

**Лекции кафедры
общественного здоровья и
здравоохранения**



КУРС
**«Медицинская
информатика»**





Лекция 4

Информационная поддержка исследовательской работы Основы медицинской статистики



Основные вопросы лекции



- Основы медицинской статистики
- Особенности компьютерной обработки данных и программного обеспечения
- Современные требования к обработке медико-биологических данных
- Основы теории вероятности, доказательная медицина
- Организация и этапы медико-биологического исследования

Основные проблемы поддержки научной работы врача

- **Широкое распространение компьютерных систем** обработки данных, которые позволяют осуществлять анализ собственного материала и при минимальной подготовки пользователя
- Особую актуальность статистические методы приобретают в условиях распространения **доказательной медицины**
- Отставание в **применении статистических методов** анализа в отечественной медицинской науке по сравнению с зарубежными публикациями
- Использование компьютерной технологии анализа данных требует не только освоения программного пакета, но и **грамотной интерпретации результатов обработки**

Современные тенденции

обработки данных с точки

зрения исследователя

- Уменьшение трудоемкости процесса обработки и анализа с использованием вычислительной техники
- Значительно увеличилось количество и сложность применяемых методов
- Необходимо иметь представление, как о математических методах обработки данных, так и о соответствующих программах

Доказательная медицина

Доказательная медицина (evidence-based medicine) - это раздел медицины, основанный на доказательствах, предполагающий поиск, сравнение, обобщение и широкое распространение полученных доказательств для использования в интересах больных (Evidence Based Medicine Working Group, 1993).

Доказательная медицина является новым направлением (технологией) сбора, анализа, обобщения и интерпретации научной информации. Она предусматривает добросовестное, объяснимое и основанное на достоверных выводах доказательства использования наилучших современных достижений для лечения пациентов.

Основной целью внедрения принципов доказательной медицины в практику здравоохранения является оптимизация качества оказания медицинской помощи с точки зрения безопасности, эффективности, стоимости и других значимых факторов.



Доказательная медицина

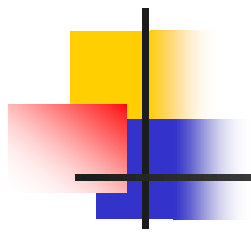
Медико-биологические исследования можно разделить на 2 типа - первичные и вторичные.

К первичным исследованиям относятся:

- а) эксперименты **на животных** или добровольцах в искусственных и контролируемых условиях;
- б) клинические испытания, при которых вмешательство, например, лекарственную терапию, проводят **на группе пациентов** с последующим ее наблюдением;
- в) исследования, при которых различные параметры (гигиенические, физиологические, эпидемиологические и т.д.) измеряют **в группах людей**.

Вторичные исследования - это:

- а) **обзоры** (несистематические - обобщение результатов первичных исследований; систематические - обобщение первичных исследований на основе жестких критериев; метаанализ - обобщение количественных данных нескольких исследований);
- б) **клинические рекомендации** - сформулированные выводы из первичных исследований, касающихся требований к действиям врача;
- в) **анализ принятия решений** - на основе первичных исследований формирует подходы к управлению здравоохранением или распределению ресурсов, позволяет строить "древо" решений, которые принимают врачи, пациенты или администраторы при выборе способов лечения или при распределении ресурсов;
- г) **экономический анализ** - на основе результатов первичных исследований делает вывод о целесообразности финансирования изучаемого метода лечения.



ОСНОВЫ МЕДИЦИНСКОЙ (САНИТАРНОЙ) СТАТИСТИКИ

Статистика

СТАТИСТИКА - это общественная наука, изучающая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной в конкретных исторических условиях.

СМОТРИТЕ!

СЛУШАЙТЕ!

ЗАПИСЫВАЙТЕ!

ВНИКАЙТЕ!



Санитарная статистика

СТАТИСТИКА САНИТАРНАЯ - это медицинская отрасль *социальной статистики*, изучающая количественные характеристики состояния здоровья населения, развития системы здравоохранения, определяет степень интенсивности влияния на них социально-экономических факторов, а также занимается применением статистических методов к обработке и анализу результатов клинических и лабораторных исследований. (БСЭ)

Состоит из:

- Статистики здоровья населения (заболеваемость, смертность и др.)
- Статистики здравоохранения (посещаемость, летальность, занятость койки и др.)
- Клинической статистики: научные и фармакологические исследования

Задача – своевременное получение и разработка достоверных данных о заболеваемости, смертности, инвалидности, физическом развитии населения в целом и отдельных его групп, о размещении, состоянии, оснащенности, медицинских кадрах учреждений здравоохранения, клинических и лабораторных исследований.

Источники информации: первичная учетная мед. документация, периодическая и годовая стат. отчетность, единовременные учеты и выборочные обследования. Расчет показателей проводят для страны в целом, городского и сельского населения и по отдельным территориям и т.д.

Этапы статистического исследования

1. Составления программы и плана статистического исследования

План исследования - это где, когда, кто и как выполняет работу. Программа исследования включает вопросы: что и в каком направлении изучать, с обозначением объекта и единиц наблюдения, учетных признаков, методов сбора, разработки и анализа материала. А также методов формирования статистической совокупности.



2. Статистическое наблюдение.

Включает сбор материала в соответствии с программой исследования.



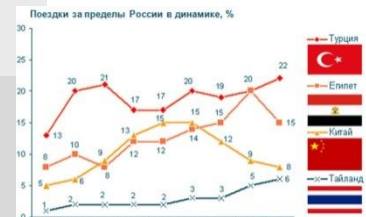
3. Группировка и разработка статистического материала.

Включает: шифровку (кодирование) статистического материала; группировку; вычисление статистических показателей (относительных и средних величин); графическое изображение показателей.

Группы взаимоотношений	Количество предприятий	
	всего	к итогу, %
Предприятия — органы государственного управления	15	0,1
Предприятия — обладатели прав на объекты интеллектуальной собственности	1075	3,3
Предприятия-инвесторы	268	0,8
Предприятия-производители	51	0,2
Предприятия-потребители	31 127	95,6
Всего	32 536	100

4. Анализ результатов исследования.

Сопоставление статистических данных; выявление закономерностей в изучаемых явлениях; обобщение результатов исследования, формулирование выводов и рекомендаций.



Источники: Статистическое агентство Республики Корея, Российский индекс Целиевый Групп 3 квартал 2011
Выборка: Россияне, 16+, совершившие путешествие за границу, проживающие в городах



1-й ЭТАП

статистического исследования

Составление плана и программы

1-й ЭТАП Составление программы и плана статистического исследования

Включает:

- Определение цели и задач исследования
- Составление плана сбора материала
- Составление программы разработки материала: форм учетных документов (бланков), в которые включены признаки, подлежащие регистрации
- Составление программы анализа собранного материала (макетов статистических таблиц)

Рабочая гипотеза



Любое исследование начинается с формулирования РАБОЧЕЙ ГИПОТЕЗЫ.

При этом целью исследования является получение данных, на основании которых выдвинутую еще до начала исследования, как говорят априори, гипотезу можно было бы принять, т.е. признать истиной, либо отвергнуть - признать ложной.

Рабочая гипотеза состоит из множества *статистических гипотез* которые подтверждаются или опровергаются методами прикладной статистики.

Например:

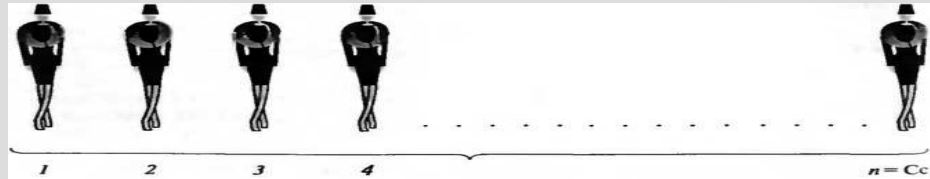
- гипотеза о форме распределения признака;
- сравнение средних двух совокупностей;
- сравнение вариабельности признака двух совокупностей и др.

Статистическая совокупность



Статистическая совокупность – это множество единиц изучаемого явления, объединенных единой качественной основой, общей связью, но отличающихся друг от друга отдельными признаками.

Объектом статистического исследования и, соответственно, наблюдения называется та совокупность, о которой должны быть собраны нужные сведения



Единица наблюдения – это составная часть объекта наблюдения, подлежащая изучению и регистрации в соответствии с программой исследования. Она является носителем признаков, подлежащих регистрации

Учетные признаки – это медико-биологические характеристики, регистрируемые у единицы наблюдения в соответствии с целями и задачами исследования

Отдельные единицы статистического наблюдения объединяются признаками сходства и отличаются между собой признаками различия.

Учетные признаки

Учитываемые (учетные) признаки – это признаки единиц наблюдения, значение которых в соответствии с целью исследования должны быть зарегистрированы исследователем.

- **Признаки сходства** – это признаки, свидетельствующие о принадлежности конкретной единицы наблюдения к изучаемой совокупности. Например, *единство времени и места исследования*.

- **Признаки различия** – это индивидуальные особенности (характеристики) каждой единицы наблюдения, являются конечным результатом статистического наблюдения, например, возраст, пульс, артериальное давление и др.

По отклику на изменение учетные признаки могут быть:
Факторными и результативными

- **Факторные** признаки вызывают действие, например, длинна доза препарата.

- **Результативные (зависимые признаки)** изменяют свое значение под влиянием факторных признаков, например, частота пульса при различной физической нагрузке.

Учетные признаки

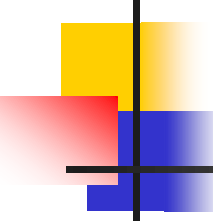
По характеру учетные признаки могут быть:
качественными и количественными

Качественные (атрибутивные) признаки - это значения которые определяются вербально (записываются словом), например, пол, образование, диагноз и др.

Количественные (цифровые) признаки - это значения которые имеют числовое выражение, например рост, масса тела, частота пульса и др.

- **Биномиальные:** Да/Нет
- **Дискретные:** выражаются несколькими заранее определенными значениями цифр, например, стадия процесса (1, 2, 3, 4)
- **Непрерывные:** выражаются числами, как правило десятичными, с любой точностью измерения

Статистические таблицы



СТАТИСТИЧЕСКАЯ ТАБЛИЦА - это форма записи изучаемой статистической совокупности, разделенной на группы в соответствии с изучаемыми признаками.

Табличное подлежащее – это то, о чем говорится в таблице, основной признак или признаки, которые, как правило, обозначены в строках таблицы.

Табличное сказуемое – это количественные признаки, характеризующие подлежащее. Они, как правило, расположены в столбцах (графах) таблицы.

Типы статистических таблиц:

- **Простые** - подлежащее характеризуется лишь одним признаком
- **Групповые** - подлежащее характеризуется двумя связанными между собой признаками
- **Комбинационные** подлежащее характеризуется тремя и более связанными между собой признаками.

Требования к составлению статистических таблиц

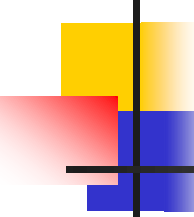


Таблица 1

Распределение числа больных по классам заболеваний

Класс заболеваний	Число больных
Болезни органов дыхания	250
Инфекционные заболевания	120
Травмы	50
Прочие болезни	30
Итого:	450

1. Заголовок, отражающий ее содержание

2. Шапка с указанием показателей или признаков и единиц измерений

3. Столбец «Всего» и строка «Итого»

4. В таблице не должно быть пустых ячеек

5. Если в документе встречается несколько таблиц, каждая из них обязательно нумеруется. Номер таблицы указывают над ней в правой части страницы

Простая статистическая таблица

Табличное подлежащее

Таблица 1

Распределение числа больных по классам
заболеваний

Класс заболеваний	Число больных
Болезни органов дыхания	
Инфекционные заболевания	
Травмы	
Прочие болезни	
Итого:	

Групповая статистическая таблица

Таблица 2

Распределение абсолютного числа больных по классам заболеваний с учетом социальных групп среди жителей района ...

Класс заболеваний	Число больных в группах			Всего
	рабочие	служащие	ИТР	
Болезни органов дыхания				
Инфекционные заболевания				
Травмы				
Прочие болезни				
Итого:				

Табличное сказуемое

Комбинационная статистическая таблица

Таблица 3

Табличное подлежащее

Распределение абсолютного числа больных по классам заболеваний с учетом возрастных и социальных групп среди жителей района ...

Классы заболеваний	Число больных															Всего				
	рабочие					служащие					ИТР					до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего
	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего	до 19 лет	20-39 лет	40-59 лет	60 лет и старше	всего					
Болезни органов дыхания																				
Инфекционные заболевания																				
Травмы																				
Прочие болезни																				
Итого:																				

Табличное сказуемое (2 признака)



2-й ЭТАП

статистического исследования

**Статистическое
наблюдение
(сбор материала)**

2-й ЭТАП Статистическое наблюдение (сбор материала)

Статистическое наблюдение – это сбор сведений посредством переписей, обследований, экспериментов и регистрации данных в учетных документах.



В процессе статистического наблюдения собирается информация, которая является основной для последующей статистической обработки и анализа.

Статистической информацией называют совокупность сведений экономического, социального, медико-биологического и иного характера, на основе которой осуществляется учет, контроль, планирование, управление и статистический анализ.

При заполнении бланков ОБЯЗАТЕЛЬНО указывают дату и сведения о человеке, заполнившего документ.

Виды статистических наблюдений

По охвату статистической совокупности исследование может быть:

- Сплошное
- Выборочное (не сплошное)

По продолжительности:

- Единовременное
- Текущее.

При сплошном статистическом исследовании группа наблюдения формируется путём полного охвата всех единиц изучаемого явления. Множество всех единиц наблюдения, охватываемых сплошным наблюдением, называется генеральной совокупностью.

На практике сплошное исследование проводится крайне редко, по причинам:

- осуществить такое наблюдение **организационно трудно или невозможно**, т.к. размер генеральной совокупности очень большой или не имеет определённых её границ.
- исследование объекта приведёт к его **уничтожению** (например, анализ качества вакцин, сывороток, медикаментов).
- сплошные исследования **во много раз дороже** выборочных.
- большие **затраты времени** на сбор данных.
- **огромные массивы данных** осложняют обработку собранных материалов.
- **результаты полученные при сплошном наблюдении будут такими - же, как результаты не сплошного наблюдения при адекватном его проведении.**

Сплошные исследования используют, как правило, для решения общегосударственных задач, например: перепись населения, информация об инфекционных заболеваниях и другие.

Выборочное статистическое наблюдение

Монографический метод – применяется для подробного описания объекта, имеющего какие-либо яркие особенности. Например: медико-социальное исследование народностей Крайнего Севера, социально-гигиеническое описание промышленного центра. Выводы, полученные при этом, относятся либо к конкретному объекту, либо могут быть распространены на очень ограниченную группу аналогичных объектов.

Метод основного массива – обследование контингентов, сосредоточенных на конкретном объекте. Например, изучение госпитализированной заболеваемости в стационаре. Данные о структуре заболеваемости, тяжести течения, прогнозе могут быть использованы только для решения частных вопросов. Судить о распространенности патологии за пределами стационара по этим данным нельзя.

Собственно выборочное исследование охватывает **выборочную совокупность** или **выборку** из генеральной совокупности.

Конечной целью выборочного исследования является получение информации о генеральной совокупности. Для этого выборочное исследование должно удовлетворять определенным требованиям. **Одно из главных условий – представительность (репрезентативность) выборки.**

**РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРКИ -
степень совпадения единиц выборочной
совокупности по изучаемым признакам
к этим признакам в генеральной
совокупности**

**ГЕНЕРАЛЬНАЯ
СОВОКУПНОСТЬ**

**ВЫБОРОЧНАЯ
СОВОКУПНОСТЬ
(выборка)**

**Единица
наблюдения**

Погрешность выборочного статистического наблюдения

Каждое выборочное исследование имеет некоторую погрешность, вытекающую из самого факта выборочности, когда результаты, полученные на выборке, переносят на всю генеральную совокупность, то есть по части судят о целом. Такую погрешность называют ошибкой репрезентативности (представительности).

Виды репрезентативности (погрешностей, ошибок):

- **Количественная репрезентативность** определяется числом **наблюдений**, гарантирующим получение статистически достоверных данных, а также соблюдением методики исследования и вариабельностью признака.
- **Качественная репрезентативность** – обозначает структурное соответствие выборочной и генеральной совокупности на основе признаков сходства.

Репрезентативность статистического наблюдения

Количественная репрезентативность зависит:

- **От числа единиц наблюдений.** Здесь действует основной постулат закона больших чисел – «чем больше наблюдений, тем результат достоверней» или «чем больше число наблюдений, тем больше значения характеристик выборки приближаются к соответствующим характеристикам генеральной совокупности».
- **Систематических ошибок наблюдения**, которые искажают результат наблюдения в одном направлении. Например, они могут возникать в результате желания опрашиваемых или исследователей представить всё в лучшем свете, чем есть на самом деле. Ошибки такого рода не могут быть устранены увеличением объёма выборки.
- **От изменчивости изучаемого признака.** Чем больше разброс значений изучаемого признака, тем больше ошибка репрезентативности. Поэтому при анализе статистических данных необходимо определять характеристики изменчивости (разброса) значений каждого учетного признака и его вероятностные характеристики распределения.

Качественная репрезентативность – это соответствие выборочной и генеральной совокупности. Определяется по структуре совокупности на основе признаков сходства. Например, если в составе генеральной совокупности 40% лиц мужского пола, то и выборочной группе их должно быть 40%.

Методы формирования выборочной совокупности

Основным правилом отбора является случайность. В силу закона больших чисел выборка будет качественно репрезентативна, если её осуществить случайно.

Случайность достигается выполнением ряда условий:

- Каждая единица генеральной совокупности должна иметь равную вероятность попасть в выборку. Например, если отбор историй болезни вести по заглавным буквам фамилий пациентов, то вероятность попасть в выборку для разных фамилий разная.
- Отбор единиц наблюдения из генеральной совокупности необходимо проводить независимо от изучаемого признака.
- Отбор должен проводиться из однородных групп.

Методы отбора:

Механический – единицы из обследуемой совокупности отбираются механически (каждый 5-й или 10-й и так далее)

Типическая, типологическая или районированная выборка – например, территория города делится на типичные районы в зависимости от степени загрязнения, в этих районах путём случайного отбора формируются группы наблюдения.

Когортный отбор - выбор лиц, объединённых моментом появления какого-то признака, например: год рождения, время начала болезни и так далее.



3-й ЭТАП

СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Обработка статистического материала

3-й ЭТАП. Обработка

статистического материала

Включает:

- контроль данных,
- шифровку,
- группировку,
- вычисление статистических показателей,
- заполнение статистических таблиц,
- графическое изображение данных.

Современное состояние

статистической обработки данных

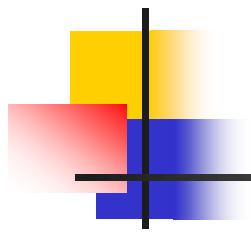
- Отказ от ручной обработки данных и использование компьютерных программ и пакетов статистического анализа.
- Изменение акцентов с требований знания математики на умение пользоваться компьютерными программами и статистическими методами.
- Возможность использовать специальные статистические методы обработки экспериментального материала, позволяющие всесторонне оценить изучаемое явление и выполнить «статистический эксперимент» врачу, не имеющему специальной математической подготовки.
- Статистический эксперимент - это применение различных методов статистической обработки данных и выбор наиболее адекватных из них, позволяющих выявить скрытые закономерности изучаемых явлений и подтвердить формулируемые выводы.

Хронология развития пакетов анализа данных

Годы	Наиболее распространённые пакеты анализа данных	Типы ЭВМ	Операционные системы
1970-1985	BMDP, SSP, SAS, SPSS	EC, CM	
1985-1995	Statgraphics, STATA, SPSS, SAS, Systat	IBM PC	DOS
1995-2000	Statgraphics, SPSS, STATISTICA, SAS, Excel	Pentium	Windows

Классификация программных средств обработки данных

Вид	Отечественные	Зарубежные	Стоимость \$
Профессиональные	Нет	SAS, BMDP	2000-10000
Универсальные	STADIA , Olimp, ARCADА	Statgraphics, SPSS, S-PLUS STATISTICA	500-5000
Специализированные	Mesosaur, DataScore, Класс-Мастер, Эвриста , САНИ и др.	Большое количество	200-1500
Табличные редакторы	МойОфис, P7-Офис	Suprecalc, Quatro PRO, Lotus 1-2-3, Excel	0 - 200



Основные принципы компьютерной обработки и анализа медико-биологических данных

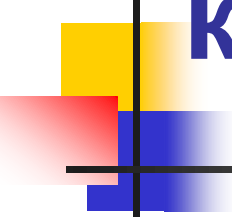
Компьютерная

обработка данных

включает:

1. Подготовку данных к компьютерной обработке
2. Предварительный анализ данных
3. Статистический эксперимент

Подготовка данных к компьютерной обработке



- **Ввод значений учетных признаков в электронную таблицу**
- **Преобразование данных:**
 - получение вторичных расчетных показателей,
 - группировка,
 - ранжирование и др.
- **Визуализацию данных с помощью диаграмм**

Создание таблицы данных в программе STATISTICA

The screenshot shows the STATISTICA - Spreadsheet2 application window. A 'Create New Document' dialog box is open, allowing the user to create a new document. The dialog has two tabs: 'Программа макроса(SVB)' and 'Крупноформатная таблица'. The 'Крупноформатная таблица' tab is active. In this tab, there are two input fields: 'Число переменных:' (Number of variables) set to 10 and 'Число регистров:' (Number of registers) set to 10. Below these fields is a 'Размещение' (Placement) section with two radio buttons: 'В новую книгу' (In new book) and 'Как автономное окно' (As separate window), with the latter selected. The background spreadsheet shows a grid with columns labeled '1 ФИО' and '2 Пол', and rows numbered 1 to 10. The status bar at the bottom shows 'C1,V1', 'Sel:OFF', 'Weight:OFF', 'CAPS', 'NUM', and 'REC'.

Пример обработки данных в таблице Excel

Microsoft Excel - Пример задачи 1

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Введите вопрос

Сохранить 10 Ж К Ч

C4

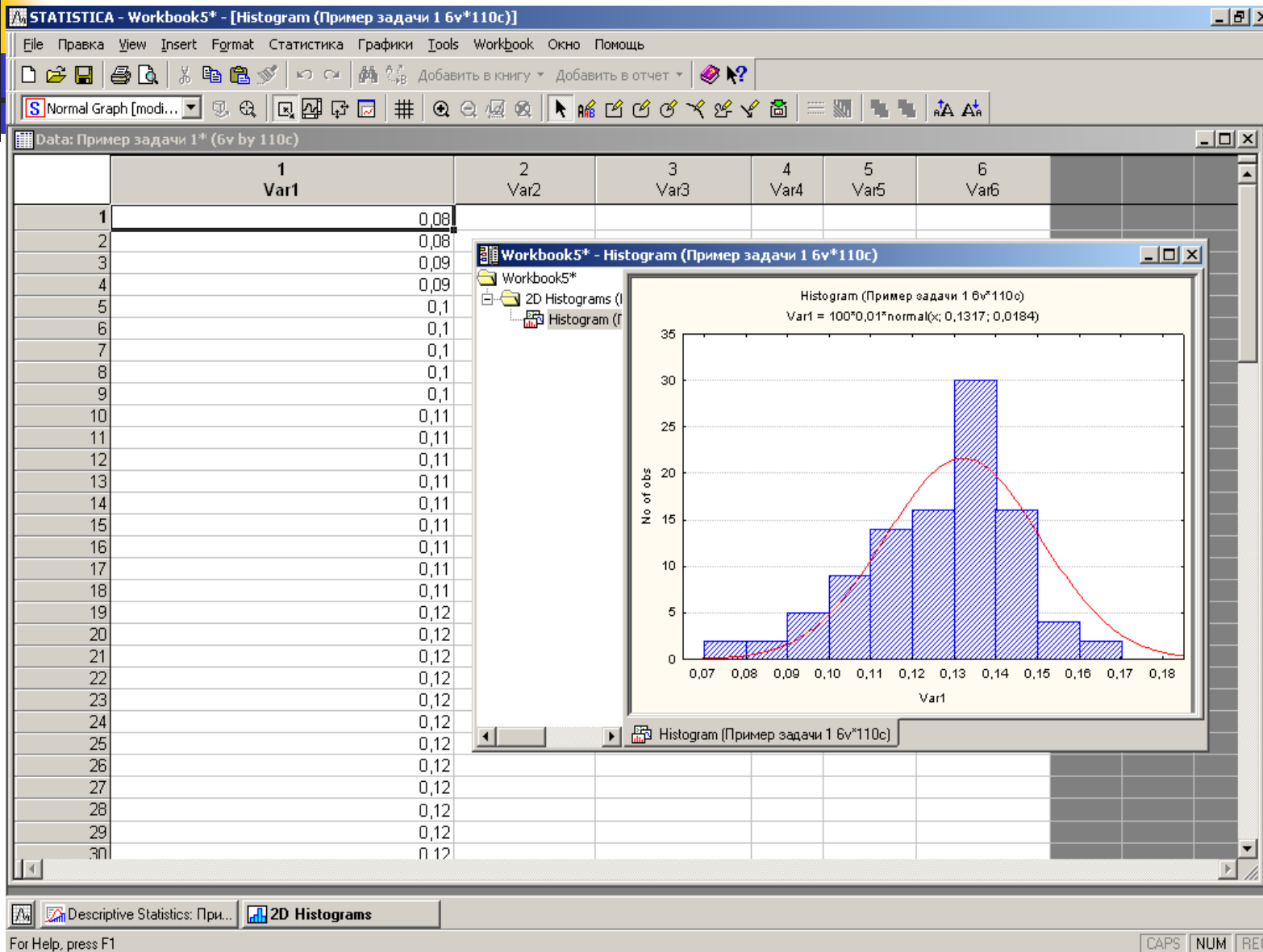
	A	B	C	D	E	F	G	H	I	J	K
1			Задача 1								
2											
3			Оценить эффективность золоуловителя								
4											
5			Средняя концентрация пыли до установки золоуловителя:								
6			0,26 мг/м ³								
7			ошибка средней								
8			0,007 мг/м ³								
9											
10			Результаты измерения уровня пыли:								Расчет критерия Стьюдента
11			0,08								17,72481
12			0,08								
13			0,09								
14			0,09								
15			0,1								
16			0,1								
17			0,1								
18			0,1								
19			0,1								
20			0,11								
21			0,11								
22			0,11								
23			0,11								
24			0,11								
25			0,11								
26			0,11								
27			0,11								
28			0,11								
29			0,12								
30			0,12								

Лист1 Лист2 Лист3

Готово NUM

Визуализация данных

(гистограмма рассеяния)



Предварительный анализ

данных включает:

- расчет относительных и средних величин описательной статистики;
- уточнение структуры данных и разделение на группы;
- выявление вероятностных законов распределения, которым подчиняются данные;
- выбор методов анализа;
- определение взаимосвязей между переменными;
- выявление различий между группами;
- визуализацию результатов обработки данных.

Расчет основных статистических характеристик – описательная статистика

The screenshot displays the STATISTICA software interface. The main window shows a data table with columns labeled '1 До', '2 После', '3 Var3', '4 Var4', '5 Var5', and '6 Var6'. A smaller window titled 'Descriptive Statistics (Пример задачи 1)' is overlaid, showing a table of statistical results for two variables: 'До' and 'После'. The 'До' variable has a mean of 0.208200 and a standard error of 0.005180. The 'После' variable has a mean of 0.131700 and a standard error of 0.001843. Red circles highlight the 'Mean' and 'Standard Error' columns in the statistics table.

Variable	Valid N	Mean	Median	Mode	Frequency of Mode	Sum	Minimum	Maximum	Std.Dev.	Standard Error
До	100	0,208200	0,220000	2200000	31	20,82000	0,100000	0,320000	0,051803	0,005180
После	100	0,131700	0,140000	1400000	30	13,17000	0,080000	0,170000	0,018426	0,001843

Выявление различий между группами – вычисление t-критерия Стьюдента

STATISTICA - Workbook20* - [T-test for Independent Samples (Пример задачи 1)]

File Правка Вид Вставка Формат Статистика Графики Инструменты Данные Workbook Окно Помощь

Аrial 10 B I U

Data: Пример задачи 1* (6v by 110c)

	1 До	2 После	3 Var3	4 Var4	5 Var5	6 Var6
1	0,1	0,08				
2	0,1	0,08				

Workbook20* - T-test for Independent Samples (Пример задачи 1)

Basic Statistics/1

Descriptive :

Describe

T-test for in

T-test for

T-test for Independent Samples (Пример задачи 1)

Note: Variables were treated as independent samples

Group 1 vs. Group 2	Mean Group 1	Mean Group 2	t-value	df	p	Valid N Group 1	Valid N Group 2	Std.Dev. Group 1	Std.Dev. Group 2	F-ratio Variances	p Variances
До vs. После	0,208200	0,131700	13,91346	198	0,00	100	100	0,051803	0,018426	7,904436	0,000000

T-test for Independent Samples (Пример задачи 1)

26	0,18	0,12
27	0,18	0,12
28	0,18	0,12
29	0,18	0,12
30	0,18	0,12

T-Test for Indepe...

For Help, press F1

C1.V1 0,2082 Set:OFF Weight:OFF CAPS NUM REC

Основные методы углубленного статистического анализа (статистический эксперимент)

- непараметрическая статистика
- дисперсионный анализ
- регрессионный анализ
- спектральный анализ
- дискриминантный анализ
- многомерная статистика
- факторный анализ
- кластерный анализ
- канонические корреляции
- дифференциальные уравнения и математическое моделирование
- методы распознавания образов

Статистические методы обработки и анализа данных, используемые в типовых медицинских исследованиях

Источник информации, задача исследования	Используемые методы обработки и анализа
Вопросники, анкеты, тесты, обследование состояния здоровья, мнение обследуемого, истории болезни, недостающая информация, выявление скрытых взаимосвязей	Составление таблиц и отчетов, корреляционный анализ, классификация, факторный анализ
Скрининговые исследования	Дискриминантный анализ, кластерный анализ, методы распознавания образов
Истории болезни, клинические обследования, медицинские записи, регистрации, исследование надежности данных, описание случаев заболеваний, сопоставление с данными прошлых лет, изучение осложнений, исследование эффективности различных процедур, изучение связи между процедурами	Составление таблиц, корреляционный анализ, дисперсионный анализ, регрессионный анализ
Медико-статистические данные, исследование заболеваемости, динамика заболеваемости, выявление периодичности	Методы анализа случайных процессов, спектральный анализ, математическое моделирование

Источник информации, задача исследования	Используемые методы обработки и анализа
Эксперименты. Лабораторные эксперименты и опыты на животных при заданных условиях	Регрессионный анализ, дисперсионный анализ, многомерный статистический анализ, методы математического моделирования
Клинические исследования. Сравнительные исследования, анализ выживаемости и наследственности с учетом принадлежности пациента к определенной группе, изучение дозировки препаратов. Разработка методов диагностики.	Дисперсионный анализ, регрессионный анализ, дискриминантный анализ, кластерный анализ, методы распознавания образов
Исследования механизмов действия патогенных факторов.	Дисперсионный анализ, регрессионный анализ, методы математического моделирования
Клинические лабораторные данные. Хранение, сбор и передача клинической информации, анализ качества и надежности лабораторных исследований.	Дисперсионный анализ, регрессионный анализ

Вычисление критерия χ^2

STATISTICA - Workbook1* - [Observed vs. Expected Frequencies (Spreadsheet1)]

File Правка Вид Вставка Формат Статистика Графики Инструменты Данные Workbook Окно Помощь

Добавить в книгу Добавить в отчет

Arial 10 B I U

Data: Spreadsheet1* (10v by 10c)

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	60	50								
2	40	50								
3										
4										
5										
6										
7										
8										
9										
10										

Workbook1* - Observed vs. Expected Frequencies (Spreadsheet1)

Observed vs. Expected Frequencies (Spreadsheet1)
Chi-Square = 4,000000 df = 1 p < 0,045501

Case	observed Var1	expected Var2	O - E	(O-E)**2 /E
C: 1	60,0000	50,0000	10,0000	2,000000
C: 2	40,0000	50,0000	-10,0000	2,000000
Sum	100,0000	100,0000	0,0000	4,000000

Observed vs. Expected...

For Help, press F1

C1,V1 60 CAPS NUM REC

U тест Манна—Уитни

STATISTICA - Workbook5* - [Mann-Whitney U Test (Пример задачи 2)]

File Правка Вид Вставка Формат Статистика Графики Инструменты Данные Workbook Окно Помощь

Добавить в книгу Добавить в отчет

Arial 10 B I U

Data: Пример задачи 2* (8v by 12c)

	1 SAP	2 HR	3 Вес	4 ЧССнагр	5 Возраст	6 Возр. группа	7 Пол	8 Показатель
Иванов	145	72	75	40	140	Средний	м	2,01388889
Сидоров	127	65	70	48	109	Средний	ж	1,95384615
Казаков	130	70	65	30	110	Младший	ж	1,85714286

Семен
Сидор

Workbook5* - Mann-Whitney U Test (Пример задачи 2)

Mann-Whitney U Test (Пример задачи 2)
By variable Пол
Marked tests are significant at p < ,05000

variable	Rank Sum м	Rank Sum ж	U	Z	p-level	Z adjusted	p-level	Valid N м	Valid N ж	2*1-sided exact p
Вес	57,00000	21,00000	0,00	2,882307	0,003948	2,592677	0,003700	6	6	0,002165

Comparing Two Grou...

For Help, press F1 C1,V1 57 CAPS NUM REC



4-й ЭТАП

СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

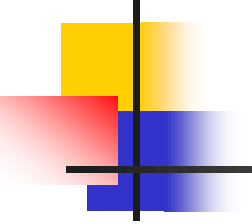
Анализ полученных данных

4-й ЭТАП. Анализ полученных данных



Элементы анализа:

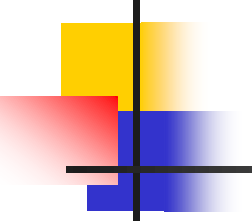
- **Интерпретация** на основе:
 - сопоставления с нормативами
 - со средними уровнями аналогичных величин
 - со стандартами
 - с данными по другим учреждениям и территориям
 - литературными данными
 - в динамике
- **Литературное оформление** работы (статьи, диссертация);
- **Выводы**
- **Предложения** для внедрения в практику;
- **Прогноз и рекомендации.**



Основные требования к описанию результатов статистической обработки данных

- **полнота и уровень описания, как самого анализа, так и его результатов**
- **адекватность выбранных методов**
- **обоснованность сформулированных
ВЫВОДОВ**

В медицинских статьях и научных работах принято представлять полученные данные с указанием средних величин, с соответствующими им показателями стандартных отклонений или стандартной ошибки.




Минимальный перечень методов статистической обработки данных

Российские научные журналы:

- Критерий t ,
- X - квадрат,
- Корреляция
- Линейная регрессия

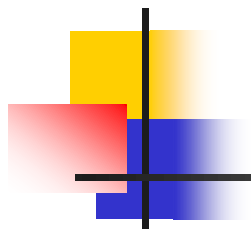
Зарубежные издания, дополнительно:

- Дисперсионный анализ



ПРИМЕР требований к статистической обработке в статьях, публикуемых в журнале «Кардиология»

«Все статьи, принятые к публикации, при необходимости будут проверяться на адекватность применения статистических методик и статистической интерпретации данных. Описывайте статистические методы настолько детально, чтобы квалифицированный читатель, имеющий доступ к оригинальным данным, смог проверить полученные Вами результаты. Представляйте, когда возможно, полученные данные в виде средних величин, с соответствующими им показателями стандартных отклонений или стандартной ошибки среднего значения. (Указывайте величину доверительного интервала, который принят за статистически достоверный в данном исследовании). При использовании более сложных статистических методов, чем расчет критериев t , X -квадрат и коэффициентов линейной регрессии, укажите, какой статистической программой Вы пользовались, и номер ее версии»



Основные теории вероятности

Наука о событиях

Теория вероятностей - это раздел математики, изучающий закономерности случайных явлений: случайные события, случайные величины, их свойства и операции над ними.



Случайная величина

Случайная величина – это количественная характеристика события демонстрирующая уровень какого-либо явления, при данном испытании.

Случайная величина в теории вероятностей, величина, принимающая в зависимости от случая те или иные значения с определёнными вероятностями. (БСЭ)

Например:

- число подтягиваний на перекладине
- время на беговой дистанции
- артериальное давление пациента, частота пульса и др.

В силу действия большого числа неконтролируемых факторов эти величины могут принимать различные значения в результате испытания. Причем до испытания невозможно предсказать их точное значение.



Классическое определение вероятности события

Вероятность какого либо события – это численное выражение возможности его наступления. Обозначается латинской буквой $P(A)$

В простейших случаях вероятности событий могут быть легко определены исходя из условий испытаний.

В общем случае:

Пусть испытание имеет n возможных несовместных исходов, т. е. отдельных событий, могущих появиться в результате данного испытания. Причем при каждом повторении испытания возможен один и только один из этих исходов. Кроме того, пусть по условиям испытания, нет никаких оснований предполагать, что один из исходов появляется чаще других, т. е. все исходы являются равновероятными.

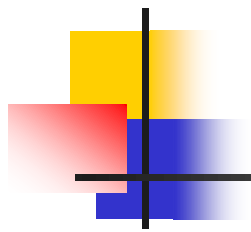
Допустим, что при n равновероятных несовместных исходах интерес представляет некоторое событие A , появляющееся при каждом из m исходов и не появляющееся при остальных $n-m$ исходах. Тогда принято говорить, что в данном испытании имеется n случаев, из которых m благоприятствуют появлению события A .

$$P(A) = \frac{m}{n}$$

Вероятность события A равна отношению числа исходов, благоприятствующих событию A , к общему числу всех равновероятных несовместных исходов опыта.
Формула Лапласа.

Условная вероятность

Теорема Байеса



Теорема Байеса (формула Байеса) — одна из основных теорем теории вероятностей, которая позволяет определить вероятность того, что произошло какое-либо событие при наличии косвенных тому подтверждений, которые могут быть неточны (вероятностны).

Формула названа в честь ее автора - Томаса Байеса. Его работа «*An Essay towards solving a Problem in the Doctrine of Chances*» впервые опубликована в 1763 году, через 2 года после смерти автора.

Полученную по формуле Байеса вероятность можно вычислять повторно и уточнять ее, принимая во внимание данные новых наблюдений. **Т.е. проводить «обучение» системы.**

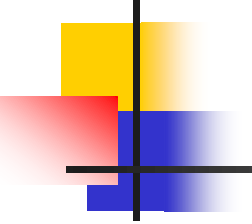
Теорема Байеса

Пусть A_1, A_2, \dots, A_n — некоторые попарно несовместимые события, хотя бы одно из которых обязательно наступает, и B — некоторое событие. Тогда, согласно теореме Байеса, условная вероятность A_k при условии, что наступило B , может быть определена по формуле

$$P(A_k | B) = P(B | A_k) P(A_k) / \sum_{k=1}^n P(B | A_k) P(A_k).$$

В теореме Байеса события A_k называют «гипотезами», вероятности $P(A_k)$ — априорными вероятностями гипотез и вероятности $P(A_k | B)$ — апостериорными вероятностями этих гипотез (при условии, что фактически наблюдается событие B).

Формула Байеса


$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

где

$P(A)$ — априорная (расчетная) вероятность гипотезы A ;

$P(A|B)$ — вероятность гипотезы A при наступлении события B (апостериорная вероятность);

$P(B|A)$ — вероятность наступления события B при истинности гипотезы A ;

$P(B)$ — вероятность наступления события B .

Формула Байеса позволяет «переставить причину и следствие»: по известному факту события вычислить вероятность того, что оно было вызвано данной причиной.

События, отражающие действие «причин», в данном случае обычно называют *гипотезами*, так как они — *предполагаемые* события, повлекшие данное. Безусловную вероятность справедливости гипотезы называют *априорной* (насколько вероятна причина вообще), а условную — с учетом факта произошедшего события — *апостериорной* (насколько вероятна причина оказалась с учетом данных о событии).

Следствие теоремы Байеса

Формула полной вероятности

Вероятность наступления события B , зависящего от ряда гипотез A_i , если известны степени достоверности этих гипотез (например, измерены экспериментально) равна

$$P(B) = \sum_{i=1}^N P(A_i)P(B|A_i)$$

Пример: Вам нужно открыть 3 двери в темноте имея 3 близких по форме ключа. В темноте ключ выбирается случайным образом. На открытие каждой из дверей тратится 5 сек. Найти вероятность того, что вы откроете все двери за 15 сек.

Решение. Пусть событие A – “открыты все двери”. Разобьем это событие на более простые. Пусть B – “открыта 1-я”, C – “открыта 2-я”, а D – “открыта 3-я”. Тогда, $A=BCD$ по определению произведения событий. Следовательно $P(A)=P(BCD)$. По теореме о вероятности произведения независимых событий $P(BCD) = P(B) * P(C) * P(D)$. Вычислим вероятности событий B , C и D . В этом примере имеется 3 равновозможных (каждый ключ выбираем из 3-х) исходов опыта. Каждому из событий B , C и D благоприятствует 1 из них, поэтому

$$P(B) = P(C) = P(D) = \frac{1}{3} \quad P(A) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = 0,037 = 3,7\%$$

Теорема Бернулли



Теорема Бернулли опубликована после его смерти, в 1713 году. Она была обобщена Пуассоном, в 1837 году, который ввел термин «закон больших чисел». Более общее понимание этого термина основано на работе П. Л. Чебышева «О средних величинах» (1867).

Пусть проводятся независимые испытания, при каждом из которых вероятность события A неизменна. Справедливо утверждение: если N достаточно велико, то с вероятностью сколь угодно близкой к единице, отличие относительной частоты $\frac{n_N}{N}$ от вероятности $P(A)$ меньше любого наперед заданного положительного числа. В математической записи:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{n_N}{N} - P(A) \right| > \varepsilon \right) = 0$$

Где: ε - произвольно малое положительное число.

Т.е. при большом числе измерений в опыте , например, много раз подбрасывая монету, “почти наверняка” в разных сериях будут получены примерно равные частоты появления герба.

Закон больших чисел

Всегда найдётся такое количество испытаний, при котором с любой заданной наперёд вероятностью относительная частота появления некоторого события будет сколь угодно мало отличаться от его вероятности.

Общий смысл закона больших чисел — совместное действие большого числа случайных факторов приводит к результату, почти не зависящему от случая.

На практике поступают следующим образом.

Пример: чтобы найти вероятность выздоровления без осложнений болезни, проверяют группу больных, например, из $N=200$ человек, и определяют количество больных, выздоровевших без осложнений. Допустим, их количество $n_N = 190$. Относительная частота $\frac{n_N}{N} = 190/200 = 0,95$. Это число и принимают приближенно за вероятность $P(A)$ выздоровления без осложнений.

Для получения большей точности можно повторить опыт, но в результате будет получено аналогичное число с незначительным отклонением от указанного.

Статистическая вероятность

Вероятностью события A в данном испытании называется число $P(A)$, около которого группируются значения относительной частоты при большом числе опытов.

Будем фиксировать число испытаний, в результате которых появилось некоторое событие A . Пусть было проведено N испытаний, в результате которых событие A появилось ровно n_N раз. Тогда число n_N называется частотой события, а отношение $\frac{n_N}{N}$ - относительной частотой события.

Экспериментальным фактом является то, что относительная частота события при большом числе повторений испытания начинает мало изменяться и стабилизируется около некоторого определенного значения, в то время как при малом числе повторений она принимает различные, совершенно случайные значения.

Поэтому ясно, что если при неограниченном повторении испытания относительная частота события будет стремиться к вполне определенному числовому значению, и это значение можно принять и качестве объективной характеристики события A . Такое число, связанное с событием A , называется статистической вероятностью события $P(A)$.

В математике неограниченное число повторений испытания записывается знаком предела (\lim) при N , стремящемся к бесконечности (∞):

$$P(A) = \lim_{N \rightarrow \infty} \frac{n_N}{N}. \quad 4.64$$

Вычисление статистической вероятности

$$P(A) = \lim_{N \rightarrow \infty} \frac{M_N}{N}$$

Пусть в 3-х сериях опытов при бросании монеты в 1-й серии из $n=4040$ раз герб выпал $m=2048$ раз. Частота появления герба в данной серии опытов равна $P(A)=2048/4040=0,5069$. При бросании той же монеты во 2-й серии $n=2000$ раз герб выпал $m=6019$ раз. Следовательно, в этой серии частота $P(A)=6019/12000=0,5016$. Наконец, в 3-й серии при $n=24000$ бросаний герб появился $m=2012$ раз с частотой $P(A)=0,5005$.

Таким образом, при большом числе бросаний монеты частота появления герба обладает устойчивостью, т. е. мало отличается от числа $0,5$. Это отклонение частоты от числа $0,5$ уменьшается с увеличением числа испытаний. Наблюдаемое в этом примере свойство устойчивости частоты является общим свойством массовых случайных событий, а именно, всегда существует такое число, к которому приближается частота появления данного события, мало отличаясь от него при большом числе испытаний. Поэтому это число называется *статистической вероятностью* события. Оно выражает объективную возможность появления события. В рассмотренном примере классическая вероятность появления герба $P(A)=m/n=1/2=0,5$ и *статистическая вероятность* равна $P(A)=0,5$ или 50%

ДОМА ПОВТОРИТЕ ЭТОТ ОПЫТ (не менее 100 подбрасываний)

Статистические величины

Абсолютные величины - это данные, значения выполненных измерений.

Относительной величиной (статистическим коэффициентом, показателем, индексом) называется отношение двух абсолютных величин (чисел), выражающих меру каких-либо явлений.

Смысл получения относительных величин заключается в нахождении общей меры и приведение к общему знаменателю, что позволяет сопоставлять уровень явления в нескольких группах.

Средняя величина - это совокупная обобщающая характеристика количественного признака.

ЧИСЛА

ИСПОЛЬЗУЕМЫЕ В СТАТИСТИКЕ



0,001 0,01 0,1 0 1 10 100 1 000 10 000 100 000

10^{-3} 10^{-2} 10^{-1} 0 1,0 10^1 10^2 10^3 10^4 10^5

Промилле	Процент	Доля (часть)	Нуль	Единица	Десять	Сто	Тысяча	Десять тысяч	Сто тысяч
----------	---------	--------------	------	---------	--------	-----	--------	--------------	-----------

$\frac{0}{00}$ $\frac{0}{0}$ $\frac{1}{n}$ 0 ————— Другие обозначаются числом —————

Абсолютные величины

Абсолютные величины являются результатом статистического наблюдения. В статистике в отличие от математики все абсолютные величины имеют размерность (единицу измерения), а также могут быть положительными и отрицательными.

Единицы измерения абсолютных величин могут быть:

- **простыми**, отражая 1 свойство (например, масса груза измеряется в тоннах)
- **сложными**, отражая несколько взаимосвязанных свойств (например, тонно-километр или киловатт-час).

Виды единиц измерения абсолютных величин:

Натуральные - применяются для исчисления величин с однородными свойствами (например, штуки, тонны, метры и т.д.).

Условно-натуральные - применяются к абсолютным величинам с однородными свойствами, но проявляющим их по-разному. Например, общее количество школьных тетрадей измеряется в у.ш.т. — условные школьные тетради размером 12 листов.

Стоимостные единицы измерения выражаются в рублях или в иной валюте, представляя собой меру стоимости абсолютной величины. Они позволяют суммировать даже разнородные величины, но их недостаток состоит в том, что при этом необходимо учитывать фактор инфляции, поэтому статистика стоимостные величины всегда пересчитывает в сопоставимых ценах.

Виды абсолютных величин

Абсолютные величины могут быть моментными или интервальными.

Моментные абсолютные величины показывают уровень изучаемого явления или процесса на определенный момент времени или дату. Например, количество денег в кошельке или число упаковок препарата в аптеке на первое число месяца.

Интервальные абсолютные величины — это итоговый накопленный результат за определенный период (интервал) времени (например, зарплата за месяц, квартал или год, число заболевших за год и др.). Интервальные абсолютные величины, в отличие от моментных, допускают последующее суммирование.

Абсолютные статистические величины **не сопоставимы** и не дают полного представления об изучаемом явлении, так как не показывают его динамику, структуру, соотношение между частями.

Для этих целей служат **относительные статистические** величины и средние величины.

Относительные величины

Относительной величиной (статистическим коэффициентом, показателем, индексом) называется отношение двух абсолютных величин (чисел), выражающих меру каких-либо явлений. Смысл получения относительных величин заключается в нахождении общей меры и приведение к общему знаменателю, что позволяет сопоставлять уровень явления в нескольких группах.

Если соотносятся абсолютные величины с одинаковой размерностью, то получаемая относительная величина будет безразмерной (размерность сократится).

На практике применяется *искусственная размерность коэффициентов*.

Она получается путем их умножения **на основание коэффициента**:

на 100 - **проценты** (%);

на 1 000 - **промилле** (‰);

на 10 000 - **продецимилле** (‱).

на 100 000 - **просантимилле** (‱).

Вместо названия *относительная статистическая величина* часто используется более краткий термин-синоним — **индекс** (от лат. *index* — показатель, коэффициент).

Виды относительных величин

- **Интенсивные:** общие и специальные
- **Экстенсивные**
- **Показатели соотношения**
- **Показатели наглядности**
- **Показатели относительной интенсивности**
- **Динамики:** коэффициент роста и темп прироста

Интенсивные коэффициенты

Интенсивные коэффициенты показывают частоту (уровень, распространённость) явления в среде которая производит это явление. Они отвечают на вопрос: как часто явление встречается в известной среде?

Основанием чаще всего является 1000 (промилле), но могут использоваться проценты, продецимилле и другие основания в зависимости от размерности.

Расчет производится через пропорцию.:

Уровень явления	-	уровень среды
P (коэффициент)	-	основание (100, 1000 или др.)

Интенсивный показатель остаётся величиной с единицами измерения (случаи заболеваний, рождений, смертей и т. д.)

$$P = \frac{\text{Уровень явления} \times 1000}{\text{Уровень среды}}$$

Общие интенсивные коэффициенты характеризуют интенсивность явления в целом.

Специальные - дают более детальную характеристику явлению.

Интенсивные коэффициенты

Например: в N районе в 2010 году проживало 60 тысяч человек, и было зарегистрировано 1800 случаев инфекционных болезней, в т. ч. 60 случаев инфекционного гепатита.

Общий интенсивный коэффициент (интенсивность явления в целом) – распространённость всех инфекционных болезней

Составляем пропорцию:

1800 случаев инфекционных заболеваний – на 60000 человек

P случаев инфекционных заболеваний – на 1000 человек (основание)

$$P = \frac{1800 \times 1000}{60000} = 30,0 \text{ ‰}$$

Т.е. 30 случаев инфекционных заболеваний на 1000 жителей за 2010 год.

Обратите внимание: случаи заболеваний и число жителей района берутся за один и тот же год.

Специальный интенсивный коэффициент (более детальная характеристика явления) – распространённость инфекционного гепатита.

$$P = \frac{60 \times 1000}{60000} = 1,0 \text{ ‰}$$

Т.е. 1 случай гепатита на 1000 жителей района в 2010 году.

Экстенсивные коэффициенты

Экстенсивные коэффициенты – отражают структуру (долю, удельный вес, распределение) явления.

Вычисляются как отношение части статистической совокупности к целой совокупности, т.е. отношение отдельного элемента к итогу.

Например, в структуре инфекционной заболеваемости жителей N района в 2010 году доля инфекционного гепатита среди всех инфекционных заболеваний составила:

$$P = \frac{\text{число случаев инф.гепатита}}{\text{число всех случаев инф.заболеваний}} \times 100\% = \frac{60 \times 100\%}{1800} = 3,3\%$$

Другие коэффициенты

Коэффициенты соотношения – относительная величина, указывающая соотношение уровня изучаемого явления с уровнем другого явления, принятого за соизмеритель. Вычисляются через пропорцию, на 100, 1000, 10000.

Например: обеспеченность населения больничными койками, врачами, соотношение средних медработников и врачей, обеспеченность жилой площадью на душу населения, и т. д.

$$\text{Обеспеченность населения койками} = \frac{\text{число коек}}{\text{численность населения}} \times 1000$$

Коэффициенты относительной интенсивности применяются, когда невозможно получить прямые интенсивные коэффициенты. Например, когда нет точных данных о составе населения.

Например, известно, что среди обратившихся за медицинской помощью в связи с травмами мужчины составили 51%, а женщины 49%. Чтобы сделать заключение о более частом травматизме у мужчин, надо рассчитать число травм на 1000 мужчин и, соответственно, на 1000 женщин в районе. В случае применения коэффициентов относительной интенсивности можно ограничиться только учётом структуры населения.

Пол	Обратились по поводу травм (%)	Состав населения (%)	Коэффициент относительной интенсивности
	А		В
Муж.	51	60	0,85
Жен.	49	40	1,23
Оба пола	100	100	-

Динамические коэффициенты

Коэффициент наглядности – используется для характеристики динамики явления, даёт более отчётливое представление о характере изменения явления во времени.

Выражаются в процентах и вычисляются от исходного уровня, принимаемого за 100%. Поскольку эти коэффициенты выражены в %, их можно использовать для сравнения числовых рядов, которые состоят из разнородных величин.

Коэффициент роста (темп роста) - показывает во сколько раз изменилось изучаемое явление или процесс во времени. Рассчитывается как отношение значения абсолютной величины в отчетный (анализируемый) период или момент времени к базисному (предыдущему):

$$i_{\mathcal{D}} = \frac{X_1}{X_0}$$

Где : 1 — отчетный (анализируемый) период, 0 — базисный (прошлый) период.

Темп изменения (темпом прироста) - $T = i_{\mathcal{D}} - 1$

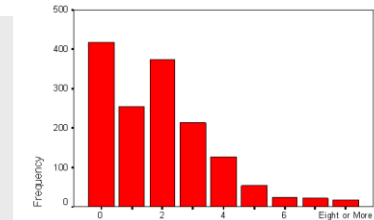
Если $T > 0$, то имеет место рост явления; $T = 0$ – стабильность, $T < 0$ — спад

Графическое изображение

ОТНОСИТЕЛЬНЫХ ВЕЛИЧИН

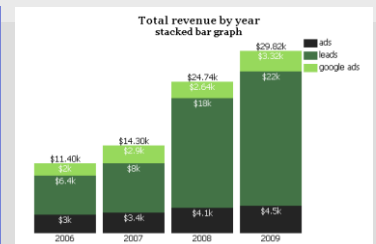
• Интенсивные показатели, соотношения, наглядности:

- Столбиковая диаграмма
- Линейная диаграмма



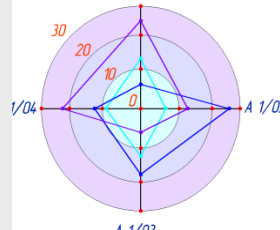
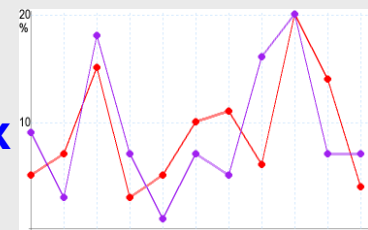
• Экстенсивные показатели:

- Внутри столбиковая диаграмма
- Секторная диаграмма

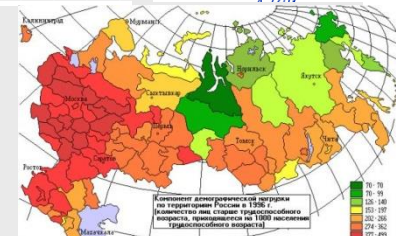


• Динамические показатели:

- Линейная диаграмма (график)
- Радиальная диаграмма (для циклических процессов)



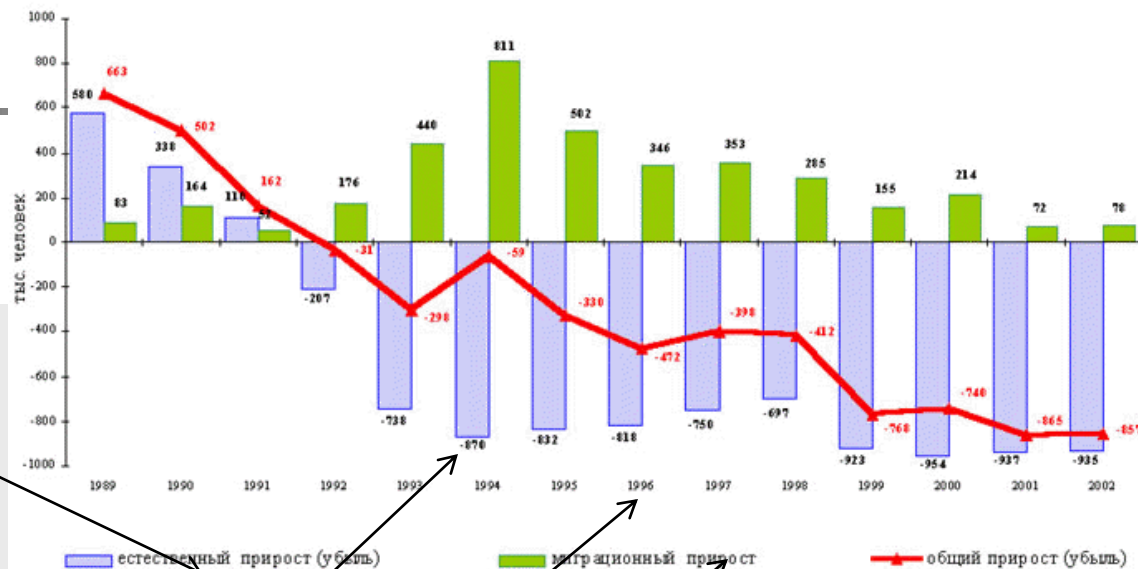
• Картограмма – это географическая карта с условными обозначениями различным цветом или штриховкой



• Картодиаграмма – это географическая карта с изображением на ней различных диаграмм



Правила графического изображение относительных величин



- Название диаграммы, располагается ниже рисунка

- Номер рисунка (если их несколько)

Рис 1. Годовой прирост населения России с 1989 г. по 2002 г.

- Наличие количественного определителя представленных явлений рядом с диаграммой.

- Обозначение осей на графиках

- Условные обозначения (легенда)

Средние величины

Средняя величина - это совокупная обобщающая характеристика количественного признака.

Обозначается буквой M в медицинской статистике
или χ в математической статистике.

К средним величинам относятся:

- Средняя арифметическая (M)
- Мода (Mo)
- Медиана (Me)

Использование средних величин:

- изучение состояния здоровья населения;
- анализ демографической ситуации;
- оценка деятельности лечебных учреждений;
- санитарно-эпидемиологические исследования;
- изучение экономических аспектов здравоохранения;
- научные медико-биологические исследования на человеке и животных.



Методика вычисления средних величин

1. Построение простого вариационного ряда
2. Ранжирование вариационного ряда
3. Группировка ряда

Вариационный ряд – это однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления.

Обозначения элементов вариационного ряда:

V - варианта

p - частота

n - общее число наблюдений

Например: Простой вариационный ряд
Измерения частоты сердечных сокращений
студентов перед экзаменом (уд/мин).

	V
1. Иванов	72
2. Сидоров	80
3. Казаков	69
4. ...	80
5. ...	76
	$n=5$

Виды вариационных рядов

1. **Простой** – каждая варианта записывается один раз
2. **Сгруппированный** – содержит число повторений вариант, частоту – p
3. **Прерывный (дискретный)** – состоит из значений выраженных только определенными числами (баллами)
4. **Непрерывный** – содержит случайные величины выраженные любыми натуральным числами
5. **Интервальный** – разбит на интервалы значений признака от и до...

	V
1.Иванов	72
2.Сидоров	80
3.Казаков	69
4. ...	80
5. ...	76
$n=5$	

Простой

	V
1.Казаков	69
2.Иванов	72
3. ...	76
4.Сидоров	80
5. ...	80
$n=5$	

Ранжированный Сгруппированный

	V	p
1.	72	1
2.	80	2
3.	69	1
4. ...	76	1
$n=5$		

Вариационный ряд:

Определение средних величин

	V	p
1.	72	1
2.	80	2
3.	69	1
4. ...	76	1
n=5		

Мода (Mo) - соответствует величине признака, который чаще других встречается в данной совокупности.

Определяется в сгруппированном ряду по наибольшей частоте вариант.

Mo=80 Уд/мин.

Медиана (Me) – величина признака, занимающая срединное значение в данной совокупности. Она определяется в ранжированном ряду и делит ряд на 2 равные части по числу наблюдений.

Me=76 Уд/мин.

Средняя арифметическая (M) – это сумма вариант делённая на число единиц наблюдения.

$$M = \frac{\sum V}{n} = (69+72+76+80+80)/5 = 75,4 \text{ Уд/мин}$$

	V
1.Казаков	69
2.Иванов	72
3. ...	76
4.Сидоров	80
5. ...	80
n=5	

Способы вычисления средней арифметической (M)

	V
1. Казаков	69
2. Иванов	72
3. ...	76
4. Сидоров	80
5. ...	80
	$n=5$

Простая средняя – вычисляется в простом вариационном ряду

$$M = (69+72+76+80+80)/5 = 75,4 \text{ Уд/мин}$$

$$M = \frac{\sum V}{n}$$

Средняя взвешенная – вычисляется в сгруппированном ряду

$$M = (72*1+80*2+69*1+76*1)/5 = 75,4 \text{ Уд/мин}$$

$$M = \frac{\sum Vp}{n}$$

По способу моментов – используется при ручных вычислениях при большом объеме совокупности.

Где: A - условная средняя (мода или медиана)
 d - отклонение от условной средней $d=V-A$

	V	p	$d=V-A$
1.	72	1	72-80=-8
2.	80	2	80-80=0
3.	69	1	69-80=-11
4.	76	1	76-80=-4
		$n=5$	

$$M = A + \frac{\sum dp}{n}$$

$$M = 80 + (-8*1+0*2-11*1-4*1)/5=75,4 \text{ Уд/мин}$$

Свойства арифметической (M)



Основными свойствами средней арифметической величины являются...

а) Сумма всех отклонений от средней равна 0;

б) При умножении (делении) всех вариантов на один и тот же множитель (делитель) средняя арифметическая умножается (делится) на тот же множитель (делитель);

в) Если прибавить (вычесть) ко всем вариантам одно и то же число, средняя арифметическая увеличивается (уменьшается) на то же число.

Вариабельность признака

Это характеристика признака, которая показывает степень его изменения в вариационном ряду.

При равных средних величинах в 2-х и более рядах их вариабельность может отличаться!

	V
1. Иванов	72
2. Сидоров	80
3. Казаков	69
4. ...	80
5. ...	76
n=5	
M=	75,4
V_{min}=	69
V_{max}=	80
Am=	11
σ=	4,88

Параметры вариабельности:

- Минимум: V_{min}
- Максимум: V_{max}
- Амплитуда: $Am = V_{max} - V_{min}$
- Размах: от V_{max} до V_{min}
- Среднее квадратическое отклонение – σ (сигма)
- Дисперсия - σ^2
- Коэффициент вариации – C

	V
1. Васильев	73
2. ...	71
3. ...	75
4. ...	79
5. ...	79
n=5	
M=	75,4
V_{min}=	71
V_{max}=	79
Am=	8
σ=	3,58

Вычисление среднего

квадратического отклонения

Среднее квадратическое отклонение характеризует степень разброса вариант вариационного ряда от средней арифметической.

Для простого ряда: $\sigma = \sqrt{\frac{\sum d^2}{n-1}}$

Для сгруппированного: $\sigma = \sqrt{\frac{\sum d^2 p}{n-1}}$

	V	d=V-M	d ²
1. Иванов	72	-6,4	41
2. Сидоров	80	-3,4	11,6
3. Казаков	69	0,6	0,36
4. ...	80	4,6	21,2
5. ...	76	4,6	21,2
n=5			Σ=95,4
M=	75,4		
σ=	4,88	$= \sqrt{\frac{95,4}{4}}$	

	V	d=V-M	d ²
1. Васильев	73	-2,4	5,76
2. ...	71	-4,4	19,4
3. ...	75	-0,4	0,16
4. ...	79	3,6	13
5. ...	79	3,6	13
n=5			Σ=51,2
M=	75,4		
σ=	3,58	$= \sqrt{\frac{51,2}{4}}$	

Вычисление коэффициента вариации

Коэффициент вариации $C = \frac{\sigma}{M} * 100\%$

- Позволяет сравнивать вариабельность различных по природе и размерности признаков.

	V
1. Иванов	72
2. Сидоров	80
3. Казаков	69
4. ...	80
5. ...	76
n=5	
M=	75,4
σ=	4,88
C=	6,5%

Степень рассеяния вариант вокруг средней арифметической, если значение коэффициента вариации является ...

Малой - до 10%,
Средней от 10 до 20%,
Сильной более 20%.

$$C = \frac{4,88}{75,4} * 100\%$$

	V
1. Васильев	73
2. ...	71
3. ...	75
4. ...	79
5. ...	79
n=5	
M=	75,4
σ=	3,58
C=	4,7%

$$C = \frac{3,58}{75,4} * 100\% = 4,7\%$$

Правило 3-х сигм

Это вероятностная зависимость между значением средней арифметической, средним квадратическим отклонением и вариантами. Применяется для прогнозирования значений средней величины в случае повторения опыта.

Согласно теории вероятности для учетных признаков, подчиняющихся нормальному закону распределения (распределение Гаусса), между значениями средней арифметической, среднеквадратическим отклонением и вариантами существует строгая зависимость .

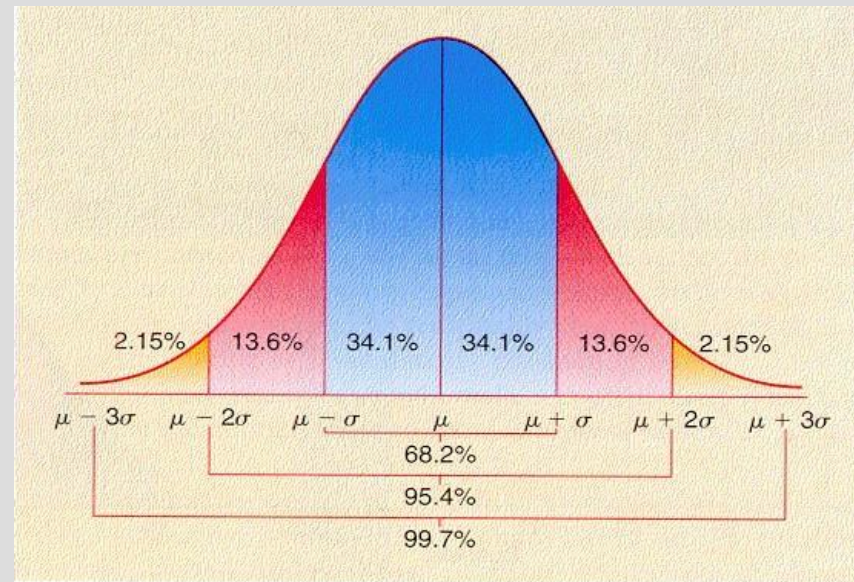
- 68,2% значений такого признака будут находиться в пределах $M \pm 1\sigma$
- 95,4% — в пределах $M \pm 2\sigma$ – степень вероятности принятый в медико-биологических экспериментах
- 99,7% — в пределах $M \pm 3\sigma$

Это значит, что проведя опыт на выборочной совокупности можно прогнозировать значения этого признака при повторных измерениях с необходимым уровнем вероятности.

Например: 95% студентов должны иметь пульс перед экзаменом

$75,4 \pm 2 \cdot 4,88$, т.е. от 66 до 85 Уд./мин.

4.88



Ошибка репрезентативности

Чтобы определить степень точности выборочного наблюдения, необходимо оценить величину ошибки, которая может случайно произойти в процессе выборки. Такие ошибки называются ошибками репрезентативности m (или средней ошибкой средней арифметической, стандартной ошибкой).

	V
1. Иванов	72
2. Сидоров	80
3. Казаков	69
4. ...	80
5. ...	76
$n=5$	
$M=$	75,4
$\sigma=$	4,88
$m=$	2,18

Определяется по формуле: $m = \frac{\sigma}{\sqrt{n-1}}$

Например: $m = \frac{4,88}{\sqrt{4}} = 2,44$

Для генеральной совокупности требуется выбрать вероятность прогноза. В научной литературе используют среднюю $\pm 2 \cdot m$ при вероятности 95,5%.

Это значит, что в генеральной совокупности у следующей группы студентов сдающих экзамен средний пульс будет составлять $75,4 \pm 2,2$ Уд./мин. при уровне вероятности 65% и $75,4 \pm 4,4$ Уд./мин. с вероятностью 95,5%

5.

Медицинская информатика



**СПАСИБО ЗА
ВНИМАНИЕ!**

ДО СВИДАНИЯ!