



ВВЕДЕНИЕ В БИОМЕДИЦИНСКУЮ СТАТИСТИКУ

МОРДОВСКИЙ ЭДГАР АРТУРОВИЧ
К.М.Н., ДОЦЕНТ

ПЛАН

- Биомедицинская статистика: назначение, задачи
- Статистическое исследование
- Типы переменных в статистике
- Относительные величины в медицинской статистике
- Анализ динамических рядов



МЕДИЦИНСКАЯ СТАТИСТИКА: НАЗНАЧЕНИЕ, ЗАДАЧИ

СТАТИСТИКА

отрасль знаний, в которой излагаются общие вопросы **сбора, измерения и анализа** массовых статистических (количественных или качественных) данных



«СТАТИСТИКА» (от лат. «status» - состояние, положение)

Впервые термин применил **Готфрид Ахенвалль** (сер. XVIII в.) при описании **состояния государства** (нем. statistik, от итал. stato - государство)

СТАТИСТИКА

изучает **КОЛИЧЕСТВЕННУЮ СТОРОНУ** массовых общественных явлений
в неразрывной связи с их **КАЧЕСТВЕННОЙ СТОРОНОЙ**

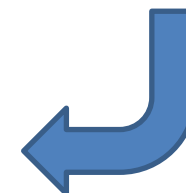


ДАННЫ

Е



ИНФОРМАЦИЯ

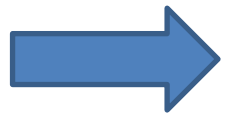


**ОБОСНОВАННЫЕ
КЛИНИЧЕСКИЕ, УПРАВЛЕНЧЕСКИЕ
РЕШЕНИЯ**

Закон больших чисел

ЗАКОН БОЛЬШИХ ЧИСЕЛ: количественные закономерности массовых явлений отчетливо проявляются лишь на достаточно большом числе единиц наблюдения

Сущность его заключается в том, что в числах, получающихся в результате массового наблюдения, выступают определенные закономерности, которые не могут быть обнаружены в небольшом числе фактов



2 следствия

Закон больших чисел

следствие 1

ЭКОЛОГИЧЕСКАЯ ОШИБКА (ecological fallacy) – результаты, полученные на выборке, нельзя экстраполировать на единицу наблюдения

Закономерности, обнаруженные на выборке, не могут являться императивом действий с конкретным человеком (пациентом)

АТОМИСТИЧЕСКАЯ ОШИБКА (atomistic fallacy) – данные, полученные на единице наблюдения, нельзя экстраполировать на выборку

Мнение одного эксперта не должно быть императивом действий на выборке

Закон больших чисел

следствие 2

Манипулируйте единицами наблюдения и получайте **КАКИЕ УГОДНО** результаты

«...цифры обманчивы, особенно когда я сам ими занимаюсь; по этому поводу справедливо высказывание, приписываемое Дизраэли:
«существует три вида лжи: ложь, наглая ложь и статистика».

СТАТИСТИЧЕСКИЕ МЕТОДЫ

Статистические методы - методы анализа статистических данных

**методы прикладной
статистики**

применяются во всех областях
научных исследований
и любых отраслях экономики
и управления

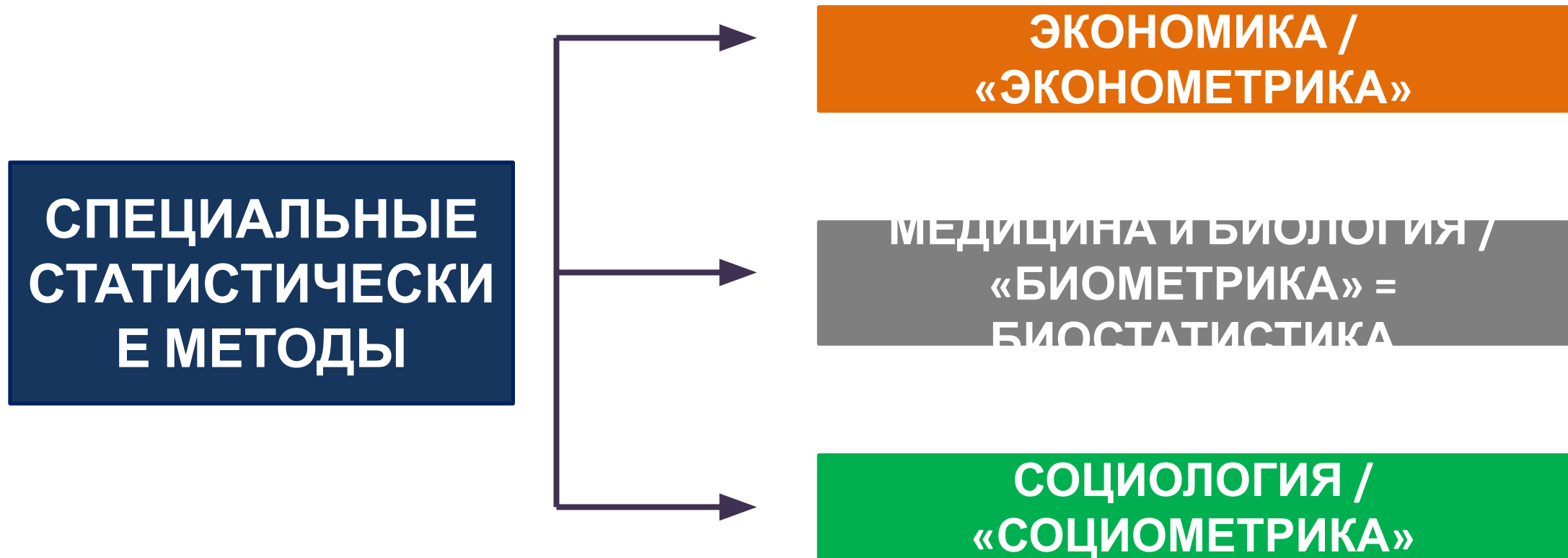
**специальные методы
статистики**

применяются в отдельных
областях научных исследований

+

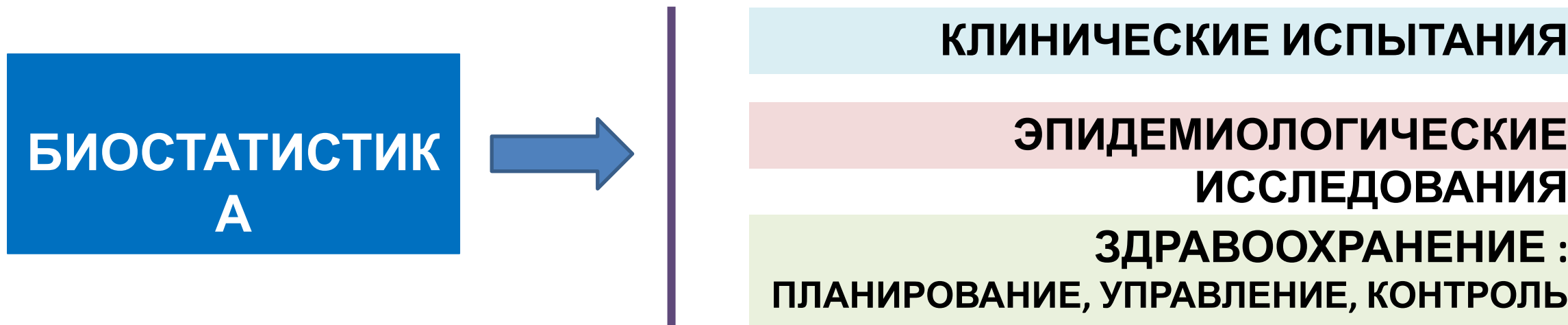
**ПЛАНИРОВАНИЕ
ЭКСПЕРИМЕНТОВ**

СТАТИСТИЧЕСКИЕ МЕТОДЫ



БИОМЕТРИКА или БИОСТАТИСТИКА

Биометрия (биологическая статистика / биостатистика) — научная отрасль на стыке биологии и вариационной статистики, связанная с **разработкой и использованием статистических методов в научных исследованиях** (как при планировании количественных экспериментов, так и при обработке экспериментальных данных и наблюдений) **в биологии, медицине, здравоохранении и эпидемиологии**



The background of the slide is a complex financial chart. It features a blue grid with a yellow line graph that rises from the bottom left and then curves downwards. A purple line graph also trends downwards from the top left. In the upper right corner, there is a candlestick chart with blue and white bars. Various numerical values are scattered across the chart, including '88.76', '56.91', '34.05', '01', '02', '03', '04', '05', '06', '07', '08', '09', '11', '1.683', and '-2.10'.

ИНФОРМАЦИОННАЯ БАЗА СИСТЕМЫ ЗДРАВООХРАНЕНИЯ

Статистическая система здравоохранения: **ИНФОРМАЦИОННАЯ БАЗА**

Федеральные
статистические
наблюдения (1)

The diagram consists of two large, dark red arrows pointing in opposite directions, connected by a white wavy line. The left arrow points left and contains the text 'Федеральные статистические наблюдения (1)'. The right arrow points right and contains the text 'Выборочные / сплошные статистические исследования (2)'.

Выборочные /
сплошные
статистические
исследования (2)

Статистическая система здравоохранения.

ИНФОРМАЦИОННАЯ БАЗА (1). ФЕДЕРАЛЬНЫЕ СТАТИСТИЧЕСКИЕ НАБЛЮДЕНИЯ

Статистическое наблюдение – это массовое, планомерное, научно-организованное наблюдение за явлениями социальной и экономической жизни, которое заключается в регистрации отобранных признаков у каждой единицы совокупности

Статистическая система здравоохранения.

ИНФОРМАЦИОННАЯ БАЗА (1). ФЕДЕРАЛЬНЫЕ СТАТИСТИЧЕСКИЕ НАБЛЮДЕНИЯ

В отечественной статистике используются *три организационные формы (типа) статистического наблюдения*:

- **ОТЧЕТНОСТЬ** (предприятий, организаций, учреждений и т. п.)
- **СПЕЦИАЛЬНО ОРГАНИЗОВАННОЕ СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ** (переписи, единовременные учеты, обследования сплошного и несплошного характера)
- **РЕГИСТРЫ**

Статистическая система здравоохранения: ИНФОРМАЦИОННАЯ БАЗА (2)



**Недооцененный
источник
статистических
данных (в России)**

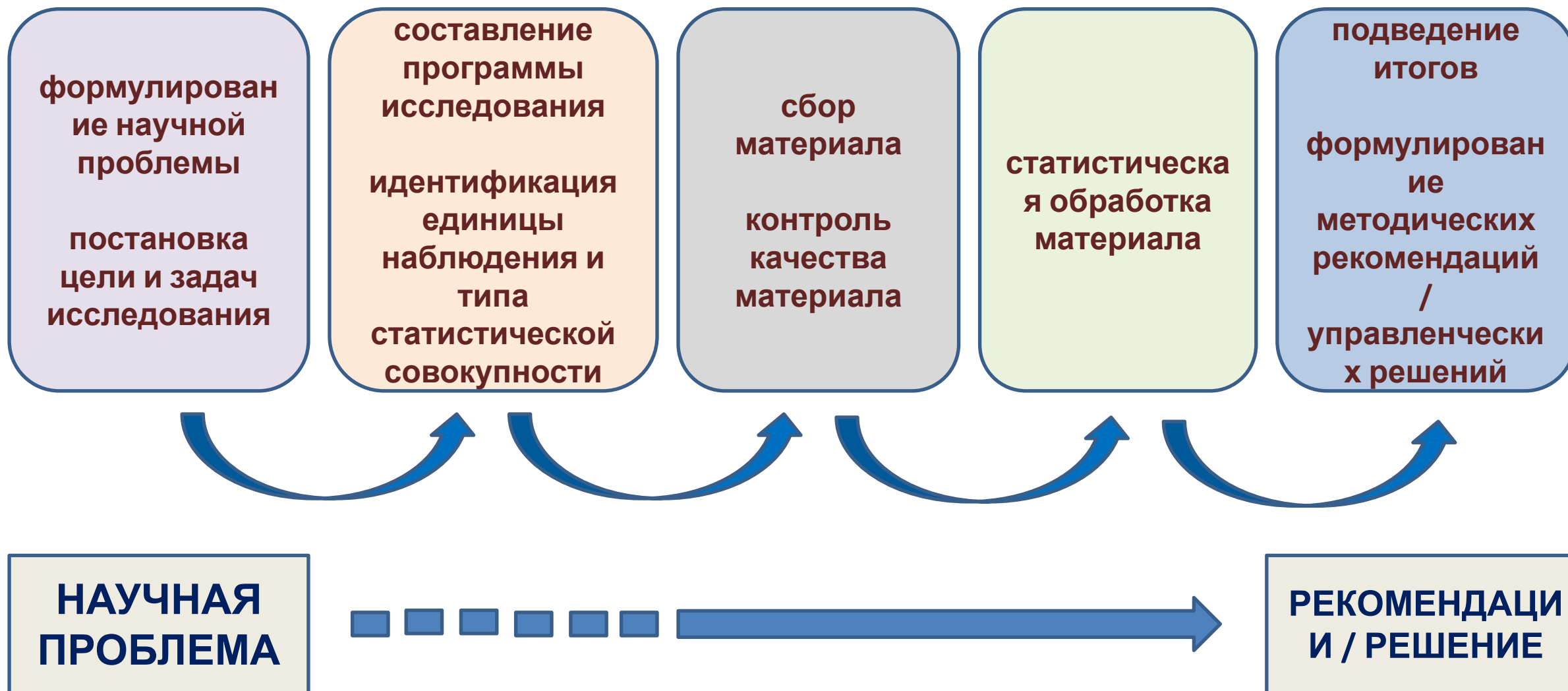
**Основной источник
статистических
данных за рубежом**

**ПРИЧИНА: дефицит
знаний, дорогие стат.
программы**

СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ

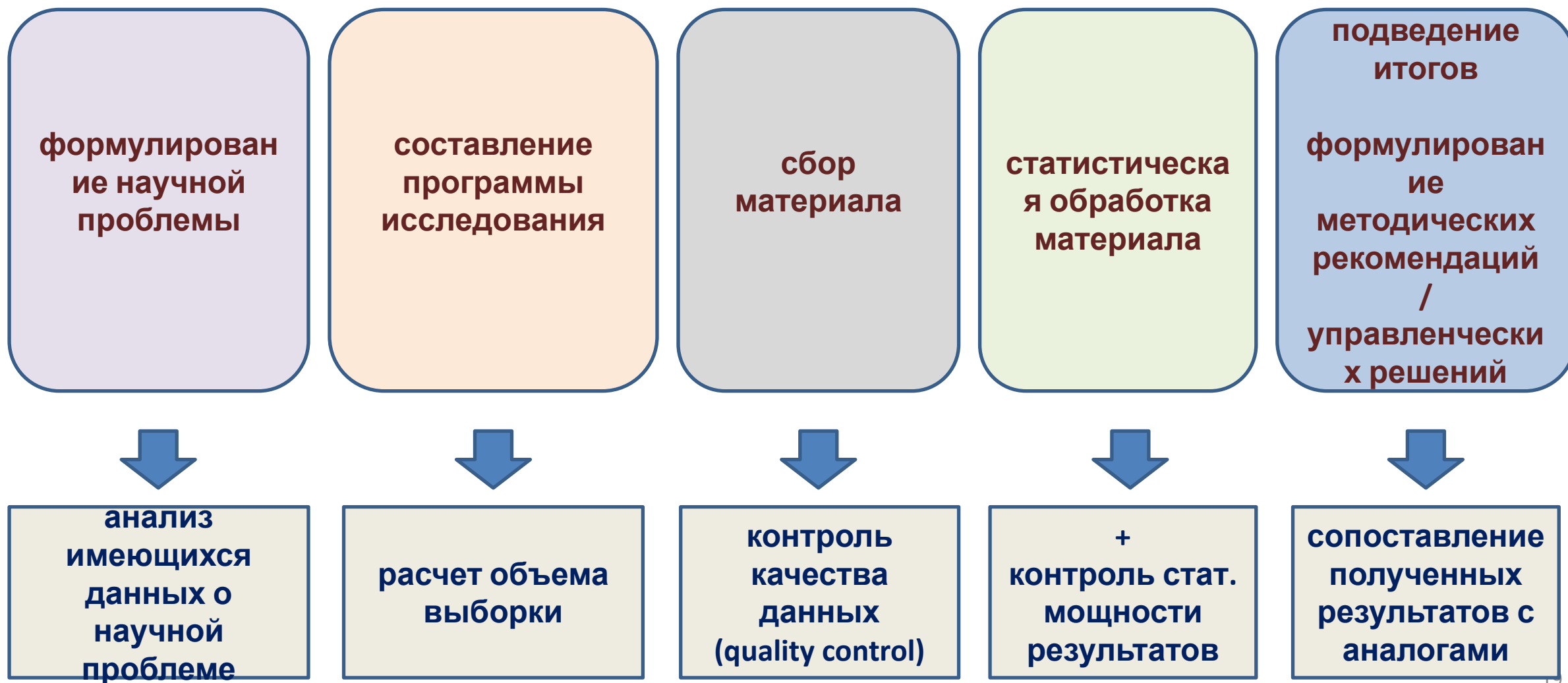
СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ — это научно организованный по единой программе сбор, сводка и анализ данных (фактов) о социально-экономических, демографических и других явлениях и процессах общественной жизни в государстве с регистрацией их наиболее существенных признаков в учетной документации

СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ



СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ:

где нужны знания статистики ???





ТИПЫ ПЕРЕМЕННЫХ В СТАТИСТИКЕ

ТИП ПЕРЕМЕННЫХ ОПРЕДЕЛЯЕТ НАБОР СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА

ПЕРЕМЕННЫЕ

КОЛИЧЕСТВЕННЫЕ

ДИСКРЕТНЫЕ
(DISCRETE)

НЕПРЕРЫВНЫЕ
(CONTINUOUS)

ЧИСЛОВОЕ ОБОЗНАЧЕНИЕ ИМЕЕТ
СОДЕРЖАТЕЛЬНЫЙ СМЫСЛ

КАТЕГОРИАЛЬНЫЕ

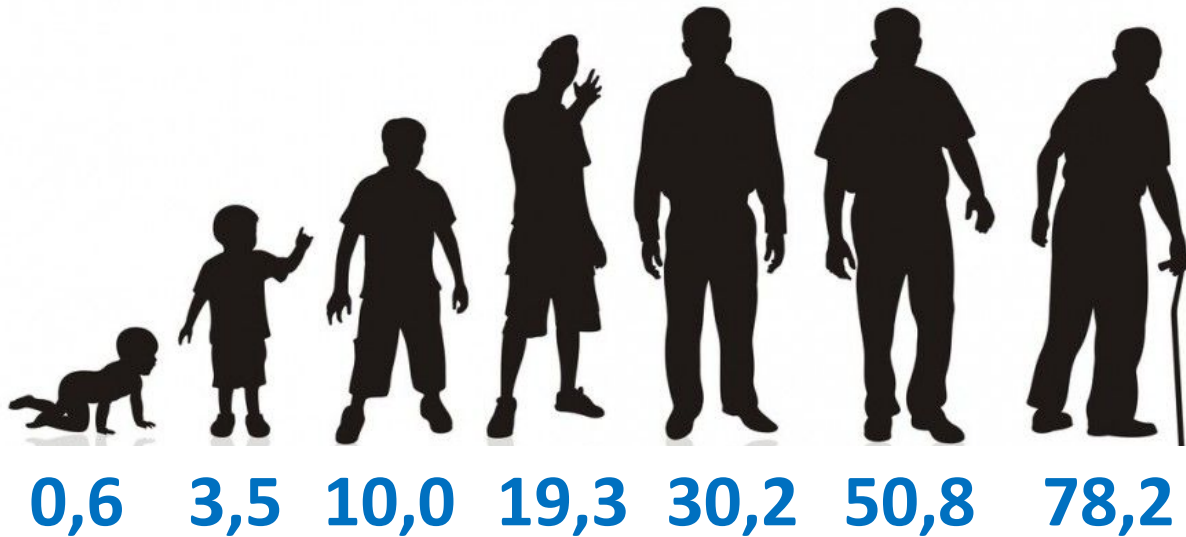
ПОРЯДКОВЫЕ
(ORDINAL)

НОМИНАЛЬНЫЕ
(NOMINAL)

ЧИСЛОВОЕ ОБОЗНАЧЕНИЕ НЕ ИМЕЕТ
СОДЕРЖАТЕЛЬНЫЙ СМЫСЛ

КОЛИЧЕСТВЕННЫЕ НЕПРЕРЫВНЫЕ ПЕРЕМЕННЫЕ

МОГУТ ПРИНИМАТЬ ЛЮБЫЕ ЗНАЧЕНИЯ [В НЕКОТОРОМ ЗАДАННОМ ИНТЕРВАЛЕ]



ПРИМЕРЫ:

- Возраст
- Масса
- Рост
- АД
- Биохимическ
ие

показатели

НАИБОЛЕЕ «ПЕРСПЕКТИВНЫЕ» :

- м.б. трансформированы в любые другие переменные
- максимальное количество статистических методов

КОЛИЧЕСТВЕННЫЕ ДИСКРЕТНЫЕ ПЕРЕМЕННЫЕ

ЯВЛЯЮТСЯ ЗНАЧЕНИЯМИ ПРИЗНАКА,
КОТОРЫЕ МОГУТ БЫТЬ ПОДСЧИТАНЫ С ПОМОЩЬЮ НАТУРАЛЬНЫХ
ЧИСЕЛ



ПРИМЕРЫ:

- Число беременностей
- Паритет
- Число детей
- Число вызовов

«МАЛОПЕРСПЕКТИВНЫЕ» :

- относительно редко используются в статистике
- можно УСЛОВНО считать непрерывными, если $N > 20$ (700)

КАТЕГОРИАЛЬНЫЕ **НОМИНАЛЬНЫЕ** ПЕРЕМЕННЫЕ

ДАННЫЕ, ДЛЯ КОТОРЫХ **НЕТ**
СОДЕРЖАТЕЛЬНОГО ИНТЕРПРЕТИРУЕМОГО ПОРЯДКА



«0»



«1»

ПРИМЕРЫ:

- Пол
- Профессия
- Расовая принадлежность

«ПЕРСПЕКТИВНЫЕ» :

- все чаще используются в статистике – идет развитие соответствующих методов, в т.ч. регрессионного анализа

КАТЕГОРИАЛЬНЫЕ ПОРЯДКОВЫЕ ПЕРЕМЕННЫЕ

ДАННЫЕ, ДЛЯ КОТОРЫХ ЕСТЬ
СОДЕРЖАТЕЛЬНЫЙ ИНТЕРПРЕТИРУЕМЫЙ ПОРЯДОК



ПРИМЕРЫ:

- Уровень
образования

«4» - не в 4 раза лучше, чем
«1»

«ПЕРСПЕКТИВНЫЕ» :

- все чаще используются в статистике – идет развитие соответствующих методов, в т.ч. регрессионного анализа

ПРАВИЛА ОФОРМЛЕНИЯ БАЗЫ ДАННЫХ

SPSS Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window

4 :

	Возраст	Пол	Образован ие	var	var
1	,60	0	1		
2	,80	0	2		
3	3,60	1	3		
4	10,60	1	2		
5	12,60	0	3		
6	12,80	1	1		
7	17,60	1	1		
8	18,30	1	1		
9	19,10	0	4		
10	19,90	0	3		
11					
12					

**1 СТРОКА = 1 ЕДИНИЦА
НАБЛЮДЕНИЯ**

**В 1 ЯЧЕЙКЕ = 1 ЧИСЛО / ЗНАЧЕНИЕ
ПЕРЕМЕННОЙ**



**СПОСОБЫ ПРЕЗЕНТАЦИИ ДАННЫХ В
СТАТИСТИКЕ (ДИСКРИПТИВНАЯ /
ОПИСАТЕЛЬНАЯ СТАТИСТИКА)**

**ТИП ПЕРЕМЕННЫХ ОПРЕДЕЛЯЕТ НАБОР СТАТИСТИЧЕСКИХ МЕТОДОВ
АНАЛИЗА**

ПЕРЕМЕННЫЕ

КОЛИЧЕСТВЕННЫЕ

**ДИСКРЕТНЫ
Е**
(DISCRETE)

**НЕПРЕРЫВН
ЫЕ**
(CONTINUOUS)

**ЧИСЛОВОЕ ОБОЗНАЧЕНИЕ ИМЕЕТ
СОДЕРЖАТЕЛЬНЫЙ СМЫСЛ**

КАТЕГОРИАЛЬНЫЕ

**ПОРЯДКОВЫ
Е**
(ORDINAL)

**НОМИНАЛЬНЫ
Е**
(NOMINAL)

**ЧИСЛОВОЕ ОБОЗНАЧЕНИЕ НЕ ИМЕЕТ
СОДЕРЖАТЕЛЬНЫЙ СМЫСЛ**

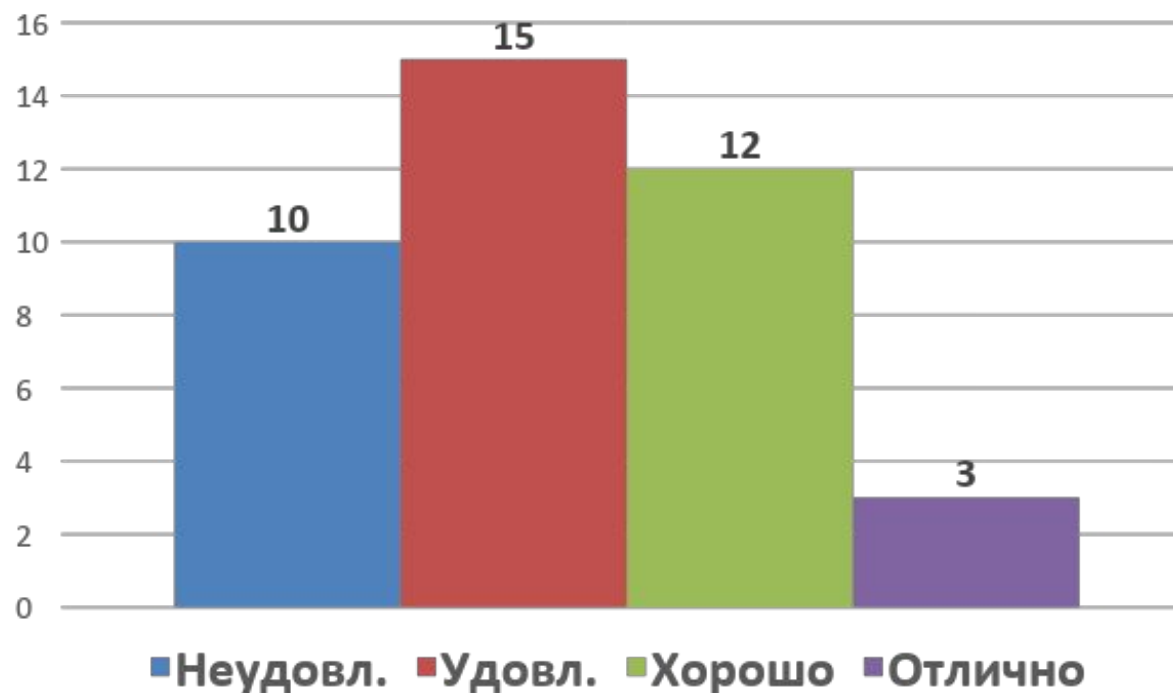
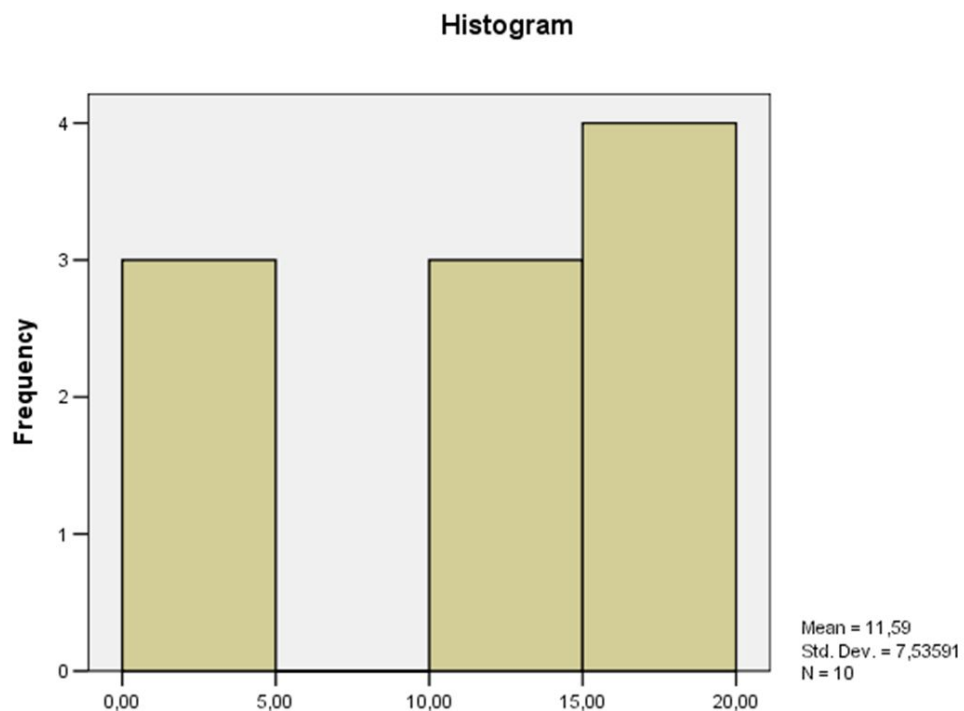
СПОСОБЫ ПРЕЗЕНТАЦИИ ДАННЫХ В СТАТИСТИКЕ (ДИСКРИПТИВНАЯ / ОПИСАТЕЛЬНАЯ СТАТИСТИКА)

ТИП ПЕРЕМЕННОЙ	ЧАСТОТНОЕ РАСПРЕДЕЛЕНИ Е	СРЕДНИЕ ВЕЛИЧИНЫ	ТАБЛИЦЫ / ГРАФИКИ
НЕПРЕРЫВНЫЕ (CONTINUOUS)	+	+	+
ДИСКРЕТНЫЕ (DISCRETE)	+	-	+
ПОРЯДКОВЫЕ (ORDINAL)	+	-	+
НОМИНАЛЬНЫЕ (NOMINAL)	+	-	+

ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ

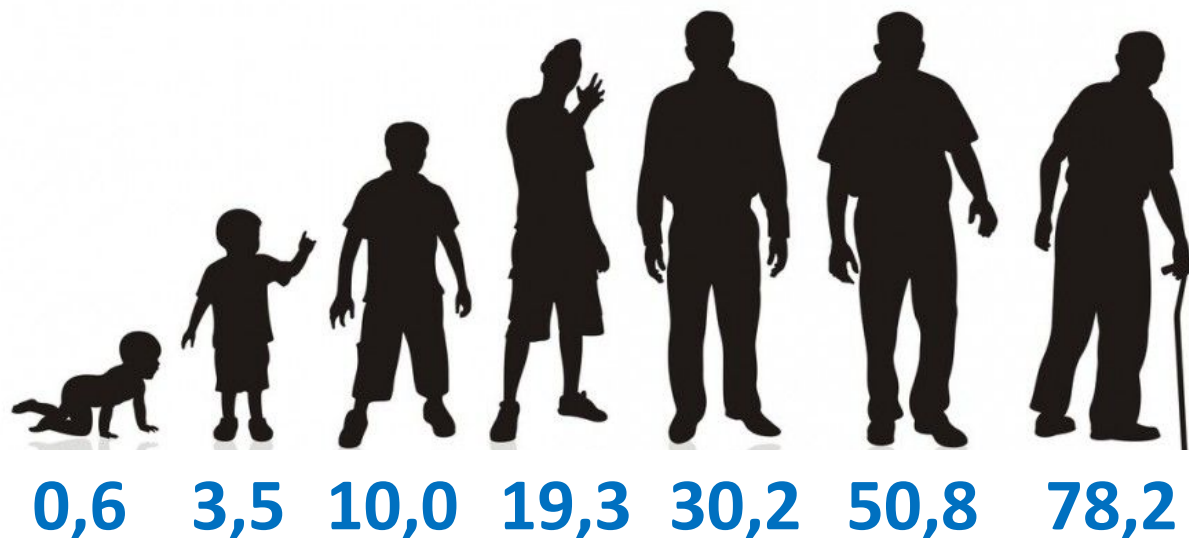
Частотное распределение переменной (frequency distribution)

ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ (*frequency distribution*) – обнаружение соответствия между значениями переменной и их вероятностями (частотой встречаемости)



ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВЕННЫХ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ

МОГУТ ПРИНИМАТЬ ЛЮБЫЕ ЗНАЧЕНИЯ *[В НЕКОТОРОМ ЗАДАННОМ ИНТЕРВАЛЕ]*



ПРИМЕРЫ:

- Возраст
- Масса
- Рост
- АД
- Биохимические показатели

ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВЕННЫХ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ

ЭТАПЫ ПОСТРОЕНИЯ ЧАСТОТНОГО РАСПРЕДЕЛЕНИЯ

1. Упорядочить (по возрастанию)
значения переменной



ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВЕННЫХ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ

ЭТАПЫ ПОСТРОЕНИЯ ЧАСТОТНОГО РАСПРЕДЕЛЕНИЯ

2. Разбить выборку на группы по равным интервалам (по формуле Стерджесса)

$$n = 1 + 3.322 \lg N$$

$$n = 1 + 3.322 \lg 10 = 4$$

n – число групп

N – количество единиц
наблюдения

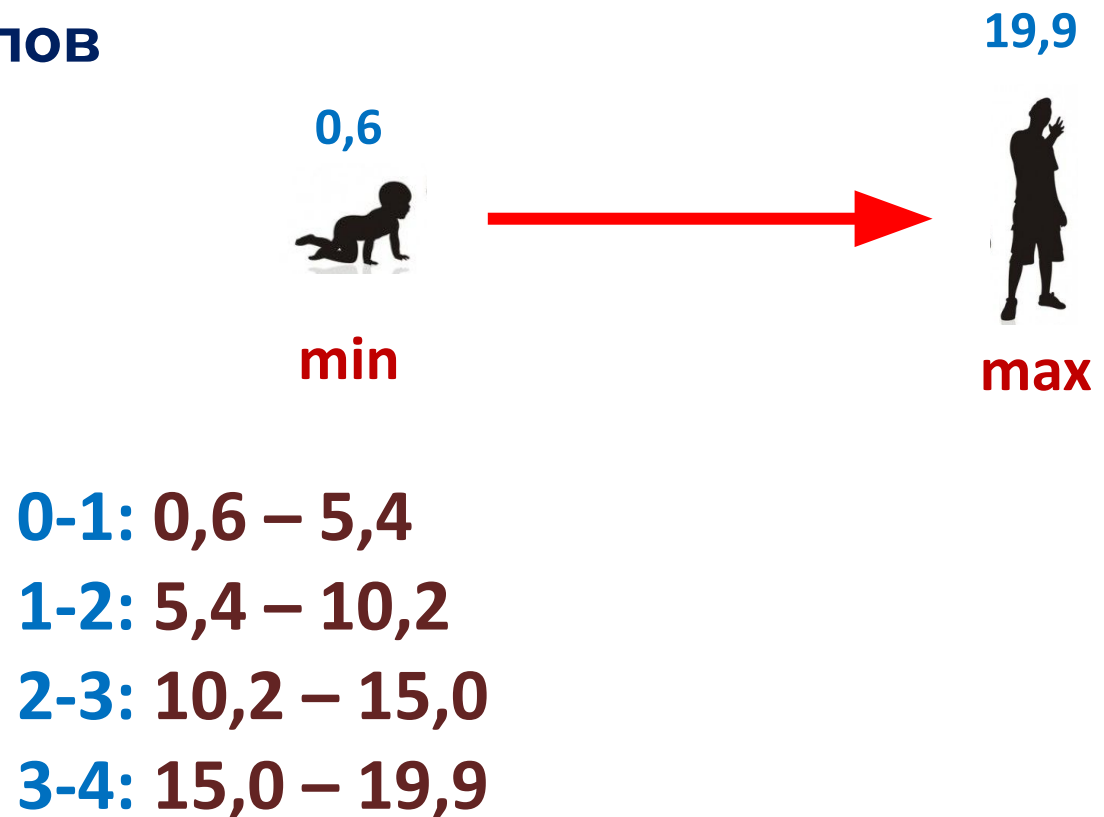
ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВЕННЫХ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ

ЭТАПЫ ПОСТРОЕНИЯ ЧАСТОТНОГО РАСПРЕДЕЛЕНИЯ

3. Рассчитать границы интервалов

$$\Delta = \frac{\max - \min}{n}$$

$$\Delta = \frac{19,9 - 0,6}{4} = 4,8$$



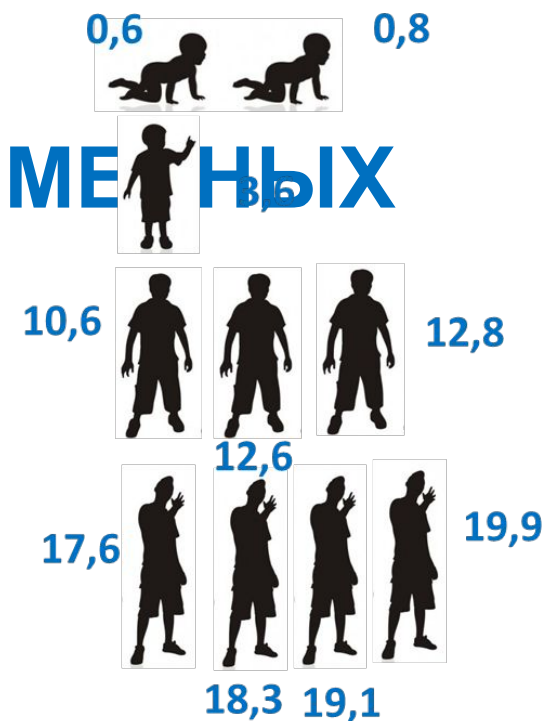
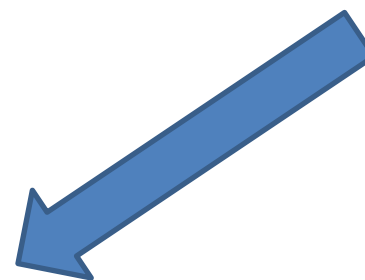
ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ КОЛИЧЕСТВЕННЫХ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ

Histogram



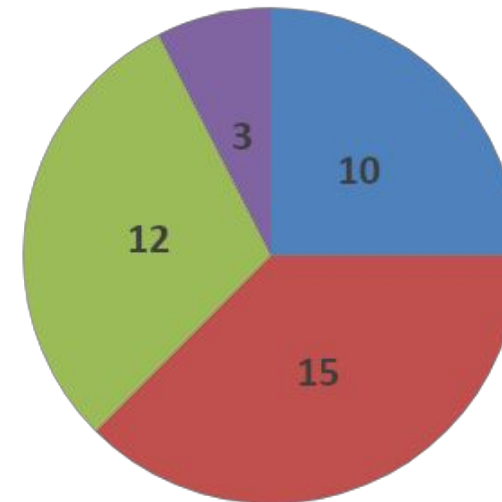
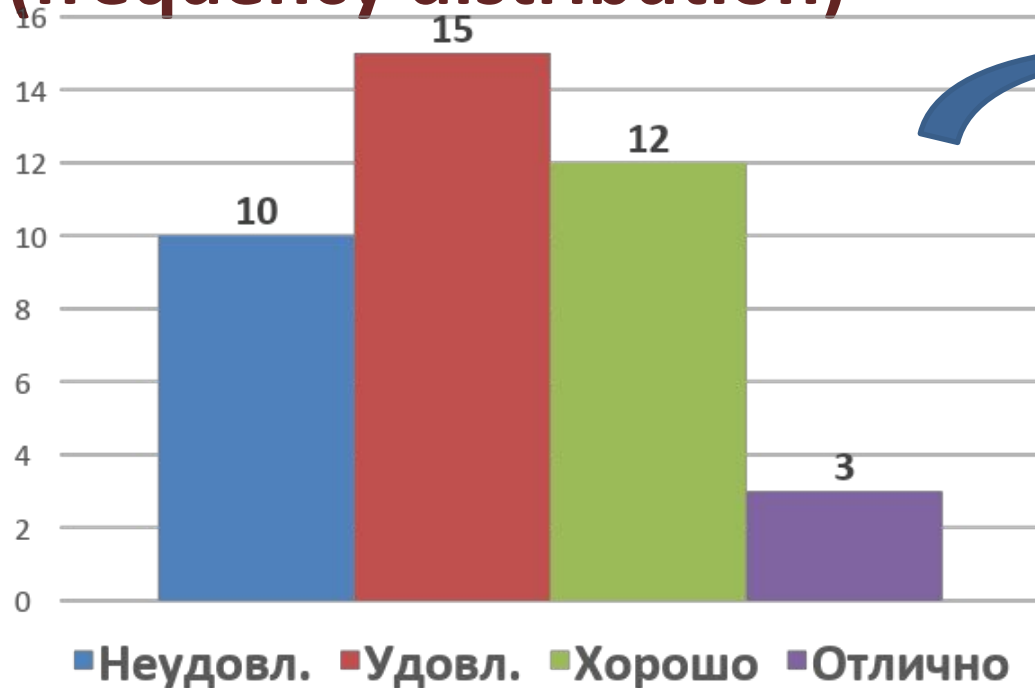
4

группы



ГИСТОГРАММА –
графическое
представление частотного
распределения, разбитого
по интервалам

Частотное распределение **ДИСКРЕТНОЙ, НОМИНАЛЬНОЙ И ПОРЯДКОВОЙ** переменных (frequency distribution)

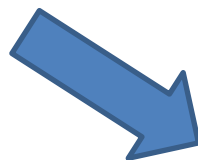


■ Неудовл. ■ Удовл. ■ Хорошо ■ Отлично

СРЕДНИЕ ВЕЛИЧИНЫ

ПРЕДСТАВЛЕНИЕ КОЛИЧЕСТВЕННЫХ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ В ВИДЕ СРЕДНИХ ВЕЛИЧИН

АНАЛИЗОМ СРЕДНИХ ВЕЛИЧИН
занимается
ВАРИАЦИОННАЯ СТАТИСТИКА



ВАРИАЦИОННАЯ СТАТИСТИКА – раздел статистики, изучающая теорию и методы изучения свойств массовых явлений, а также **вычисления и анализа их количественных характеристик**

ОСНОВНЫЕ ТЕРМИНЫ ВАРИАЦИОННОЙ СТАТИСТИКИ

- **ВАРИА(Н)ТА = ЕДИНИЦА НАБЛЮДЕНИЯ** - каждый первичный элемент статистической совокупности
- **СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ** - группа, состоящая из всех относительно однородных элементов (единиц наблюдения) отобранная в соответствии с характеристиками поставленной цели
- **(УЧЕТНЫЙ) ПРИЗНАК** - предмет исследования, которым обладают варианты

СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ может быть образована по одному или по нескольким **УЧЕТНЫМ ПРИЗНАКАМ**

ВАРИАЦИОННАЯ СТАТИСТИКА – помогает изучить свойства массовых явлений, вычислить и проанализировать их количественные характеристики

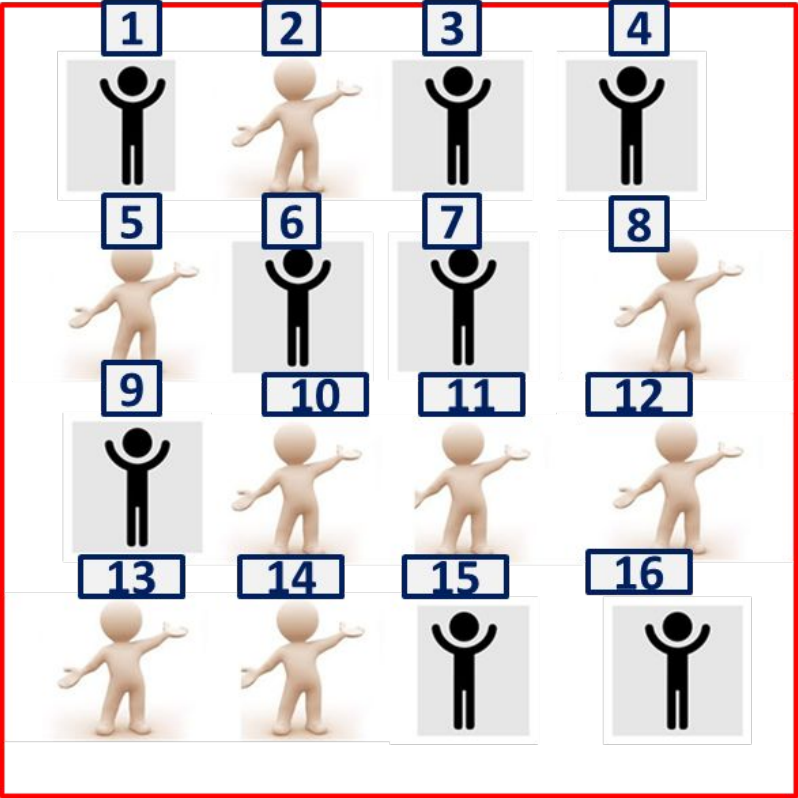


ВАРИАЦИОННЫЙ РЯД

Для того, чтобы изучить свойства статистической совокупности, необходимо построить:

ВАРИАЦИОННЫЙ РЯД — статистический ряд чисел в данной выборке или совокупности, состоящий из последовательно расположенных вариа(н)т и частот (повторяемости)

Как построить вариационный ряд?



- 1. Есть 16 единиц (человек), которые формируют статистическую совокупность
- 2. Они обладают разным цветом кожи

№	Цвет
1	ЧЕРНЫЙ
2	БЕЛЫЙ
3	ЧЕРНЫЙ
4	ЧЕРНЫЙ
5	БЕЛЫЙ
6	ЧЕРНЫЙ
7	ЧЕРНЫЙ
8	БЕЛЫЙ
9	ЧЕРНЫЙ
10	БЕЛЫЙ
11	БЕЛЫЙ
12	БЕЛЫЙ
13	БЕЛЫЙ
14	БЕЛЫЙ
15	ЧЕРНЫЙ
16	ЧЕРНЫЙ

ВАРИАЦИОННЫЙ РЯД — статистический ряд чисел в данной выборке или совокупности, состоящий из последовательно расположенных вариат и их частот (повторяемости)

Варианта	Частота
ЧЕРНЫЙ	8
БЕЛЫЙ	8

ВАРИАЦИОННЫЙ РЯД позволяет представить данные компактно

НАЗНАЧЕНИЕ ВАРИАЦИОННОГО РЯДА

Вариационный ряд необходим для определения

СРЕДНЕЙ ВЕЛИЧИНЫ ПРИЗНАКА

КРИТЕРИЕВ РАЗНООБРАЗИЯ ПРИЗНАКА

«ФОРМЫ» РАСПРЕДЕЛЕНИЯ ПРИЗНАКА, подлежащего изучению

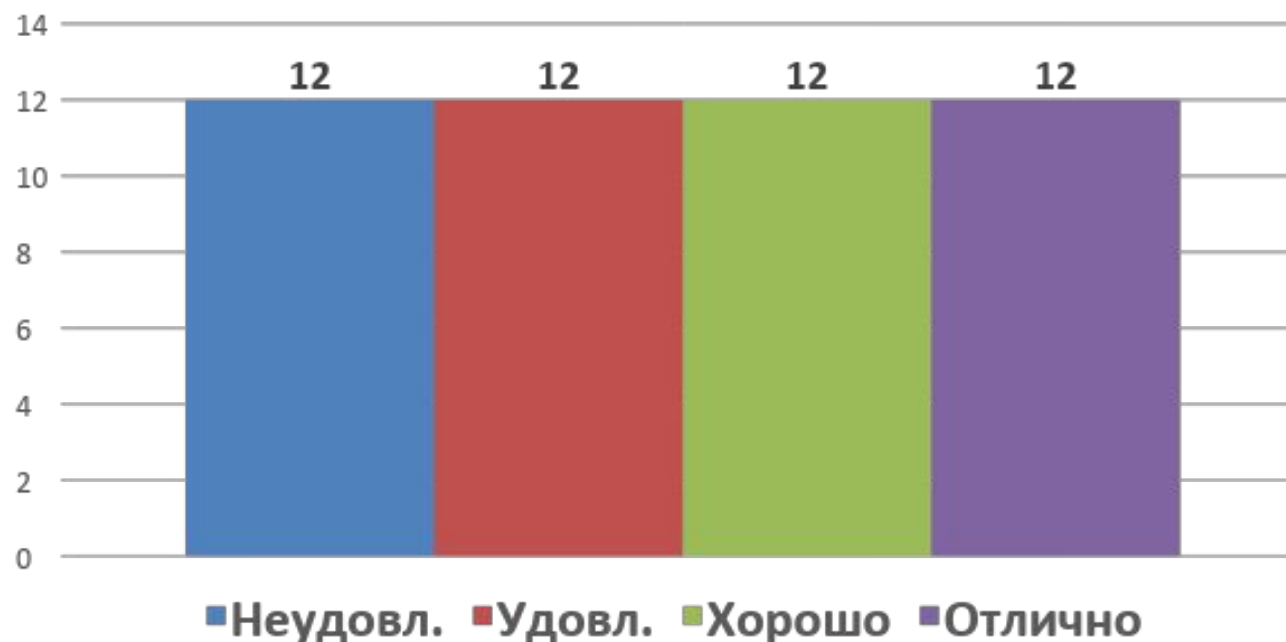
СРЕДНЯЯ ВЕЛИЧИНА - это обобщающая характеристика размера изучаемого признака; **позволяет одним числом количественно охарактеризовать качественно однородную совокупность**



ВИДЫ СРЕДНИХ ВЕЛИЧИН

МОДА - значение во множестве наблюдений, которое встречается **НАИБОЛЕЕ ЧАСТО**

В статистической совокупности м.б.
НИ ОДНОЙ МОДЫ



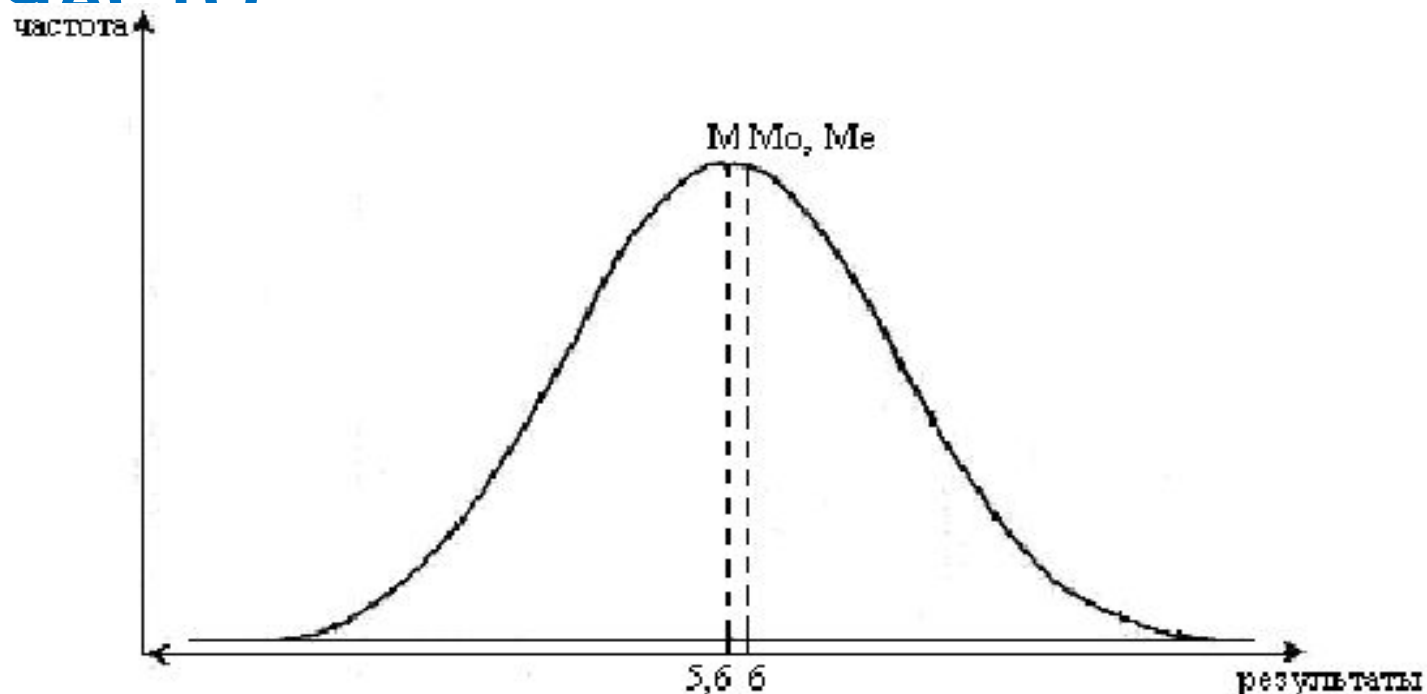
**АМОДАЛЬНОЕ
РАСПРЕДЕЛЕНИЕ**

ВИДЫ СРЕДНИХ ВЕЛИЧИН

МОДА - значение во множестве наблюдений, которое встречается **НАИБОЛЕЕ ЧАСТО**

В статистической совокупности м.б.

ОДНА МОДА

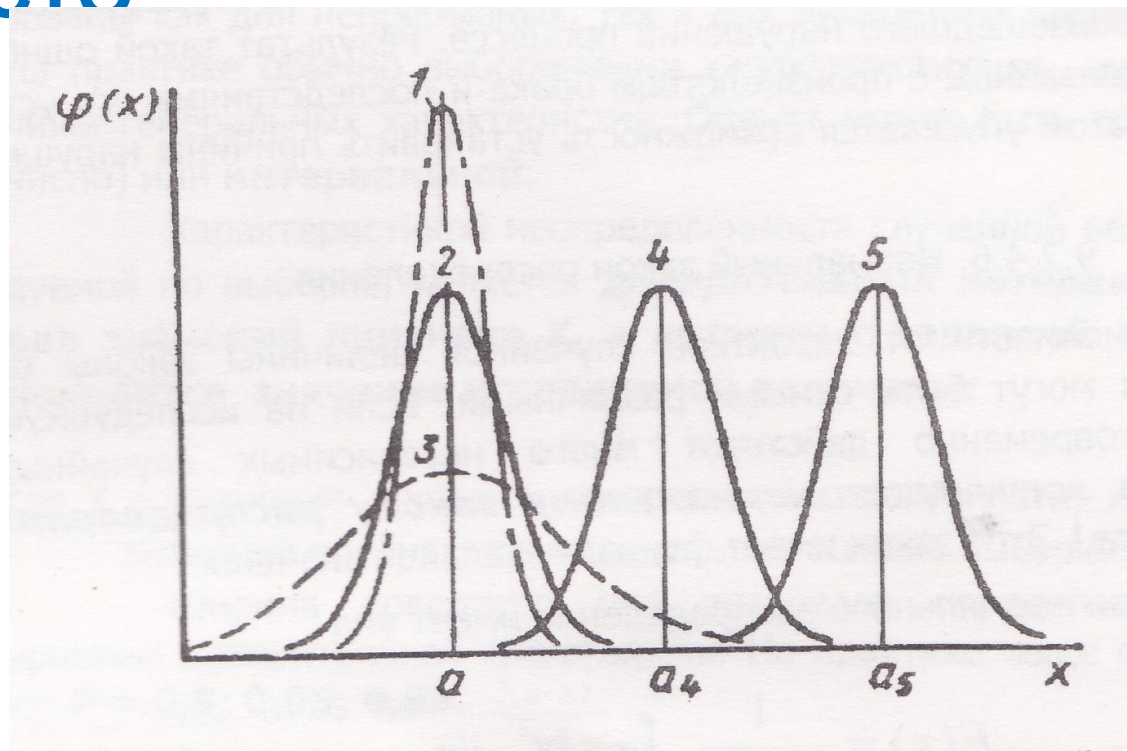


**УНИМОДАЛЬНОЕ
РАСПРЕДЕЛЕНИЕ**

ВИДЫ СРЕДНИХ ВЕЛИЧИН

МОДА - значение во множестве наблюдений, которое встречается **НАИБОЛЕЕ ЧАСТО**

В статистической совокупности м.б.
НЕСКОЛЬКО МОД



**МУЛЬТИМОДАЛЬНОЕ
РАСПРЕДЕЛЕНИЕ**

ВИДЫ СРЕДНИХ ВЕЛИЧ

Если имеется чётное количество случаев и два средних значения различаются, то медианой может служить **любое число между ними**



МЕДИАНА – варианта, делящая вариационный ряд на две равные половины ; остальные 50% - меньше медианы)



50% единиц наблюдения имеют значение признака, меньшее медианы

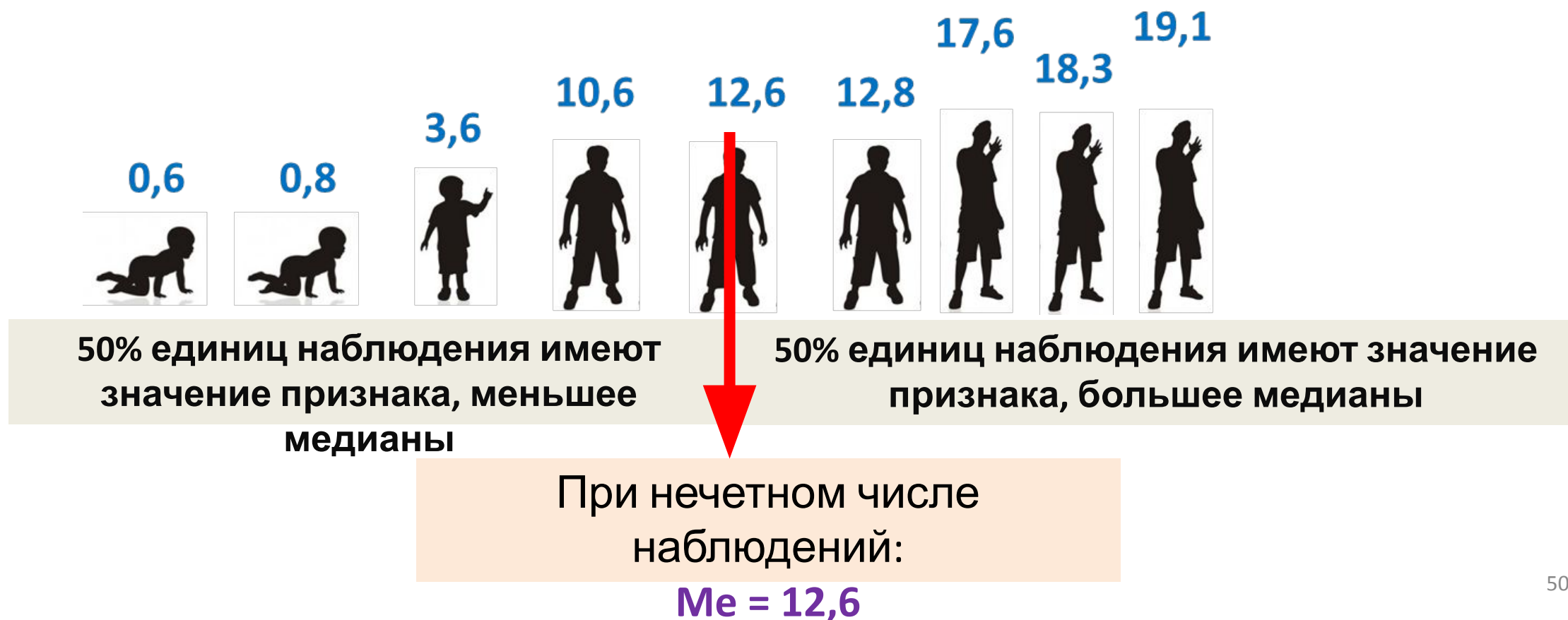
50% единиц наблюдения имеют значение признака, большее медианы

При четном числе наблюдений:

$$Me = (12,6 + 12,8) / 2 = 12,7$$

ВИДЫ СРЕДНИХ ВЕЛИЧИН

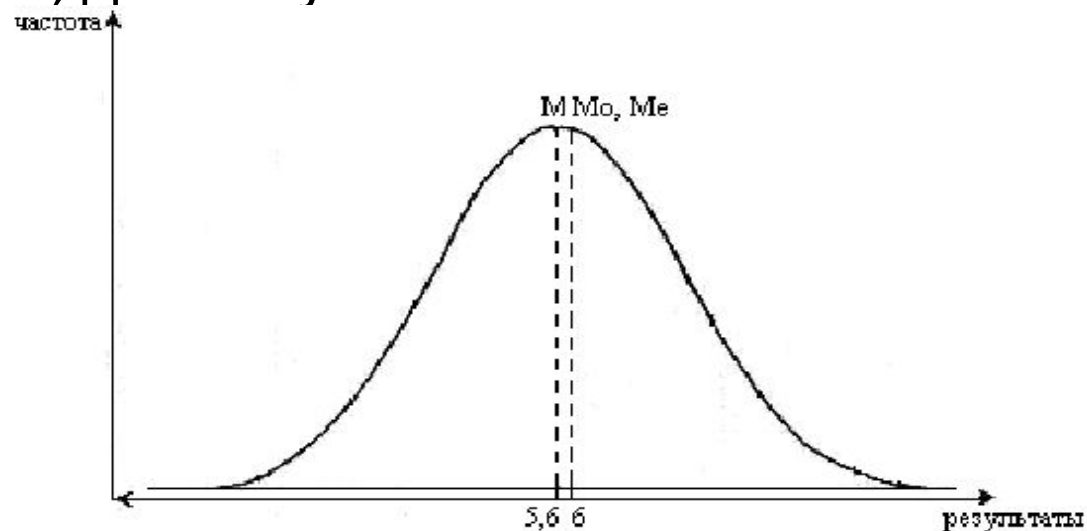
МЕДИАНА – варианта, делящая вариационный ряд на две равные половины ; остальные 50% - меньше медианы)



ВИДЫ СРЕДНИХ ВЕЛИЧИН

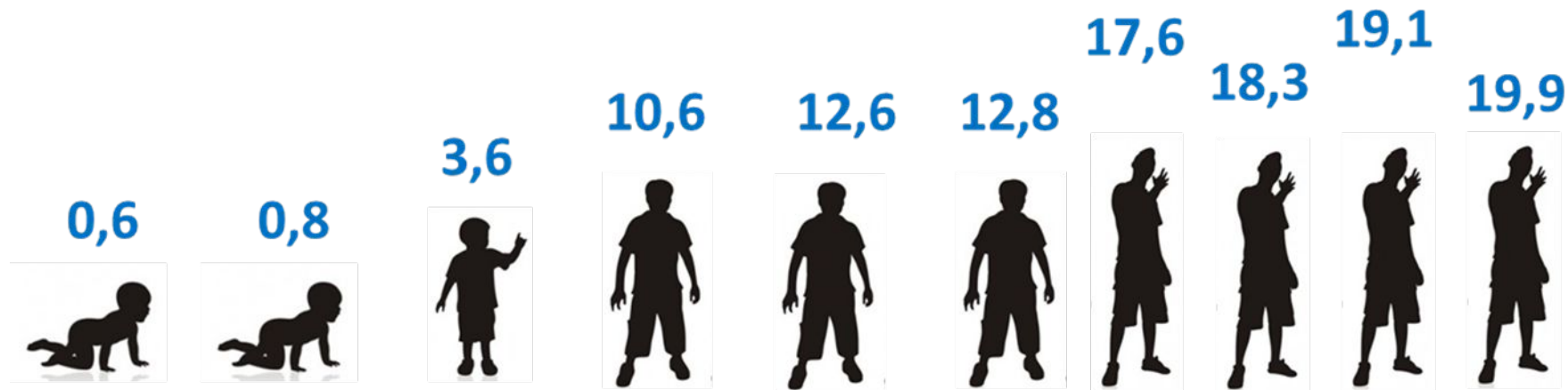
СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЕЛИЧИНА (x) – одна из наиболее распространённых мер центральной тенденции, представляющая собой сумму всех зафиксированных значений, делённую на их количество

$$\overline{X}_{арифм} = \frac{\sum x_i}{n},$$



СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЕЛИЧИНА (x) в отличие от моды и медианы опирается на все произведенные наблюдения

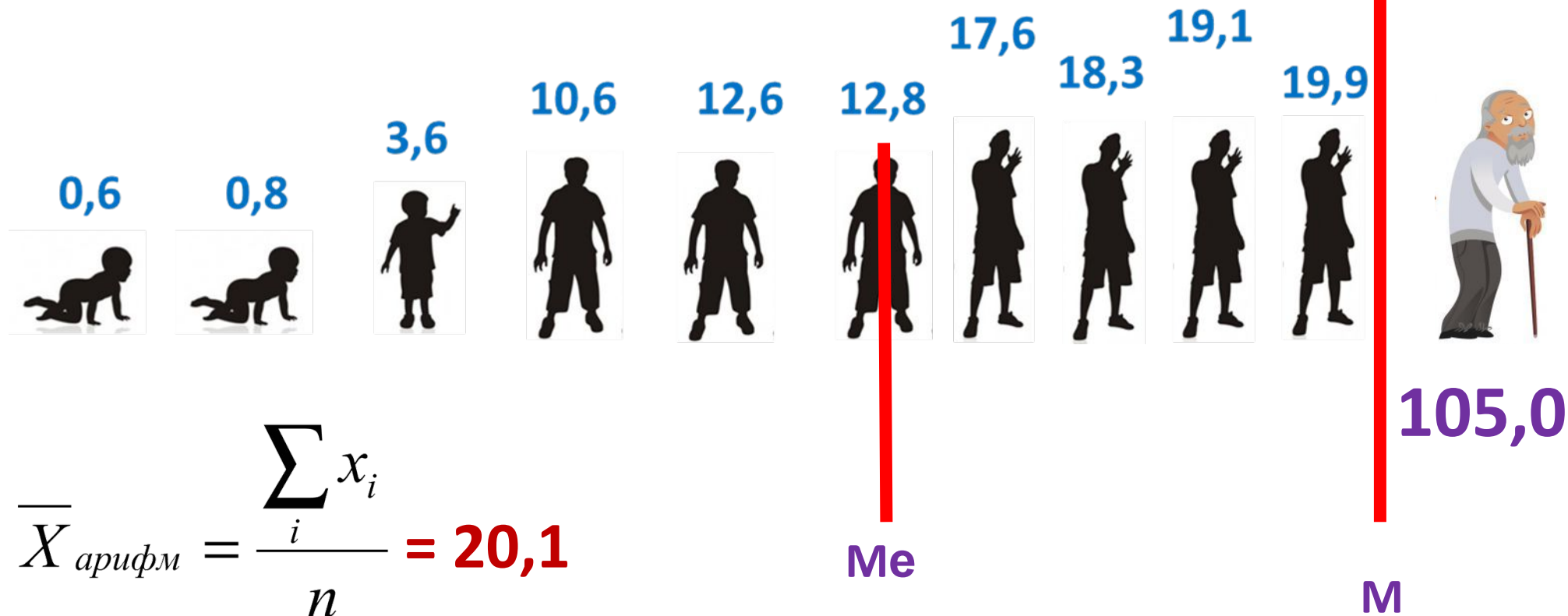
«ПРОБЛЕМА» СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ



$$\overline{X}_{арифм} = \frac{\sum_i x_i}{n} = \mathbf{11,6}$$

$$Me = (12,6 + 12,8) / 2 = 12,7$$

«ПРОБЛЕМА» СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ



$$Me = (12,6 + 12,8) / 2 = 12,8$$

10 / 11 единиц наблюдения имеют значение признака (возраст) меньше, чем средняя арифметическая

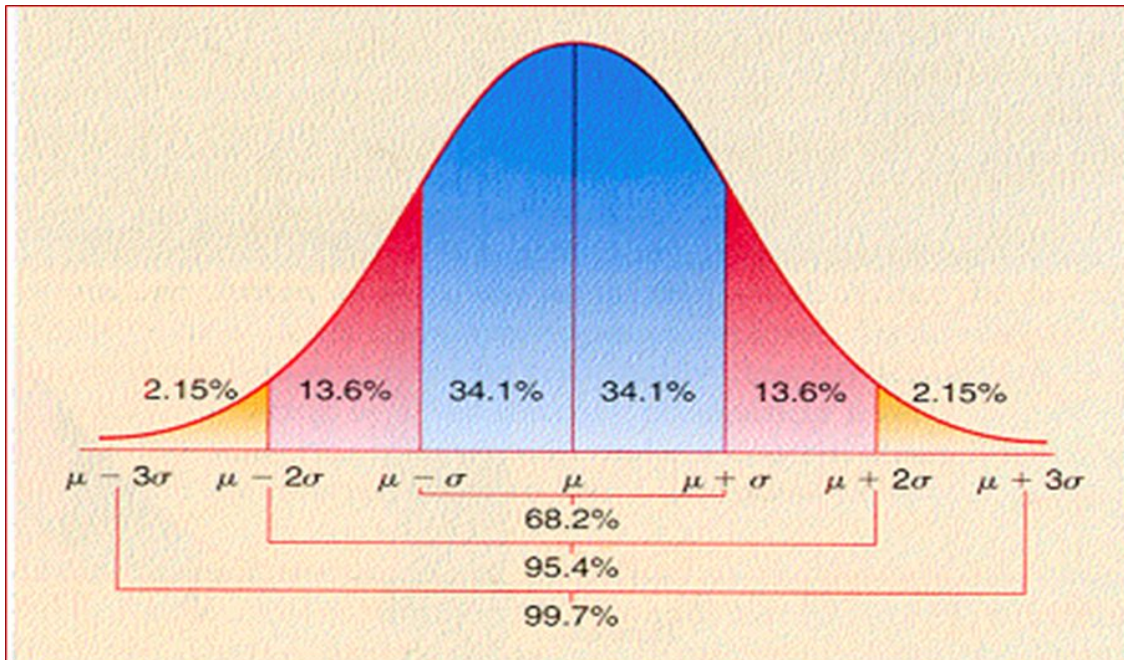
ПРИМЕНЕНИЕ СРЕДНИХ ВЕЛИЧИН

СРЕДНЯЯ ВЕЛИЧИНА ЛИШЬ ОПИСЫВАЕТ ЯВЛЕНИЕ

ПОМИМО САМОЙ СРЕДНЕЙ ВЕЛИЧИНЫ НЕОБХОДИМО
ВЫБРАТЬ И РАССЧИТАТЬ КРИТЕРИЙ ЕЕ ТОЧНОСТИ

ВИДЫ СРЕДНИХ ВЕЛИЧИН И МЕРА ИХ ТОЧНОСТИ

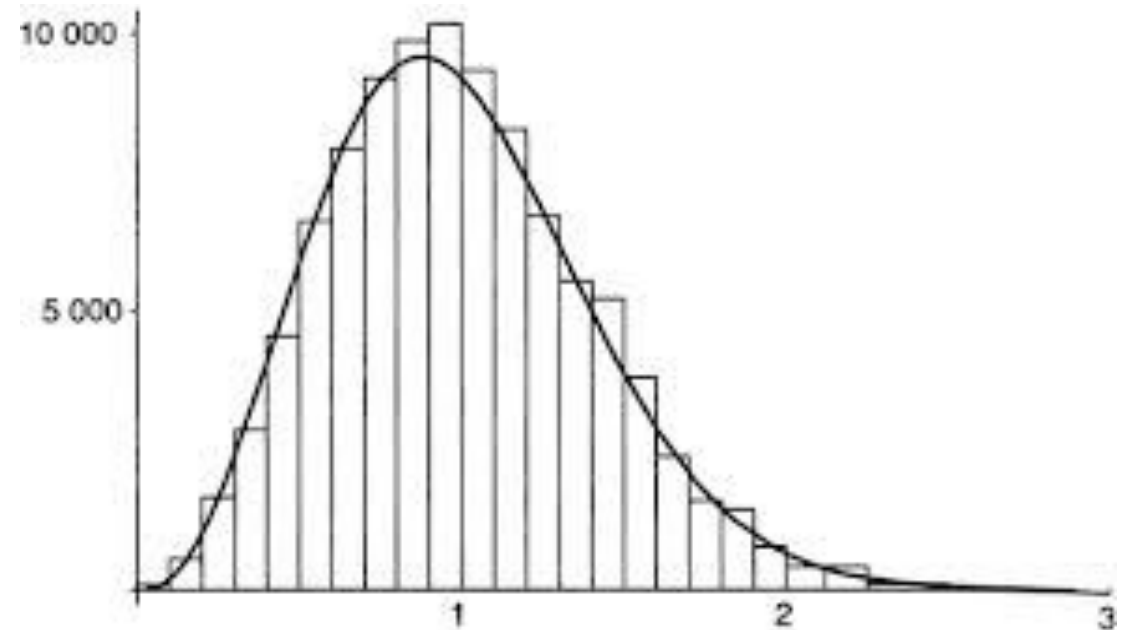
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ



ИСПОЛЬЗУЕМ:

- СРЕДНЮЮ АРИФМЕТИЧЕСКУЮ
- СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ

СКОШЕННОЕ РАСПРЕДЕЛЕНИЕ



ИСПОЛЬЗУЕМ:

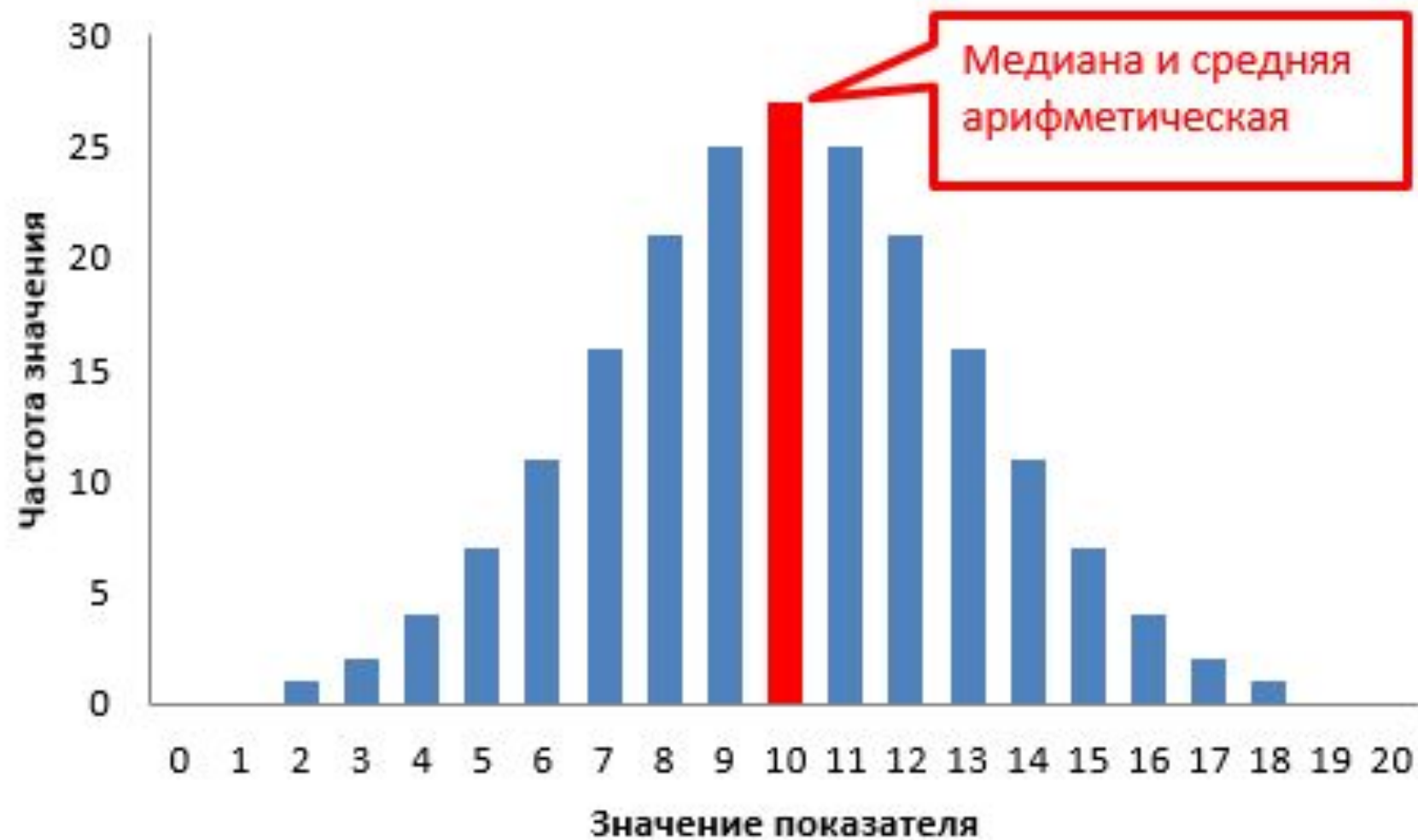
- МЕДИАНУ (Q_2)
- МЕЖКВАРТИЛЬНЫЙ ИНТЕРВАЛ ($Q_1 - Q_3$)

Как определить, подчиняется ли вариационный ряд закону нормального распределения или нет ?

- ТЕСТ ШАПИРО-УИЛКА
- ТЕСТ КОЛМОГорова-СМИРНОВА

**ТЕСТЫ – В БОЛЬШИНСТВЕ СТАТИСТИЧЕСКИХ ПРОГРАММ
(IBM SPSS, STATA, STATISTICA, PASW, R)**

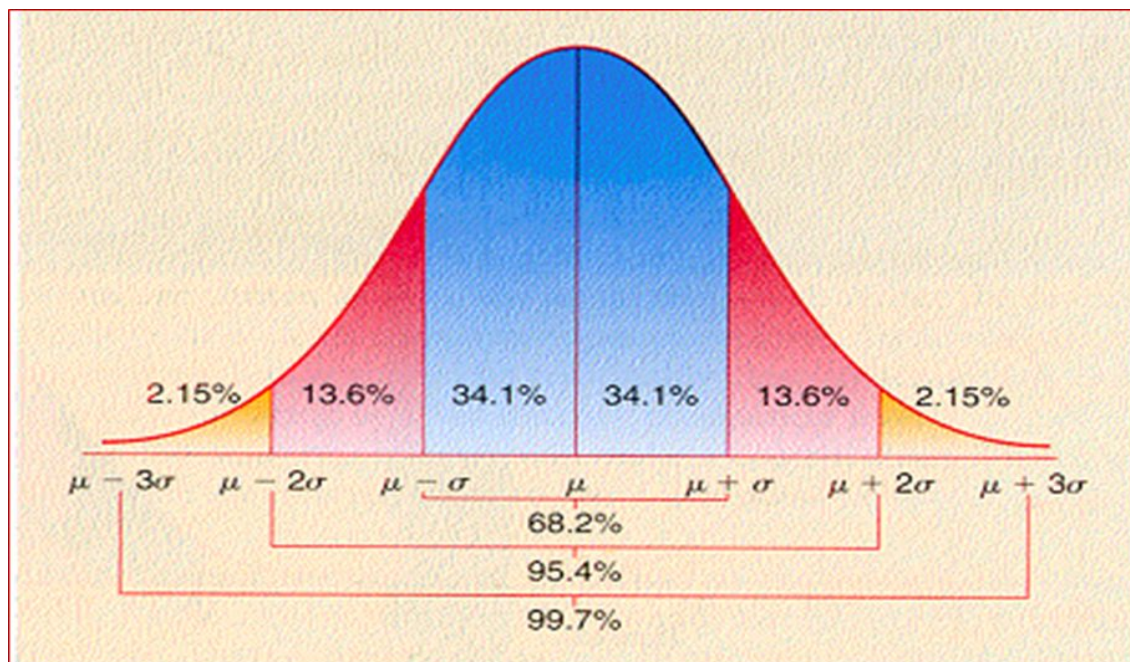
УСЛОВНЫЕ ПРИЗНАКИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ



$$M \approx Me \approx Mo$$

ЕСЛИ РАСПРЕДЕЛЕНИЕ «НОРМАЛЬНОЕ»

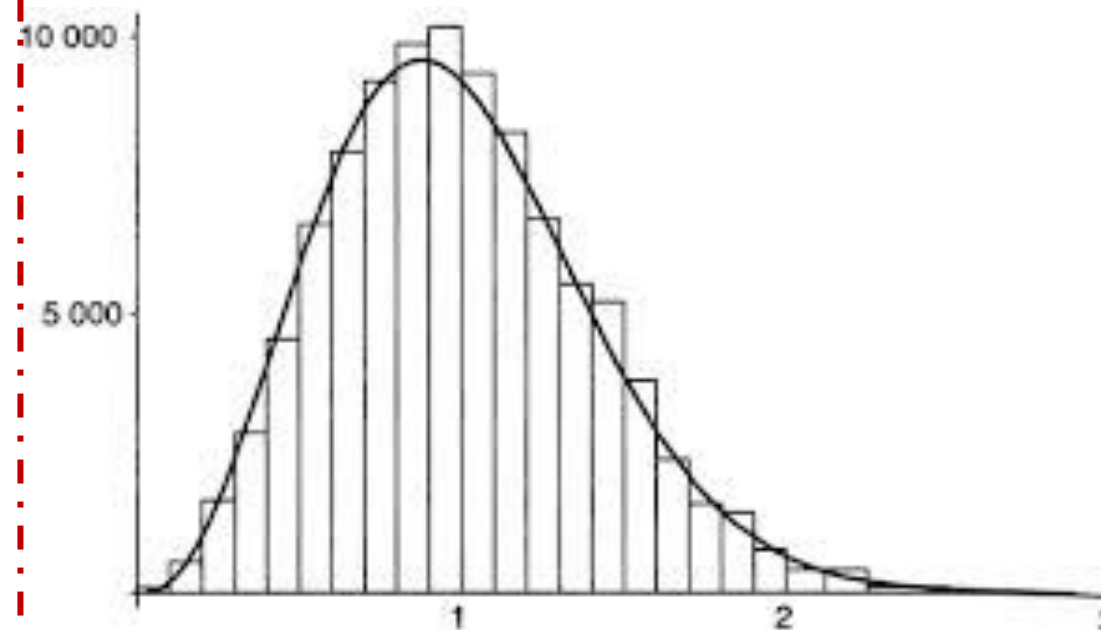
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ



ИСПОЛЬЗУЕМ:

- СРЕДНЮЮ АРИФМЕТИЧЕСКУЮ
- СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ

СКОШЕННОЕ РАСПРЕДЕЛЕНИЕ



ИСПОЛЬЗУЕМ:

- МЕДИАНУ (Q_2)
- МЕЖКВАРТИЛЬНЫЙ ИНТЕРВАЛ ($Q_1 - Q_3$)

МЕТОДИКА РАСЧЕТА СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ

V_{см}	P	d=(V-M)	d²
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
M = 122,9	n=9		$\sum d^2 = 188,8$

МЕТОДИКА РАСЧЕТА СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ

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

$$\delta = \pm \sqrt{\frac{\sum d^2}{n-1}}$$

**$N > 30$ (50/100/500/5000) («большая
выборка»)**

**$N \leq 30$ («малая
выборка»)**

$$\delta = \pm \sqrt{\frac{188,8}{8}} = \pm 4,8$$

**В нашем примере $N=9 (\leq 30)$ – малая
выборка**

МЕТОДИКА РАСЧЕТА СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ

$$\delta = \pm \sqrt{\frac{188,8}{8}} = \pm 4,8$$

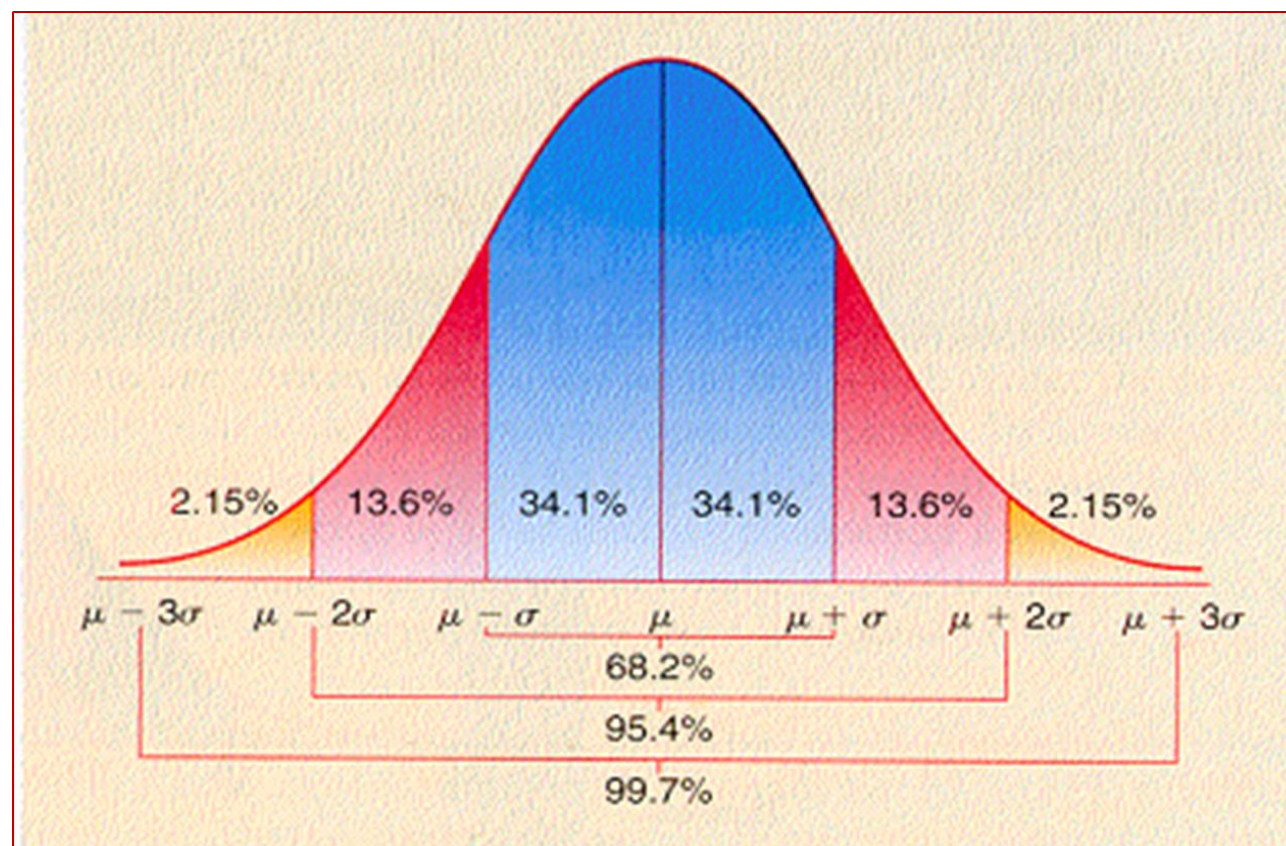
$$SD = 4,8$$

SD^2 = ДИСПЕРСИЯ (VARIANCE)

**ДИСПЕРСИЯ = ОСНОВА ДЛЯ ГРУППЫ СТАТИСТИЧЕСКИХ ПРИЕМОВ:
Т.Н. «ДИСПЕРСИОННОГО АНАЛИЗА» (ANOVA)**

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ СРЕДНЕЙ ВЕЛИЧИНЫ

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ — интервал, который покрывает неизвестный параметр с заданной надёжностью



ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ СРЕДНЕЙ ВЕЛИЧИНЫ

Построение доверительного интервала для математического ожидания генеральной совокупности при известном стандартном отклонении

$$(1) \quad \bar{X} \pm Z \frac{\sigma}{\sqrt{n}} \quad \text{или} \quad \bar{X} - Z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z \frac{\sigma}{\sqrt{n}}$$

РАСЧЕТ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА (95% - го)

РАССЧИТАТЬ ДИ С 95% ТОЧНОСТЬЮ:

$$n = 25$$

$$X \text{ (сред.арифм.)} = 10$$

$$\sigma = 0,5$$

$$(1) \quad \bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

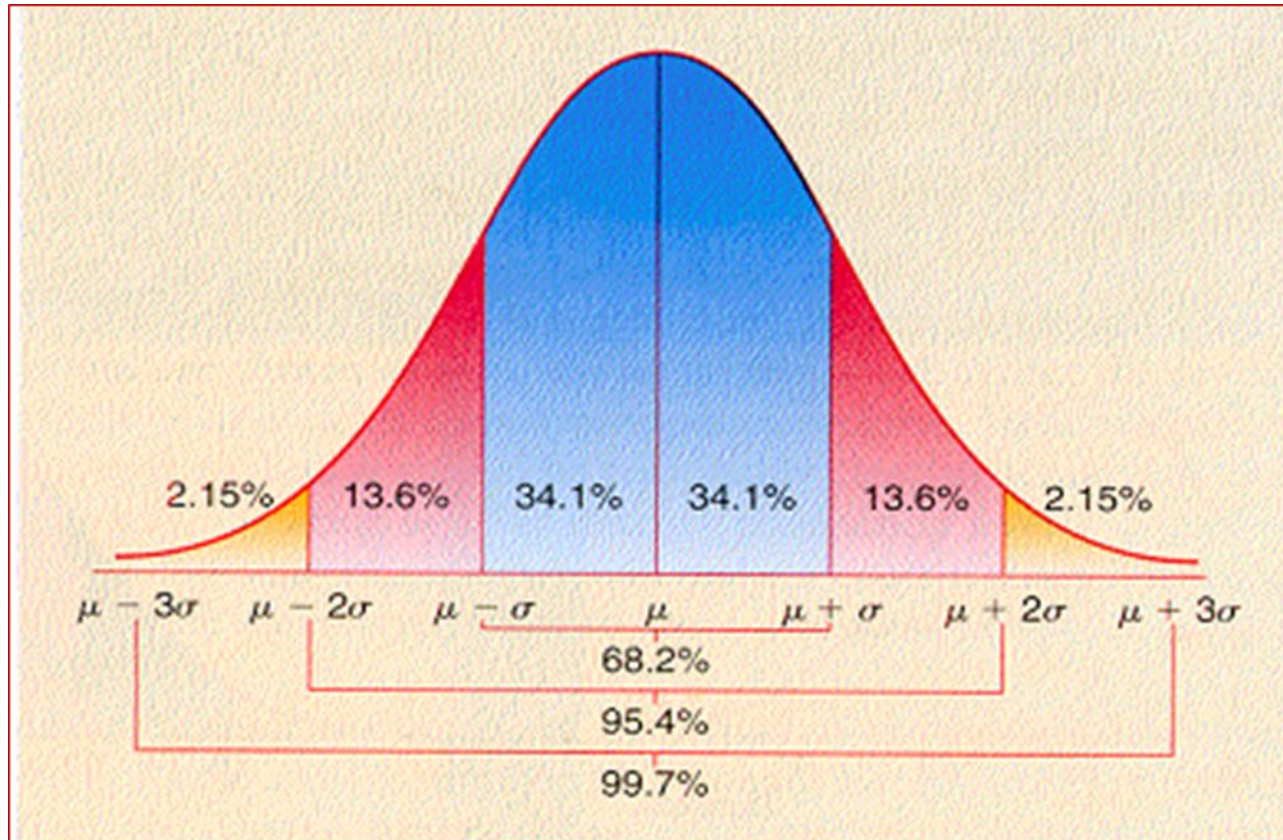
РЕШЕНИЕ:

$$\text{Верхняя граница: } 10 + 1,96 \times 0,5 / \sqrt{25} = 10,196$$

$$\text{Нижняя граница: } 10 - 1,96 \times 0,5 / \sqrt{25} = 9,804$$

Соответственно: $X = 10,000$ с 95% ДИ (9,804; 10,196)

РАСЧЕТ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА



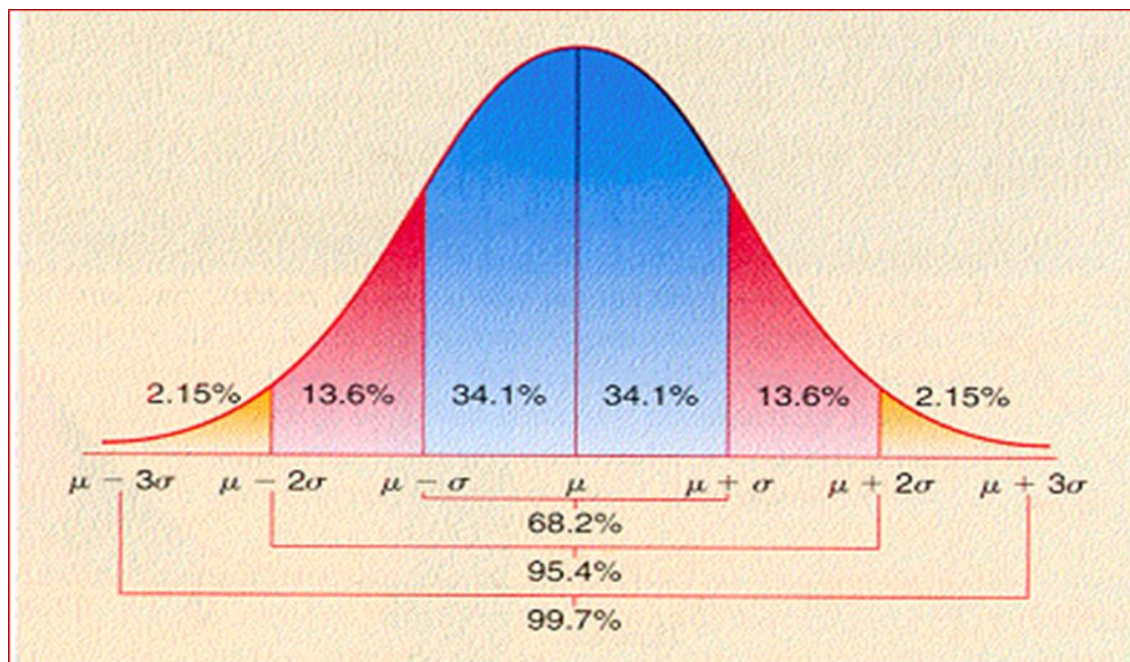
$$(1) \quad \bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

**ЧЕМ ТОЧНЕЕ
ДОВЕРИТЕЛЬНЫЙ
ИНТЕРВАЛ, ТЕМ ОН
ШИРЕ**

**ЛУЧШЕ ИСПОЛЬЗОВАТЬ БЕСПЛАТНЫЕ СТАТ.
ПРОГРАММЫ: WINPERI**

ЕСЛИ РАСПРЕДЕЛЕНИЕ «СКОШЕННОЕ»

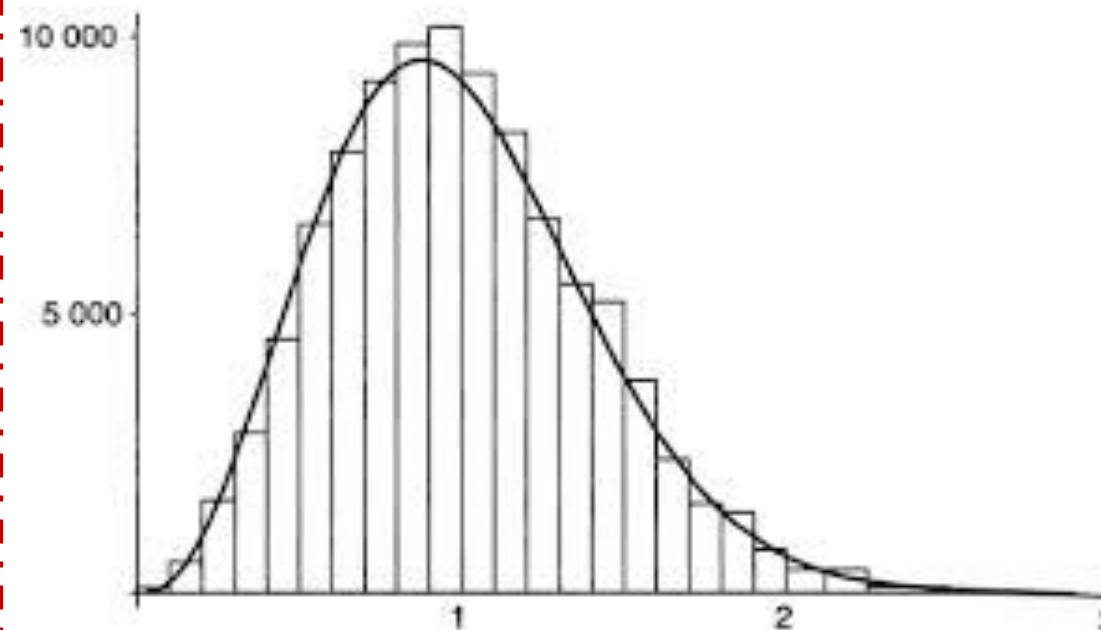
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ



ИСПОЛЬЗУЕМ:

- СРЕДНЮЮ АРИФМЕТИЧЕСКУЮ
- СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ

СКОШЕННОЕ РАСПРЕДЕЛЕНИЕ

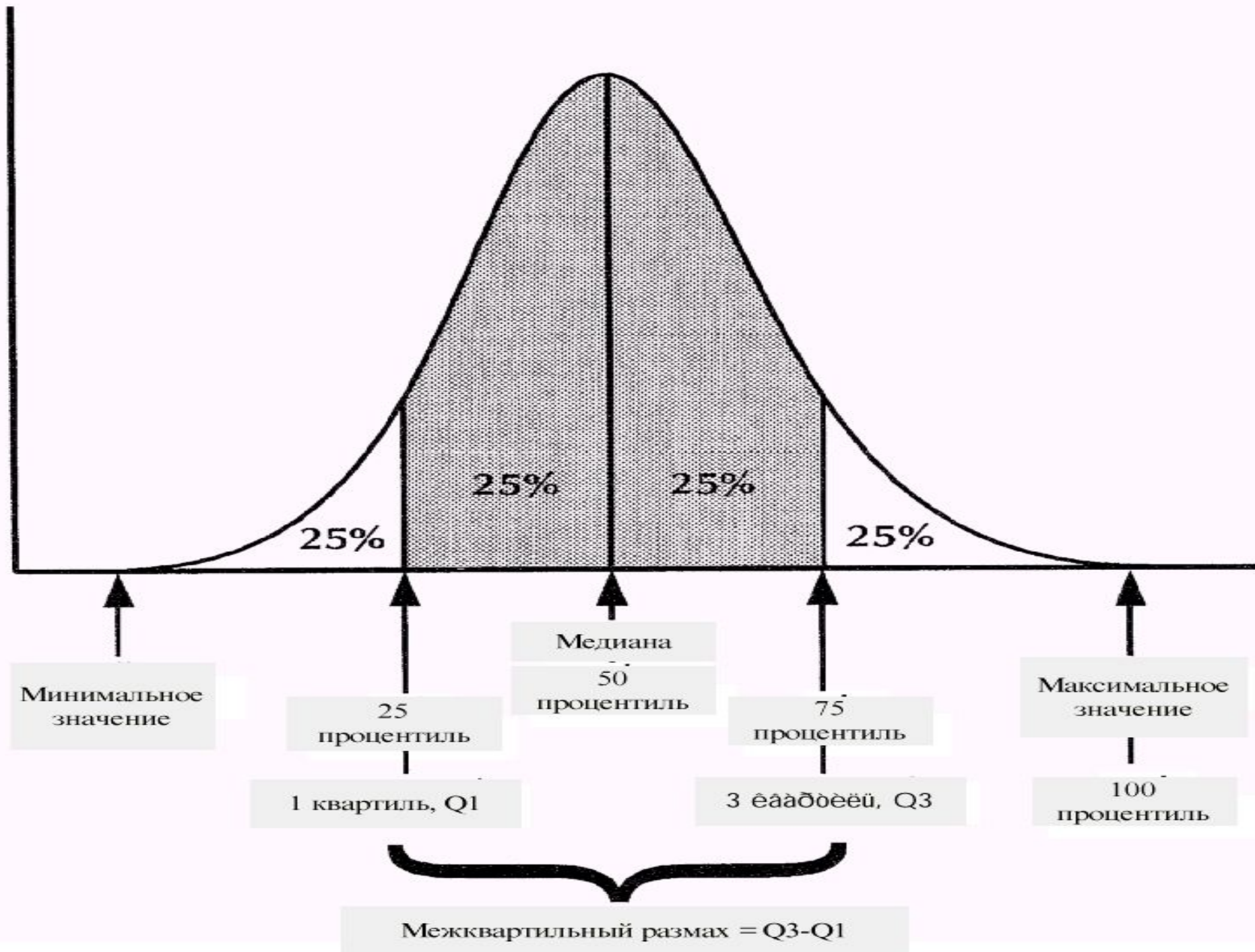


ИСПОЛЬЗУЕМ:

- МЕДИАНУ (Q_2)
- МЕЖКВАРТИЛЬНЫЙ ИНТЕРВАЛ ($Q_1 - Q_3$)

КВАРТИЛИ (4), ДЕЦИЛИ (10) и ПРОЦЕНТИЛИ (100)





ПРЕДСТАВЛЕНИЕ ДАННЫХ

ДЛЯ СРЕДНЕЙ АРИФМЕТИЧЕСКОЙ: 10,000 с 95% ДИ (9,804; 10,196)

не рекомендуется указывать SD или V

ДЛЯ МЕДИАНЫ: Q2 (Q1; Q3)



ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ В МЕДИЦИНСКОЙ СТАТИСТИКЕ

Статистические величины

- Для оценки изучаемых явлений, составляющую статистическую совокупность, используют **статистические величины**:
 - абсолютные;
 - относительные;
 - средние.

Относительные величины

Относительные величины применяют:

- для характеристики распределения признаков в совокупности;
- для сравнения совокупностей.

Виды относительных величин:

- Интенсивные;
- Экстенсивные;
- Показатели соотношения;
- Показатели наглядности.

ЭКСТЕНСИВНЫЕ КОЭФФИЦИЕНТЫ

ЭКСТЕНСИВНЫЕ КОЭФФИЦИЕНТЫ – характеризуют распределение явления на его составные части, его **внутреннюю структуру** или отношение частей к целому (удельный вес)

$$ЭК = \frac{\text{ЧАСТЬ}}{\text{ЦЕЛОЕ}} \times 100\%$$

При анализе экстенсивный показатель следует помнить, что им пользуются только для характеристики состава (структуры) явления **в данный момент времени и в данном месте**

Примеры: структура населения по полу, возрасту, социальному положению; структура заболеваний по нозологии; структура причин смерти

ИНТЕНСИВНЫЕ КОЭФФИЦИЕНТЫ

ИНТЕНСИВНЫЕ КОЭФФИЦИЕНТЫ – характеризуют силу, частоту (степень интенсивности, уровень) распространения явления в среде, **в которой оно происходит** и с которой оно непосредственно связано

$$\text{ИК} = \frac{\text{ВЕЛИЧИНА ЯВЛЕНИЯ}}{\text{ВЕЛИЧИНА СРЕДЫ, КОТОРАЯ ПРОДУЦИРУЕТ ЯВЛЕНИЕ}} \times 100 / 1000 / 10000 / 100000$$

Выбор числового основания (100; 1000; 10000... и т.д.) зависит от распространенности явления - чем реже встречается изучаемое явление, тем большее основание

ИНТЕНСИВНЫЕ КОЭФФИЦИЕНТЫ

ПРИМЕР РАСЧЕТА ИНТЕНСИВНОГО КОЭФФИЦИЕНТА РОЖДАЕМОСТИ

В г. «А» общее число родившихся живыми за год составило 6.700 чел. Среднегодовая численность населения: 570.000 чел.

$$\begin{array}{rcl} 6.700 & - & 570.000 \\ X & - & 1.000 \end{array}$$

$$X = (6.700 \times 1.000) / 570.000 = 11,75$$

Т.о., коэффициент рождаемости составил 11,75 на 1000 нас.

КОЭФФИЦИЕНТЫ СООТНОШЕНИЯ

КОЭФФИЦИЕНТЫ СООТНОШЕНИЯ – характеризуют численное соотношение двух, не связанных непосредственно между собой, независимых совокупностей, сопоставляемых только логически

$$ИК = \frac{\text{ВЕЛИЧИНА ЯВЛЕНИЯ "А"}}{\text{ВЕЛИЧИНА СРЕДЫ, КОТОРАЯ НЕ ПРОДУЦИРУЕТ ЯВЛЕНИЕ}} \times 100 / 1000 / 10000 / 100000$$

Если интенсивные коэффициенты характеризуют частоту явлений, происходящих в данной среде (т.е. непосредственно связанных с этой средой), то коэффициенты соотношения указывают на числовое соотношение двух явлений, непосредственно между собой не связанных

КОЭФФИЦИЕНТЫ НАГЛЯДНОСТИ

КОЭФФИЦИЕНТЫ НАГЛЯДНОСТИ – применяются с целью более наглядного и доступного сравнения рядов абсолютных, относительных или средних величин. Они представляют технический прием преобразования цифровых показателей

Этот коэффициент получают путем преобразования ряда величин по отношению к одной из них – базисной (любой, не обязательно начальной). Эта базисная величина принимается за 1; 100; 1000 и т.п., а остальные величины ряда, при помощи обычной пропорции, пересчитываются по отношению к ней

Рождаемость в России за 1997 и 2000 гг. (на 1000 нас.)

Годы	Интенсивный коэффициент	Коэффициенты наглядности (уровень рождаемости в 1997 году принят за 100%)
1997	8,6	100,0%
2000	8,3	96,5%



АНАЛИЗ ДИНАМИЧЕСКИХ РЯДОВ

Медицинская статистика. ЗАДАЧИ

- ★ разработка специальных методов исследования массовых процессов и явлений в медицине и здравоохранении;
- ★ выявление наиболее существенных закономерностей и тенденций в здоровье населения в целом и в различных его группах.

С УЧЕТОМ конкретных условий и образа жизни, А ТАКЖЕ **динамики развития сети**, деятельности учреждений здравоохранения и медицинских кадров.

6



Зачем медицинская статистика реализует эти задачи?

- АНАЛИЗ произошедших событий / явлений
- ПРОГНОЗ событий / явлений

Т.е. медицинская статистика изучает / анализирует произошедшие события, чтобы спрогнозировать и организовать деятельность системы здравоохранения на перспективу

7

Медицинская статистика

ВРАЧУ НЕОБХОДИМО:

анализировать происходящие **ВО ВРЕМЕНИ** изменения в состоянии здоровья отдельных групп населения, в деятельности медицинских учреждений, в экспериментальных исследованиях

Выявление основной тенденции изучаемого явления вне влияния "случайных" факторов позволяет определять **ЗАКОНОМЕРНОСТИ** изменений явления и на этой основе осуществлять **ПРОГНОЗИРОВАНИЕ**

Динамический ряд

ДИНАМИЧЕСКИЙ РЯД — это ряд однородных статистических величин, показывающих изменение явления во времени

Год	Рождаемость	Смертность
2009	12,3	14,1
2010	12,5	14,2
2011	12,6	13,5
2012	13,3	13,3
2013	13,3	13,1

Величины, составляющие динамический ряд, называются **УРОВНЯМИ РЯДА**



Коэффициенты рождаемости и смертности в России в 2009-2013 гг. (на 1000 нас.)

Уровни динамического ряда

УРОВНИ ДИНАМИЧЕСКОГО РЯДА МОГУТ БЫТЬ ПРЕДСТАВЛЕНЫ:

А) АБСОЛЮТНЫМИ ВЕЛИЧИНАМИ

Численность населения России в
2009-2013 гг.

Год	Население
2009	142,785,344
2010	142,849,472
2011	142,960,908
2012	143,201,700
2013	143,502,097

Уровни динамического ряда

УРОВНИ ДИНАМИЧЕСКОГО РЯДА МОГУТ БЫТЬ ПРЕДСТАВЛЕНЫ:

Б) ОТНОСИТЕЛЬНЫМИ ВЕЛИЧИНАМИ

Год	Рождаемость	Смертность
2009	12,3	14,1
2010	12,5	14,2
2011	12,6	13,5
2012	13,3	13,3
2013	13,3	13,1

Не рекомендуется сравнивать в динамике **ЭКСТЕНСИВНЫЕ ВЕЛИЧИНЫ** (удельный вес), так как величина их изменения зависит от соотношения



Коэффициенты рождаемости и смертности в России в 2009-2013 гг. (на 1000 нас.)

Уровни динамического ряда

УРОВНИ ДИНАМИЧЕСКОГО РЯДА МОГУТ БЫТЬ ПРЕДСТАВЛЕНЫ:

В) СРЕДНИМИ ВЕЛИЧИНАМИ

Средняя продолжительность жизни населения России в
2009-2013 гг., лет

Год	Средняя продолжительность жизни, лет
2009	68,8
2010	68,9
2011	69,8
2012	70,2
2013	70,8

ПОКАЗАТЕЛИ ДИНАМИЧЕСКИХ РЯДОВ

АБСОЛЮТНЫЙ ПРИРОСТ представляет собой разность между последующим и предыдущим уровнем. Измеряется в тех же единицах, в которых представлены уровни ряда

Год	Средняя продолжительность жизни, лет	АБСОЛЮТНЫЙ ПРИРОСТ
2009	68,8	-
2010	68,9	$= 68,9 - 68,8 = 0,1$
2011	69,8	$= 69,8 - 68,9 = 0,9$
2012	70,2	$= 70,2 - 69,8 = 0,4$
2013	70,8	$= 70,8 - 70,2 = 0,6$

ПОКАЗАТЕЛИ ДИНАМИЧЕСКИХ РЯДОВ

ТЕМП РОСТА – процентное отношение последующего уровня к предыдущему

СРАВНЕНИЕ МАСШТАБОВ ЯВЛЕНИЯ

Год	Средняя продолжительность жизни, лет	ТЕМП РОСТА
2009	68,8	-
2010	68,9	$= (68,9 / 68,8) \times 100\% = + 100,14\%$
2011	69,8	$= (69,8 / 68,9) \times 100\% = + 101,31\%$
2012	70,2	$= (70,2 / 69,8) \times 100\% = + 100,57\%$
2013	70,8	$= (70,8 / 70,2) \times 100\% = + 100,85\%$

ПОКАЗАТЕЛИ ДИНАМИЧЕСКИХ РЯДОВ

ТЕМП ПРИРОСТА – процентное отношение абсолютного прироста к предыдущему уровню

ДИНАМИКА ЯВЛЕНИЯ

Год	Средняя продолжительность жизни, лет	АБСОЛЮТНЫЙ ПРИРОСТ	ТЕМП ПРИРОСТА
2009	68,8	-	-
2010	68,9	$= 68,9 - 68,8 = 0,1$	$= (0,1 / 68,8) \times 100\% = + 0,14\%$
2011	69,8	$= 69,8 - 68,9 = 0,9$	$= (0,9 / 68,9) \times 100\% = + 1,31\%$
2012	70,2	$= 70,2 - 69,8 = 0,4$	$= (0,4 / 70,2) \times 100\% = + 0,57\%$
2013	70,8	$= 70,8 - 70,2 = 0,6$	$= (0,6 / 70,8) \times 100\% = + 0,85\%$

ПОКАЗАТЕЛИ ДИНАМИЧЕСКИХ РЯДОВ

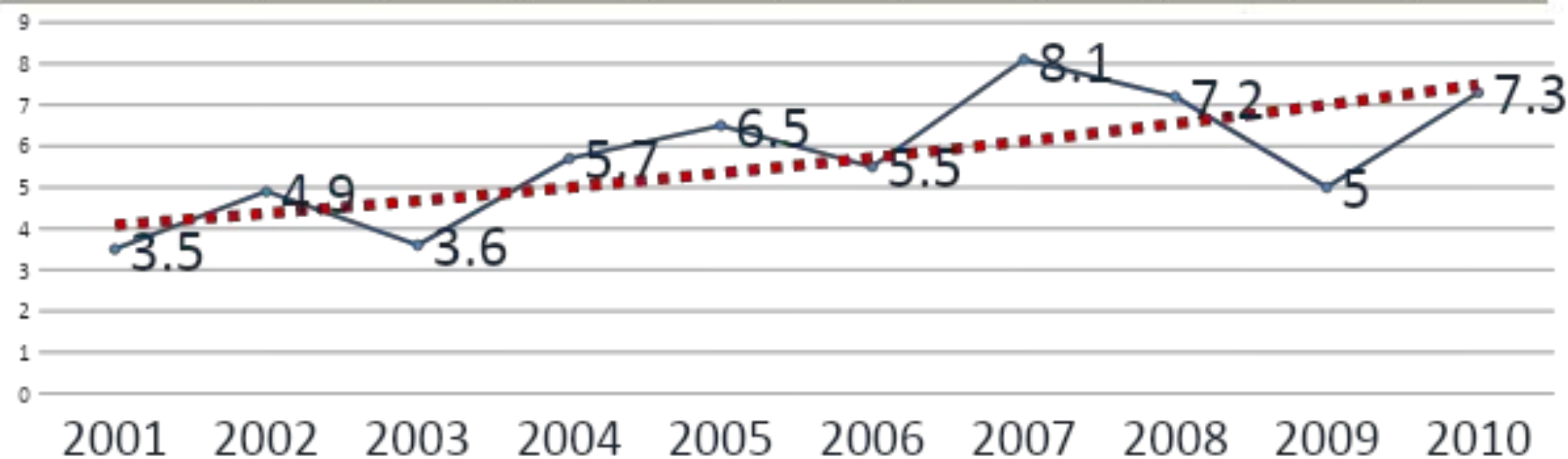
СОДЕРЖАНИЕ 1% ПРИРОСТА – отношение абсолютного прироста к темпу прироста

ПЛАНОВЫЕ ПОКАЗАТЕЛИ

Год	Средняя продолжительность жизни, лет	АБСОЛЮТНЫЙ ПРИРОСТ	ТЕМП ПРИРОСТА	СОДЕРЖАНИЕ 1% ПРИРОСТА
2009	68,8	-		-
2010	68,9	0,1	0,14%	$= (0,1 / 0,14\%) = 0,71$
2011	69,8	0,9	1,31%	$= (0,9 / 1,31\%) = 0,69$
2012	70,2	0,4	0,57%	$= (0,4 / 0,57\%) = 0,70$
2013	70,8	0,6	0,85%	$= (0,6 / 0,85\%) = 0,71$

Таблица. Заболеваемость населения Н-ского района ветряной оспой за 10 лет
(на 10 000 населения)

Годы	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Показатель	3,5	4,9	3,6	5,7	6,5	5,5	8,1	7,2	5,0	7,3



ВЫРАВНИВАНИЕ ДИНАМИЧЕСКОГО РЯДА

Если числовые значения признака динамического ряда имеют значительные колебания, что затрудняет выявление закономерностей развития явления, производится **ВЫРАВНИВАНИЕ ДИНАМИЧЕСКОГО РЯДА**

МЕТОДЫ ВЫРАВНИВАНИЯ ДИНАМИЧЕСКОГО РЯДА

- 1) **УКРУПНЕНИЕ ИНТЕРВАЛА** путем суммирования уровней ряда за смежные периоды
- 2) **ВЫЧИСЛЕНИЕ ГРУППОВОЙ СРЕДНЕЙ** (суммируются смежные уровни соседних периодов, а затем полученную сумму делят на число слагаемых)
- 3) **ВЫЧИСЛЕНИЕ СКОЛЬЗЯЩЕЙ СРЕДНЕЙ** (позволяет каждый уровень заменить на среднюю величину данного уровня и двух соседних с ними)
- 4) **МЕТОД НАИМЕНЬШИХ КВАДРАТОВ**

МЕТОД УКРУПНЕНИЯ ИНТЕРВАЛОВ (ПЕРИОДОВ)

Укрупнение периодов применяется, когда явление в интервальном ряду выражено в **АБСОЛЮТНЫХ ВЕЛИЧИНАХ**, уровни которых суммируются по более крупным периодам. Применение возможно при **КРАТНОМ ЧИСЛЕ ПЕРИОДОВ**

Годы	Число случаев на 100 рабочих	Укрупнение интервала
2008	57,0	
2009	58,0	$57+58=115$
2010	64,0	
2011	48,0	$64+48=112$
2012	36,0	
2013	57,0	$36+57=93$
2014	44,0	
2015	49,0	$44+49=93$

Таблица 1

Первичная
заболеваемость
болезнью N на
комбинате

в 2008 – 2015 гг. (на 100
работающих)

ВЫЧИСЛЕНИЕ ГРУППОВОЙ СРЕДНЕЙ

Вычисление групповой средней применяется, когда уровни интервального ряда выражены в **АБСОЛЮТНЫХ, СРЕДНИХ ИЛИ ОТНОСИТЕЛЬНЫХ ВЕЛИЧИНАХ**, которые суммируются, а затем делятся на число слагаемых. Способ применяется при **КРАТНОМ ЧИСЛЕ ПЕРИОДОВ**

Годы	Число случаев на 100 рабочих	Групповая средняя
2008	57,0	
2009	58,0	$(57+58)/2=57,5$
2010	64,0	
2011	48,0	$(64+48)/2=56,0$
2012	36,0	
2013	57,0	$(36+57)/2=46,5$
2014	44,0	
2015	49,0	$(44+49)/2=46,5$

Таблица 1

Первичная
заболеваемость
болезнью N на
комбинате

в 2008 – 2015 гг. (на 100
работающих)

ВЫЧИСЛЕНИЕ СКОЛЬЗЯЩЕЙ СРЕДНЕЙ

Расчет скользящей средней применяется, когда уровни явлений любого ряда выражены в **АБСОЛЮТНЫХ, СРЕДНИХ ИЛИ ОТНОСИТЕЛЬНЫХ ВЕЛИЧИНАХ**. Данный метод применяется при наличии **НЕКРАТНОГО ЧИСЛА ВРЕМЕННЫХ ПЕРИОДОВ** (7, 11, 13, 17, 19) достаточно **ДЛИННОГО ДИНАМИЧЕСКОГО РЯДА**. Путем вычисления групповой средней значений 3 периодов, а в последующем переходя на определенный уровень и два соседних с ним, осуществляется "скольжение" по периодам. Каждый уровень заменяется на среднюю величину (из данного уровня и двух соседних с ним)

Годы	Число случаев на 100 рабочих	Скользящая средняя
2008	57,0	—
2009	58,0	$(57+58+64)/3=57,5$
2010	64,0	$(58+64+48)/3=57,0$
2011	48,0	$(64+48+36)/3=49,0$
2012	36,0	$(48+36+57)/3=47,0$
2013	57,0	$(36+57+44)/3=45,7$
2014	44,0	$(57+44+49)/3=50,0$
2015	49,0	—

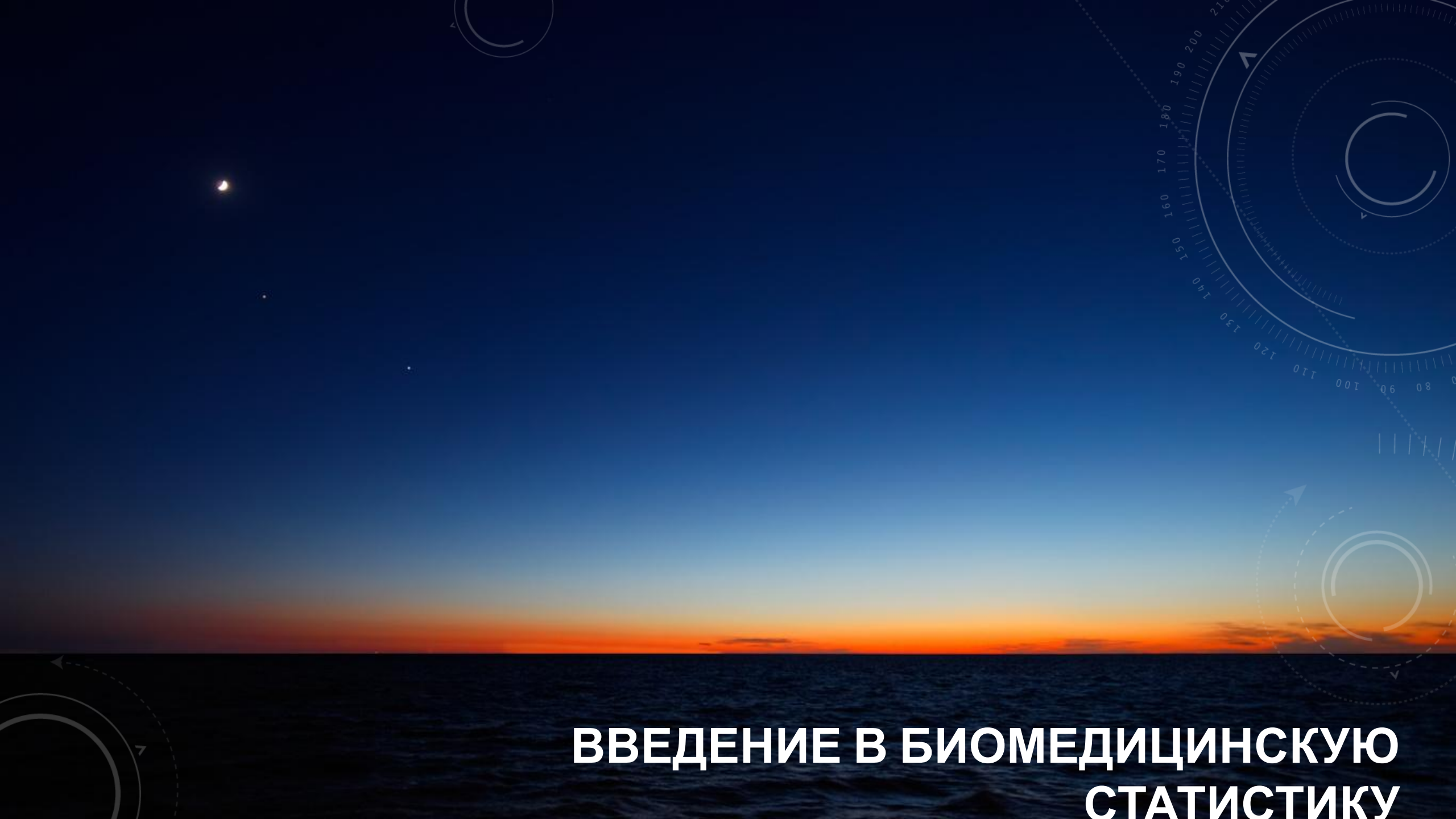
Таблица 1

Первичная
заболеваемость
болезнью N на
комбинате
в 2008 – 2015 гг. (на 100
работающих)

Анализ динамического ряда предусматривает:

- 1) выравнивание динамического ряда (при необходимости)
- 2) расчет показателей динамического ряда
- 3) графическое изображение показателей динамического ряда
- 4) анализ полученных результатов

**С РАЗВИТИЕМ СТАТИСТИЧЕСКОГО ПРОГРАММНОГО
ОБЕСПЕЧЕНИЯ (STATA, SPSS, STATISTICA, R ...) ТЕХНИКА
«РУЧНОГО» АНАЛИЗА ДИНАМИЧЕСКОГО РЯДА БОЛЕЕ НЕ
ИСПОЛЬЗУЕТСЯ ВВИДУ ВОЗМОЖНОСТИ ПОЛУЧИТЬ
ИСКАЖЕННЫЕ РЕЗУЛЬТАТЫ**

The background features a serene sunset over a dark, textured ocean. The sky transitions from a deep blue at the top to a bright orange glow at the horizon. Several technical, semi-transparent overlays are present: a large circular scale with degree markings (90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210) and arrows is in the upper right; a smaller circular diagram with concentric circles and arrows is in the lower right; and a partial circular diagram is visible in the bottom left corner.

ВВЕДЕНИЕ В БИОМЕДИЦИНСКУЮ СТАТИСТИКУ