

**Технология цифровой обработки данных.
Компьютерный анализ медицинских данных.
Использование широких возможностей редактора
электронных таблиц Microsoft Excel.**

ЛЕКЦИЯ 1

ПЛАН ЛЕКЦИИ

1

Основные элементы электронной таблицы.

2

Формулы и функции. Операции с формулами

3

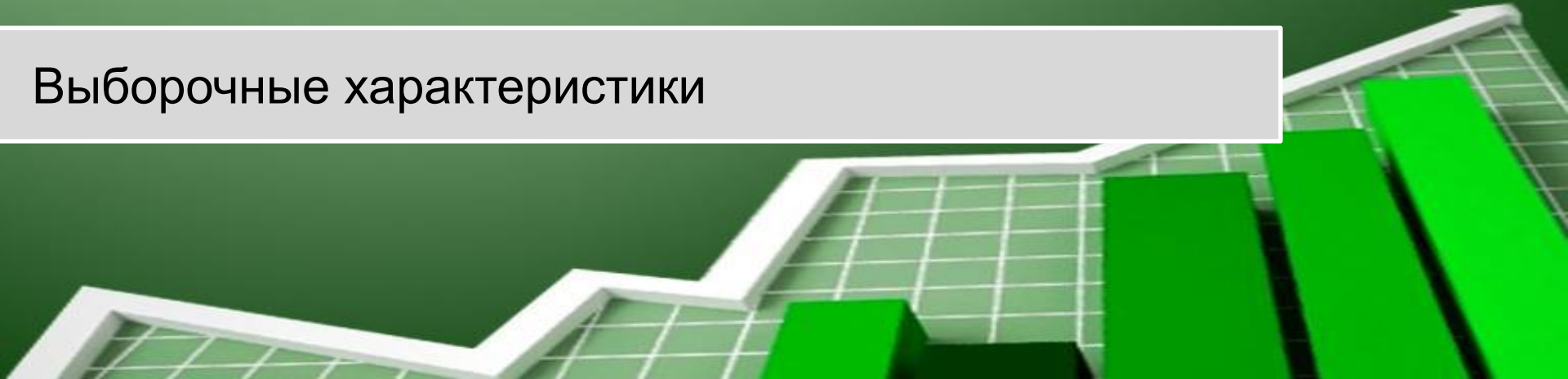
Предмет и содержание медицинской статистики

4

Основные понятия математической статистики

5

Выборочные характеристики



Назначение электронных таблиц

Для представления данных в удобном виде используют **таблицы**. Компьютер расширяет возможности использования таблиц за счет того, что позволяет не только представлять их в электронной форме, но и обрабатывать входящие в них данные.

Класс программ, используемый для этой цели, называется **табличными процессорами**, или, проще, **электронными таблицами (ЭТ)**.



Определение ЭТ. Программные средства

Электронные таблицы – это машинные модели обычных таблиц, состоящие из строк и столбцов, обрабатываемые и сохраняемые в памяти компьютера.

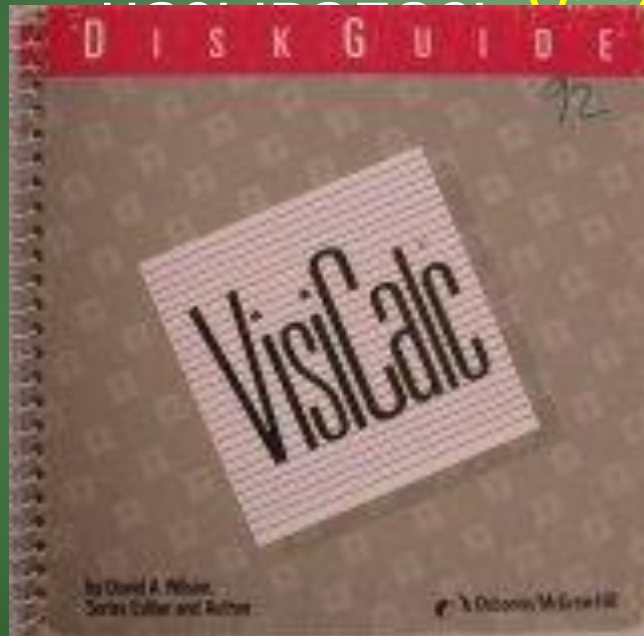
Первая программа была разработана в США в 1979 г.

Наиболее популярны такие программы, как **Excel**, **QuattroPro**, **Lotus 1-2-3**, **Works**



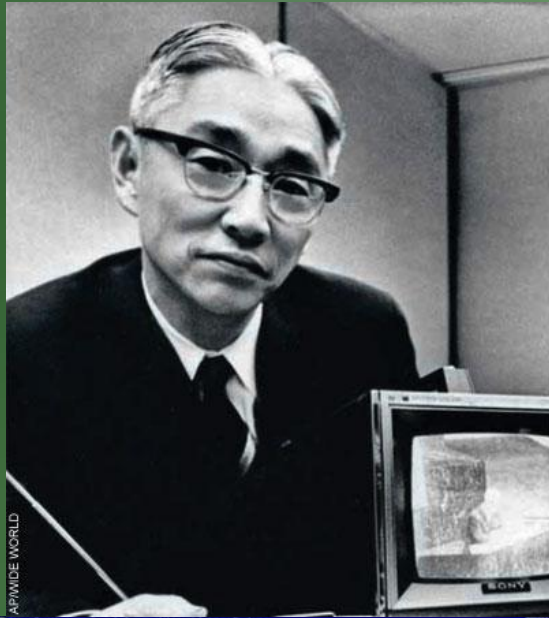
Программные средства ЭТ. VisiCalc

Первая программа для работы с электронными таблицами — табличный процессор, была создана в 1979 году, предназначалась для компьютеров типа Apple II и

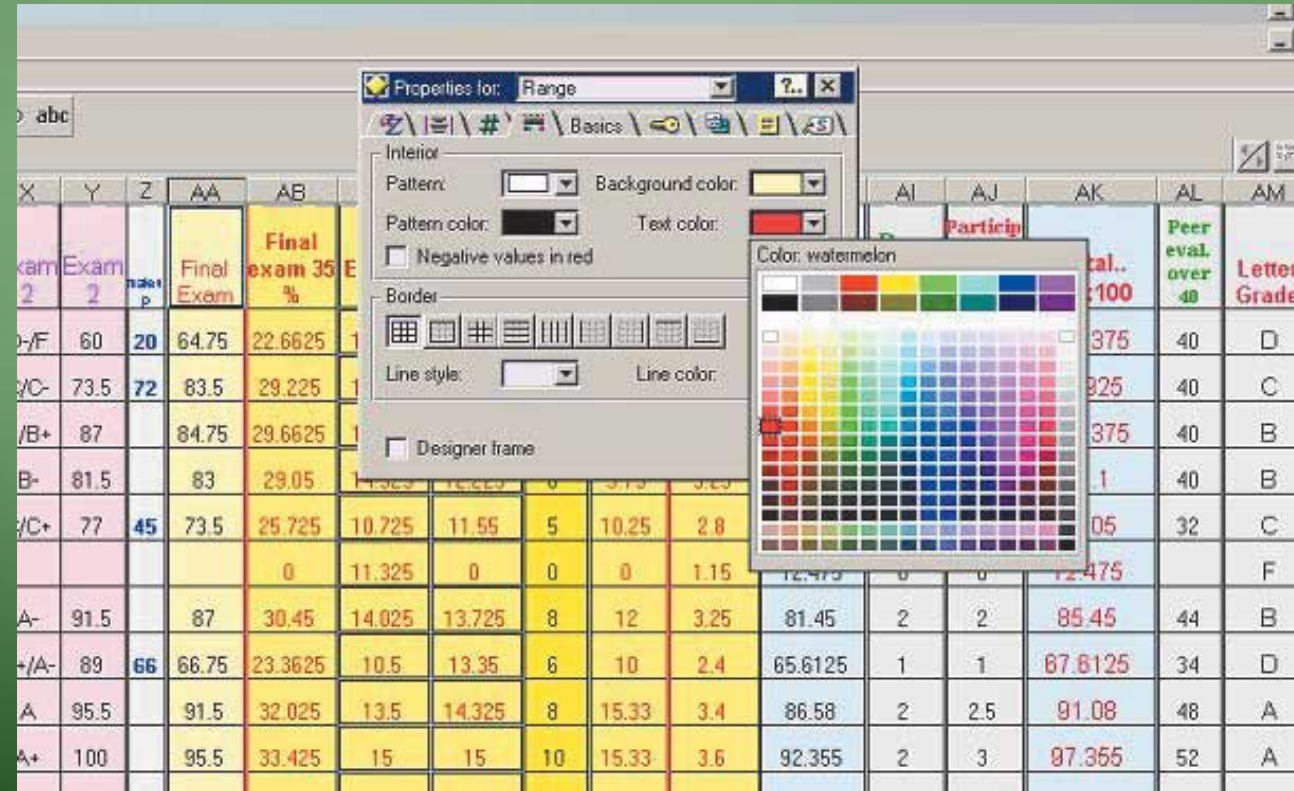
The screenshot shows a terminal window with a spreadsheet. The title bar at the top reads '120 (0) +R20+12' and '19'. The spreadsheet content is as follows:

HOME BUDGET, 1979			
MONTH	NOV	DEC	TOTAL
SALARY	2500.00	2500.00	30000.00
OTHER			
INCOME	2500.00	2500.00	30000.00
FOOD	400.00	400.00	4800.00
RENT	350.00	350.00	4200.00
HEAT	110.00	120.00	575.00
REC.	100.00	100.00	1200.00
TAXES	1000.00	1000.00	12000.00
ENTERTAIN	100.00	100.00	1200.00
MISC	100.00	100.00	1200.00
CAR	300.00	300.00	3600.00
EXPENSES	2460.00	2470.00	28775.00
REMAINDER	40.00	30.00	1225.00
SAVINGS	30.00	30.00	330.00

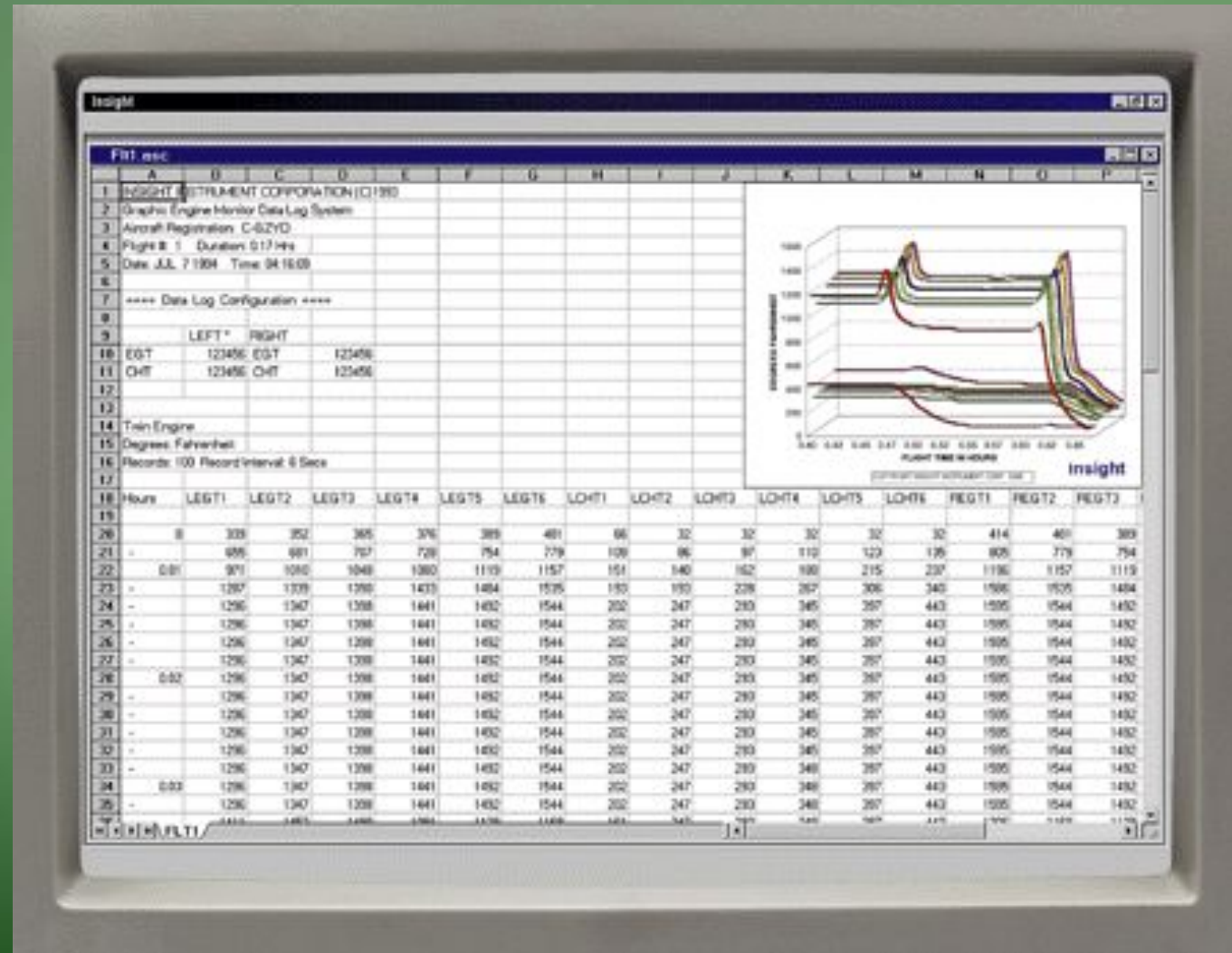
Программные средства ЭТ. Lotus 1-2-3



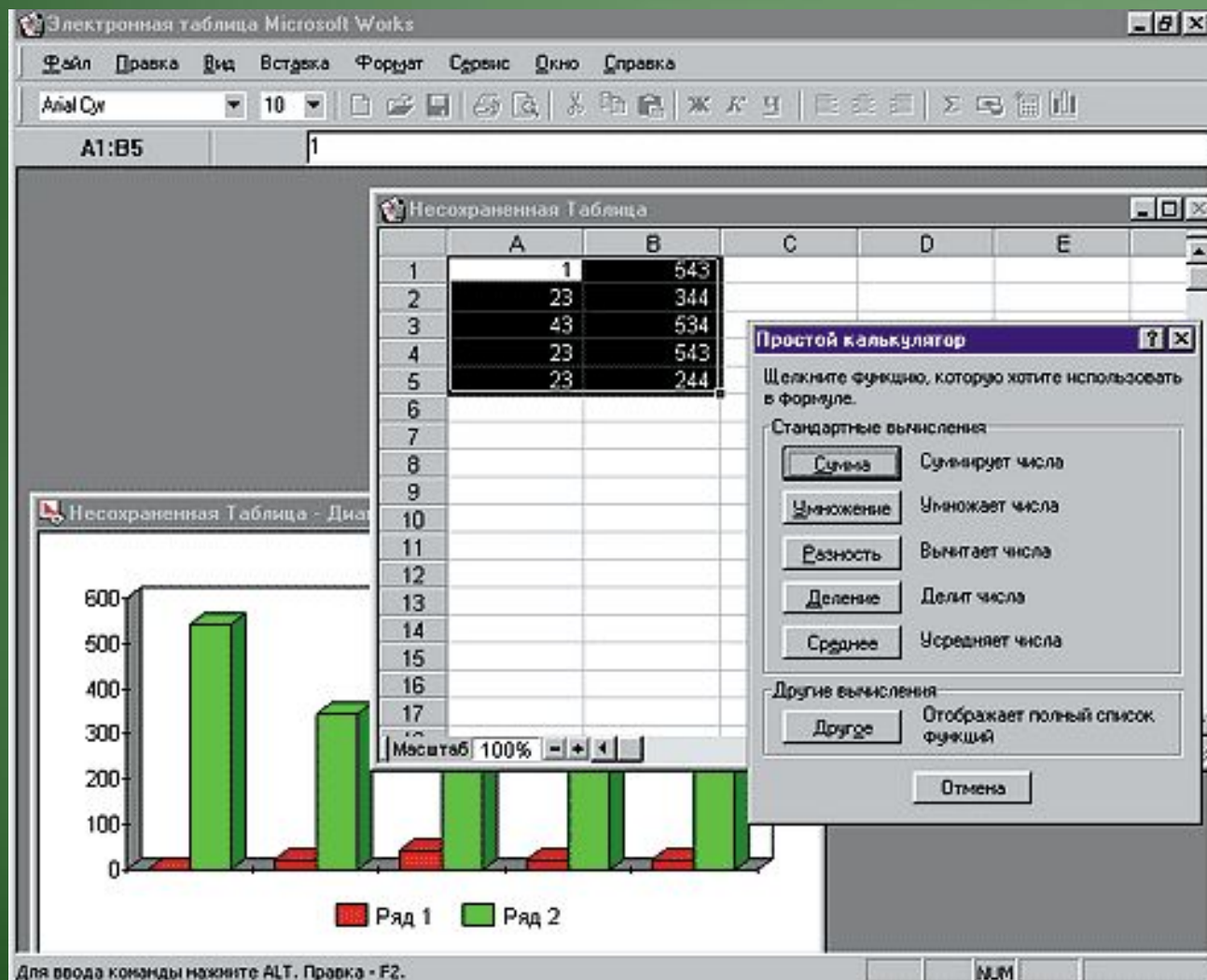
Lotus 1-2-3 появляется в 1982 году. Предназначался для IBM PC. Lotus объединял в себе вычислительные возможности электронных таблиц, деловую графику и функции реляционной СУБД.



Программные средства ЭТ. QuattroPro

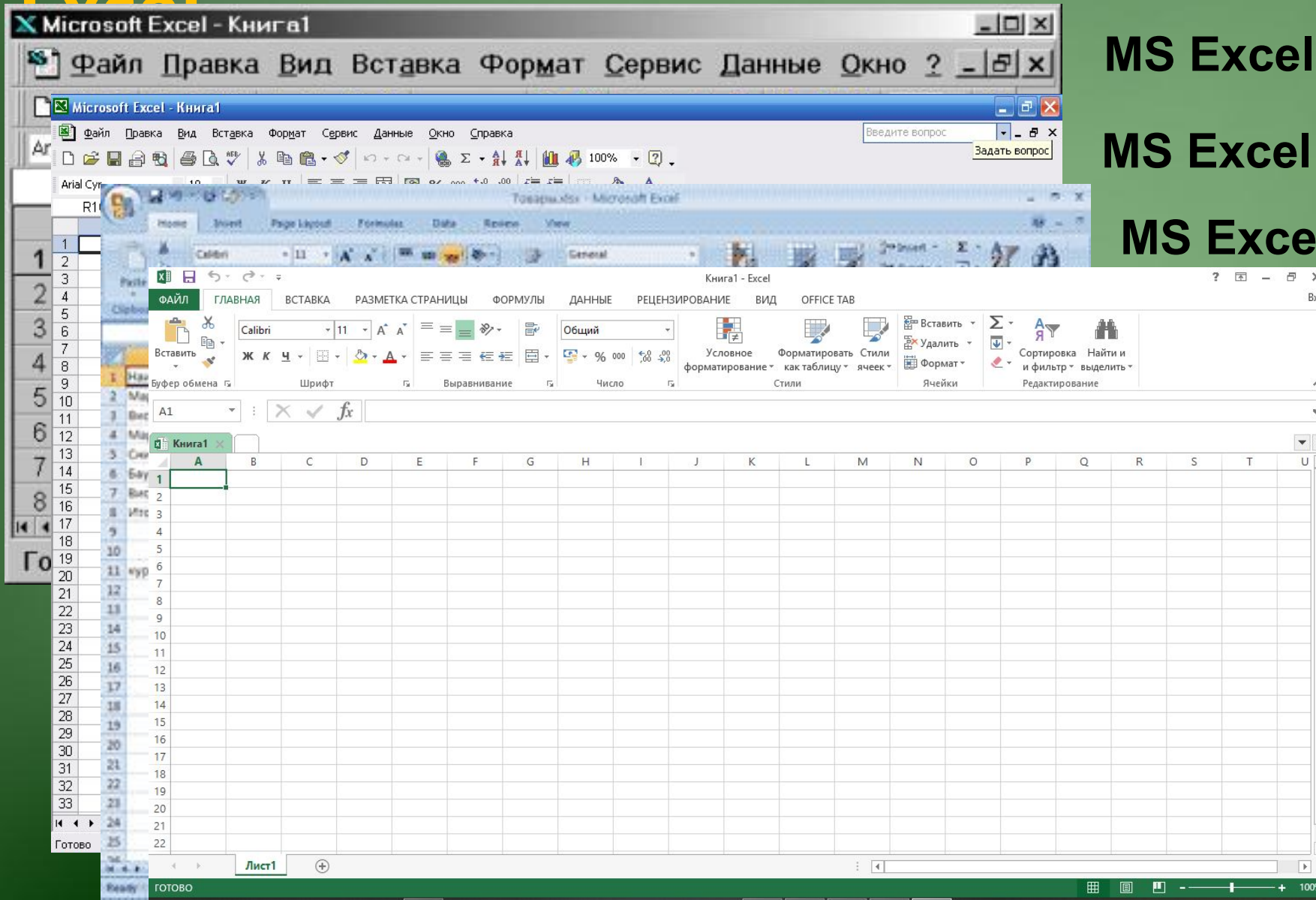


Программные средства ЭТ. Works



Пакет программ **MicroSoft Works** представляет собой интегрированную среду, включающую в себя текстовый процессор, электронные таблицы, систему управления базами данных и средство коммуникаций (систему компьютерной связи).

Программные средства ЭТ. Microsoft Excel



MS Excel 97

MS Excel 2003

MS Excel 2007

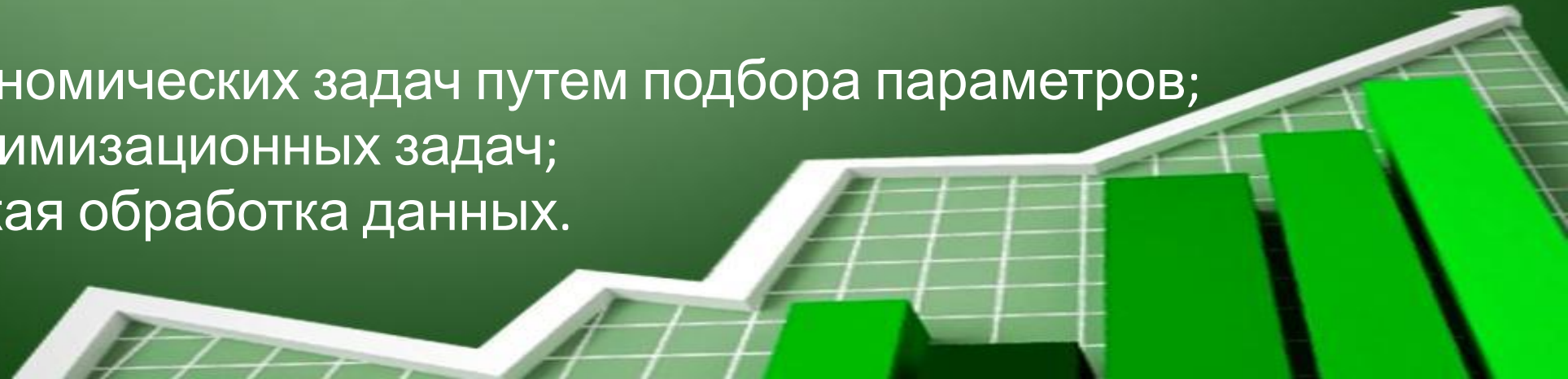


MS Excel 2013



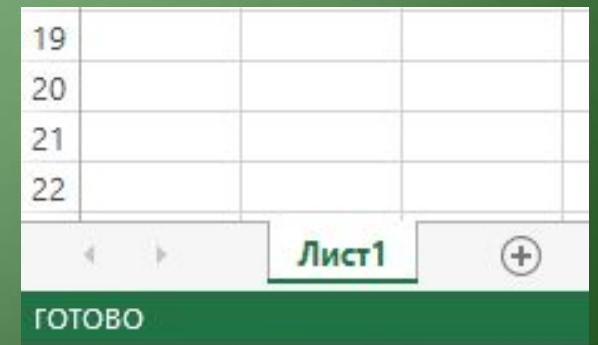
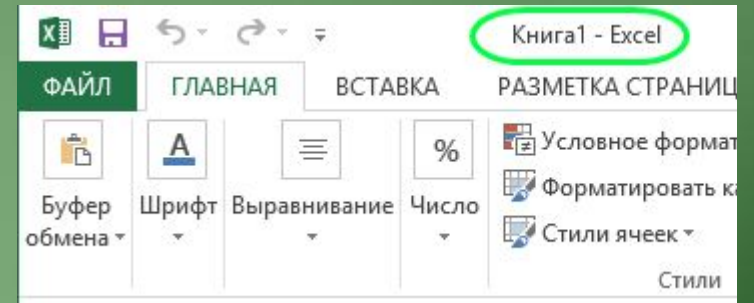
Функции табличных процессоров

- создание и редактирование электронных таблиц;
- оформление и печать электронных таблиц;
- создание многотабличных документов, объединенных формулами;
- построение диаграмм, их модификация и решение экономических задач графическими методами;
- работа с электронными таблицами как с базами данных: сортировка таблиц, выборка данных по запросам;
- создание итоговых и сводных таблиц;
- использование при построении таблиц информации из внешних баз данных;
- решение экономических задач путем подбора параметров;
- решение оптимизационных задач;
- статистическая обработка данных.



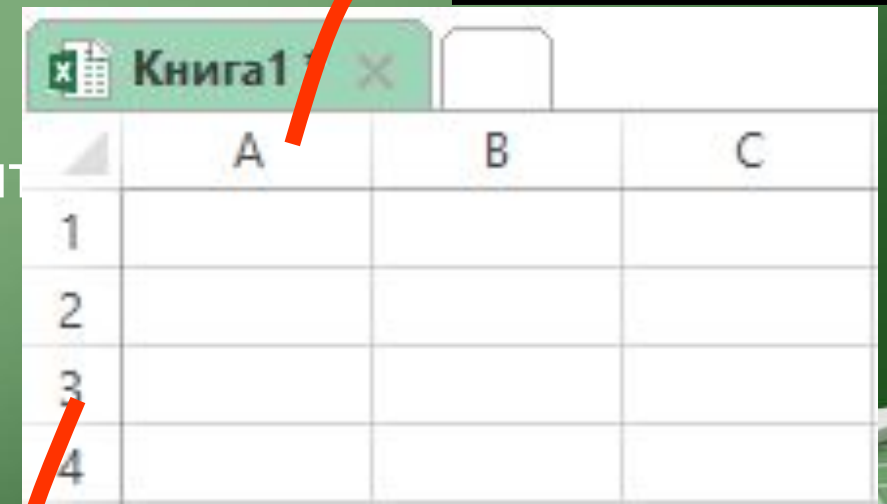
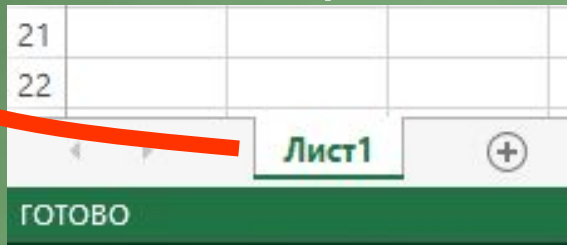
Основные элементы электронной таблицы.

- Документ электронной таблицы называется рабочей книгой.
- Рабочая книга представляет собой набор рабочих листов, каждый из которых имеет табличную структуру и может содержать одну или несколько таблиц. В окне документа отображается только текущий рабочий лист активной рабочей книги, с которым и ведется работа.



Основные элементы электронной таблицы.

- Каждый рабочий лист имеет название, которое отображается на **ярлычке листа**, отображаемом в его нижней части. С помощью ярлычков можно переключаться между рабочими листами, входящими же рабочую книгу.

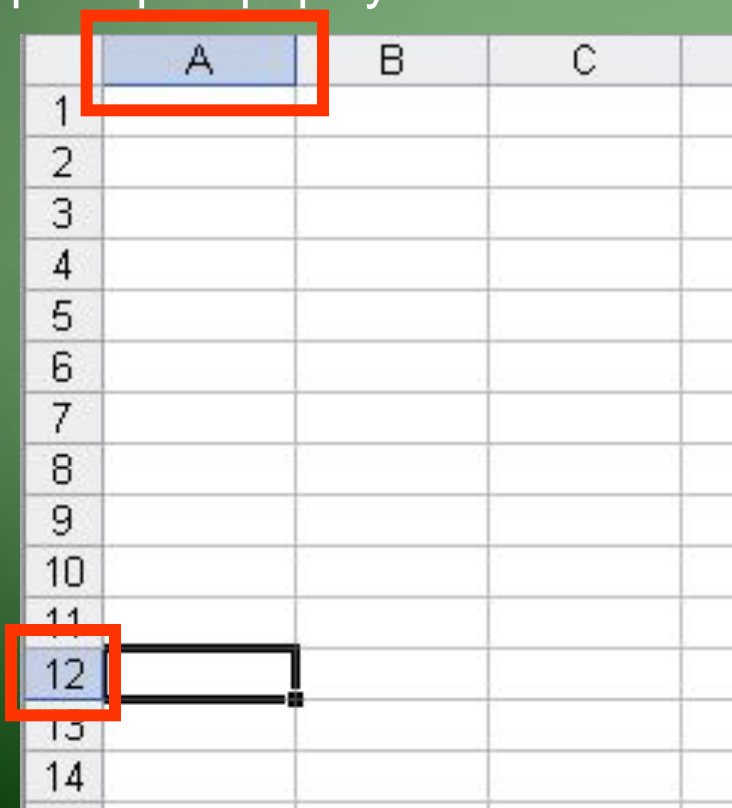


- Табличное пространство рабочего листа состоит из строк и столбцов.
- Столбцы** озаглавлены прописными латинскими буквами и, далее, двухбуквенными комбинациями. Всего рабочий лист может содержать до 16 384 столбцов, пронумерованных от A до XFD.
- Строки** последовательно нумеруются числами от 1 до 1 048 576

Основные элементы электронной таблицы.

На пересечении столбцов и строк образуются ячейки таблицы. Они являются минимальными элементами хранения данных.

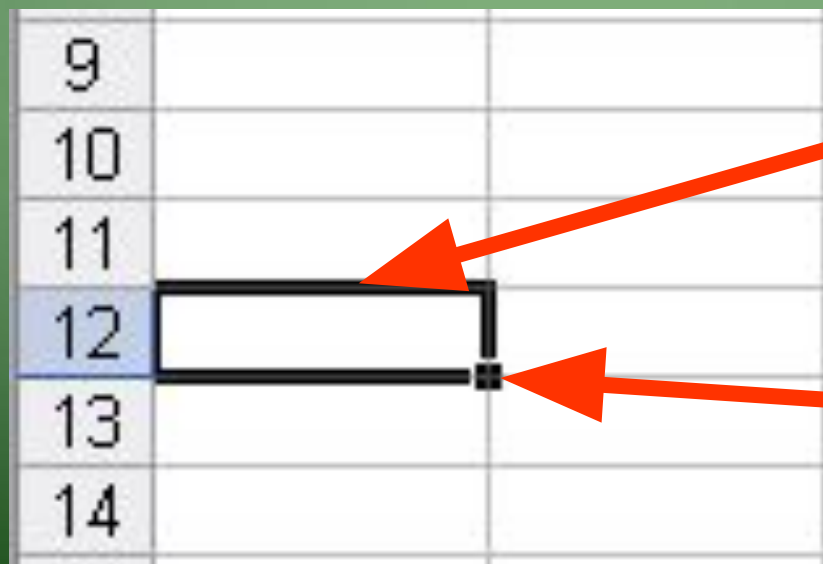
Обозначение отдельной ячейки сочетает в себе номер столбца и строки, на пересечении которой она расположена, например: A12 или AB21. Обозначение ячейки выполняет адресную функцию. Адреса ячеек используют при необходимости ссылки на нее, например в формулах.



имя ячейки



Одна из ячеек всегда активна и выделена **рамкой**. Эта рамка играет роль курсора. Операции ввода и редактирования всегда осуществляются в активной ячейке.



The image shows a portion of a spreadsheet grid. The rows are numbered 9 through 14. The cell at row 12, column 2 is highlighted with a thick black border. A small black square, known as a fill handle, is located at the bottom-right corner of this cell. Two red arrows point from labels to these elements: one to the border and one to the fill handle.

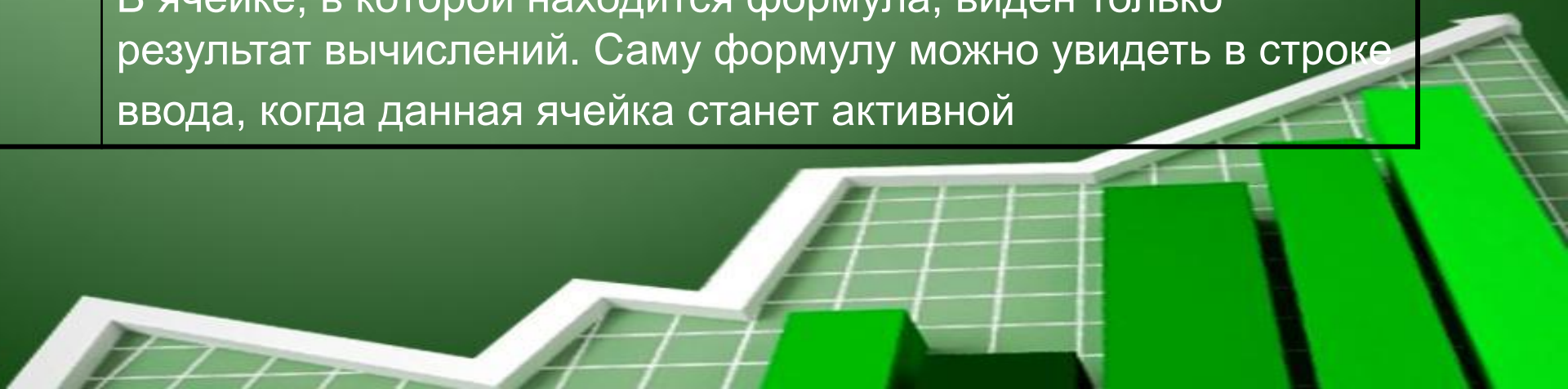
9		
10		
11		
12		
13		
14		

рамка

маркер
автозаполнения

Типы данных в электронных таблицах

Текст	Любая последовательность символов. Данные текстового типа используются для заголовков таблиц, заголовков строк и столбцов
Число	Это числовая константа
Формула	<p>Это выражение, состоящее из числовых величин и арифметических операций. Кроме числовых величин, в формулу могут входить в качестве аргументов адреса ячеек, функции и другие формулы.</p> <p>Пример формулы: =A5/H8.</p> <p>В ячейке, в которой находится формула, виден только результат вычислений. Саму формулу можно увидеть в строке ввода, когда данная ячейка станет активной</p>

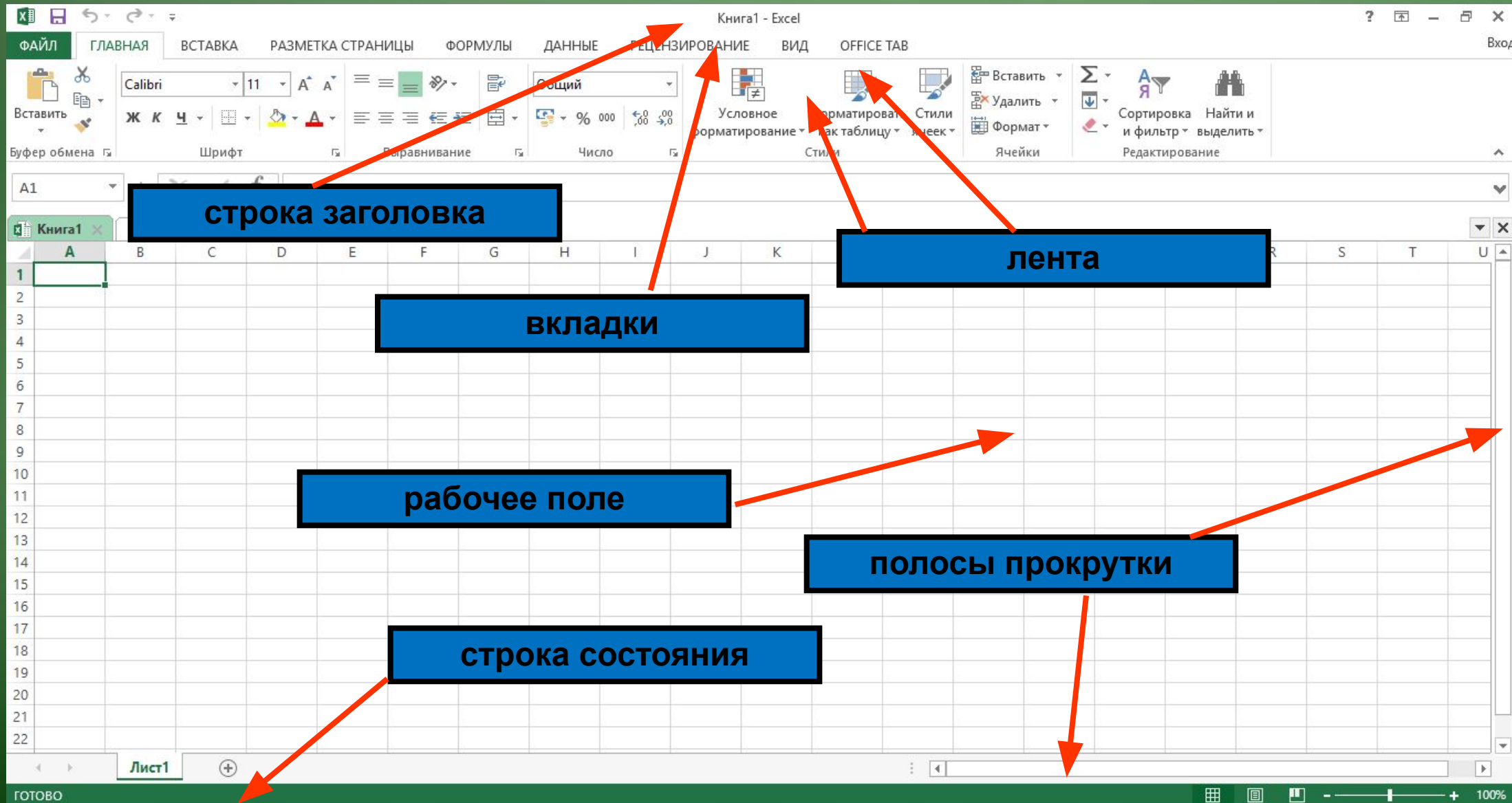


Типы данных в электронных таблицах

Функции	Это запрограммированные формулы, позволяющие проводить часто встречающиеся последовательности вычислений. Например, функция автосуммирования может быть представлена следующим образом: = СУММ (A1:A4)
Дата	Может быть представлена в различных форматах и с которой можно выполнять различные арифметические и логические операции



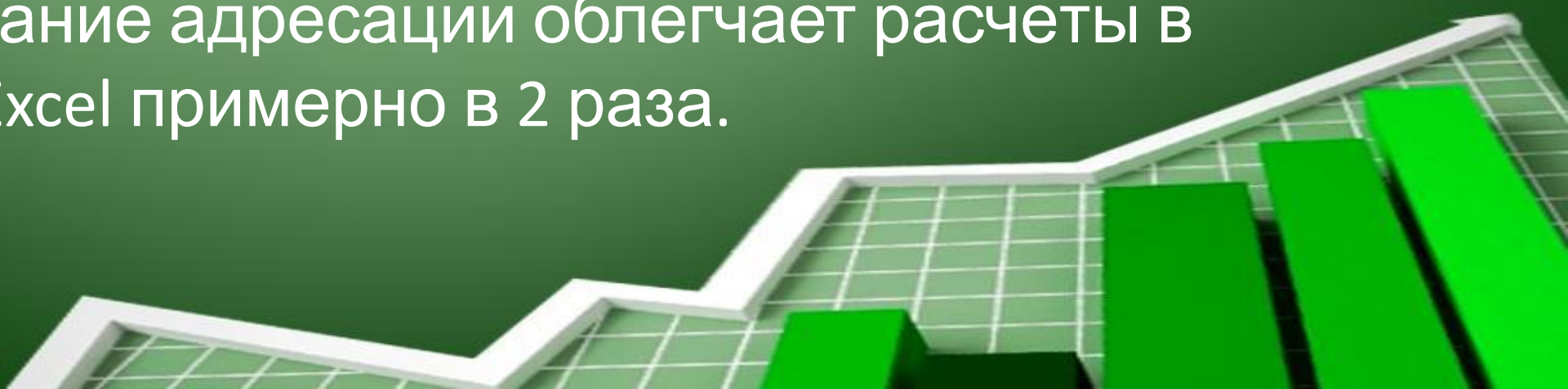
Внешний вид окна MS Excel



Адресация в Excel



- **Адресация** - это указание на данные находящиеся в определённой ячейке.
- Использование адресации облегчает расчёты в таблицах Excel примерно в 2 раза.



Адресация в Excel

- **Относительная** - это адресация, при которой формула читается относительно текущего положения ячейки.
- $=a1*v1$ – данную формулу, находящуюся в ячейке C1, ЭВМ для «себя» читает следующим образом: содержимое ячейки, находящейся на два столбца слева в той же строке, перемножить с содержимым ячейки находящейся на один столбец слева в той же строке.
- Если эту формулу скопировать из ячейки C1 в ячейку C2, то ее «понимание для ЭВМ» остается точно таким же. Т.е. она возьмет ячейку, находящуюся на 2 столбца слева (а это будет ячейка a2), и перемножит ее с ячейкой находящейся на 1 столбец слева (это будет ячейка b2). Формула в ячейке C2 примет вид $=a2*v2$
- Если эту формулу скопировать в ячейку C3, то она примет вид $=a3*v3$

	A	B	C
1	5	7	$=a1*v1$
2	10	15	$=a2*v2$
3			
4			

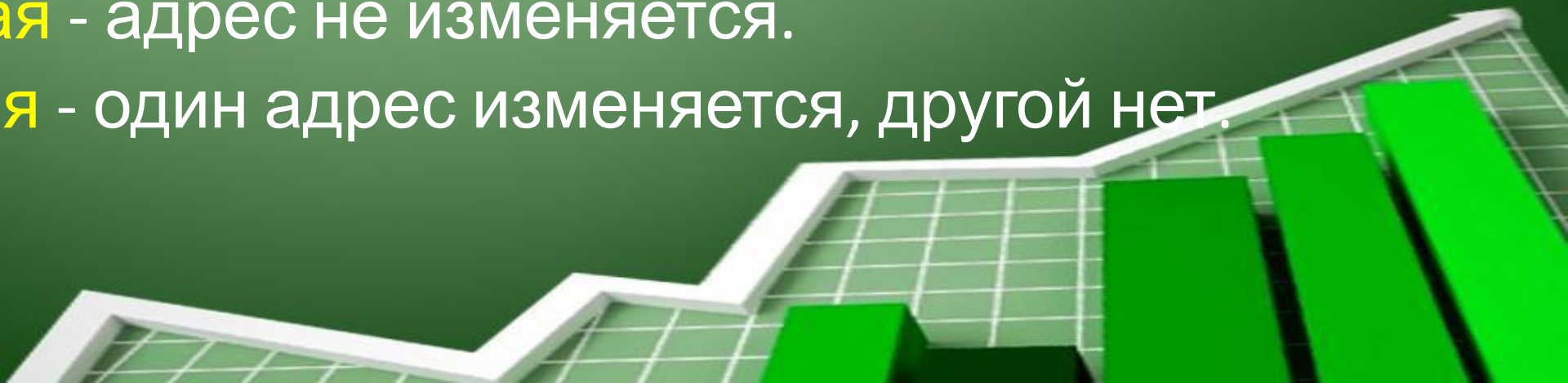
Адресация в Excel

- **Абсолютная** – это адресация, при которой идёт указание на конкретную ячейку, адрес которой не изменяется.
- \$ - признак абсолютной адресации. \$ - комбинация кнопок (Shift + 4) в англ. языке.
- Знак \$ ставится в двух местах - и перед буквой столбца и перед номером строки.

	A	B	C	D	E
1	наименование	Цена у.е.	Цена рублей	Курс	
2	Cooler 1	10	=B2*\$D\$2	35.45	
3	Cooler 2	15	=B3*\$D\$2		
4	Cooler 3	20			

Адресация в Excel

- **Смешанная** – адресация, при которой идёт изменение только одного адреса. Знак \$ ставится только в одном месте, или перед буквой столбца или перед номером строки.
- При копировании формул:
- **Относительная** - адреса изменяются, относительно текущего положения ячейки.
- **Абсолютная** - адрес не изменяется.
- **Смешанная** - один адрес изменяется, другой нет.



Формулы и функции. Операции с формулами

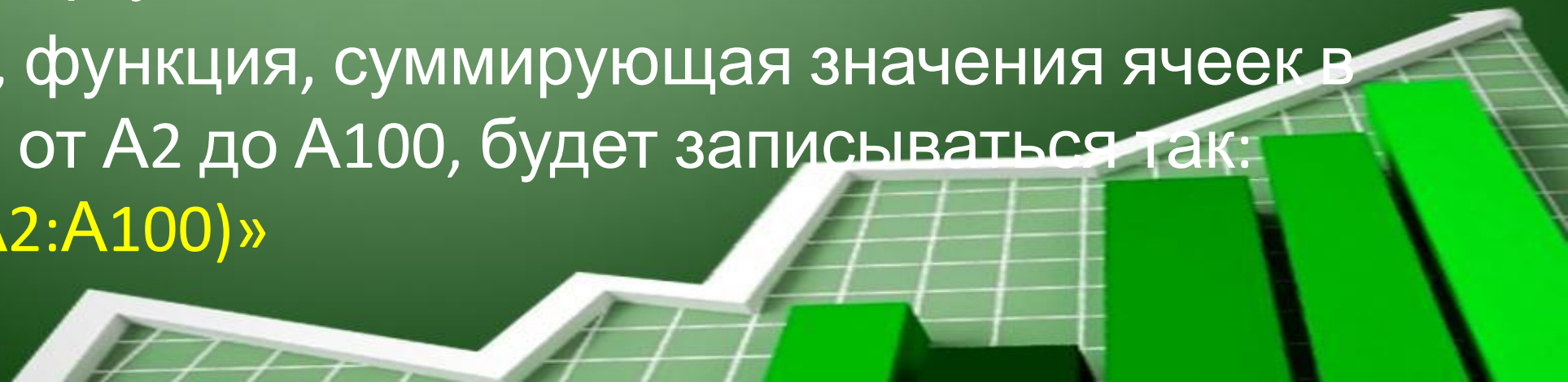
Возможность использования формул и функций является одним из важнейших свойств программы обработки электронных таблиц. Текст формулы, которая вводится в ячейку таблицы, должен начинаться со знака **равенства (=)**, чтобы программа Excel могла отличить формулу от текста. После знака равенства в ячейку записывается **математическое выражение**, содержащее аргументы, операции и функции.

В качестве аргументов в формуле обычно используются **числа** и **адреса ячеек**.



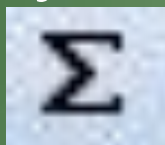
Понятие функции

- Для выполнения более сложных вычислений и облегчения многих расчетов используются **функции**.
- **Функцией** в **Microsoft Excel** называется сокращенная запись некоторых predetermined формул, которые выполняют вычисления, используя указанные значения – аргументы.
- Например, функция, суммирующая значения ячеек в диапазоне от A2 до A100, будет записываться так:
«=СУММ(A2:A100)»



Функция суммирования

Для суммирования нескольких ячеек используется кнопка

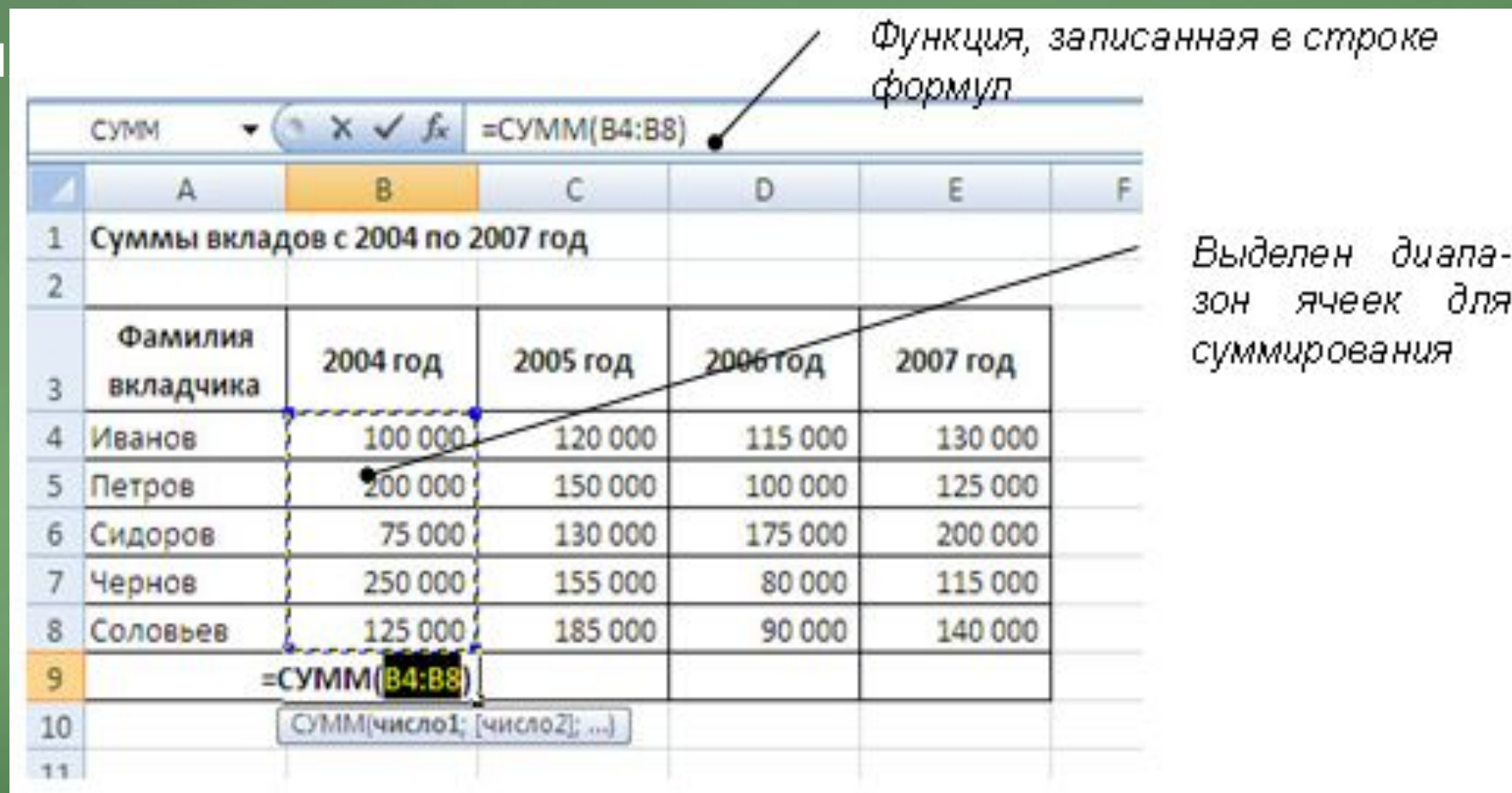


Функция, записанная в строке формул

	A	B	C	D	E	F
1	Суммы вкладов с 2004 по 2007 год					
2						
3	Фамилия вкладчика	2004 год	2005 год	2006 год	2007 год	
4	Иванов	100 000	120 000	115 000	130 000	
5	Петров	200 000	150 000	100 000	125 000	
6	Сидоров	75 000	130 000	175 000	200 000	
7	Чернов	250 000	155 000	80 000	115 000	
8	Соловьев	125 000	185 000	90 000	140 000	
9		=СУММ(B4:B8)				
10						
11						

Выделен диапазон ячеек для суммирования

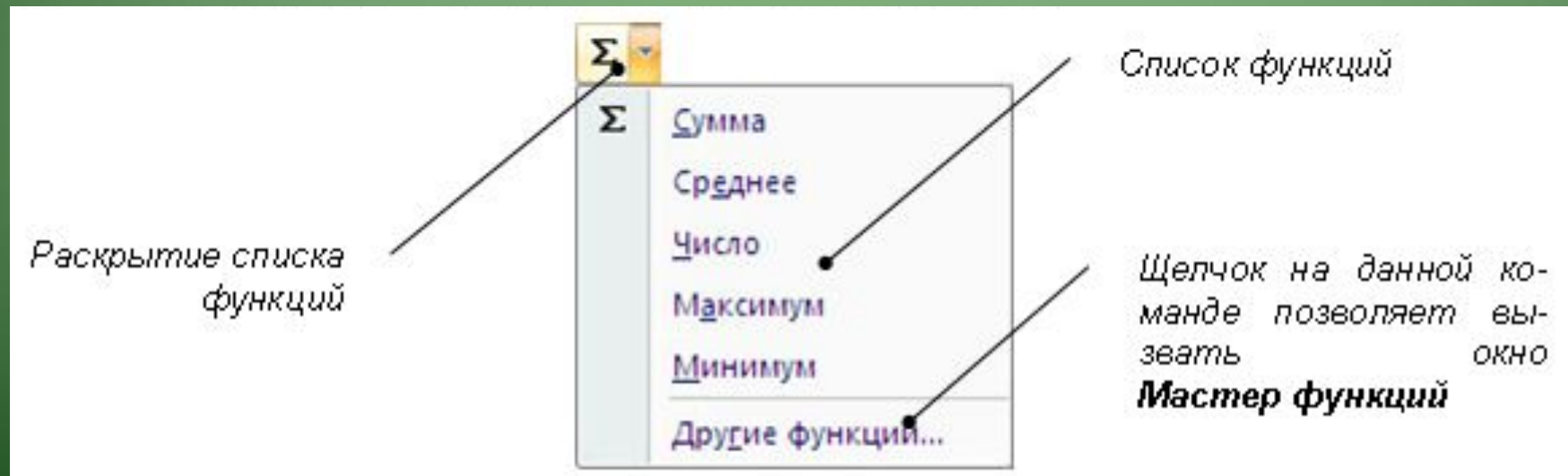
СУММ(число1; [число2]; ...)

The image shows a screenshot of an Excel spreadsheet. The formula bar at the top displays '=СУММ(B4:B8)'. A dashed blue border highlights the range of cells B4:B8, which contains the values for the years 2004 through 2007 for five different surnames. A tooltip at the bottom of the spreadsheet shows the syntax for the SUM function: 'СУММ(число1; [число2]; ...)'. A 3D bar chart is visible in the bottom right corner of the slide.

Статистические функции

Кроме функции суммирования часто используются функции *Среднее (СРЗНАЧ)*, *Минимум (МИН)* и *Максимум (МАКС)*, которые относятся к категории *Статистических функций*.

Для вставки любой из этих функций необходимо из списка кнопки *Сумма* выбрать название нужной функции:

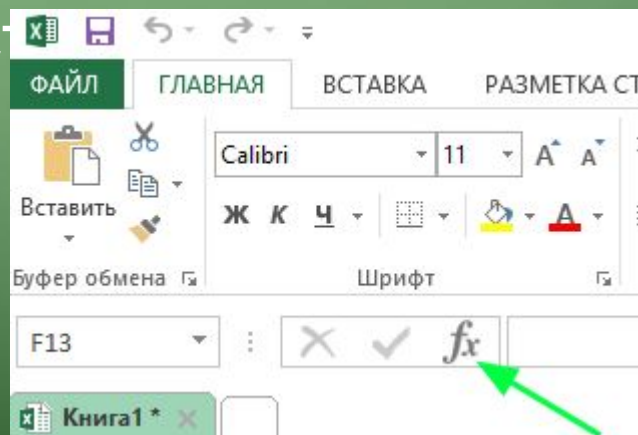


Использование мастера функций

Для использования других функций, содержащихся в программе Microsoft Excel, можно вызвать диалоговое окно **Мастер функций**.

В данном окне все функции разбиты по категориям: математические, статистические, финансовые и другие.

Для вызова данного окна требуется из списка кнопки **Сумма** Σ выбрать команду **Другие функции...** или щелкнуть fx на кнопке **Вставить функцию** в С



Диалоговое окно Мастер функций:

В данном поле можно ввести название функции для ее поиска

Список функций, принадлежащих к выбранной категории

Мастер функций - шаг 1 из 2

Поиск функции:

Введите краткое описание действия, которое нужно выполнить, и нажмите кнопку "Найти"

Найти

Категория: Математические

Выберите функцию:

- РИМСКОЕ
- РЯД.СУММ
- СР.СРЕЖДУ
- СЛ.ЧИС
- СТЕПЕНЬ
- СУММ**
- СУММЕСЛИ

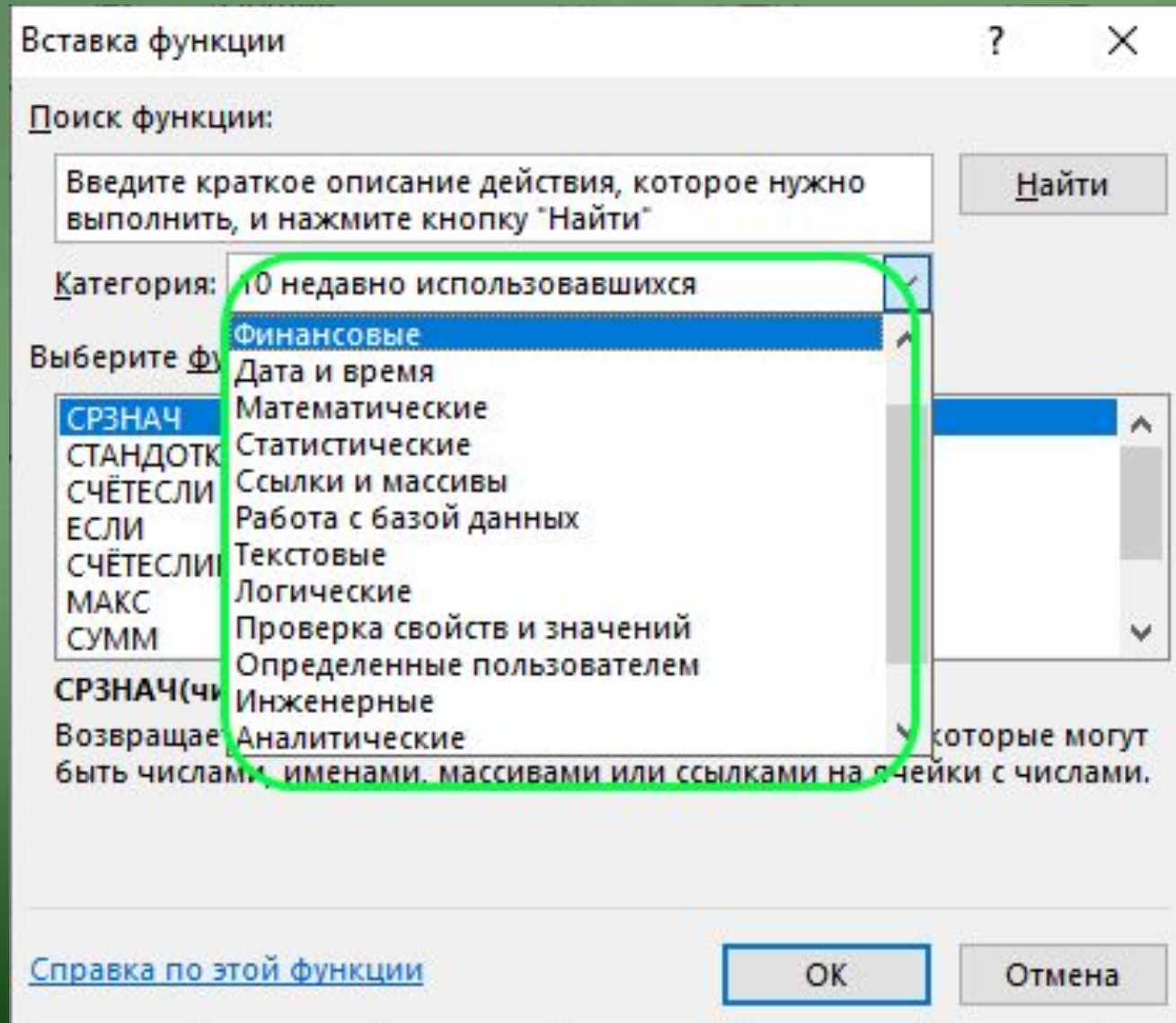
СУММ(число1;число2;...)
Суммирует аргументы.

[Справка по этой функции](#)

ОК Отмена

Список, позволяющий выбрать категорию функции (математические, статистические, логические и др.)

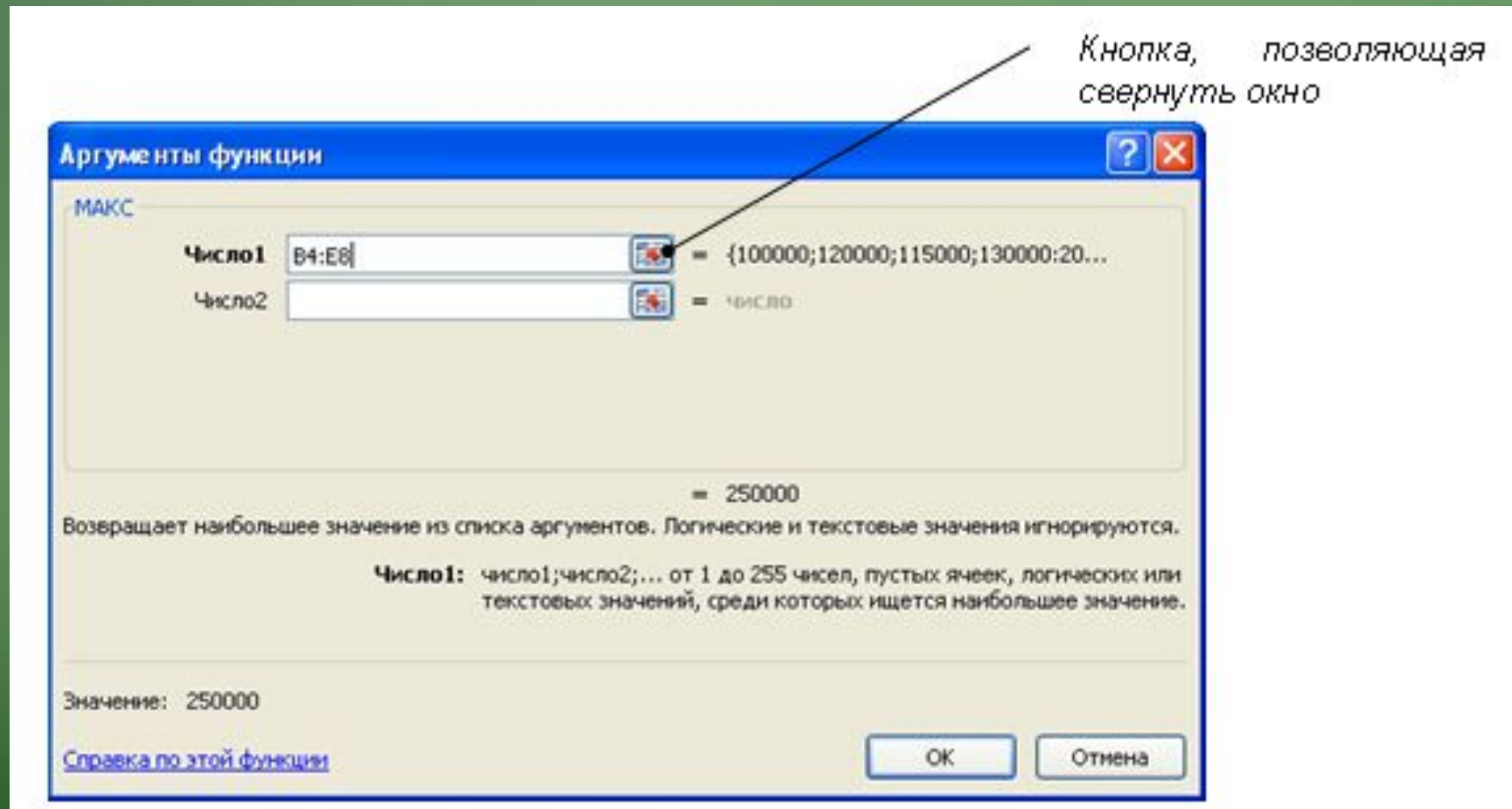
Перечень категорий функций



Пример вставки функции с использованием Мастера функций

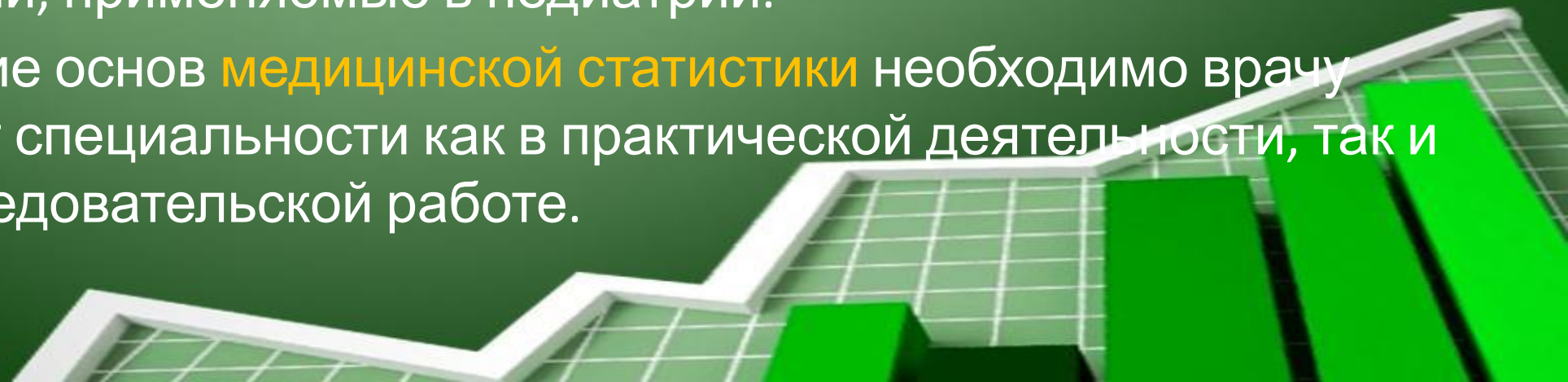
B11		fx				
	A	B	C	D	E	F
1	Суммы вкладов с 2004 по 2007 год					
2						
3	Фамилия вкладчика	2004 год	2005 год	2006 год	2007 год	
4	Иванов	100 000	120 000	115 000	130 000	
5	Петров	200 000	150 000	100 000	125 000	
6	Сидоров	75 000	130 000	175 000	200 000	
7	Чернов	250 000	155 000	80 000	115 000	
8	Соловьев	125 000	185 000	90 000	140 000	
9	Итого:	750 000	740 000	560 000	710 000	
10						
11	Максимальный вклад					
12						

Диалоговое окно Аргументы функции



Предмет и содержание медицинской статистики

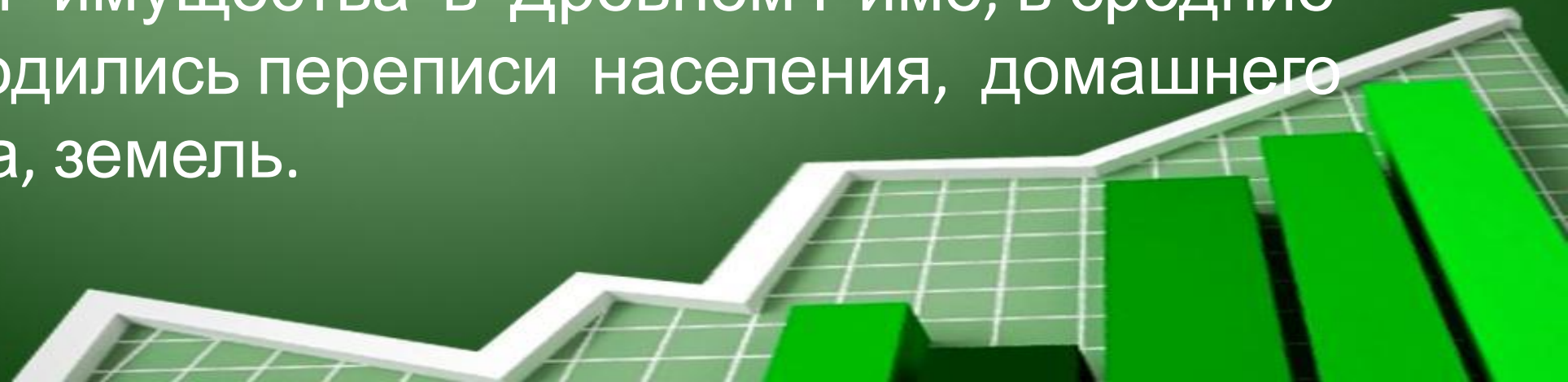
- В клинических и экспериментальных медицинских исследованиях **статистические методы** способствуют установлению закономерностей, присущих патологическому процессу, позволяют применять количественные критерии для оценки отдельных видов лечения. Все это требует значительного расширения числа врачей, владеющих статистической методикой в такой степени, чтобы самостоятельно проводить исследования, анализировать медико-статистические данные касающиеся не только здоровья населения, но и показатели, применяемые в педиатрии.
- Поэтому знание основ **медицинской статистики** необходимо врачу независимо от специальности как в практической деятельности, так и в научно-исследовательской работе.



- Начало формирования статистики положено зарождением хозяйственного учета и связано с образованием государства. Становление **статистики** определялось социально-экономическими потребностями общества и государства. **Статистический метод** формировался постепенно по мере накопления учетных данных, опыта их обработки и анализа. Достижения социально-экономической статистики являются результатом исторического развития. Термин **«статистика»** (лат. **status** – положение, состояние явлений). Первым его употребил немецкий ученый Г. Ахенваль в 1749 году.



- Прежде чем стать наукой в ее современном понимании **статистика** прошла многовековую историю развития.
- Числовые данные, относящиеся к тем или иным явлениям, начали применяться уже в глубокой древности. Так, известно, что еще за 5 тысяч лет до нашей эры проводился подсчёт населения в Китае, вёлся учет имущества в Древнем Риме, в средние века проводились переписи населения, домашнего имущества, земель.



- **Статистика** – наука, изучающая количественные закономерности материальных явлений в неразрывной связи с их качественной стороной.
- **Медицинская статистика** - раздел статистики, изучающий состояние здоровья населения и общественное здравоохранение.



**Направления
статистики:**

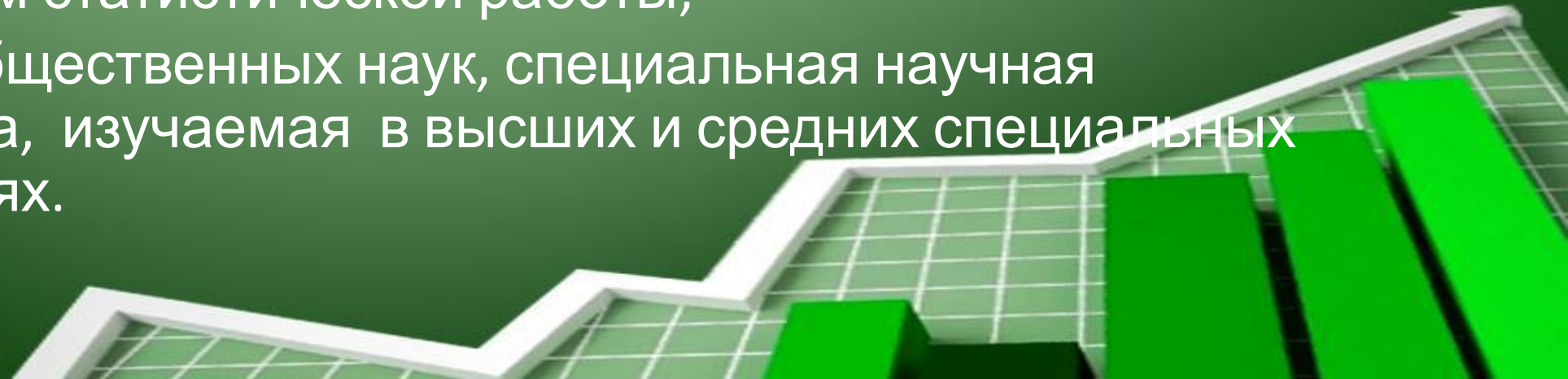
Описательная.

Экономическая.



Термин “статистика” употребляется в трёх значениях:

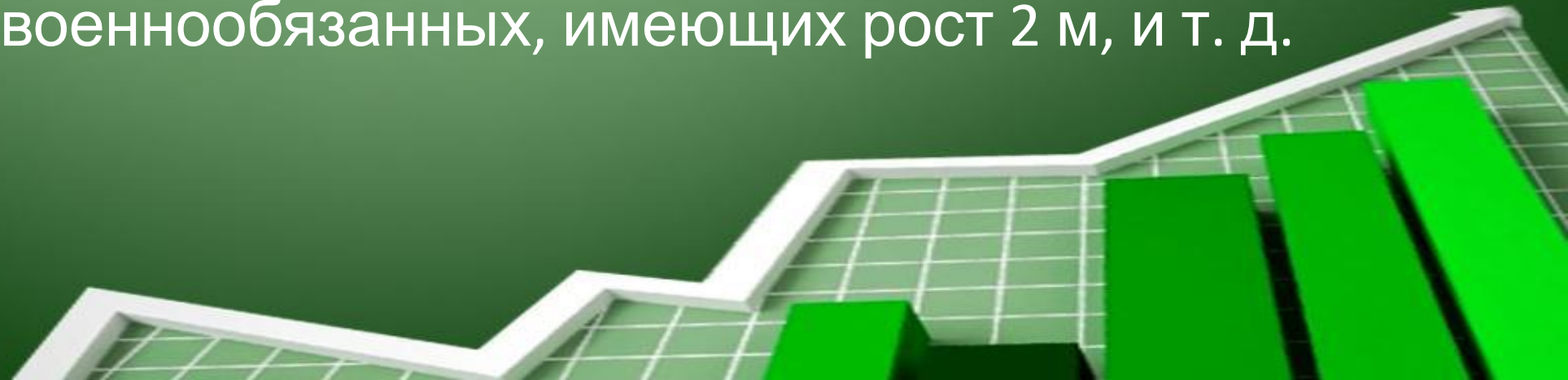
- 1) Отрасль практической деятельности (“статистический учёт”) по сбору, обработке, анализу и публикации массовых цифровых данных о самых различных явлениях и процессах общественной жизни.
- 2) Совокупность цифровых сведений, статистические данные, предоставляемые в отчетности предприятий, организаций, отраслей экономики, а также публикуемые в сборниках, справочниках, периодической прессе, которые являются результатом статистической работы;
- 3) Отрасль общественных наук, специальная научная дисциплина, изучаемая в высших и средних специальных учреждениях.



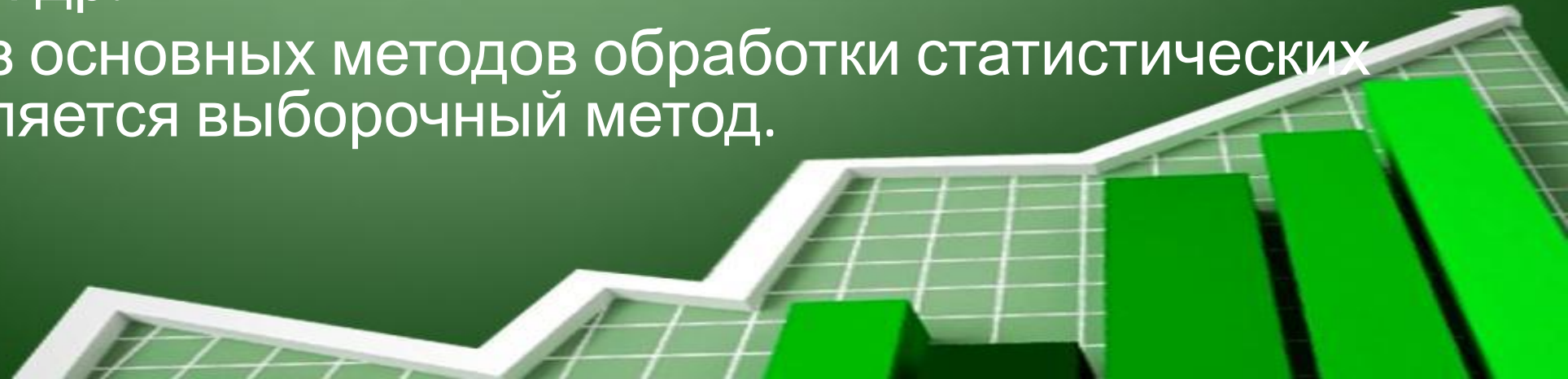
- **Математическая статистика** возникла в XVII веке. Решения первых задач математической статистики содержатся в сочинениях основоположников теории вероятностей Я. Бернулли, П. Лапласа и С. Пуассона.
- В **математической статистике** рассматриваются приближенные методы отыскания законов распределения и числовых характеристик по результатам экспериментов.



- В **математической статистике** разрабатываются теории и методы обработки информации о массовых явлениях.
- Исходным материалом статистического исследования являются статистические данные.
- Под **статистическими данными** понимают сведения о числе объектов обширной совокупности признаков.
- Например, сведения:
 - а) о числе отличников в каждом курсе данного факультете;
 - б) о числе военнообязанных, имеющих рост 2 м, и т. д.



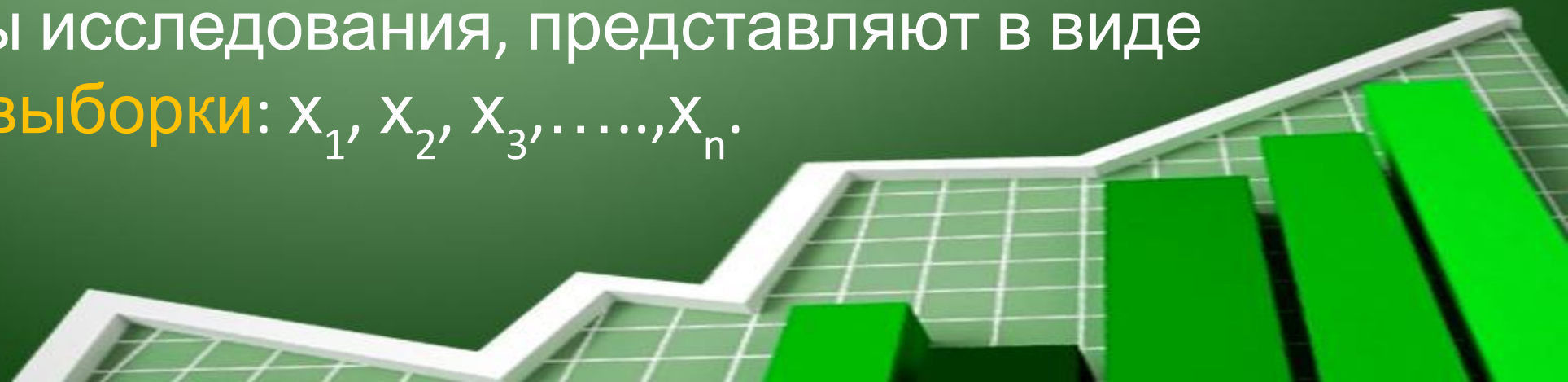
- На основании **статистических данных** часто можно делать вполне определенные научно обоснованные выводы, представляющие большую ценность для науки и практики.
- Для этого статистические данные должны быть предварительно определенным образом систематизированы и обработаны.
- Методы математической статистики широко применяются в самых различных областях знаний - в физике, астрономии, экономике, геологии, гидрологии, биологии, медицине и др.
- Одним из основных методов обработки статистических данных является выборочный метод.



- **Генеральный совокупность**- совокупность всех исследуемых объектов.
- **Объем выборки** - из всей совокупности отбирают случайным образом определенное число объектов.
- **Например**, для контроля качества растворов в ампулах для инъекций на отсутствие в них механических загрязнений из серии 5000 ампул отбирают 150 ампул.
- Здесь $N=5000$ - объем генеральной совокупности
- $n=150$ - объем выборки.



- Для того чтобы свойства выборки достаточно хорошо отражали свойства генеральной совокупности, выборка должна быть репрезентативной (представительной).
- Согласно закону больших чисел, можно утверждать, что выборка будет репрезентативной, если ее осуществить случайно. Для статистической обработки результаты исследования, представляют в виде **числовой выборки**: $x_1, x_2, x_3, \dots, x_n$.



- **Вариационный ряд**
- Выборку, представляющую собой возрастающую последовательность чисел, называют вариационным рядом.
- Любую числовую выборку можно записать в виде вариационного ряда.
- Например, записав значения выборки
- $1, 10, -2, 1, 0, 1, 10, 7, -2, 10, 10, 7$
- в виде возрастающей последовательности, получим вариационный ряд
- $-2, -2, 0, 1, 1, 7, 7, 10, 10, 10, 10.$



- Пусть при исследовании некоторой генеральной совокупности получена числовая выборка объема n , причем значение x_1 встретилось в выборке n_1 раз, значение x_2 - n_2 раз, ..., x_k - n_k раз.
- Числа $n_1, n_2, n_3, \dots, n_k$ называют **частотами**,
- а их отношения к объему выборки,
- $n_1/n; n_2/n; \dots, n_k/n$ **относительными частотами**



- Очевидно, что сумма частот равна объему выборки $n_1+n_2+n_3+\dots+n_k = n$,
- а сумма относительных частот равна единице,
- $n_1/n + n_2/n + \dots + n_k/n = 1$
- Статистический ряд
- Последовательность пар число с частотой встречаемости $(x_1;n_1), (x_2;n_2), \dots, (x_k;n_k)$ называют **статистическим рядом**.
- Обычно **статистический ряд** записывают в виде таблицы:

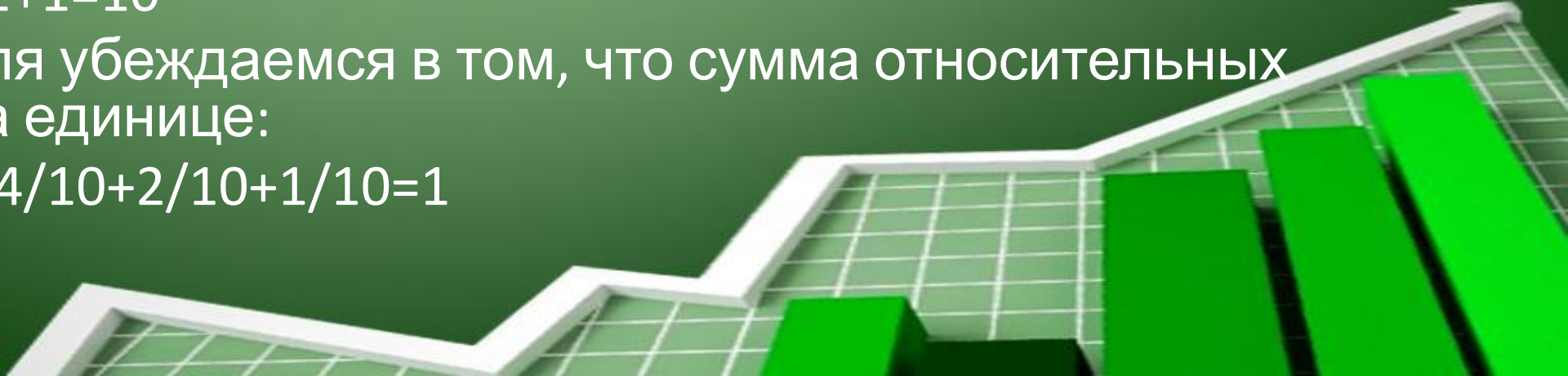
x_i	x_1	x_2	x_3	x_k
n_i	n_1	n_2	n_3	n_k

- Пример. Для выборки
- 3, 8, -1, 3, 0, 5, 3, -1, 3, 5
- 1) Определить объем и размах.
- 2) Записать выборку в виде вариационного ряда и в виде статистического ряда.
- Объем выборки $n=10$, ее размах равен
- $8 - (-1) = 9$
- Записав значения выборки в виде возрастающей последовательности, получим вариационный ряд
- -1, -1, 0, 3, 3, 3, 3, 5, 5, 8.



- Статистический ряд можно записать в виде последовательности пар чисел
- $(-1;2),(0;1),(3;4),(5;2),(8;1)$
- или в виде таблицы

x_i	-1	0	3	5	8
n_i	2	1	4	2	1

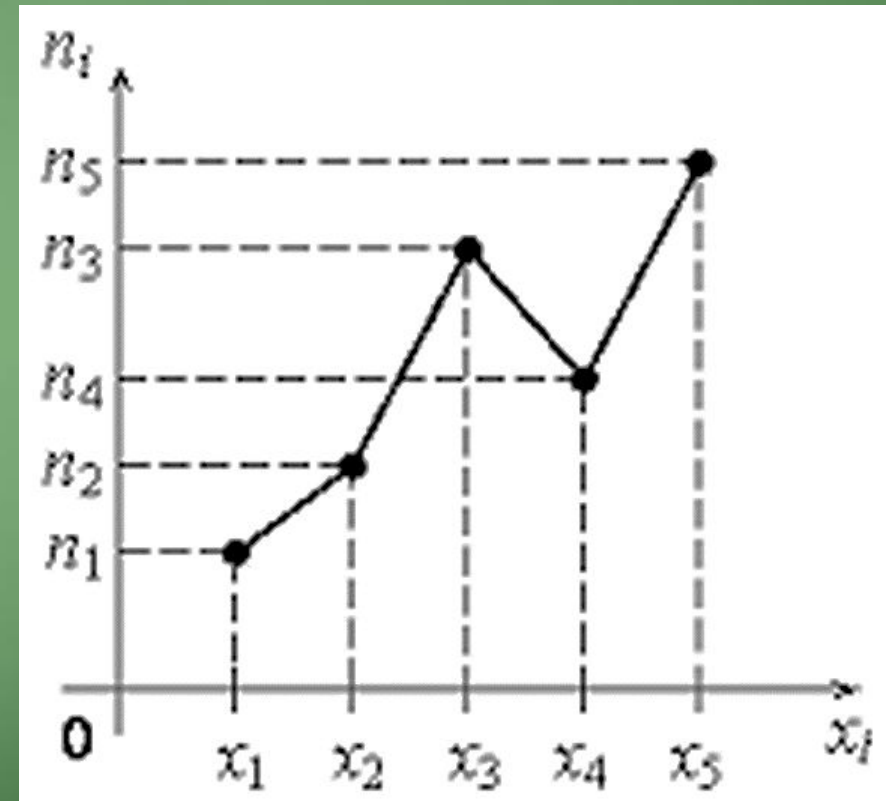
- Для контроля находим сумму частот, которая должна быть равной объему выборки
 - $2+1+4+2+1=10$
 - Для контроля убеждаемся в том, что сумма относительных частот равна единице:
 - $2/10+1/10+4/10+2/10+1/10=1$
- 

Графические изображения выборки: Полигон и гистограмма

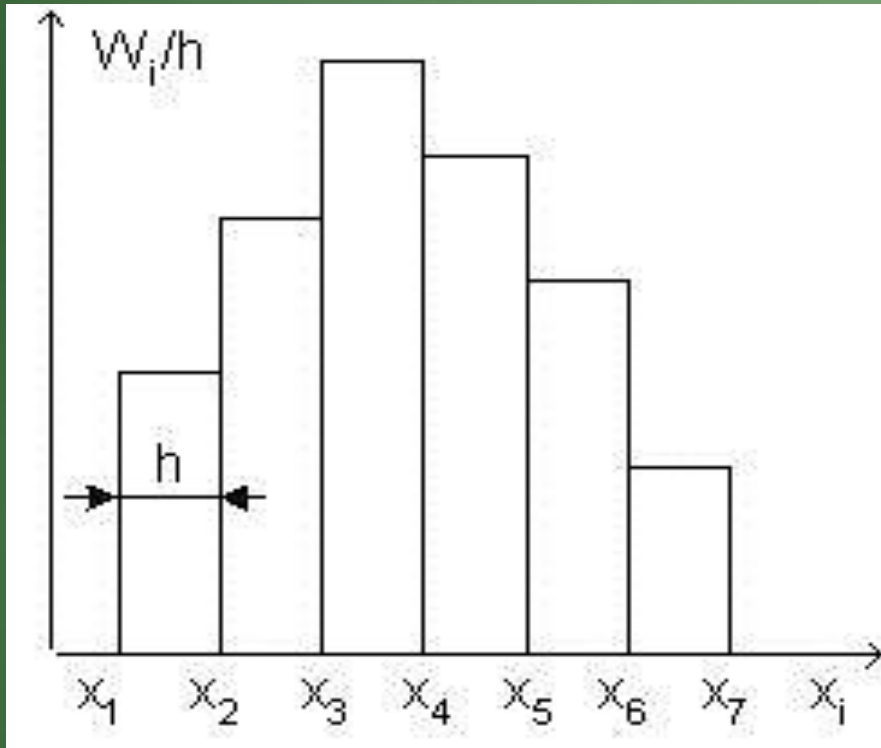
Полигоном частот называют ломаную линию с вершинами в точках $(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k)$

Полигон относительных частот получается из полигона частот сжатием вдоль оси ординат в n раз.

$(x_1; n_1/n), (x_2; n_2/n), \dots, (x_k; n_k/n)$



Полигон и гистограмма



- **Гистограммой** частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высоты равны отношению n_i/h (плотность частоты)

Выборочные характеристики

Среднее арифметическое значение.

Пусть имеется некоторая выборка объема n :

$$x_1, x_2, \dots, x_n.$$

Сумма объема выборки определяется с помощью формулы

$$S = x_1 + x_2 + \dots + x_n$$

Выборочное **среднее** для вариационного ряда определяется формулой

$$M_B = (x_1 + x_2 + \dots + x_n) / n$$

$$\sum x_i$$

$$M_B = \frac{\sum x_i}{n}$$

n



Среднее квадратичное отклонение

- Этот показатель характеризует степень разброса результатов наблюдений вокруг средней арифметической величины, а также дает сведения об устойчивости и не устойчивости изучаемого процесса.
- Для вариационного ряда **среднее квадратичное отклонение** вычисляется по формуле

$$\sigma_B = \sqrt{((x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2) / (n - 1)}$$

$$\sigma_B = \sqrt{\frac{\sum (x_i - M)^2}{n - 1}}, \quad i = 1, n$$



- Для **статистического ряда** среднее квадратичное отклонение вычисляется по формуле

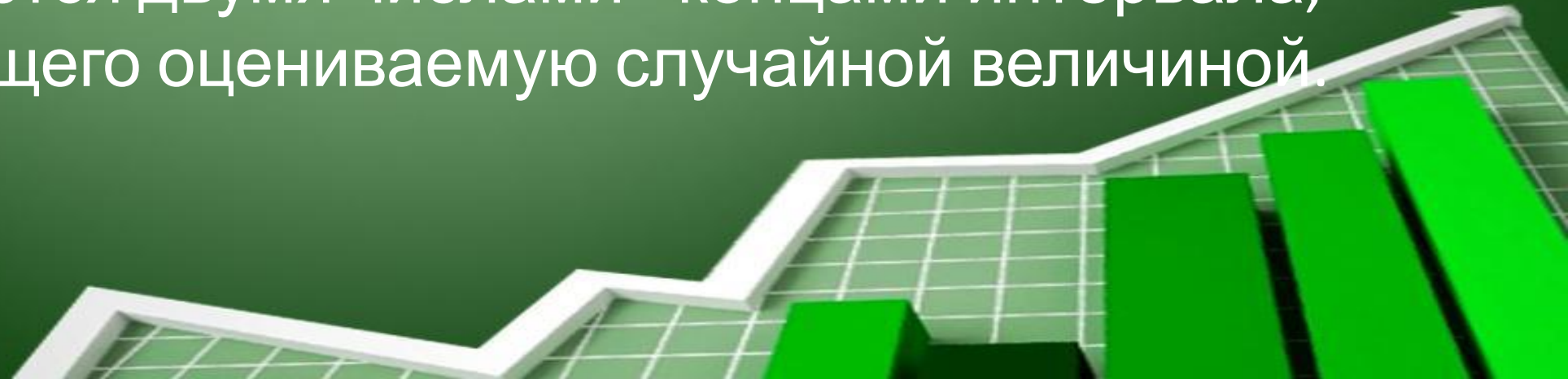
$$\sigma_{\text{ст}} = \sqrt{(n_1(x_1 - M)^2 + n_2(x_2 - M)^2 + \dots + n_k(x_k - M)^2) / (n - 1)}$$

$$\sigma_{\text{ст}} = \sqrt{\frac{\sum n_i (x_i - M)^2}{n - 1}}, \quad i = 1, k$$



Стандартная ошибка

- При выборке малого объема, оценка арифметического значения может значительно отличаться от истинного значения случайной величины и приводить к грубым ошибкам.
- Поэтому при небольшом объеме наблюдений используются интервальные оценки, которые определяются двумя числами - концами интервала, покрывающего оцениваемую случайной величиной.



Стандартная ошибка

среднего арифметического вычисляется по формуле

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

Доверительный интервал ($M \pm m$)

Величина стандартная ошибка, показывает те пределы в которых заключается истинное значение случайной величины

$$M \pm m.$$

А) Верхний предел $M + m$

Б) Нижний предел $M - m$

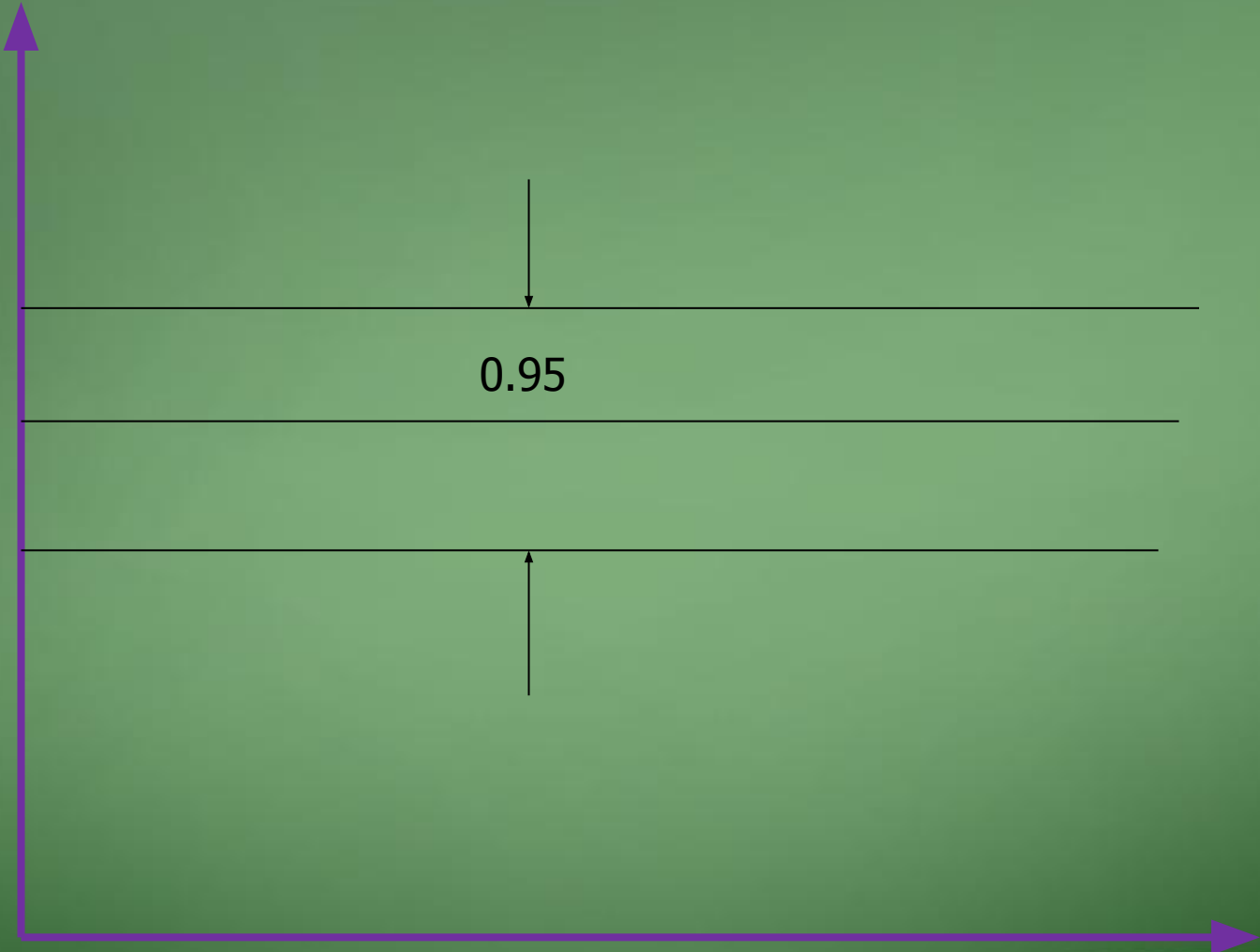


$M+m$

M

$M-m$

0.95



Показатель точности результатов (Критерий достоверности)

С помощью среднего арифметического значения M и средней квадратичное ошибки m , можно оценить точности производимых наблюдений.

Для этого вычисляется показатель точности наблюдений, который обозначим R_t .

$$R_t = m / M * 100\%$$




▣ Если $R \leq 5\%$, то тогда 95% результатов исследований считается достоверными.

(Вероятность и доверительный интервал результатов исследования, подчиняющихся нормальному закону распределения

$P = 95\% (M + \sigma)$, $P = 99\% (M + 2\sigma)$, $P = 99,9 (M + 3\sigma)$).



$$T = \frac{|M_1 - M_2|}{\sqrt{m_1^2 + m_2^2}}$$

- где, **T** - критерий Стьюдента;
 - **M₁** - среднее арифметическое значение результата наблюдений группы контрольных опытов;
 - **M₂** - среднее арифметическое значение результата других групп наблюдений;
 - **m₁, m₂** - стандартные ошибки контрольных и наблюдаемых групп;
- 

Далее на основании вычисленных значений $T_{\text{эксп}}$ и числа k - степень свободы, по таблице Стьюдента определяется вероятность различий.

При сравнении двух групп (рядов) степень свободы вычисляется по формуле

$$k = n_1 + n_2 - 2$$

n_1 - число наблюдений в контрольных группах (опытах).

n_2 - число наблюдений в других группах (опытах).



Критерий Фишера

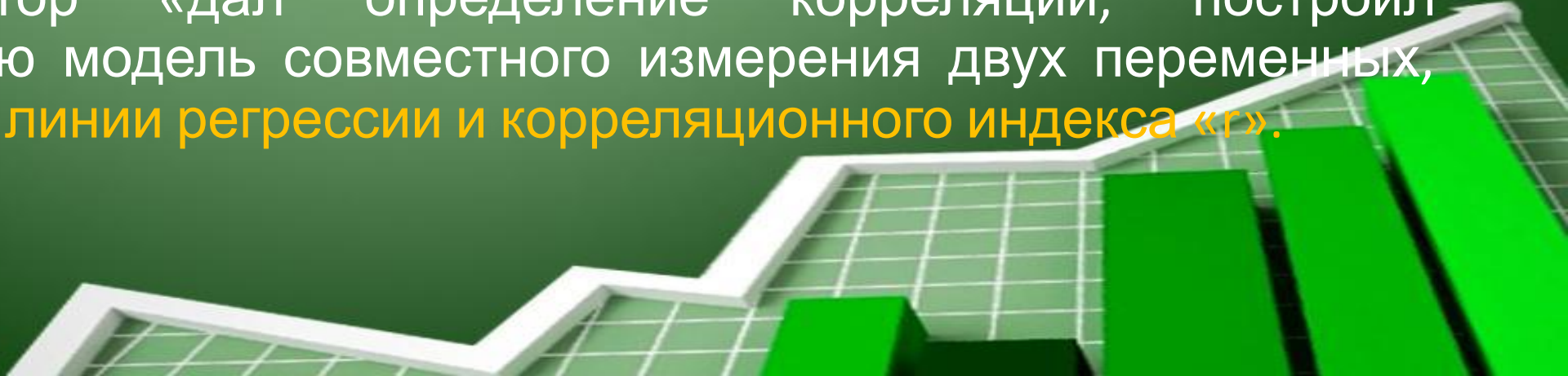
- Для проверки гипотезы о равенстве основных дисперсий сумм ($\sigma_1^2 = \sigma_2^2$) вычисление критерия Стьюдента при нормальном распределении может быть недостаточным, особенно, при малых числах измерений. Было предложено рассмотреть отношение дисперсий и обозначить буквой F:

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$



Развитие корреляционного и регрессионного анализа

Начало корреляционного и регрессионного анализа относится ко второй половине XIX века и связано с именем двоюродного брата Ч. Дарвина – Френсисом Гальтоном (1822–1911). Он ввел понятие «**закона регрессии**», связав его со средним снижением роста сыновей по сравнению с ростом отцов (1899 г). Ему же принадлежит введение числовой меры, оценивающей силу связи показателей (**корреляцию**). Поэтому началом разработки корреляционно-регрессионного анализа ученые-статистики считают статью Ф. Гальтона «Регрессия, наследственность и панмиксия» (1896 г), в которой автор «дал определение корреляции, построил теоретическую модель совместного измерения двух переменных, ввел **понятие линии регрессии и корреляционного индекса «r»**».



В развитии методов корреляционного и регрессионного анализа особо следует отметить заслуги таких ученых, как К- Пирсон, Х. Спирмэн, А. Бравэ, Г. У. Юл, А. А. Чупров, С. М. Бартлет, М. Г. Кендэл, С. Коллер, М. Езекиэл и многих других.



Виды связей

В медико - биологических исследованиях большое внимание уделяется установлению **связей между различными параметрами**, признаками или явлениями.

1. Причинно-следственная связь

Причинно-следственные отношения – это связь явлений и процессов, когда изменение одного из них – причины – ведет к изменению другого – следствия.



2. Функциональные и стохастические СВЯЗИ

Между различными явлениями и их признаками необходимо прежде всего выделить **2 типа связей**: **функциональную** (жестко детерминированную) и **статистическую** (стохастически детерминированную).

Стахостическая связь – это связь между величинами, при которой одна из них, случайная величина y , реагирует на изменение другой величины x или других величин $x_1, x_2 \dots x_n$ (случайных или неслучайных) изменением закона распределения.



Однако в медико - биологических экспериментах функциональные зависимости встречаются далеко не всегда, а чаще одному значению признака соответствует несколько значений другого признака.

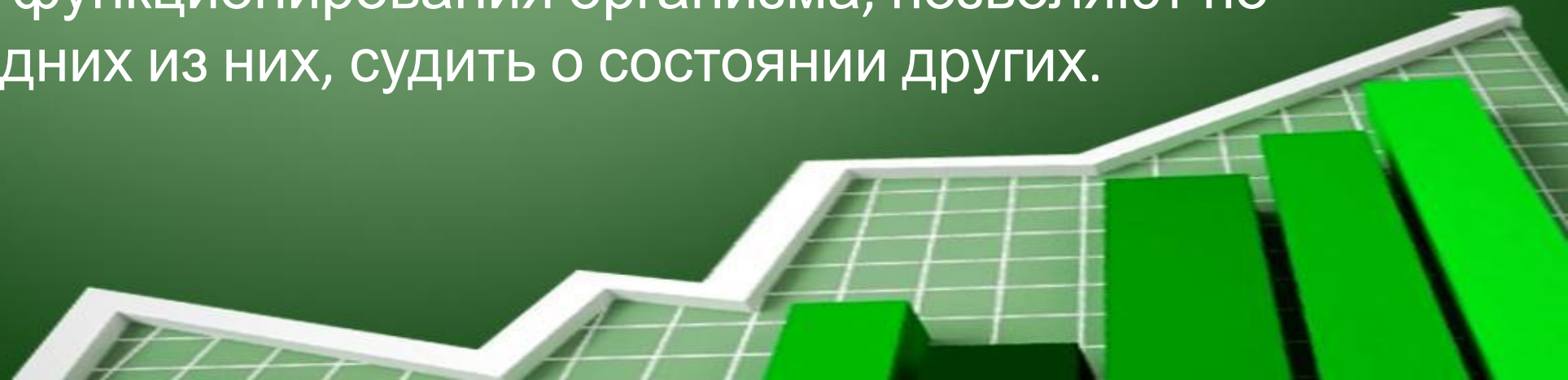
Например, при одном и том же росте, вес различных индивидуумов может быть различен.

Корреляционный анализ

Раздел математической статистики, занимающейся установлением взаимосвязей между случайными величинами, называется **корреляционно - регрессионным** анализом.

С помощью корреляционного анализа определяется **теснота и направление связи** между признаками.

Установление взаимосвязи между различными признаками и показателями функционирования организма, позволяют по изменениям одних из них, судить о состоянии других.

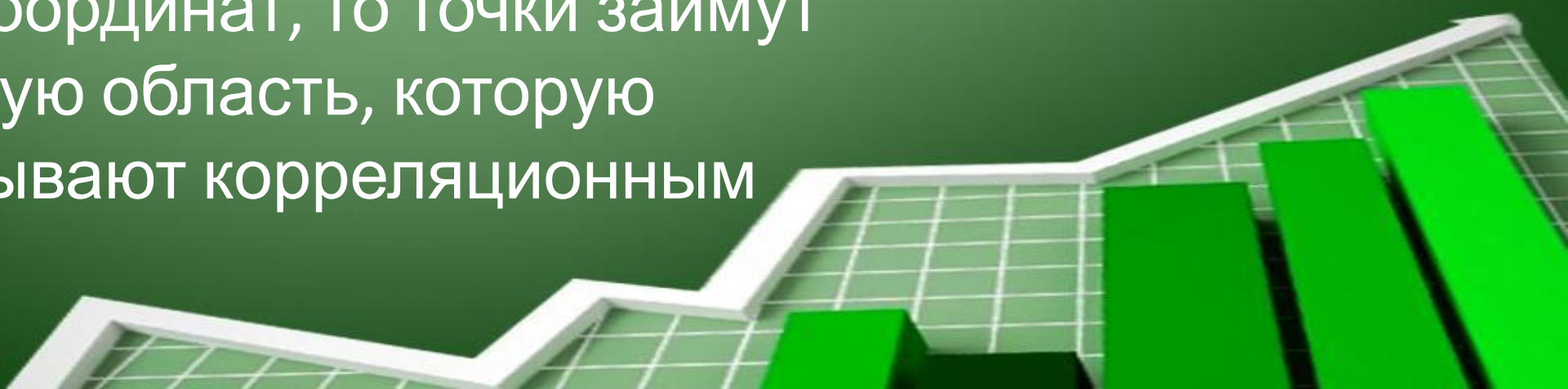
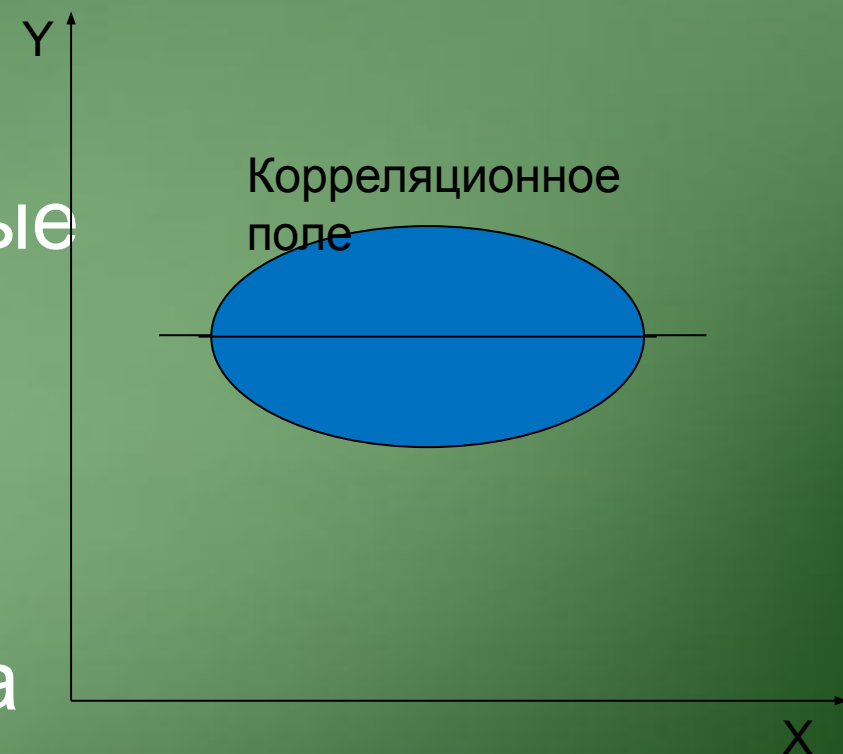


Пусть в результате медико - биологических исследований получены экспериментальные данные изменения двух показателей X и Y .

$$X: x_1; x_2; \dots; x_n$$

$$Y: y_1; y_2; \dots; y_n$$

Если полученные данные нанести на плоскость координат, то точки займут определенную область, которую обычно называют корреляционным полем.




Наиболее часто в экспериментальных исследованиях используется **линейная корреляционная зависимость**.

При этом подразумевается, что генеральные совокупности для X и Y имеют **нормальный закон распределения** или близкий к нему.

В этом случае определение точности линейной связи производится по формуле

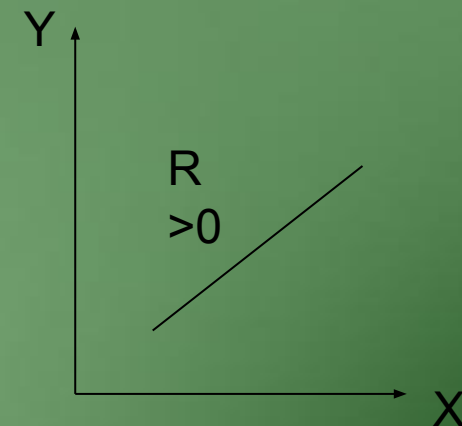
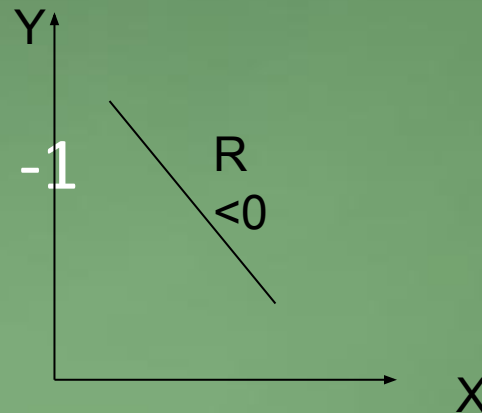
$$r = \frac{(X * Y)_{cp} - X_{cp} * Y_{cp}}{\sigma_x * \sigma_y}$$

где r - коэффициент корреляции σ_x и σ_y - средние квадратичное отклонения величин X и Y

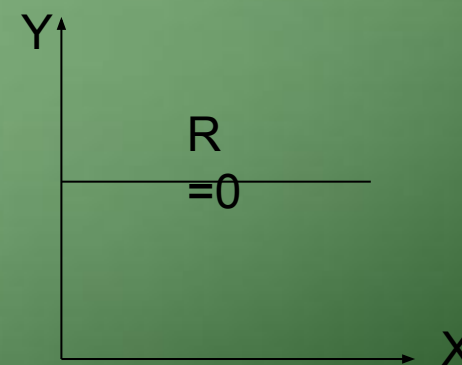


Коэффициент корреляции лежит в пределах от **-1 до +1** ($-1 \leq R_{x/y} \leq +1$):

1) Отрицательные значения R от 0 до -1 указывают на **обратную зависимость** между переменными.



2) Положительные значения R от 0 до +1 характеризуют **прямую связь** между признаками.



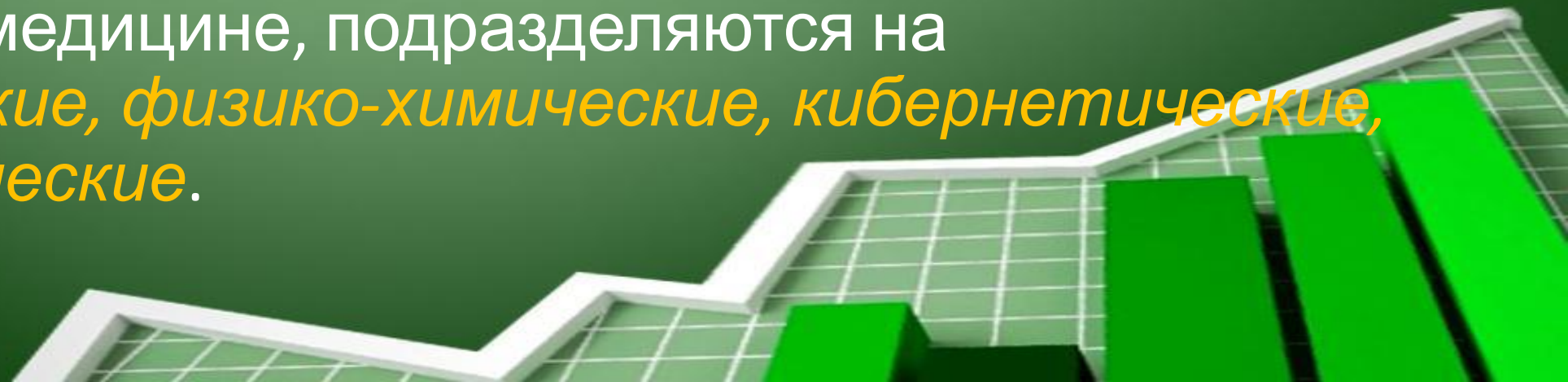
3) Когда $R=0$, это говорит об **отсутствии статистической линейной взаимосвязи** между признаками.

МОДЕЛИРОВАНИЕ

Моделирование - это метод исследования, при котором объект исследования заменяется другим, который называется моделью.

В настоящее время существуют самые разнообразные типы моделей, которые можно классифицировать различными способами.

В соответствии с этим все модели, используемые в биологии и медицине, подразделяются на *биологические, физико-химические, кибернетические, математические.*



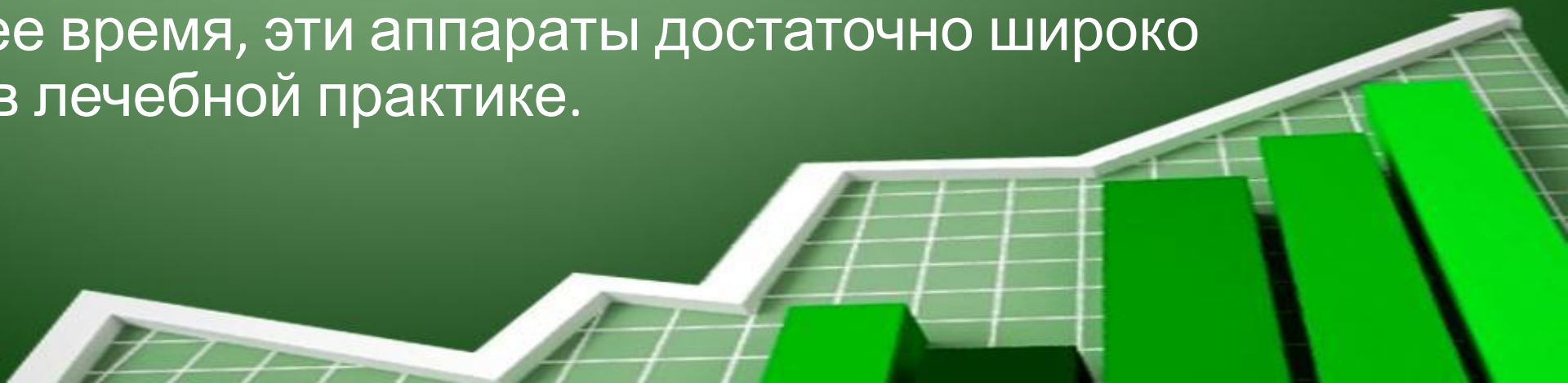
Физические модели

Физические (аналоговые) модели - это физические системы или устройства, которые путем специального конструирования имитируют реальный объект.

К физическим моделям относятся **электромеханические устройства**, заменяющие органы и системы живого организма:

- аппарат искусственного дыхания (моделируют легкое),
- аппарат искусственного кровообращения (модель сердца),
- аппарат гемодиализа (модель почки) и т.д.

В настоящее время, эти аппараты достаточно широко используется в лечебной практике.



Математические модели

Это совокупность уравнений, устанавливающих количественную зависимость между параметрами, которые описывают свойства изучаемого объекта, явления или процесса.

Особенности математических моделей:

- математическая модель позволяет судить о поведении объекта в таких условиях, которые трудно создать в эксперименте или клинике;
- математическая модель позволяет прогнозировать течение болезни и подобрать оптимальные варианты лечения;
- математическая модель биологических систем, позволяет уменьшить время исследования и сократить число животных, необходимых для исследования.

Регрессионный анализ

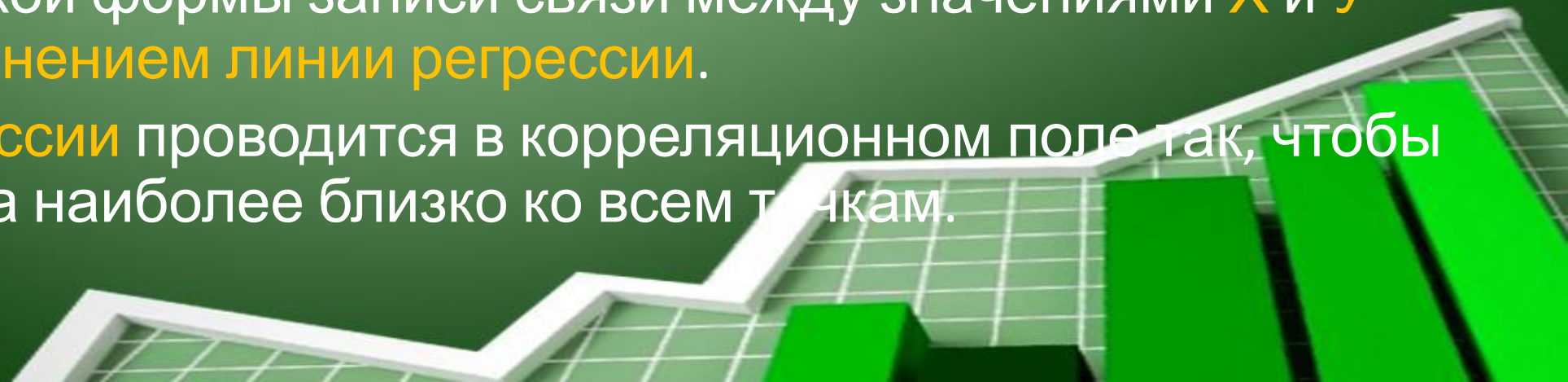
Функция, позволяющая по величине одного признака $-X$, находить средние значения другого признака $-Y_x$, называется *регрессией*.

Регрессионный анализ позволяет описать форму зависимости между случайными переменными с помощью уравнений регрессии (линейных, квадратичных, показательных и т. д.).

Регрессионный и корреляционный анализы тесно связаны между собой.

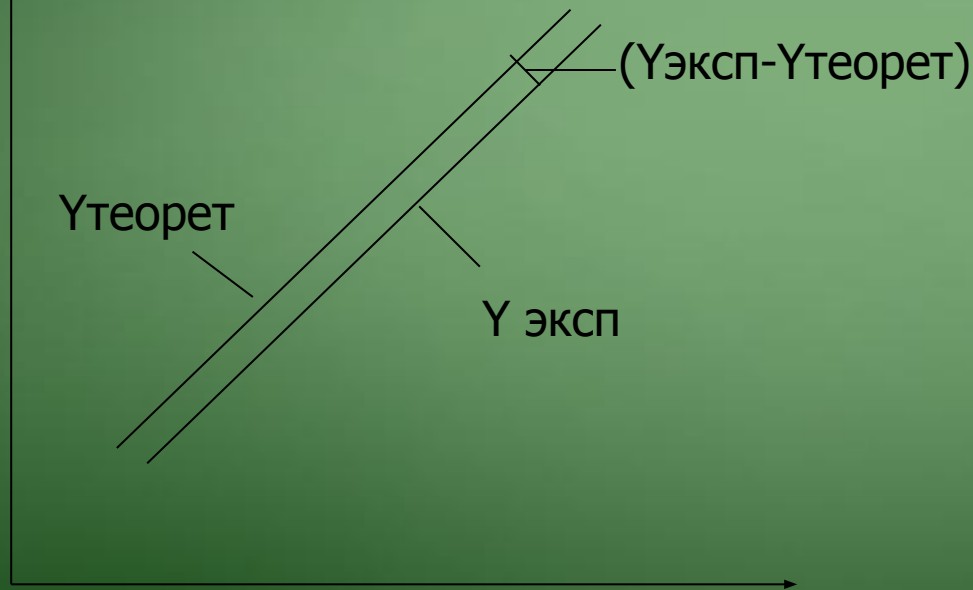
Аналитической формы записи связи между значениями X и Y является *уравнением линии регрессии*.

Линия регрессии проводится в корреляционном поле так, чтобы она проходила наиболее близко ко всем точкам.



В качестве критерия близости обычно используется - **МНК**
«метод наименьших квадратов».

МНК- минимум суммы квадратов отклонения экспериментальных точек от линии регрессии. $\sum (Y_{\text{экс}} - Y_{\text{теор}})^2 \Rightarrow \min (i=1, n)$.



Рассмотрим линейную регрессию, которая выражается простым уравнением линейной за $Y_x = a * X + b$

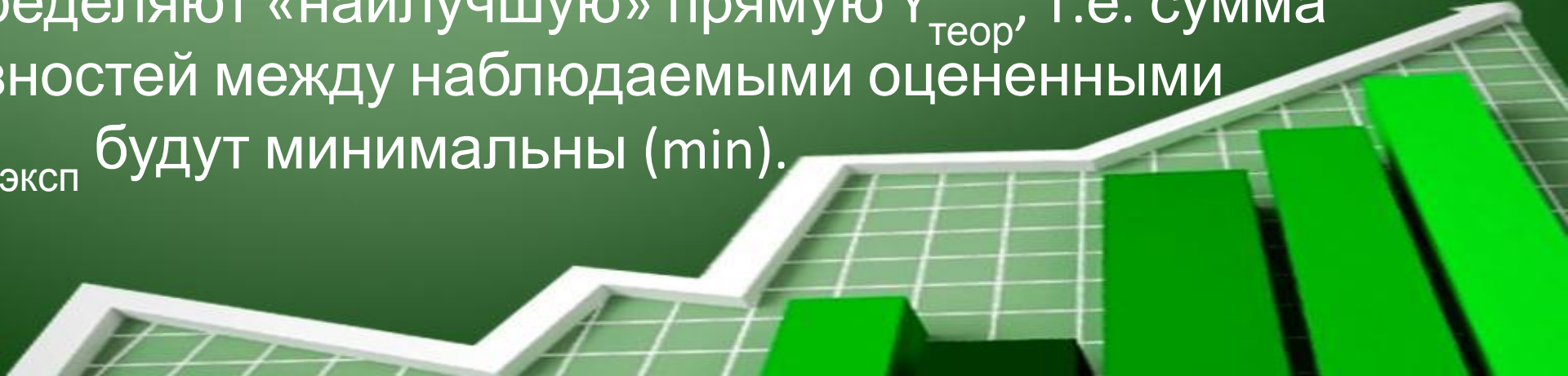
Здесь, Y_x - средняя арифметическая значение переменной Y , соответствующее заданному значению X ; a и b - параметры уравнения;

$$a = \frac{(X * Y)_{cp} - X_{cp} * Y_{cp}}{(X^2)_{cp} - (X_{cp})^2}$$

$$b = Y_{cp} - a * X_{cp}$$

b - свободный член, a - коэффициент регрессии.

Вычисленные по этим формулам параметры уравнения регрессии определяют «наилучшую» прямую $Y_{теор}$, т.е. сумма квадратов разностей между наблюдаемыми оцененными значениями $Y_{эксп}$ будут минимальны (min).



Прогнозирование


Прогнозирование осуществляется путем экстраполяции сформировавшегося процесса (математическая модель) к настоящему времени в будущее.

В таблице приведена зависимость атмосферного давления от высоты подъема.

Давление P (мм Hg ст.)	750	740	730	720	710	700	690
Высота H (км)	1	2	3	4	5	6	7

На основании этих данных строится математическая модель в виде уравнении регрессии $y=ax+b$, которая полностью описывает эту таблицу. На основе полученной математической модели определяется атмосферное давление $Y_{\text{прогн}}$ на высоте $X_{\text{прогноз}}$.

В формулу $Y=-10x+760$ в место x ставится $X_{\text{прогноз}}$ (прогнозируемая высота) и определяем прогнозируемую величину давления $Y_{\text{прогн}}$.



Do you have
any questions?

