинством критерия  $U$  является то, что его можно использовать при очень малом числе наблюдений. Так, этот критерий применим уже при  $n_1 = n_2 = 3$ . В качестве примера рассмотрим результаты 3 замеров артериального давления (АД) у рабочих основной и 4 замеров – контрольной

группы, работающих в особо опасных условиях труда с психоэмоциональным напряжением.

Опытная группа	95 мм рт. ст.
	85 мм рт. ст.
Контрольная группа	95 мм рт. ст.
	75 мм рт. ст.
	80 мм рт. ст.
	80 мм рт. ст.
	90 мм рт. ст.

Необходимо сравнить эти две независимые выборки для определения статистической достоверности выявленных различий.

Объединяем выборки *O* и *K* и упорядочиваем их в одном вариационном ряду. Присваиваем значениям ранги:

Значения:	75	80	80	85	90	95	95
Ранги:	1	2,5	2,5	4	5,0	6,5	6,5
Выборки:	К	К	К	О	К	О	О

Определяем сумму рангов в выборках *A* и *B* ( $R_1$  и  $R_2$ ):

$$\Sigma R_1 = 4 + 6,5 + 6,5 = 17;$$

$$\Sigma R_2 = 1 + 2,5 + 2,5 + 5 = 11.$$

Вычисляем:

$$U_1 = 3 \times 4 + \frac{3 \times (3 + 1)}{2} - 17 = 1;$$

$$U_2 = 3 \times 4 + \frac{4 \times (4 + 1)}{2} - 11 = 11.$$

В данном случае есть основание использовать односторонний критерий, так как для умственно-эмоционального напряжения характерно повышение, а не понижение давления. И вопрос здесь ставится односторонне: не возникло ли статистически значимое повышение АД у лиц основной группы в связи с условиями труда?

Минимальное значение  $U$  определено для  $U_1$ , оно равно 1. По табл. 3 приложения критическое (пороговое) значение наименьшей  $U$  для одностороннего критерия при  $n_1=3$  и  $n_2=4$  составляет 0. Полученное значение (1) больше порогового ( $1 > 0$ ). Различия не достоверны ( $p > 0,05$ ).

Расчет критерия Манна-Уитни можно проводить в модификации Е.В. Гублера и А.А. Генкина (1973). Метод основывается на подсчете ин-



версий (нарушений в расположении чисел, когда число одного вариационного ряда стоит впереди числа другого вариационного ряда).

**Пример.** После проведения летней оздоровительной кампании уровень гемоглобина крови (г/л) детей, отдохавших в санаториях Черноморского побережья, составил 144, 148, 151, 147, 134, 135, 129, 131, 140, 136 (10 детей основной группы). Аналогичные показатели были у детей, отдохавших в Санкт-Петербурге: 128, 130, 126, 132, 125, 138, 150, 125 (8 детей контрольной группы). Достоверно ли различие показателей в основной и контрольной группах?

Расположим все данные в порядке возрастания в двух графах  $X$  (основная группа) и  $Y$  (контрольная группа) так, чтобы в одной строке был только один результат:

$X$	$Y$	Число инверсий
-	125	0
-	125	0
-	126	0
-	128	0
129	-	-
-	130	1
131	-	-
-	132	2
134	-	-
135	-	-
136	-	-
-	138	5
140	-	-
144	-	-
147	-	-
148	-	-
-	150	9
151	-	-

Определим число инверсий, когда число ряда  $X$  стоит впереди числа ряда  $Y$ .

Перед числами 125, 125, 126 и 128 нет чисел ряда  $X$ , поэтому инверсия нулевая. Перед числом 130 – одна инверсия (129), перед числом 132 – 2 инверсии (129 и 131), перед числом 138 – 5 (123, 131, 134, 135 и 136), перед числом 150 – 9 инверсий (123, 131, 134, 135, 136, 140, 144, 147, 148).

Таким образом,  $U = 1 + 2 + 5 + 9 = 17$ .

Найденную величину сравниваем с критическим значением по табл. 3 приложения. При  $n_1 = 10$  и  $n_2 = 8$  находим максимальное значение  $U$ , при котором можно считать различия в сравниваемых группах достоверными. Оно равно 17 при  $p=0,05$ . Следовательно, различие уровней гемоглобина крови детей, проводших лето на Черноморском побережье и в Санкт-Петербурге, статистически достоверно ( $p = 0,05$ ).

## **11.2. Методы сравнения двух зависимых выборок.**

### **Критерий Вилкоксона ( $T$ )**

Критерий является весьма чувствительным для подтверждения систематических различий параллельных определений. Ранговый критерий (парный критерий Вилкоксона, критерий  $T$ ) рассматривается как наиболее мощный непараметрический критерий для сравнения двух выборок с распределением величин, значительно отличающихся от нормального (Урбах В.Ю., 1963; Каминский Л.С., 1964; Поляков Л.Е., 1971; Лакин Г.Ф., 1980; Сепитлиев Д., 1968; Закс Л., 1976). При этом могут сравниваться также вариационные ряды и с нормальным распределением признаков.

Методика расчета  $T$  требует попарного расположения значений признака в двух сравниваемых совокупностях (одна и та же группа из 6 – 25 единиц наблюдения до и после эксперимента). Первоначально необходимо найти разность величин каждого из параллельных определений. Отбрасываются пары, имеющие нулевое значение, с соответствующим уменьшением  $n$ . Затем эти разности надо ранжировать по восходящей градации и пронумеровать в естественном порядке числа, начиная с единицы. Если у нескольких разностей одна и та же величина, все они нумеруются числом, равным средней величине номеров, которыми следует их пометить. Знаки при ранжировании величин во внимание не принимаются. Так, например, величинам +2 и -2 присваивается равный ранг.

Каждый ранг отмечается знаком "+" или "-" в зависимости от значения полученной разности – отрицательной или положительной. Суммируются положительные и отрицательные ранги. Меньшая сумма рангов сравнивается с табличными значениями (табл. 5 приложения) для данного числа наблюдений и заданного порога надежности (для  $p = 0,05$  или  $p =$

0,01) при одностороннем или двустороннем критериях. Если меньшая сумма рангов меньше табличного значения, различия сравниваемых вариационных рядов достоверны.

**Пример.** Исследовалось систолическое артериальное давление у рабочих (табл. 13).

Таблица 13

**Показатели систолического артериального давления у рабочих ( $n = 8$ )**

<i>АД, мм. рт. ст.</i>		<i>Разность</i>	<i>Ранговый номер разности</i>
<i>до работы</i>	<i>после работы</i>		
110	120	-10	-4
130	160	-30	-7
120	125	-5	-2
125	130	-5	-2
130	130	0	-
115	140	-25	-6
140	135	5	2
140	120	20	5

*Порядок расчета:*

- определяем разность значений признака "до" и "после";

- не обращая внимания на знак, располагаем разности в восходящем порядке и нумеруем их. Первые три из них одинаковы по абсолютной величине -5. Они должны получить номера от 1 до 3. Согласно указанному выше правилу, эти разности получают один средний номер. Затем перед полученными таким образом номерами ставим алгебраический знак. Знак этот совпадает со знаком разностей;

- складываем номера с положительным знаком ( $2 + 5 = 7$ ). Меньший из итогов (7) оцениваем в соответствии с числом наблюдений по табл. 8 приложения. Табличное значение  $T$ , при котором для  $n = 8$  и одностороннего критерия различия достоверности ( $p = 0,05$ ), равно 6, а для двустороннего критерия – 4. Полученное значение (7) больше табличного. При данном числе наблюдений различия нельзя считать статистически значимыми ( $p < 0,05$ ).

### 11.3. Критерий согласия ( $\chi^2$ )

Статистическое сравнение двух выборок, изучаемый признак в которых представлен частотой этого признака в абсолютных, а не в относительных величинах, производится с помощью критерий согласия ( $\chi^2$ ) по формуле:

$$\chi^2 = \frac{(ad - bc) * (a + b + c + d)}{(a + b) * (c + d) * (a + c) * (b + d)}, \text{ где}$$

a, b, c, d – условные обозначения каждой клетки (пола) табл. 14.

Таблица 14

#### Результаты применения Тимогена с целью профилактики

Группы пациентов	Число обследованных		
	болели	не болели	итого
Принимали препарат	56 <i>a</i>	675 <i>b</i>	731 <i>a + b</i>
Не принимали препарат	272 <i>c</i>	11 396 <i>d</i>	11 668 <i>c + d</i>
Всего	329 <i>a + c</i>	18 155 <i>b + d</i>	18 488 <i>a + b + c + d = n</i>

Подставим в формулу числа из табл. 20. После соответствующих расчетов получим  $\chi^2 = 56,2$ .

При сравнении двух выборок с одним изучаемым в них признаком число степеней свободы равно 1. Вероятность ( $P$ ) для критерия согласия  $\chi^2$ , т.е. вероятность ошибочного (безошибочного) прогноза событий в зависимости от величины  $\chi^2$ , определяется по табл. 6 приложения.

В нашем примере ( $\chi^2 = 56,2$ ) выявлена существенная эффективность применения тимогена в профилактических целях, то есть мы тем самым отвергаем «нулевую» гипотезу – предложение об отсутствии связи между приемом этого препарата и снижением заболеваемости ( $p < 0,001$ ).

### 11.4. Точный метод Фишера

Простейшее альтернативное распределение, состоящее из двух градаций, часто встречается в медико-биологических исследованиях. В таких случаях широко используется точный метод Фишера (ТМФ), практически не имеющий ограничений при малых выборках.

Два альтернативных распределения с двумя градациями размещаются в таблице из 4 клеток. Она называется четырехпольной. Для оценки

различий в таких таблицах можно также использовать критерий согласия ( $\chi^2$ ). Однако мощность последнего ограничена при числе опытов в каждом из сравниваемых распределений менее 10. ТМФ для четырехпольной таблицы лишен этого ограничения, однако он требует громоздких вычислений. Чтобы их избежать, В.С.Генес (1964) разработал специальные таблицы. Общее число наблюдений ( $n$ ) в первом распределении может колебаться от 2 до 20, во втором – от 2 до 16.

Значительно уменьшенная трудоемкость ТМФ при использовании таблиц позволила применять его в любых двух независимых выборках. Для этого необходимо составить общий упорядоченный ряд из членов обеих выборок и найти его середину, по обе стороны от которой будет находиться одинаковое число членов. При четном числе членов середина будет находиться между двумя членами, при нечетном она придется на одно из чисел. Далее следует в каждой половине определить число членов, относящихся к первой и второй выборкам. Если по табл. 7 приложения эти соотношения окажутся существенно различными, можно сделать вывод о различиях между взятыми выборками в их средних тенденциях.

**Пример.** Частота сердечных сокращений после физической нагрузки у 4 спортсменов-разрядников составляла 72, 74, 78, 90, а у 6 физкультурников – 86, 94, 110, 115, 120, 124 уд/мин. Спрашивается, можно ли по такому небольшому числу наблюдений сделать вывод о существенных различиях в частоте сердечных сокращений у этих двух групп лиц.

*Составим общий упорядоченный ряд:*

72, 74, 78,	90	
86,	94, 110, 115, 120, 124	

Найдем середину ряда (она отмечена вертикальной чертой). В первой выборке ни один член не попал в правую половину упорядоченного ряда, в которой находятся 5 или 6 членов второй выборки. По табл. 17 приложения определяем, что эти различия значимы ( $p = 0,05$ ).

С помощью ТМФ можно выявить различия в форме распределений при отсутствии различий в средних тенденциях.

**Пример.** Пульс у студента до экзаменов составлял 88, 70, 58, 92, 90, 53, 80, 73, 60, 54, 99, а через час после экзамена: 83, 61, 62, 59, 83, 55, 57, 63, 66, 67 уд/мин.

*Составим ранжированный ряд:*

53, 54, 58, 60, 70, 73, 80, 88, 90, 92, 99  
55, 57, 59, 61, 62, 63, 65, 66, 67, 83, 83

Видно, что наиболее существенные различия между выборками состоят в ширине распределений: во второй выборке ни один из 11 ее членов не выходит за пределы, ограниченные числами 55 и 83, а в первой выборке 6 из 12 ее членов находятся вне этих пределов. По табл. 17 приложения определяем, что такие различия значимы ( $p = 0,05$ ).

### **11.5. Степени свободы**

Число степеней свободы отражает число наблюдений, которые свободны изменяться после определенных ограничений, наложенных на данные, оставляя при этом определенные важные характеристики выборки без изменений. Ограничения обуславливаются природной структурой данных.

Число степеней свободы для любой статистики равно количество независимых величин, используемых при вычислении статистики (объем выборки) минус количество параметров статистики, которые оцениваются с помощью них.

Число степеней свободы для дисперсии на 1 меньше объема выборки. Это соответствует количеству информации доступной в наборе данных для оценки генеральной дисперсии. Любая статистика, для которой не делается корректировка с учетом потери независимости или свободы, является смещенной оценкой соответствующего параметра генеральной совокупности. Это происходит потому, что законы теории вероятностей предполагают независимость наблюдений, т.е. допускают свободу в изменениях.

#### ***Методика вычисления:***

- в критериях согласия с  $n$  наблюдениями,  $df = n-1$ ;
- в критериях согласия с биномиальным или пуассоновским распределением с  $n$  наблюдениями,  $df = n-2$ ;
- в критериях согласия с нормальным распределением с  $n$  наблюдениями,  $df = n-3$ ;

- в критериях взаимосвязи при таблице сопряженности из  $m$  строк и  $n$  столбцов,  $df = (n-1)(m-1)$ .

**Вопросы для самоподготовки:**

1. Изучите основные понятия темы (вариационный ряд, средняя величина, среднее квадратическое отклонение, коэффициент вариации и др.).
2. Изучите виды вариационных рядов, методику их формирования.
3. Изучите методику расчета средних величин и критериев разнообразия вариационного ряда ( $M$ ,  $\sigma$ ,  $Cv$ ).
4. Изучите методику анализа средних величин: значение среднеквадратического отклонения и коэффициента разнообразия для оценки вариабельности изучаемого признака и типичности средней величины.
5. Что такое нормальное распределение вариационного ряда и его значение для оценки общественного здоровья и организации медицинской помощи?
6. Какова область применения характеристик вариационного ряда ( $M$ ,  $\sigma$ ,  $Cv$ )?
7. В каком случае применяют параметрические способы оценки достоверности результатов исследования?
8. Каковы условия применения параметрического способа оценки достоверности результатов исследования?
9. Методика определения ошибки репрезентативности средней величины и интенсивного показателя?
10. Дайте понятие о критерии « $t$ », его выбор в способе определения доверительных границ и оценку в способе достоверности разности результатов исследования.
11. Каковы условия применения непараметрических методов?
12. Какие методы непараметрической статистики применяются для сравнения двух независимых групп?
13. Какие методы непараметрической статистики применяются для сравнения двух зависимых групп?
14. В каком случае применяется точный метод Фишера?

## 12. МАТЕРИАЛЫ ДЛЯ САМОПОДГОТОВКИ

### 12.1. Тестовые задания

Предложенные тесты направлены на формирование следующих компетенций: ПК-4 и ПК-20 – у обучающихся по специальностям 31.05.01 «Лечебное дело» и 31.05.02 «Педиатрия»; ПК-4 и ПК-17 – по специальности 31.05.03 «Стоматология»; ОПК-7 – по специальности 32.05.01 «Медико-профилактическое дело».

*Выберите один правильный ответ*

**Тема: «Организация статистического исследования»**

#### 1. ПОД СТАТИСТИКОЙ ПОНИМАЮТ

- 1) самостоятельную общественную науку, изучающую количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной
- 2) сбор, обработку и хранение информации, характеризующей количественные закономерности общественных явлений
- 3) анализ массовых количественных данных с использованием статистических методов
- 4) анализ массовых количественных данных с использованием математических методов

#### 2. ПОД МЕДИЦИНСКОЙ СТАТИСТИКОЙ ПОНИМАЮТ

- 1) отрасль статистики, изучающей здоровье населения
- 2) совокупность статистических методов, необходимых для анализа деятельности медицинских организаций
- 3) отрасль статистики, изучающей вопросы, связанные с медициной, гигиеной, санитарией и здравоохранением
- 4) отрасль статистики, изучающей вопросы, связанные с медициной и социальной гигиеной

#### 3. ПРЕДМЕТОМ ИЗУЧЕНИЯ МЕДИЦИНСКОЙ СТАТИСТИКИ ЯВЛЯЕТСЯ

1. Информация о здоровье населения
2. Информация о влиянии факторов окружающей среды на здоровье человека
3. Информация о кадрах, деятельности медицинских организаций и служб здравоохранения



4. Информация о результатах клинических и экспериментальных исследований в медицине

- 1) верно 1,2,3
- 2) верно 1,3
- 3) верно 2,4
- 4) верно все перечисленное

4. ПЕРВЫЙ ЭТАП МЕДИКО-СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ

1. Группировку материала
2. Составление программы.
3. Составление плана исследования
4. Вычисление статистических показателей.
5. Анализ результатов
6. Выводы на основе результатов исследования

- 1) верно 1,4
- 2) верно 2,3
- 3) верно 5,6
- 4) верно все перечисленное

5. ВТОРОЙ ЭТАП МЕДИКО-СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ

- 1) составление программы и плана исследования
- 2) сбор данных
- 3) вычисление статистических показателей
- 4) анализ результатов
- 5) выводы на основе результатов исследования

6. ТРЕТИЙ ЭТАП МЕДИКО-СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ

1. Группировку материала
2. Составление программы
3. Составление плана исследования
4. Вычисление статистических показателей
5. Анализ результатов
6. Выводы на основе результатов исследования
7. Сводка данных в статистические таблицы

- 1) верно 2,3
- 2) верно 5,6
- 3) верно 1,4,7
- 4) верно все перечисленное

#### 7. ЧЕТВЕРТЫЙ ЭТАП МЕДИКО-СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ

1. Группировку материала
2. Составление программы
3. Составление плана исследования
4. Вычисление статистических показателей
5. Анализ результатов
6. Выводы на основе результатов исследования
7. Сводка данных в статистические таблицы
  - 1) верно 2,3
  - 2) верно 5,6
  - 3) верно 1,4,7
  - 4) верно все перечисленное

#### 8. ПРОГРАММА СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ

1. Определение единиц наблюдения
2. Статистическую регистрацию
3. Разработку статистического документа
4. Составление макетов статистических таблиц
5. Определения объема наблюдений
  - 1) верно 1,3,4
  - 2) верно 2
  - 3) верно 1,5
  - 4) верно все перечисленное

#### 9. ПЛАН СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ ВКЛЮЧАЕТ

1. Определение единиц наблюдения
2. Определение сроков проведения исследования
3. Разработку статистического документа
4. Составление макетов статистических таблиц
5. Определения объема наблюдений
  - 1) верно 1,3,4

- 2) верно 2,5
- 3) верно 1,5
- 4) верно все перечисленное

#### 10. К ОТНОСИТЕЛЬНЫМ ВЕЛИЧИНАМ ОТНОСЯТСЯ ПОКАЗАТЕЛИ

- 1. Интенсивные
- 2. Соотношения
- 3. Экстенсивные
- 4. Наглядности
- 5. Индексы

- 1) верно 1,2,3,4
- 2) верно 1,2,5
- 3) верно 4,5
- 4) верно все перечисленное

#### 11. ИНТЕНСИВНЫЕ ПОКАЗАТЕЛИ ОТРАЖАЮТ

- 1. Частоту
- 2. Уровень
- 3. Распространенность.
- 4. Удельный вес
- 5. Долю части по отношению к целому

- 1) верно 1,2, 3
- 2) верно 3,4
- 3) верно 1,2,3
- 4) верно все перечисленное

#### 12. ЭКСТЕНСИВНЫЕ ПОКАЗАТЕЛИ ОТРАЖАЮТ

- 1. Частоту
- 2. Динамику
- 3. Удельный вес
- 4. Долю части по отношению к целому

- 1) верно 1,2
- 2) верно 3,4
- 3) верно 1,2,3
- 4) верно все перечисленное

#### 13. К ПОКАЗАТЕЛЯМ СООТНОШЕНИЯ ОТНОСЯТСЯ

- 1. Уровень заболеваемости

2. Структура коечного фонда
3. Обеспеченность населения койками
4. Обеспеченность населения врачами
5. Динамика численности населения
  - 1) верно 1,2
  - 2) верно 3,4
  - 3) верно 1,2,3
  - 4) верно все перечисленное

14. ИНТЕНСИВНЫМИ ПОКАЗАТЕЛЯМИ ЯВЛЯЮТСЯ ПОКАЗАТЕЛИ

1. Материнской смертности
2. Рождаемости
3. Структуры материнской смертности
4. Обеспеченности кадрами
  - 1) верно 1,2
  - 2) верно 3
  - 3) верно 2,4
  - 4) верно все перечисленное

15. ОСНОВАНИЕМ ИНТЕНСИВНОГО ПОКАЗАТЕЛЯ МОЖЕТ БЫТЬ

1. 100
2. 1000
3. 10000
4. 100000
  - 1) верно 1
  - 2) верно 2,3
  - 3) верно 4
  - 4) верно все перечисленное

## 12.2. Ситуационные задачи

Предложенные для самоподготовки ситуационные задачи направлены на формирование следующих компетенций: ПК-4 и ПК-20 – у обучающихся по специальностям 31.05.01 «Лечебное дело» и 31.05.02 «Педиатрия»; ПК-4 и ПК-17 – по специальности 31.05.03 «Стоматология»; ОПК-7 – по специальности 32.05.01 «Медико-профилактическое дело».

**Задача 1.** В городе "Н" с численностью населения 50 000 человек, численность лиц в возрасте 70-79 лет составила 9845 человек; из их числа в течение года умерло 784 человека, в том числе 452 от ишемической болезни сердца (ИБС). В городе работает 190 врачей.

Рассчитать:

1. Повозрастной коэффициент (показатель) смертности у населения данной возрастной группы.
2. Долю лиц, умерших от ишемической болезни сердца.
3. Показатель обеспеченности врачами на 10 000 населения.

**Задача 2.** Рассчитать динамику рождаемости в показателях наглядности в городе Н. за 2012-2016гг. исходные данные представлены в таблице.

Динамика показателя рождаемости в городе Н. за 2012-2016гг.

Показатели	Годы				
	2012	2013	2014	2015	2016
Рождаемость (‰)	14,0	13,7	14,6	14,6	14,8

**Задача 3.** Вычислить стандартизированные показатели заболеваемости гепатитом рабочих двух цехов и, сравнив их с интенсивными показателями, оформить таблицу и сделать соответствующий вывод. Стандартизацию провести прямым методом. За стандарт принять состав рабочих по полу в обоих цехах.

Распределение больных гепатитом среди работников в цехах А и Б по полу

Пол	Цех А		Цех Б	
	Число рабочих	Из них больных	Число рабочих	Из них больных
Мужчины	50	1	170	4
Женщины	200	10	30	3
Оба пола	250	11	200	7

**Задача 4.** Определить размер и направление взаимосвязи между уровнем концентрации фтора в питьевой воде и числом лиц, пораженных флюорозом (в процентах к числу обследованных) (табл.) по коэффициенту корреляции Спирмена ( $\rho$  или  $r_s$ ).

Таблица

Распространенность флюороза среди населения, употребляющего воду с различным содержанием фтора

Содержание фтора в воде, мг/л	Процент пораженных флюорозом
x	y
Низкое 0,5	0,0
Оптимальное 1,0	3,0
Условно-допустимое 1,5	15,0
Повышенное 3,5	98,0
Недопустимое 5,0	98,0

**Задача 5.** Требуется определить, имеется ли зависимость между температурой тела и частотой пульса (условные данные) (табл.) по коэффициенту корреляции Пирсона ( $r_{xy}$ ).

Таблица

Зависимость между температурой тела и частотой сердечных сокращений

Температура тела, x	Частота пульса в минуту, y	$d_x$	$d_y$	$d_x * d_y$	$d_x^2$	$d_y^2$
36	60	-2	-20	40	4	400
36	70	-2	-10	20	4	100
38	80	0	0	0	0	0
40	90	+2	+10	20	4	100
40	100	+2	+20	40	4	100
$\Sigma x=190$ $M_x=38$	$\Sigma y=400$ $M_y=80$	$\Sigma d_x=0$	$\Sigma d_y=0$	$\Sigma d_x d_y =120$	$\Sigma d_x^2=16$	$\Sigma d_y^2=1000$

### 12.3. Эталоны ответов на тестовые задания

1.	1	6.	3	11.	1
2.	3	7.	2	12.	2
3.	4	8.	1	13.	2
4.	2	9.	2	14.	1
5.	2	10.	1	15.	4

### 12.4. Эталоны ответов к ситуационным задачам

#### Задача 1:

1. Коэффициент смертности населения в возрастной группе 70-79 лет рассчитывается:

$$\frac{784 \cdot 1000}{9845} = 79,6 \text{ ‰}$$

2. Доля лиц, умерших от ишемической болезни сердца среди умерших в возрастной группе 70-79 лет составит:

$$\frac{452}{784} \cdot 100 = 57,7\%$$

3. Расчет показателя обеспеченности врачами на 10 000 населения:

$$\frac{190 \times 10000}{50000} = 38,0 \text{ ‰}$$

Вывод. Коэффициент смертности населения в возрастной группе 70-79 лет составил 79,6‰. Доля лиц, умерших от ишемической болезни сердца, среди лиц, умерших в возрастной группе 70-79 лет составила 57,7%. Показатель обеспеченности врачами на 10 000 населения в городе составил 38.

**Задача 2.** Показатели наглядности (темп роста или снижения) по годам:

$$\begin{aligned} \text{в 2013г.} - \frac{13,7}{14,0} \cdot 100 = 97,9\% & \quad \text{в 2014г.} - \frac{14,6}{14,0} \cdot 100 = 104,3\% \\ \text{в 2015г.} - \frac{14,6}{14,0} \cdot 100 = 104,3\% & \quad \text{в 2016г.} - \frac{14,8}{14,0} \cdot 100 = 105,7\% \end{aligned}$$

Таблица

Динамика показателя рождаемости в городе Н. за 2012-2016гг.

Показатели	Годы				
	2012	2013	2014	2015	2016
Рождаемость (‰)	14,0	13,7	14,6	14,6	14,8
Показатель наглядности (%)	100,0	97,9	104,3	104,3	105,7

Вывод. В динамике за период наблюдения отмечался рост уровня рождаемости в городе Н. на 5,7%, с 14,0 ‰ в 2012г. до 14,8 ‰. В 2016г., показатель наглядности в 2016г. составил 105,7%.

**Задача 3.** Прямой метод, который состоит из 5 этапов.

I этап – расчет общих и специальных (по каждой группе – половой, возрастной и др.) интенсивных показателей (или средних величин) для двух сравниваемых совокупностей.

II этап – выбор и расчет стандарта.

III этап – расчет «ожидаемых величин» для каждой группы стандарта.

IV этап – определение стандартизованных показателей.

V этап – сравнение групп по общим интенсивным (или средним) и стандартизованным показателям. Выводы.

Целесообразно все эти расчетные операции представить в виде этапов стандартизации и оформить их в виде таблицы (табл.).

Таблица

Распределение больных гепатитом среди работников в цехах А и Б по полу

Пол	Цех А		Цех Б		I этап		II этап	III этап	
	Число рабочих	Из них больных	Число рабочих	Из них больных	% больных гепатитом		Стандарт (графа 1+графа 3)	«ожидаемое число» больных в группе стандарта	
					А	Б		А	Б
М	50	1	170	4	2,0	2,3	220	4,4	5,06
Ж	200	10	30	3	5,0	10,0	230	11,5	23,0
Итого	250	11	200	7	4,4	3,3	450	15,9	28,06
					IV этап		100	3,5%	6,2%

На I этапе требуется определить, как часто среди рабочих мужчин и женщин цеха А и цеха Б встречаются лица, страдающие гепатитом.



Для этого вычисляют процент больных гепатитом мужчин и женщин отдельно. Расчет ведется следующим образом: число больных надо разделить на соответствующее число рабочих и умножить на 100.

В цехе А больных гепатитом мужчин  $\frac{1}{50} \cdot 100 = 2,0\%$ , а больных гепатитом женщин  $\frac{10}{200} \cdot 100 = 5,0\%$ . Всего же больных гепатитом среди рабочих обоего пола было  $-\frac{11}{250} \cdot 100 = 4,4\%$ . Подобные же расчеты надо сделать для цеха Б.

Следует проанализировать интенсивные показатели по цеху А и по цеху Б. Из табл. 1 (I этап) видно, что у мужчин и женщин цеха А реже встречается гепатит, чем у мужчин и женщин цеха Б, хотя в целом по строке («итого») в цехах А и Б отмечено противоположное соотношение показателей: общий процент больных больше в цехе А, а не в цехе Б. Этот факт объясняется тем, что на общие показатели по каждому цеху оказал влияние разный по полу состав рабочих в этих цехах. Так, в цехе А в составе рабочих было больше женщин (а среди женщин чаще встречаются больные гепатитом, чем среди мужчин). В цехе Б большую часть рабочих составляли мужчины, которые реже страдают этим заболеванием. Следовательно, в данном случае различие в общих итоговых интенсивных показателях, полученных по цехам, может быть связано с неоднородностью полового состава работающих в этих цехах.

*II этап.* Выбор и расчет стандарта. За стандарт следует принимать тот состав совокупностей, в котором отразились бы все особенности состава сравниваемых групп. В нашем примере за такой стандарт принята суммарная численность работавших в цехах А и Б (графа 1 + графа 3). Стандартом для группы мужчин будет 220 (50 + 170), а для женщин 230 (200+30). Общее число рабочих в двух цехах составит 450. Зная интенсивные показатели (I этап) и стандарт (II этап), можно перейти к III этапу – определению «ожидаемого числа» больных гепатитом в каждой группе стандарта.

*III этап.* Определение «ожидаемых чисел» больных в каждой группе стандарта. Порядок вычисления можно показать на следующем примере. В цехе А мужчин, больных гепатитом, 2%. Сколько было бы больных

мужчин среди 220 человек при данном уровне заболеваемости? Составим пропорцию и произведем расчет:

$$\frac{100 - 2\%}{220 - x} = \frac{2 \cdot 220}{100} = 4,4 \quad \text{больного}$$

Таким образом, 4,4 – «ожидаемое число» мужчин, больных гепатитом, из 220 мужчин группы стандарта, уровень заболеваемости которых такой, как у рабочих цеха А

Аналогичные расчеты произведем по цеху Б. По таблице видно, что «ожидаемое» число мужчин, больных гепатитом в цехе Б с характерным для них уровнем заболеваемости, если их будет 220 человек составил 5,06 человек. Такой же принцип расчета «ожидаемых чисел» больных в группе стандарта для женщин цехов А и Б. После расчета всех ожидаемых чисел переходим к IV этапу

*IV этап.* Определение стандартизованных показателей. Просуммировав абсолютные «ожидаемые» числа больных гепатитом мужчин и женщин по каждому цеху отдельно, полученные суммы относят к общей численности рабочих, указанной в итоговой строке стандарта. Это отношение выражают в процентах и получают стандартизованные показатели. Так, в цехе А сумма «ожидаемых» чисел больных составит:

$$4,4 + 11,5 = 15,9.$$

Стандартизованный показатель заболеваемости в цехе А равен:

$$\frac{15,9 \cdot 100}{450} = 3,5\% .$$

Аналогично рассчитывается стандартизованный показатель по цеху Б.

*V этап.* Сравнение уровней заболеваемости гепатитом рабочих цеха А и цеха Б по стандартизованным и общим интенсивным показателям. Для этого следует выписать стандартизованные и интенсивные показатели в отдельную аналитическую таблицу (табл.), сопоставить эти данные и сделать общий вывод.

Таблица

Сопоставление уровней заболеваемости гепатитом рабочих в цехах А и Б (в %)

Показатели	Цех А	Цех Б	Результаты сравнения
Стандартизованные	3,5	6,2	А<Б
Общие интенсивные (грубые)	4,4	3,3	А>Б

Вывод: сравнение стандартизованных показателей, рассчитанных для работавших в цехах А и Б, позволяет сделать заключение, что при одинаковом составе рабочих по полу в цехах А и Б показатель заболеваемости гепатитом рабочих в цехе А значительно ниже, чем в цехе Б. При сравнении общих (грубых) интенсивных показателей результаты получились противоположными (А>Б) в связи с тем, что на общие интенсивные показатели оказал влияние разный состав рабочих по полу в этих цехах. Произошло это от того, что у женщин значительно выше заболеваемость гепатитом, чем у мужчин, а в цехе А преобладали женщины, тогда как в цехе Б – мужчины.

#### Задача 4.

Распространенность флюороза среди населения, употребляющего воду с различным содержанием фтора

Содержание фтора в воде, мг/л	Процент пораженных флюорозом	Порядковый номер (ранг)		Разность рангов	Квадрат разности рангов
		$x_1$	$y_1$		
$x$	$y$	$x_1$	$y_1$	$d (x_1 - y_1)$	$d^2$
Низкое 0,5	0,0	1	1	0	0,0
Оптимальное 1,0	3,0	2	2	0	0,0
Условно-допустимое 1,5	15,0	3	3	0	0,0
Повышенное 3,5	98,0	4	4,5	-0,5	0,25
Недопустимое 5,0	98,0	5	4,5	+0,5	0,25
					$\Sigma d^2 = 0,5$

1. Присваиваем каждому значению X и Y ранг в порядке возрастания.
2. Находим разность рангов –  $d (x_1 - y_1)$ .
3. Вычисляем квадрат разности рангов –  $d^2$  и суммируем –  $\Sigma d^2$
4. Рассчитываем коэффициент корреляции Спирмена по формуле:

$$\rho = 1 - \frac{6 \cdot \Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 0,5}{5(5^2 - 1)} = 1 - \frac{3,0}{120} = +0,975 \text{ (связь сильная и прямая);}$$

$$m_p = \sqrt{\frac{1 - \rho^2}{n - 2}} = \sqrt{\frac{1 - 0,975^2}{5 - 2}} = \pm 0,128;$$

$$t = \frac{\rho}{m_p} = \frac{0,972}{0,128} = 7,62, \text{ что соответствует вероятности достоверности}$$

более 99%

Вывод: между уровнем концентрации фтора в питьевой воде и числом лиц, пораженных флюорозом, наблюдается статистически значимая

прямая и сильная связь, так как коэффициент ранговой корреляции составил  $r = +0,975$ , уровень значимости  $p < 0,01$ . Таким образом, с увеличением содержания фтора в воде увеличивается распространенность флюороза среди населения.

### Задача 5.

Зависимость между температурой тела и частотой сердечных сокращений

Температура тела, x	Частота пульса в минуту, y	$d_x$	$d_y$	$d_x \cdot d_y$	$d_x^2$	$d_y^2$
36	60	-2	-20	40	4	400
36	70	-2	-10	20	4	100
38	80	0	0	0	0	0
40	90	+2	+10	20	4	100
40	100	+2	+20	40	4	100
$\Sigma x = 190$ $M_x = 38$	$\Sigma y = 400$ $M_y = 80$	$\Sigma d_x = 0$	$\Sigma d_y = 0$	$\Sigma d_x d_y = 120$	$\Sigma d_x^2 = 16$	$\Sigma d_y^2 = 1000$

1. Построить вариационные ряды из парных признаков  $x$  и  $y$ .
2. Определить их средние величины –  $M_x$  и  $M_y$ .
3. Найти  $d$  – отклонение каждой варианты от средней для ряда  $x$  ( $d_x = x - M_x$ ) и для ряда  $y$  ( $d_y = y - M_y$ ).
4. Полученные отклонения перемножить ( $d_x \cdot d_y$ ) и просуммировать ( $\Sigma d_x \cdot d_y$ ).
5. Каждое отклонение возвести в квадрат и суммировать по ряду  $x$  –  $\Sigma d_x^2$  и по ряду  $y$  –  $\Sigma d_y^2$ .
6. Определить произведение  $\Sigma d_x^2 \cdot \Sigma d_y^2$  и из произведения

$$\text{Рассчитать } r_{xy} \text{ по формуле: } r_{xy} = \frac{\Sigma d_x \cdot d_y}{\sqrt{\Sigma d_x^2 \cdot \Sigma d_y^2}} = \frac{120}{\sqrt{16 \cdot 1000}} = \frac{120}{126,5} = 0,952.$$

$$m_r \sqrt{\frac{1 - 0,952^2}{5 - 2}} = \pm 0,19$$

$$t = \frac{0,952}{0,19} = 5,0 \text{ что соответствует вероятности достоверности более}$$

99%.

Вывод: между температурой тела и частотой сердечных сокращений выявлена статистически значимая сильная прямая корреляционная связь, так как коэффициент корреляции составил  $r = + 0,952$ , уровень значимости  $p < 0,01$ . Таким образом, с увеличением температуры тела увеличивается частота сердечных сокращений.

## 13. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

### *Основная*

1. Лисицын, Ю.П. Общественное здоровье и здравоохранение [Электронный ресурс] : учебник для студ. мед. вузов / Ю.П. Лисицын. – 3-е изд., испр. и доп. – Электрон. текстовые дан. – М.: Гэотар Медиа, 2015. – 544 с. – Режим доступа: <http://www.studmedlib.ru/ru/book/ISBN9785970432914.html>
2. Медик, В.А. Общественное здоровье и здравоохранение [Электронный ресурс] : учебник / В.А. Медик, В.И. Лисицын. – 4-е изд., перераб. и доп. – Электрон. текстовые дан. – М. : Гэотар Медиа, 2016. – on-line. – Режим доступа:  
<http://www.studmedlib.ru/book/ISBN9785970437018.html>

### *Дополнительная*

1. Новосёлов, М. Медицинская статистика, 2013. – Режим доступа: <http://cyberleninka.ru/article/n/meditsinskaya-statistika>
2. Омельченко, В. П., Демидова А. А. Медицинская информатика. – М.: ГЭОТАР-МЕДИА, 2016. – 527 с.
3. Электронная учебная библиотека ФГБОУ ВО БГМУ Минздрава России  
<http://library.bashgmu.ru>

### *Использованная литература*

1. Банерджи, А. Медицинская статистика понятным языком: вводный курс: [пер. с англ.]; под ред. В.П. Леонова. – М.: Практическая медицина, 2014. – 287 с.
2. Трухачёва, Н.В. Медицинская статистика: учеб. Пособие / Н.В. Трухачёва. – Ростов-на-Дону: Феникс, 2017. – 324 с.
3. Петри, А. Наглядная медицинская статистика: учеб. пособие / А.Петри, К. Сэбин. – М.: ГЭОТАР-Медиа, 2015. – 215 с.

## 14. Приложения

Таблица 1


### Стандартные значения критерия $t$ (Критерий Стьюдента)

n	0,05	0,01	0,001	n	0,05	0,01	0,001
1	12,71	63,66		21	2,08	2,83	3,82
2	4,30	9,93	31,60	22	2,07	2,82	3,79
3	3,18	5,84	12,94	23	2,07	2,81	3,77
4	2,78	4,60	8,61	24	2,06	2,80	3,75
5	2,57	4,03	6,86	25	2,06	2,79	3,73
6	2,45	3,71	5,96	26	2,06	2,78	3,71
7	2,36	3,50	5,41	27	2,05	2,77	3,69
8	2,31	3,36	5,04	28	2,05	2,76	3,67
9	2,26	3,25	4,78	29	2,04	2,76	3,66
10	2,23	3,17	4,59	30	2,04	2,75	3,65
11	2,20	3,11	4,44	40	2,02	2,70	3,55
12	2,18	3,06	4,32	50	2,01	2,68	3,50
13	2,16	3,01	4,22	60	2,00	2,66	3,46
14	2,15	2,98	4,14	80	1,99	2,64	3,42
15	2,13	2,95	4,07	100	1,98	2,63	3,39
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,90	3,97	200	1,97	2,60	3,34
18	2,10	2,88	3,92	500	1,96	2,59	3,31
19	1,96	2,86	3,88	$\infty$	1,96	2,58	3,29
20	2,09	2,85	3,85				
	5%	1%	0,1%		5%	1%	0,1%

Таблица 2а

**Стандартизация коэффициентов смертности от злокачественных новообразований  
в городах "А" и "Б" ("прямой метод")**

Возрастные группы	I этап						III этап
	Город "А"			Город "Б"			"ожидаемое" число умерших при фактической по возрастной смертности в городе "Б" и возрастном составе в городе "А" (стандарт) на 100000 населения
	численность населения (тыс.)	число умерших	смертность от злокачественных новообразований	численность населения (тыс.)	число умерших	смертность от злокачественных новообразований	
До 30 лет	350	14	4.0	750	45	6.0	$\frac{6 \cdot 350000}{100000} = 21.0$
30-39 лет	100	25	25.0	120	36	30.0	$\frac{30 \cdot 100000}{100000} = 30.0$
40-49 лет	95	114	120.0	125	175	140.0	$\frac{140 \cdot 95000}{100000} = 133.0$
50-59 лет	75	240	320.0	95	361	380.0	$\frac{380 \cdot 75000}{100000} = 285.0$
60 лет и старше	80	544	680.0	110	825	750.0	$\frac{750 \cdot 80000}{100000} = 600.0$
ВСЕГО	700	937	133.9	1200	1442	120.0	1069.0

 – стандарт

IV этап:  
1069.0 – 700000

$x_{ст} = 100000$

$$x_{ст} = \frac{1069.0 \cdot 100000}{700000} = 152.7\% \text{ ооо}$$

**Стандартизация коэффициентов смертности от злокачественных новообразований в городах N и M.  
Косвенный метод (числа условные)**

Возрастные группы	I этап		II этап		
	Смертность населения в гор. "С" на 100000 чел. (стандарт)	Численность населения		"Ожидаемые" числа умерших от злокачественных новообразований	
		Город N	Город M	Город N	Город M
до 30 лет	5.0	280000	275000	$\frac{280000 \cdot 5.0}{100000} = 14.0$	$\frac{275000 \cdot 5.0}{100000} = 13.8$
30-39 лет	32.0	90000	78000	$\frac{90000 \cdot 32.0}{100000} = 28.8$	$\frac{78000 \cdot 32.0}{100000} = 24.9$
40-49 лет	130.0	75000	56000	$\frac{75000 \cdot 130.0}{100000} = 97.5$	$\frac{56000 \cdot 130.0}{100000} = 72.8$
50-59 лет	360.0	70000	51000	$\frac{70000 \cdot 360.0}{100000} = 252.0$	$\frac{51000 \cdot 360.0}{100000} = 183.6$
60 лет и старше	730.0	65000	40000	$\frac{65000 \cdot 730.0}{100000} = 474.0$	$\frac{40000 \cdot 730.0}{100000} = 292.0$
ВСЕГО	125.0	580000	500000	-	-
Умерло от злокачественных новообразований		754	590	866.8 III этап	587.1 III этап
Смертность на 100000 населения		130.0	118.0	$\frac{754 \cdot 125}{866.6} = 108.7$	$\frac{590 \cdot 125}{587.1} = 125.6$



**Стандартизация коэффициентов смертности от злокачественных новообразований в городе N.  
Обратный метод (числа условные)**

Возрастные группы	I этап	II этап			
	Повозрастные коэффициенты смертности от злокачественных новообразований на 100000 населения, принятого за стандарт	2006г.		2016 г.	
		Число умерших от злокачественных новообразований в данном возрасте	"Ожидаемая" численность населения	Число умерших от злокачественных новообразований в данном возрасте	"Ожидаемая" численность населения
До 30 лет	4.0	21	525000	18	450000
30-39 лет	35.0	44	125714	36	102857
40-49 лет	132.0	156	110606	181	136364
50-59 лет	354.0	221	62469	278	78523
60 лет и старше	722.0	482	66759	558	72280
<b>ВСЕГО</b>	<b>121.0</b>	<b>924</b>	<b>890548</b>	<b>1071</b>	<b>840024</b>

## Критическое значение для критерия Манна и Уитни (по Л. Заксу, 1976)

Односторонний критерий:  $p = 0,05$ двусторонний критерий:  $p = 0,1$ 

n <sub>1</sub>	n <sub>2</sub>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	-	-																		
3	-	-	0																	
4	-	-	0	1																
5	-	0	1	2	4															
6	-	0	2	3	5	7														
7	-	0	2	4	6	8	11													
8	-	1	3	5	8	10	13	15												
9	-	1	4	6	9	12	15	18	21											
10	-	1	4	7	11	14	17	20	24	27										
11	-	1	5	8	12	16	19	23	27	31	34									
12	-	2	5	9	13	17	21	26	30	34	38	42								
13	-	2	6	10	15	19	24	28	33	37	42	47	51							
14	-	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	-	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	-	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	-	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	-	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

**Односторонний критерий:  $p = 0,05$**   
**двусторонний критерий:  $p = 0,1$**

n <sub>1</sub>	n <sub>2</sub>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	0	5	11	19	26	34	41	49	57	65	73	81	89	97	105	113	121	130	138	146
22	0	5	12	20	28	36	44	52	60	68	77	85	94	102	111	119	128	136	145	154
23	0	5	13	21	29	37	46	54	63	72	81	90	98	107	116	125	134	143	152	161
24	0	6	13	22	30	39	48	57	66	75	85	94	103	113	122	131	141	150	160	169
25	0	6	14	23	32	41	50	60	69	79	89	98	108	118	128	137	147	157	167	177
26	0	6	15	24	33	43	53	62	72	82	92	103	113	123	133	143	154	164	174	185
27	0	7	15	25	35	45	55	65	75	86	96	107	117	128	139	149	160	171	182	192
28	0	7	16	26	36	46	57	68	78	89	100	111	122	133	144	156	167	178	189	200
29	0	7	17	27	38	48	59	70	82	93	104	116	127	138	150	162	173	185	196	208
30	0	7	17	28	39	50	61	73	85	96	108	120	132	144	156	168	180	192	204	216

Продолжение табл. 3

**Односторонний критерий:  $p = 0,025$**

**двусторонний критерий:  $p = 0,05$**

n <sub>1</sub>	n <sub>2</sub>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	-	-																		
4	-	-	-	0																
5	-	-	0	1	2															
6	-	-	1	2	3	5														
7	-	-	1	3	5	6	8													
8	-	0	2	4	6	8	10	13												
9	-	0	2	4	7	10	12	15	17											
10	-	0	3	5	8	11	14	17	20	23										
11	-	0	3	6	9	13	16	19	23	26	30									
12	-	1	4	7	11	14	18	22	26	29	33	37								
13	-	1	4	8	12	16	20	24	28	33	37	41	45							
14	-	1	5	9	13	17	22	26	31	36	40	45	50	55						
15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64					
16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75				
17	-	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87			
18	-	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99		
19	-	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	
20	-	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
21	-	3	8	15	22	29	36	43	50	58	65	73	80	88	96	103	111	119	126	134
22	-	3	9	16	23	30	38	45	53	61	69	77	85	93	101	109	117	125	133	141

Односторонний критерий:  $p = 0,025$ двусторонний критерий:  $p = 0,05$ 

n <sub>1</sub>	n <sub>2</sub>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
23	-	3	9	17	24	32	40	48	56	64	73	81	89	98	106	115	123	132	140	149
24	-	3	10	17	25	33	42	50	59	67	76	85	94	102	111	120	129	138	147	156
25	-	3	10	18	27	35	44	53	62	71	80	89	98	107	117	126	135	145	154	163
26	-	4	11	19	28	37	46	55	64	74	83	93	102	112	122	132	141	151	161	171
27	-	4	11	20	29	38	48	57	67	77	87	97	107	117	127	137	147	158	168	178
28	-	4	12	21	30	40	50	60	70	80	90	101	111	122	132	143	154	164	175	186
29	-	4	13	22	32	42	52	62	73	83	94	105	116	127	138	149	160	171	182	193
30	-	5	13	23	33	43	54	65	76	87	98	109	120	131	143	154	166	177	189	200
31	-	5	14	24	34	45	56	67	78	90	101	113	125	136	148	160	172	184	196	208
32	-	5	14	24	35	46	58	69	81	93	105	117	129	141	153	166	178	190	203	215
33	-	5	15	25	37	48	60	72	84	96	108	121	133	146	159	171	184	197	210	222
34	-	5	15	26	38	50	62	74	87	99	112	125	138	151	164	177	190	203	217	230
35	-	6	16	27	39	51	64	77	89	103	116	129	142	156	169	183	196	210	224	237
36	-	6	16	28	40	53	66	79	92	106	119	133	147	161	174	188	202	216	231	245
37	-	6	17	29	41	55	68	81	95	109	123	137	151	165	180	194	209	223	238	252
38	-	6	17	30	43	56	70	84	98	112	127	141	156	170	185	200	215	230	245	259
39	0	7	18	31	44	58	72	86	101	115	130	145	160	175	190	206	221	236	252	267
40	0	7	18	31	45	59	74	89	103	119	134	149	165	180	196	211	227	243	258	274

Односторонний критерий:  $p = 0,005$ двусторонний критерий:  $p = 0,01$ 

n <sub>1</sub>	n <sub>2</sub>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	-	-																		
3	-	-	-																	
4	-	-	-	-																
5	-	-	-	-	0															
6	-	-	-	0	1	2														
7	-	-	-	0	1	3	4													
8	-	-	-	1	2	4	6	7												
9	-	-	0	1	3	5	7	9	11											
10	-	-	0	2	4	6	9	11	13	16										
11	-	-	0	2	5	7	10	13	16	18	21									
12	-	-	1	3	6	9	12	15	18	21	24	27								
13	-	-	1	3	7	10	13	17	20	24	27	31	34							
14	-	-	1	4	7	11	15	18	22	26	30	34	38	42						
15	-	-	2	5	8	12	16	20	24	29	33	37	42	46	51					
16	-	-	2	5	9	13	18	22	27	31	36	41	45	50	55	60				
17	-	-	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70			
18	-	-	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81		
19	-	0	3	7	12	17	22	28	33	39	45	51	57	63	69	74	81	87	93	
20	-	0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105
21	-	0	3	8	14	19	25	32	38	44	51	58	64	71	78	84	91	98	105	112
22	-	0	4	9	14	21	27	34	40	47	54	61	68	75	82	89	96	104	111	118

**Односторонний критерий:  $p = 0,005$**

**двусторонний критерий:  $p = 0,01$**

n <sub>1</sub>	n <sub>2</sub>																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
23	-	0	4	9	15	22	29	35	43	50	57	64	72	79	87	94	102	109	117	125
24	-	0	4	10	16	23	30	37	45	52	60	68	75	83	91	99	107	115	123	131
25	-	0	5	10	17	24	32	39	47	55	63	71	79	87	96	104	112	121	129	138
26	-	0	5	11	18	25	33	41	49	58	66	74	83	92	100	109	118	127	135	144
27	-	1	5	12	19	27	35	43	52	60	69	78	87	96	105	114	123	132	142	151
28	-	1	5	12	20	28	36	45	54	63	72	81	91	100	109	119	123	138	148	157
29	-	1	6	13	21	29	38	47	56	66	75	85	94	104	114	124	134	144	154	164
30	-	1	6	13	22	30	40	49	58	68	78	88	98	108	119	129	149	150	160	170

**Граничные значения  $t$  (критерия Стьюдента)  
(по Л. Заксу, 1976)**

Число степеней свободы (k)	Вероятность ошибки (p) для двустороннего критерия				
	0,10	0,05	0,02	0,01	0,002
1	6,314	12,706	31,821	63,657	318,309
2	2,920	4,303	6,965	9,925	22,327
3	2,353	3,182	4,541	5,841	10,214
4	2,132	2,776	3,747	4,604	7,173
5	2,015	2,571	3,365	4,032	5,893
6	2,943	2,447	3,143	3,707	5,208
7	2,895	2,365	2,908	3,499	4,785
8	2,860	2,306	2,896	3,355	4,501
9	2,833	2,262	2,821	3,250	4,397
10	2,812	2,228	2,764	3,169	4,144
11	2,796	2,201	2,718	3,106	4,025
12	2,782	2,179	2,681	3,055	3,930
13	2,771	2,160	2,650	3,012	3,852
14	2,761	2,145	2,624	2,977	3,787
15	2,753	2,131	2,602	2,947	3,733
16	2,746	2,120	2,583	2,921	3,686
17	2,740	2,110	2,567	2,898	3,646
18	2,734	2,101	2,552	2,878	3,610
19	2,729	2,093	2,539	2,861	3,579
20	2,725	2,086	2,528	2,845	3,552
Число степеней свободы (k)	0,05	0,025	0,01	0,005	0,001

*Вероятность ошибки (p) для одностороннего критерия*



Продолжение табл. 4

Число степеней свободы (k)	Вероятность ошибки (p) для двустороннего критерия				
	0,10	0,05	0,02	0,01	0,002
21	1,721	2,080	2,518	2,831	3,527
22	1,717	2,074	2,508	2,814	3,505
23	1,714	2,069	2,500	2,807	3,485
24	1,711	2,064	2,492	2,797	3,467
25	1,708	2,060	2,485	2,787	3,450
26	1,706	2,056	2,479	2,779	3,435
27	1,703	2,052	2,473	2,771	3,421
28	1,701	2,048	2,467	2,763	3,408
29	1,699	2,045	2,462	2,756	3,396
30	1,697	2,042	2,457	2,750	3,385
35	1,690	2,030	2,438	2,724	3,340
40	1,684	2,021	2,423	2,704	3,307
45	1,679	2,014	2,412	2,690	3,281
50	1,676	2,009	2,403	2,678	3,261
60	1,671	2,000	2,390	2,660	3,232
70	1,667	1,994	2,381	2,648	3,211
80	1,664	1,990	2,374	2,639	3,195
90	1,662	1,987	2,368	2,632	3,183
100	1,660	1,984	2,364	2,626	3,174
120	1,658	1,980	2,358	2,617	3,160
200	1,653	1,972	2,345	2,601	3,131
500	1,648	1,965	2,334	2,586	3,107
$\infty$	1,645	1,960	2,236	2,576	3,090
Число степеней свободы	0,05	0,025	0,01	0,005	0,001

Вероятность ошибки (p) для одностороннего критерия

**Граничные значения критерия Т (Вилкоксона)  
(по I.Adem, 1971)**

n	Вероятность ошибки (p) для двустороннего критерия			
	0,1	0,05	0,02	0,01
5	1	-	-	-
6	3	1	-	-
7	4	3	1	-
8	6	4	2	1
9	9	6	4	2
10	11	9	6	4
11	14	11	8	6
12	18	14	10	8
13	22	18	13	10
14	26	22	16	13
15	31	26	20	16
16	36	30	24	20
17	42	35	29	24
18	48	41	33	28
19	54	47	38	33
20	61	53	44	38
21	68	59	50	43
22	76	66	57	49
23	84	74	63	55
24	92	82	70	62
25	101	90	77	69
n	0,05	0,025	0,01	0,005

*Вероятность ошибки (p) для одностороннего критерия*

Граничные значения  $\chi^2$  для одной степени свободы.При одностороннем критерии  $p$  – значения уменьшить вдвое (по Л. Заксу, 1976)

$\chi^2$	$p$	$\chi^2$	$p$	$\chi^2$	$p$	$\chi^2$	$p$	$\chi^2$	$p$
0	1,00000	2,1	0,14730	4,0	0,04550	6,0	0,01431	8,0	0,00468
0,1	0,75183	2,2	0,13801	4,1	0,04288	6,1	0,01352	8,1	0,00443
0,2	0,65472	2,3	0,12937	4,2	0,04042	6,2	0,01278	8,2	0,00419
0,3	0,58388	2,4	0,12134	4,3	0,03811	6,3	0,01207	8,3	0,00396
0,4	0,52709	2,5	0,11385	4,4	0,03594	6,4	0,01141	8,4	0,00375
0,5	0,47950	2,6	0,10686	4,5	0,03389	6,5	0,01079	8,5	0,00355
0,6	0,43858	2,7	0,10035	4,6	0,03197	6,6	0,010,20	8,6	0,00336
0,7	0,40278	2,8	0,09429	4,7	0,03016	6,7	0,00964	8,7	0,00318
0,8	0,37109	2,9	0,8858	4,8	0,02846	6,8	0,00912	8,8	0,00301
0,9	0,34278	3,0	0,08326	4,9	0,02686	6,9	0,00862	8,9	0,00285
1,0	0,31731	3,1	0,07829	5,0	0,02535	7,0	0,00815	9,0	0,00270
1,1	0,29427	3,2	0,07364	5,1	0,02393	7,1	0,00771	9,1	0,00256
1,2	0,27332	3,3	0,06928	5,2	0,02259	7,2	0,00729	9,2	0,00242
1,3	0,25421	3,3	0,06928	5,3	0,02133	7,3	0,00690	9,3	0,00229
1,4	0,23672	3,4	0,06320	5,4	0,02014	7,4	0,00652	9,4	0,00217
1,5	0,22067	3,5	0,06137	5,5	0,01902	7,5	0,00617	9,5	0,00205
1,6	0,20590	3,6	0,05778	5,6	0,01796	7,6	0,00584	9,6	0,00195
1,7	0,19229	3,7	0,05441	5,7	0,01697	7,7	0,00552	9,7	0,00184
1,8	0,17971	3,8	0,05125	5,8	0,01603	7,8	0,00522	9,8	0,00174
1,9	0,16808	3,9	0,04829	5,9	0,01514	7,9	0,00494	9,9	0,00165
2,0	0,15730	4,0	0,04550	6,0	0,01431	8,0	0,00468	10,0	0,00157

## Точный метод Фишера (ТМФ).

Таблицы для малого числа наблюдений ( $n_1 \leq 20, n_2 \leq 16$ )

Число однонаправленных эффектов  $K$  к первой ( $K_1$ ) и второй ( $K_2$ ) выборках,  
при которых различия между ними можно считать существенными (односторонний критерий)

n <sub>1</sub>	K <sub>1</sub>	n <sub>2</sub>																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		K <sub>2</sub>																		
		p=0,05																		
2	0	-	-	-	5	6	7	8	9	9	10	11	12	12	13	14	15	16	16	17
3	0	-	3	4	5	5	6	7	7	8	9	9	10	11	11	12	12	13	14	14
	1	-	-	-	-	-	-	-	9	10	11	12	13	14	15	15	16	17	18	19
4	0	-	3	4	4	5	5	6	6	7	8	8	9	9	10	10	11	11	12	12
	1	-	-	-	5	6	7	8	9	9	10	11	12	12	13	14	15	15	16	17
	2	-	-	-	-	-	-	-	-	-	-	12	13	14	15	16	17	18	19	20
5	0	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	9	10	10	11
	1	-	-	4	5	6	6	7	8	8	9	10	10	11	12	12	13	14	14	15
	2	-	-	-	-	-	7	8	9	10	11	12	12	13	14	15	16	16	17	18
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	17	18	19	20
6	0	2	3	3	4	4	4	5	5	6	6	6	7	7	8	8	8	9	9	10
	1	-	3	4	5	5	5	6	7	8	8	9	9	10	11	11	12	12	13	13
	2	-	-	-	5	6	6	8	8	9	10	11	11	12	13	13	14	15	16	16
	3	-	-	-	-	-	-	-	9	10	11	12	13	14	15	15	16	17	18	19

n <sub>1</sub>	K <sub>1</sub>	n <sub>2</sub>																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		K <sub>2</sub>																		
		p=0,05																		
7	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	19	20
	0	2	3	3	3	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9
	1	-	3	4	4	5	5	6	7	7	8	8	9	9	10	10	11	11	12	12
	2	-	-	4	5	6	6	7	8	8	9	10	10	11	12	12	13	14	14	15
	3	-	-	-	-	6	7	8	9	10	10	11	12	13	13	14	15	16	17	17
8	4	-	-	-	-	-	-	-	-	-	11	12	13	14	15	16	17	17	18	19
	0	2	3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8
	1	-	3	4	4	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11
	2	-	-	4	5	5	6	7	7	8	9	9	10	10	11	11	12	13	13	14
	3	-	-	-	5	6	7	8	8	9	10	10	11	12	13	13	14	15	15	16
9	4	-	-	-	-	-	-	8	9	10	11	12	12	13	14	15	16	16	17	18
	5	-	-	-	-	-	-	-	-	-	12	13	14	15	16	17	18	19	20	20
	0	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7	7
	1	-	3	4	4	4	5	5	6	6	7	7	7	8	8	9	9	10	10	10
	2	-	3	4	5	5	6	6	7	7	8	9	9	10	10	11	11	12	12	13
10	3	-	-	4	5	6	7	7	8	9	9	10	10	11	12	12	13	14	14	15
	4	-	-	-	-	6	7	8	9	9	10	11	12	12	13	14	15	15	16	17
	5	-	-	-	-	-	-	-	9	10	11	12	13	14	14	15	16	17	18	19
	6	-	-	-	-	-	-	-	-	-	-	-	-	14	15	16	17	18	19	20
	0	2	2	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7	7
	1	2	3	3	4	4	5	5	5	6	6	7	7	7	8	8	9	9	9	10
	2	-	3	4	4	5	5	6	7	7	8	8	9	9	10	10	11	11	12	12
	3	-	-	4	5	6	6	7	7	8	9	9	10	10	11	12	12	13	13	14
4	-	-	-	5	6	7	8	8	9	10	10	11	12	12	13	14	14	15	16	
5	-	-	-	-	-	7	8	9	10	11	11	12	13	14	14	15	16	17	17	
6	-	-	-	-	-	-	-	-	10	11	12	13	14	15	16	16	17	19	19	

Продолжение табл. 7

n <sub>1</sub>	K <sub>1</sub>	n <sub>2</sub>																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		K <sub>2</sub>																		
		p=0,05																		
11	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	17	18	19	20
	0	2	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7
	1	2	3	3	4	4	4	5	5	6	6	6	7	7	7	8	8	8	9	9
	2	-	3	4	4	5	5	6	6	7	7	8	8	9	9	9	10	10	11	11
	3	-	-	4	5	5	6	7	7	8	8	9	9	10	10	11	12	12	13	13
	4	-	-	-	5	6	7	7	8	9	9	10	10	11	12	12	13	14	14	15
	5	-	-	-	-	6	7	8	9	9	10	11	11	12	13	14	14	15	16	17
	6	-	-	-	-	-	-	8	9	10	11	12	12	13	14	15	16	16	17	18
12	7	-	-	-	-	-	-	-	-	-	11	12	13	14	15	16	17	18	18	19
	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	19	20
	0	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6
	1	2	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9
	2	-	3	4	4	5	5	5	6	6	7	7	8	8	9	9	9	10	10	11
	3	-	3	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12
	4	-	-	4	5	6	6	7	8	8	9	9	10	11	11	12	12	13	14	14
	5	-	-	-	5	6	7	8	8	9	10	10	11	12	12	13	14	14	15	16
13	6	-	-	-	-	6	7	8	9	10	10	11	12	13	13	14	15	16	16	17
	7	-	-	-	-	-	-	-	9	10	11	12	13	13	14	15	16	17	18	18
	8	-	-	-	-	-	-	-	-	-	-	12	13	14	15	16	17	18	19	19
	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	19	20
	0	2	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6
	1	2	3	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8
	3	-	3	4	4	5	5	6	6	7	7	8	8	9	9	0	10	11	11	12
	4	-	-	4	5	5	6	7	7	8	8	9	9	10	11	11	12	12	13	13

Продолжение табл. 7

n <sub>1</sub>	K <sub>1</sub>	n <sub>2</sub>																			
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
		K <sub>2</sub>																			
		p=0,05																			
14	5	-	-	-	5	6	7	7	8	9	9	10	10	11	12	12	13	14	14	15	
	6	-	-	-	-	6	7	8	8	9	10	11	11	12	13	13	14	15	16	16	
	7	-	-	-	-	-	7	8	9	10	11	11	12	13	14	14	15	16	17	18	
	8	-	-	-	-	-	-	-	-	9	10	11	12	13	14	15	15	16	17	18	19
	9	-	-	-	-	-	-	-	-	-	-	-	-	13	14	15	16	17	18	19	20
	0	2	2	2	3	3	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6
	1	2	3	3	3	4	4	4	4	5	5	5	5	6	6	6	7	7	7	8	8
	2	2	3	3	4	4	5	5	5	6	6	7	7	7	8	8	8	8	9	9	10
	3	-	3	4	4	5	5	6	6	7	7	8	8	9	9	9	9	10	10	11	11
	4	-	-	4	5	5	6	6	7	7	8	9	9	10	10	11	11	12	12	13	13
15	5	-	-	4	5	6	6	7	8	8	9	9	10	11	11	12	12	13	14	14	
	6	-	-	-	5	6	7	7	8	9	10	10	11	12	12	13	14	14	15	16	
	7	-	-	-	-	6	7	8	9	9	10	11	12	12	13	14	15	15	16	17	
	8	-	-	-	-	-	-	8	9	10	11	12	12	13	14	15	16	16	17	18	
	9	-	-	-	-	-	-	-	-	10	11	12	13	14	15	16	16	17	18	19	
	10	-	-	-	-	-	-	-	-	-	-	-	-	-	14	15	16	17	18	19	20
	0	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5
	1	2	3	3	3	4	4	4	4	4	5	5	5	6	6	6	6	7	7	7	7
	2	2	3	3	4	4	4	4	5	5	6	6	6	7	7	7	8	8	8	9	9
	3	-	3	4	4	5	5	6	6	6	6	7	7	8	8	9	9	9	10	10	11
4	-	3	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	12	
5	-	-	4	5	6	6	7	7	8	8	9	10	10	11	11	12	12	13	14	14	
6	-	-	-	5	6	7	7	8	9	9	10	10	11	12	12	13	14	14	15	15	
7	-	-	-	-	6	7	8	8	9	10	11	11	12	13	13	14	15	15	16	16	
8	-	-	-	-	-	7	8	9	10	10	11	12	13	13	14	15	16	16	17	17	
9	-	-	-	-	-	-	8	9	10	11	12	13	13	14	15	16	17	17	18	18	

Продолжение табл. 7

n <sub>1</sub>	K <sub>1</sub>	n <sub>2</sub>																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		K <sub>2</sub>																		
		p=0,05																		
16	10	-	-	-	-	-	-	-	-	-	11	12	13	14	15	16	17	17	18	19
	11	-	-	-	-	-	-	-	-	-	-	-	-	-	15	16	17	18	19	20
	0	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5
	1	2	2	3	3	3	4	4	4	4	5	5	5	6	6	6	6	7	7	7
	2	2	3	3	4	4	4	4	5	5	5	6	6	6	7	7	7	8	8	9
	3	-	3	4	4	4	4	5	5	6	6	7	7	7	8	8	9	9	9	10
	4	-	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12
	5	-	-	4	5	5	6	6	7	8	8	9	9	10	10	11	11	12	12	13
	6	-	-	4	5	6	6	7	8	8	9	9	10	11	11	12	12	13	14	14
	7	-	-	-	5	6	7	7	8	9	9	10	11	11	12	13	13	14	15	15
	8	-	-	-	-	6	7	8	9	9	10	11	12	12	13	14	14	15	16	16
	9	-	-	-	-	-	7	8	9	10	11	11	12	13	14	14	15	16	17	18
	10	-	-	-	-	-	-	-	9	10	11	12	13	14	14	15	16	17	18	19
11	-	-	-	-	-	-	-	-	-	-	12	13	14	15	16	17	18	19	19	
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	18	19	20	



**Методика расчетов основных показателей рядов динамики**

Основные показатели	2012 г.	2013 г.	2014 г.	2015 г.	2016 г.	2017 г.
Число родившихся в РБ (в тыс.) $Y_i$	53,6	46,8	47,4	48,5	44,9	43,8
<b>Темп роста %</b>						
1. базисный метод		$\frac{46,8}{53,6} \cdot 100 = 87,3$	$\frac{47,4}{53,6} \cdot 100 = 88,4$	90,5	83,8	81,7
2. цепной метод	-	$\frac{46,8}{53,6} \cdot 100 = 87,3$	$\frac{47,4}{46,8} \cdot 100 = 101,3$	102,3	92,6	97,6
<b>Абсолютный прирост (тыс. чел.)</b>						
1. базисный метод	-	$46,8 - 53,6 = -6,8$	$47,4 - 53,6 = -6,2$	-5,1	-8,7	-9,8
2. цепной метод	-	$46,8 - 53,6 = -6,8$	$47,4 - 46,8 = 0,6$	1,1	-3,6	-1,1
<b>Темп прироста %</b>						
1. базисный метод	-	$-6,8 : 53,6 \cdot 100 = -12,7$	$-6,2 : 53,6 \cdot 100 = -11,6$	-9,5	-16,2	-18,3
2. цепной метод	-	$-6,8 : 53,6 \cdot 100 = -12,7$	$0,6 : 46,8 \cdot 100 = 1,3$	2,3	-7,4	-2,4

## 15. СЛОВАРЬ ТЕРМИНОВ

**Альфа-ошибка (a-error) или ошибка I рода (Type I error)** – вероятность ошибочного отклонения нулевой гипотезы. В клинических исследованиях это вероятность того, что одно лечение будет считаться эффективнее другого, когда в действительности это различие отсутствует.

**Анализ в зависимости от назначенного лечения (intention to treat analysis)** – способ анализа данных в контролируемом испытании. Анализ проводится в зависимости от того, к какой группе – экспериментальной или контрольной – был отнесен пациент при рандомизации независимо от того, получил он в действительности исследуемую терапию или нет.

**Анализ дожития (survival analysis, time-to-event analysis)** – статистические методы анализа времени, прошедшего от заданного начального момента до момента наступления определенного исхода (обычно – смерти). Метод позволяет описать любой исход, дихотомический по своему характеру и имеющий место лишь однажды в процессе наблюдения, например, время до развития инфаркта миокарда или до рецидива злокачественного новообразования. К способам анализа времени наступления события относятся анализ Каплана-Мейера (Kaplan-Meier analysis), регрессионная модель пропорционального риска Кокса (Cox proportional hazards regression model), лог-ранк тест (log-rank test) и другие.

**Анализ решений (decision analysis)** – статистический подход к принятию решений, осуществляемый путем сопоставления альтернатив с учетом экономических затрат и последствий для больного в заданных условиях.

**Анализ чувствительности (sensitivity analysis)** – метод, используемый в мета-анализе, фармакоэкономике и анализе решений для оценки влияния различных параметров на конечный результат. Анализ «наилучший вариант – наихудший вариант» (best case / worst case analysis) представляет собой специальный вид анализа чувствительности.

**Бета-ошибка (B-error), или ошибка II рода (Type II error)** – вероятность ошибочного принятия нулевой гипотезы. В клинических исследованиях это вероятность того, что оба лечения будут считаться равно эффективными, когда в действительности одно лечение лучше, чем другое.

**Величина  $p$  ( $p$  value)** – вероятность того, что полученный результат абсолютно случаен. Величина  $p$  может изменяться от 1 (результат точно случаен) до 0 (результат точно не случаен). Величина  $p$ , меньшая или равная заданному уровню альфа-ошибки (например, 0,05), свидетельствует о статистической значимости полученного различия. В то же время, она не характеризует клиническую значимость результата.

**Величина  $p_a$**  – вероятность альфа-ошибки. Обычно, когда пишут  $p$ , имеют в виду именно  $p_a$ .

**Величина  $p_b$**  – вероятность бета-ошибки.

**Вероятностная выборка (probability sample)** – это такая выборка, вероятность попадания в которую для каждого индивидуума известна заранее.

**Вмешивающийся фактор, или конфаундер, конфаундинг-фактор, неучтенный фактор (confounding factor)** – фактор, связанный с известным исследуемым фактором, и некоторым образом влияющий на результат (исход). Например, известно, что у мужчин чаще, чем у женщин возникает ИБС. Однако это может быть связано не с исследуемым фактором (пол), а, допустим, с тем, что мужчины чаще курят, в значительной степени подвергаются стрессам, потребляют больше пищи, богатой холестерином. Влияние вмешивающихся факторов приводит к возникновению систематической ошибки (confounding bias).

**Выборка (sample)** – это часть популяции, полученная путем отбора. Клинические исследования обычно выполняются на выборках. Оценку характеристик популяции на практике приходится осуществлять путем оценки этих характеристик по выборке.

**Выборка вероятностная (probability sample)** – выборка, вероятность попадания в которую для каждого индивидуума известна заранее.

**Выборка случайная (random sample)** – такая выборка, вероятность попадания в которую для каждого индивидуума в популяции одинакова.

**Выборка смещенная (biased sample)** – это такая выборка, которая вследствие систематической ошибки отличается от популяции, представляющей предмет исследования, или от популяции, по отношению к которой должны применяться результаты исследования. Например, лица, добровольно участвующие в программах профилактики заболеваний, отличаются от прочих людей в популяции и представляют собою смещенную выборку из популяции.

**Выборки, размер (sample size)** – число больных, которое необходимо включить в исследование для невелирования случайности в качестве объяснения полученного результата. Размер выборки зависит от четырех характеристик исследования: величины различия в частоте исходов между группами,  $p_a$ ,  $p_b$  и природы данных. Размер выборки следует учитывать исследователю, планирующему эксперимент, и читателю, решающему, следует ли доверять опубликованным результатам.

**Группа контрольная (control group), или группа сравнения** – группа испытуемых, получающих обычное лечение, или не получающих лечения, или получающих плацебо. Результаты измерений в контрольной группе сравниваются с показателями в экспериментальной группе для оценки эффекта исследуемого метода лечения.

**Группа экспериментальная (experimental group)** – это группа, подвергающаяся вмешательству (лечению) в ходе исследования. Иначе она еще называется *группой лечения или группой вмешательства* (treatment group или intervention group).

**Данные качественные (nominal data)** – это такие признаки, которые нельзя выразить количественно. Например: диагноз, группа крови, страна проживания. Частным случаем качественных данных являются *дихотомические данные*.

**Данные количественные (numerical, or interval data)** – это такие величины, которым присущ естественный порядок расположения с равными интер-

валами между последовательными значениями, независимо от их места на шкале (масса тела, концентрация глюкозы в крови).

**Данные дискретные (discrete data)** – это количественные данные, которые выражаются целыми числами. Например: число беременностей; число рождений живых младенцев; число судорожных припадков у больного в течение месяца.

**Данные дихотомические (dichotomous data)** – признаки, которые могут иметь только два значения (присутствует, отсутствует, да-нет, жив-умер).

**Данные непрерывные (continuous data)** – это количественные данные, которые могут принимать любое значение на непрерывной шкале. Примеры: масса тела, артериальное давление, парциальное давление кислорода в артериальной крови.

**Данные порядковые (ordinal data)** – это величины, которые могут быть расположены в естественном порядке (ранжированы), например, от малого до большого или от хорошего до плохого, но размер интервала между такими категориями не может быть выражен количественно (например: стадии болезни; оценки «высокий, средний, низкий» или «отсутствует, слабый, умеренный, тяжелый»).

**Добавочная доля популяционного риска (population attributable risk fraction)** – доля заболеваемости (смертности, инвалидности) в ПОПУЛЯЦИИ, связанная с данным фактором риска; рассчитывается путем деления добавочного популяционного риска на общую заболеваемость (смертность, инвалидность) в популяции.

**Добавочный, или атрибутивный риск (attributable risk)** – определяется как заболеваемость для лиц, подвергавшихся воздействию фактора риска, минус заболеваемость для лиц, не подвергавшихся этому воздействию. Добавочный риск – это дополнительные случаи развития заболевания, обусловленные воздействием фактора риска. Учитывая способ вычисления добавочного риска, его также называют *разницей рисков (risk difference)*, или *непосредственным риском*.

**Доверительный интервал (confidence interval)** – статистический показатель, позволяющий оценить, в каких пределах может находиться истинное значение параметра в популяции; диапазон колебаний истинных значений. Величины, полученные в исследованиях на выборке больных, отличаются от истинных величин в популяции вследствие влияния случайности. Так, 95% доверительный интервал означает, что истинное значение величины с вероятностью в 95% лежит в его пределах. Доверительные интервалы помогают сориентироваться, соответствует ли данный диапазон значений представлениям читателя о клинической значимости эффекта и каких результатов можно ожидать, применив описанную методику на сходной группе больных. Величина доверительного интервала характеризует степень доказательности данных, в то время как величина *p* указывает на вероятность ошибочного отклонения нулевой гипотезы (см. Величина *p*).

**Достоверность (validity)** – характеристика, показывающая, в какой мере результат измерения соответствует истинной величине. Достоверность исследования определяется тем, в какой мере полученные результаты справедливы в отношении данной выборки (*internal validity*). Это внутренняя характеристика, она касается именно данной группы больных и не обязательно распространяется на другие группы.

**Исследование «до-после» (before-after study)** – неконтролируемое клиническое исследование, описывает течение заболевания в одной группе пациентов, подвергающейся изучаемому вмешательству. Подход основан на предположении, что любое улучшение, наблюдаемое после лечения, обусловлено именно лечением. Это предположение может оказаться ложным, что делает данный метод весьма уязвимым.

**Исследование распространенности (prevalence study)** – вариант описательного исследования, проводимого в определенный момент времени для оценки распространенности заболевания или исхода, изучения течения заболевания и т.п.

**Исследование серий случаев (case series)** – описательное исследование, представляющее собой количественный анализ группы больных.

**Исследование случай – контроль (case control study)** – ретроспективное исследование, в котором по архивным данным или воспоминаниям, суждениям пациентов производится сравнение двух групп, в одну из которых отобраны пациенты с определенной патологией, а в другую – лица без нее. При этом сначала осуществляется отбор группы пациентов с изучаемым заболеванием и сходной по остальным признакам группы лиц без этого заболевания. Затем ретроспективно оценивается частота воздействия возможного фактора риска в обеих группах. Полученные данные позволяют рассчитать относительный риск развития заболевания в связи с изучаемым фактором.

**Исследование совокупного риска (aggregate risk study), или экологическое исследование (ecological study)** – исследование, в котором фактор риска характеризуется средним воздействием на группу; его задача – оценить влияние на людей факторов окружающей среды в целом.

**Исторический, или непараллельный контроль (historical or nonconcurrent control)** – способ клинического исследования, при котором результаты современного лечения сравниваются с данными наблюдения аналогичных пациентов в прошлом. Противоположность *параллельному контролю* (*concurrent control*), при котором контрольная группа набирается одновременно и по тем же правилам, что и экспериментальная группа

**Историческое когортное исследование (historical cohort study)** – *когорта* выделена по архивным документам и прослежена до настоящего времени.

**Клиническая эпидемиология (clinical epidemiology)** – это наука, разрабатывающая методы клинических исследований, которые дают возможность делать справедливые заключения, контролируя влияние систематических и случайных ошибок.

**Клинические испытания (clinical trials)** – специальный вид когортных исследований для оценки результатов разных вмешательств, условия проведения которых (отбор групп вмешательства, характер вмешательства, организация наблюдения и оценка исходов) направлены на устранение влияния систематических ошибок на получаемые результаты.

**Когорта (cohort)** – группа лиц, изначально объединенных каким-либо общим признаком (например, здоровые лица, или больные на определенной стадии заболевания) и наблюдаемых в течение определенного периода времени, чтобы проследить, что с ними произойдет в дальнейшем.

**Когорта дожития (survival cohort)** – когорта, в которую включаются пациенты, имеющие заболевание и доступные для наблюдения – например, пациенты специализированной клиники. Другой термин для таких групп – *доступные когорты (available patients cohort)*.

**Когортное исследование (cohort study)** – исследование, в котором определенная когорта пациентов прослеживается в течение некоторого периода времени. Когортные исследования называют также *продольными* или *лонгитудинальными* (longitudinal study). Термином *проспективные* (prospective study) подчеркивается, что пациенты прослеживаются во времени – имеется в виду, что группа сформирована в настоящее время и будет прослежена в будущем. Термином *исследованиями заболеваемости* (incidence study) обращается внимание на то, что основным способом оценки является регистрация новых случаев заболевания в течение определенного срока.

**Мета-анализ (meta-analysis)**. Методология объединения разнородных и выполненных различными авторами исследований, относящихся к одной теме, для повышения достоверности оценок одноименных результатов. Количественный анализ объединенных результатов нескольких клинических испытаний одного и того же вмешательства. Такой подход обеспечивает большую статистическую мощность, чем в каждом отдельном испытании за счет увеличения размера выборки. Используется для обобщенного представления результатов многих испытаний и для увеличения доказательности результатов испытаний. *Кумулятивный мета-анализ* — метод расчета обобщенной точечной оценки изучаемого эффекта, при котором исследования добавляются по одному в определенном порядке. *Мета-регрессионный анализ* – регрессионный анализ, в котором в качестве единицы наблюдения рассматриваются отдельные исследования.

**Многофакторное моделирование (multivariable modeling)** – метод, обеспечивающий математическое выражение сочетанного влияния многих переменных. Метод называется *многофакторным*, так как он оценивает эффект многих переменных. Термин «*моделирование*» означает, что это математическая конструкция, которая рассчитывается из полученных данных, но основана она на упрощающих допущениях о характеристиках этих данных (например, что все переменные имеют нормальное распределение и одинаковую дисперсию).

**Многофакторный анализ (multivariable analysis)** – это совокупность статистических методов, которые одновременно рассматривают влияние многих переменных на какой-либо один фактор. Если после устранения влияния этих переменных действие фактора сохраняется, его воздействие считается независимым. Кроме того, эти методы применяются для выделения из большого числа признаков малого подмножества, которое вносит независимый и существенный вклад в конечный результат (исход), что используется для ранжирования переменных по силе их воздействия на исход и для других целей.

**Мощность, статистическая (statistical power)** – вероятность того, что в исследовании будет найдено статистически достоверное различие, когда это различие действительно существует. Статистическая мощность равняется  $1 - P_{\text{B}}$  (см. *Величина  $P_{\text{B}}$* ). Мощность исследования аналогична чувствительности диагностического теста.

**Обобщаемость (external validity or generalizability)** – внешняя характеристика, она определяется тем, в какой мере результаты данного исследования применимы к другим группам больных.

**Описание случая (case report, case study)** – простое изложение материала о нескольких больных.

**Открытое исследование (open, open-label, unmasked study)** – исследование, в котором не применялся слепой метод.

**Относительный риск (relative risk), или отношение рисков (risk ratio)** – отношение заболеваемости среди лиц, подвергавшихся и не подвергавшихся воздействию факторов риска. Относительный риск не несет информации о величине абсолютного риска (заболеваемости). Даже при высоких значениях относительного риска абсолютный риск может быть совсем небольшим, если заболевание редкое. Относительный риск показывает силу связи между воздействием и заболеванием.

**Отношение шансов (odds ratio)** – определяется как отношение шансов события в одной группе к шансам события в другой, или как отношение шансов того, что событие произойдет, к шансам того, что событие не произойдет. В исследованиях случай контроль отношение шансов используется для оценки относительного риска.

**Ошибка, систематическая, смещение (systematic error, bias)** – это неслучайное, однонаправленное отклонение результатов от истинных значений. Систематическая ошибка может возникать вследствие отбора (при создании выборки) – *sampling*, или *assembling bias*, вследствие измерений – *measurement bias*, при воздействии неучтенных факторов – *confounding bias* и во многих других случаях. О систематической ошибке говорят также, имея в виду предвзятость при публикации положительных результатов исследований и отклонении отрицательных – *publication bias*. Для борьбы с систематическими ошибками и получения достоверных данных используются организационные методы (например, рандомизация, «слепой метод» и т.п.), а также внесение поправок, учитывающих величину смещения.

**Ошибка случайная или вариабельность (random variation)** – отклонение результата отдельного наблюдения (измерения) от его истинного значения, обусловленное исключительно случайностью.

**Ошибка экологическая (ecological fallacy)** – систематическая ошибка, связанная с тем, что заболевшие лица в группе обследования на самом деле могли не испытывать влияния факторов риска. Возникает, когда группу выделяют по общим косвенным признакам (проживание на одной территории, принадлежность к одной профессии), а не по наличию фактора риска у отдельных субъектов (например, по потреблению алкоголя).

**Плацебо (placebo)** – лекарственная форма, неотличимая от исследуемого препарата по внешнему виду, цвету, вкусу и запаху, но не оказывающая специфического действия (например, таблетки глюкозы или инъекции изотонического раствора хлорида натрия), или иное безразличное вмешательство, используемое в медицинских исследованиях для имитации лечения с целью устранения систематической ошибки, связанной с *плацебо-эффектом*.

**Плацебо – эффект (placebo effect)** – изменение состояния пациента (отмечаемое самим пациентом или лечащим врачом), связанное с фактом лечения, а не с биологическим действием препарата.

**Популяция (population)** – это совокупность индивидуумов, из которой отбирается выборка и на которую могут быть распространены результаты, полученные для этой выборки. Популяция может представлять собой все население (обычно таковы популяции в эпидемиологических исследованиях причин заболеваний), или же состоять из пациентов, госпитализированных в определенную клинику или из пациентов с определенным заболеванием (что чаще имеет место в клинических исследованиях). Таким образом, можно говорить об общей популяции, госпитальной популяции или популяции пациентов с конкретным заболеванием. Эпидемиологическое определение популяции отличается от биологического (экологического).

**Претестовая вероятность** – вероятность наличия заболевания до применения диагностического теста.

**Проверка гипотез (hypothesis testing) о равенстве средних** – метод статистического доказательства гипотезы о равенстве средних величин. Нулевая гипотеза ( $H_0$ ) заключается в том, что различий между двумя выборками, оцененными по их средним значениям, нет. Положительная гипотеза ( $H_1$ ) утверждает обратное. Принятие нулевой гипотезы означает, что обнаруженная разница случайна, отклонение нулевой гипотезы – что разница не случайна. Вероятность ошибочного отклонения нулевой гипотезы называется альфа-ошибкой, уровень статистической значимости обнаруженных различий обозначается как величина  $p$ . Вероятность ошибочного понятия нулевой гипотезы называется бета-ошибкой.

**Прогностическая ценность (predictive value)** – вероятность наличия заболевания при известном результате диагностического теста. Прогностическая ценность называется также апостериорной (или посттестовой) вероятностью,



поскольку это вероятность наличия (или отсутствия) болезни после того, как стали известны результаты теста.

**Прогностическая ценность отрицательного результата (*negative predictive value*)** – вероятность отсутствия заболевания при отрицательном ("нормальном") результате теста.

**Прогностическая ценность положительного результата (*positive predictive value*)** – это вероятность заболевания при положительном ("ненормальном") результате теста.

**Профилактика, вторичная (*secondary prevention*)** – это вмешательства, направленные на то, чтобы замедлить или остановить развитие заболевания у больных.

**Профилактика, первичная (*primary prevention*)** – вмешательства, направленные на предупреждение возникновения болезни, обычно за счет устранения ее причин.

**Профилактика, третичная (*tertiary prevention*)** – мероприятия, направленные на предотвращение ухудшения течения или осложнения заболевания после того, как болезнь проявилась.

**Рандомизация (*randomization*)** – процедура, обеспечивающая случайное распределение больных в группы вмешательства и контроля. Случайным распределением достигается отсутствие различий между двумя группами и, таким образом, снижается вероятность систематической ошибки в клинических исследованиях вследствие различия групп по каким – либо признакам.

**Распространенность (*prevalence*)** – это частота некоторого состояния в группе. Засчитывается как отношение числа лиц, у которых на момент обследования наблюдается изучаемое состояние (болезнь или исход), к числу всех лиц в группе (популяции, обследованным, работникам предприятия).

**Риск** – вероятность нанесения ущерба здоровью или причинения дискомфорта.

**Риск, добавочный (*атрибутивный*)** – см. добавочный риск.

**Риск, относительный** – см. относительный риск.

**Риск, популяционный добавочный (*атрибутивный*) (*population attributable risk*)** – рассчитывается как произведение добавочного риска на распространенность фактора риска в популяции. Этот показатель отражает дополнительную заболеваемость в популяции, связанную с фактором риска.

**Скрининг (*screening*)** – массовое обследование лиц, не считающих себя больными, для выявления скрыто протекающих заболеваний или других состояний (факторов риска будущих заболеваний). Обычно проводится с использованием дешевых, простых, неинвазивных диагностических процедур, имеющих высокую чувствительность.

**Слепой (или маскированный) метод (*blinding, or masking*)** – процедура, обеспечивающая отсутствие информации о том, к какой группе вмешательства или контрольной отнесен каждый испытуемый. При простом слепом методе информация отсутствует только у пациента, при двойном слепом – у пациента и исследователя, при тройном слепом – у пациента, исследователя и лиц, прово-

дящих статистическую обработку результатов исследования. Применяется для поиска систематической ошибки в клинических исследованиях.

**Смещение к среднему (*regression to the mean*)** – закономерность в биометрических исследованиях, выражающаяся в том, что величины, отклоняющиеся от среднего значения, при последующих измерениях оказываются ближе к среднему.

**Специфичность диагностического теста (*specificity*)** – вероятность отрицательного результата диагностического теста при отсутствии болезни.

**Точка разделения (*cut-off point, cutpoint*)** – величина, используемая для разделения ряда величин на две части. В диагностическом процессе точка разделения отделяет "нормальные" значения показателя от "ненормальных". Результаты теста, лежащие в области "нормы", называются "отрицательным результатом", а лежащие в области "патологии" – "положительным результатом".

**Точность теста (*test accuracy*)** – доля правильных результатов теста (истинно положительных и истинно отрицательных) в общем количестве полученных результатов.

**Фактор риска (*risk factor*)** – особенность организма или внешнее воздействие, приводящие к увеличению риска возникновения заболевания или иному неблагоприятному исходу.

**Фармакоэпидемиология.** Наука о применении эпидемиологических знаний, методов и доводов при изучении благоприятных и нежелательных эффектов для рационального использования в различных человеческих популяциях. Предметом исследования являются результаты наблюдений, собранных при повседневном клиническом обслуживании больных, в отличие от рафинированных данных рандомизированных, двойных слепых клинических испытаний. Цель исследований – определить реальную эффективность лекарственных средств в клинической практике.

**Фармакоэкономика** – анализ стоимости лекарственной терапии для системы здравоохранения и общества с целью выявить наиболее выгодные препараты (виды лечения) для финансирующих организаций, пациентов и производителей. На результатах фармако-экономического анализа основываются решения о планах финансирования различных программ здравоохранения, о приоритетах среди возможных альтернатив, о возможностях пациента использовать оптимальный для него вариант лечения.

**Частота новых случаев (*incidence*)** – это отношение числа лиц, у которых в течение определенного времени развилось изучаемое состояние, ко всем обследованным в группе, где исходно этого состояния никто не имел. Частота новых случаев болезни в популяции, где это заболевание исходно отсутствовало, называется заболеваемостью. Частота новых случаев исходов при определенном заболевании указывает соответственно на инвалидность или летальность при этом заболевании. Для оценки частоты новых случаев отбирают группу лиц без заболевания (или без изучаемого исхода) и периодически обследуют ее, подсчитывая число новых случаев, появившихся за определенный период.

**Частотное распределение (frequency distribution)** – величины, расположенные в порядке возрастания, и частота, с которой каждая величина встречается в популяции или выборке. Обычно представляется в виде графика, на котором по горизонтальной оси откладываются наблюдаемые величины, а по вертикальной – частоты, с которыми эти величины наблюдаются.

**Число больных, которым необходимо лечить (number needed to treat, NNT)** – способ оценки относительной эффективности двух методов лечения. Показывает, какое количество больных необходимо подвергнуть лечению исследуемым методом для предотвращения одного случая изучаемого исхода. Например, если для предотвращения одного случая инфаркта миокарда надо лечить антигипертензивным средством 33 пациента с артериальной гипертензией в течение 5 лет, то NNT равно 33 за 5 лет. Этот показатель является величиной, обратной снижению абсолютного риска.

**Чувствительность диагностического теста (sensitivity)** – вероятность положительного результата диагностического теста при наличии болезни.

**Шансы (odds)** – отношение вероятности того, что событие произойдет, к вероятности того, что событие не произойдет. Шансы и вероятности содержат одну и ту же информацию, но по-разному выражают ее. Если вероятность того, что событие произойдет обозначить  $P$ , то шансы этого события будут равны  $P/(1 - P)$ . Например, если вероятность выздоровления – 0,3, то шансы выздороветь равны  $0,3/(1 - 0,3) = 0,43$ . Шансы удобнее использовать для некоторых расчетов, чем вероятности.

**Экспериментальное исследование (experimental study)** – сравнительное исследование, заранее спланированное и посвященное изучению влияния, по крайней мере, одного вмешательства. Сравнения могут производиться между двумя и более группами или внутри одной группы до и после назначения вмешательства.

**Эффект ярлыка (labeling)** – психологическое воздействие результатов теста или диагноза на пациента – участника скрининга.

Составители:

*Шарафутдинова Назира Хамзиновна* – зав. кафедрой общественного здоровья и организации здравоохранения с курсом ИДПО, д.м.н., профессор;

*Киреева Эльза Фаритовна* – доцент кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, к.м.н.;

*Николаева Ирина Евгеньевна* – зав. кафедрой рентгенэндоваскулярных диагностики и лечения, к.м.н.

*Павлова Марина Юрьевна* – доцент кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, к.м.н.;

*Халфин Рауль Магруфович* – доцент кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, к.б.н.;

*Шарафутдинов Марат Амирович* – профессор кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, д.м.н.;

*Борисова Марина Владимировна* – доцент кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, к.м.н.;

*Латыпов Айрат Борисович* – доцент кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, к.м.н.;

*Шамилавна Галикеева Ануза* – профессор кафедры общественного здоровья и организации здравоохранения с курсом ИДПО, д.м.н.

## **Статистические методы в медицине и здравоохранении**

### **Учебное пособие**

Лицензия № 0177 от 10.06.96 г.

Подписано к печати 19.06.2018 г.

Отпечатано на ризографе с готового оригинал-макета,  
представленного авторами.

Формат 60x84 <sup>1</sup>/<sub>16</sub>. Усл.-печ. л. 7,61.

Тираж 110 экз. Заказ № 45

450008, г. Уфа, ул. Ленина, 3,

Тел.: (347) 272-86-31

ФГБОУ ВО БГМУ Минздрава России