

# Таблица «Excel»

Применение ЭТ «Excel»  
в статистике

# Цель урока

- Показать практическое использование электронной таблицы “Excel” в различных отраслях знаний, в частности, в статистике

# Задачи урока

- Познакомить с понятием «Статистика».
- Познакомить и научить применять метод наименьших квадратов (МНК) при обработке статистических данных.
- Нарботать практические знания и умения при применении МНК для обработки статистических данных.

# Постановка задачи и гипотезы

- **Задача:** в нашем городе предполагается строительство завода гиганта ( без учета розы ветров и вредных выбросов).
- **Независимые эксперты (это вы) должны провести анализ предположительных последствий данного строительства и влияние увеличивающихся выбросов вредных примесей в атмосферу на резкое ухудшение экологии и здоровья жителей города и принять решение «Да/нет».**
- **Гипотеза. Влияние выбросов на жителей города реально существует. Докажем, что особый вред наносит здоровью жителей угарный газ, который влияет на рост такого заболевания как бронхиальная астма...**

# Введение

## ■ Задачи экспертов

- Установить при какой концентрации окиси углерода резко возрастает число больных бронхиальной астмой.
- Используя экспериментальные данные и методы науки статистики доказать данную зависимость.
- Найти функцию, которая бы максимально близко описала зависимость количества болеющих бронхиальной астмой ( $P$ ) от содержания в воздухе окиси углерода ( $C$ ).
- Дать рекомендации о необходимости строительства данного завода.

## ■ Исходная информация

- Для доказательства закономерности зависимости роста заболеваний от концентрации окиси углерода в воздухе необходимо собрать данные в различных городах о концентрации в воздухе окиси углерода и о количестве больных бронхиальной астмой.
- Какая зависимость существует в данном случае пока неизвестно, но она явно прослеживается. И вот здесь нам и поможет такая наука как статистика

# Статистика

- Для подтверждения или опровержения гипотезы воспользуемся статистическими данными.

## Для экспертов:

- Статистика - это наука о сборе, измерении и анализе массовых количественных данных.
- Используем данные медицинской статистики...

## Дополнительная информация

- Различают медицинскую статистику, экономическую статистику, социальную статистику и другие.
- Математический аппарат статистики разрабатывается специальным разделом статистики «Математической статистикой»

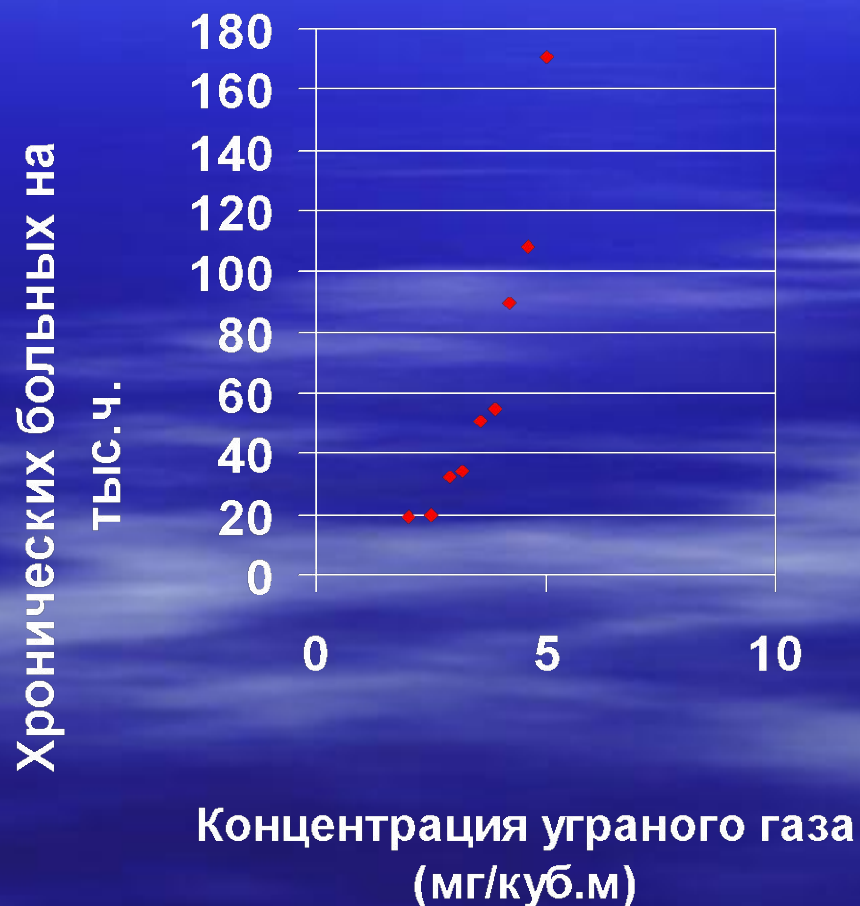
# Статистические данные

(исходная информация медицинской статистики)

Зависимость в таблице

Концентрация С, мг/куб.м	Количество больных Р, бол./тыс.
2	19
2,5	20
2,9	32
3,2	34
3,6	51
3,9	55
4,2	90
4,6	108
5	171

Построен точечный график данной  
зависимости

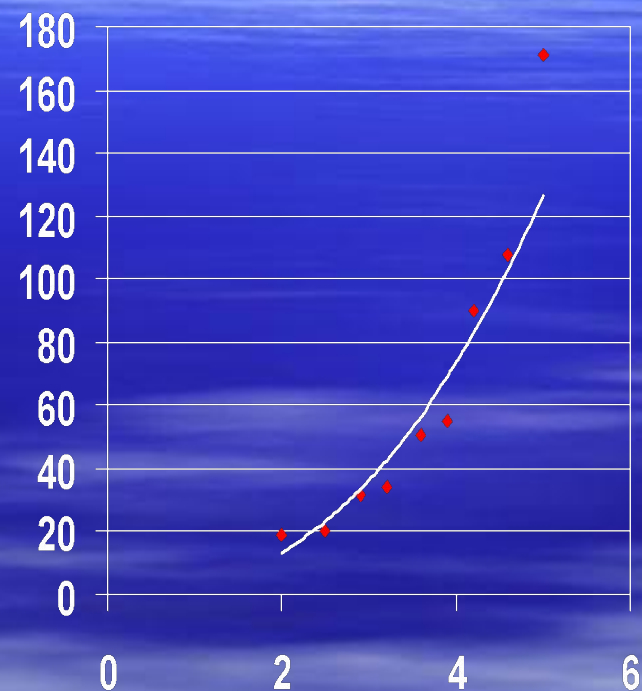


# Математическая модель

- Теперь начнем строить математическую модель, т.е. найдем зависимость  $P$  от  $C$  или  $P(C)$ , для этого:
- Постройте график, наиболее близко проходящий к точкам (данным). Это можно сделать вручную, но тогда мы не найдем формулу зависимости.
- Хотя вид функции неизвестен, но поиск её происходит методом подбора в таблице Excel, используя метод наименьших квадратов.

## Внимание !!

- При этом соблюдаем следующие требования к функции:
  1. Простота функции.
  2. График должен проходить вблизи экспериментальных точек, отклонения должны быть минимальны и равномерны.



Построение графической зависимости.

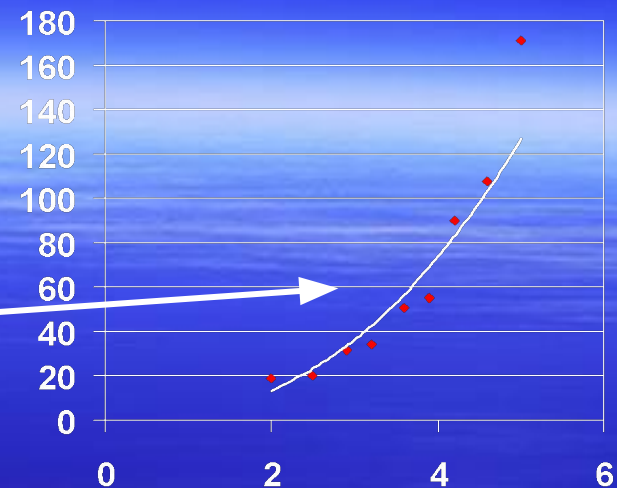
Рис.1.



- Функцию, полученную на рис.1 называют **регрессионной моделью** и данная модель подтверждает нашу гипотезу

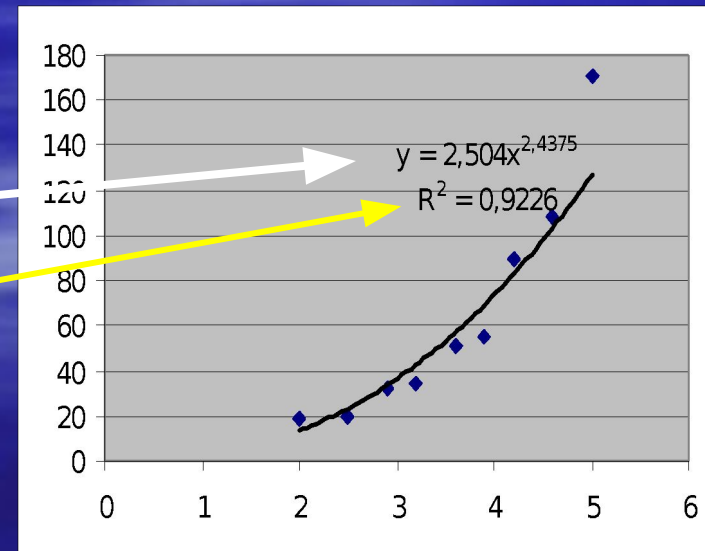
- Регрессионная модель-это функция, описывающая зависимость между количественными характеристиками сложных систем.

- При этом достаточно задать описание функции и коэффициент детерминированности ( $R^2$ )



Построение графической зависимости.

Рис.1.



## ■ Выводы:

- Хотя статистические данные носят приблизительный, усредненный характер, и получаются путем многократных измерений, в данном случае видна тенденция роста зависимости количества больных бронхиальной астмой от концентрации окиси углерода.

- Регрессионная модель –  
это функция,  
описывающая  
зависимость  
между  
количественными  
характеристиками  
сложных систем.

# Заключение

(продолжите предложения сами)

- Выдвинутая гипотеза .....
- ..... функция, описывающая зависимость количества больных  $P$  от концентрации  $C$ .
- Строительство завода окажет ..... воздействие на здоровье населения и строительство завода на территории города.....

# Метод Наименьших Квадратов (метод построения регрессивных моделей)

- Построение регрессионной модели происходит в два этапа:
  1. Подбор вида функции.
  2. Вычисление параметров функции.

Первый этап не имеет конкретного, строгого решения, а зависит от опыта и интуиции. Чаще всего выбирают среди следующих функций:

$y=ax + b$ -линейной функции;

$y=ax^2+bx + c$ - квадратичной функции;

$y=a \ln(x) + b$  – логарифмическая функция;

$y=ae^{bx}$  – экспоненциальная функция;

$y=ax^b$  – степенная функция.



# Метод Наименьших Квадратов (МНК)

- Теперь главным будет найти параметры функций так, чтобы она располагалась как можно ближе к экспериментальным данным.
- Решение данной проблемы было предложено К.Гауссом в XVIII в. Это решение и стало называться методом наименьших квадратов (МНК).

# МНК

- МНК заключается в том, что сумма квадратов отклонений  $y$ -координат всех экспериментальных точек от  $y$ -координат графика подобранной функции должен быть минимальной.
- Теперь вы знаете о данном методе, который широко используется в статистической обработке данных.

# Важно!!!

- Используя МНК, по данному набору экспериментальных точек можно построить любую функцию. А будет ли она нас удовлетворять?
- Давайте рассмотрим три функции построенные в “Excel” с использованием метода наименьших квадратов по ранее использованной таблице.
- Запишите функции в тетрадь!

	A	B
1	2	19
2	2,5	20
3	2,9	32
4	3,2	34
5	3,6	51
6	3,9	55
7	4,2	90
8	4,6	108
9	5	171

Построим таблицу в Excel и по ней построим точечную диаграмму



Линия тренда

Тип Параметры

Построение линии тренда (аппроксимация)

Линейная
  Логарифмическая

Степенная
  Экспоненциальная

Укажем параметры

Добавим к графику линию тренда

Мастер диаграмм (шаг 2 из 4): источник данных диаграммы

Диапазон данных: Ряд

Диапазон: =Лист1!\$A\$1:\$B\$9

Ряды в:  строках  столбцах

Нажмем ОК

Тип Параметры

Название аппроксимирующей (сглаженной) кривой

автоматическое: Линейный (Ряд1)
  другое:

Прогноз

вперед на: 0 единиц

назад на: 0 единиц

пересечение кривой с осью Y в точке: 0

показывать уравнение на диаграмме

поместить на диаграмму величину достоверности аппроксимации (R^2)



Посмотрите на линейный график: точки достаточно далеки от основного графика. Коэффициент детерминированности всего в пределах 0,8...

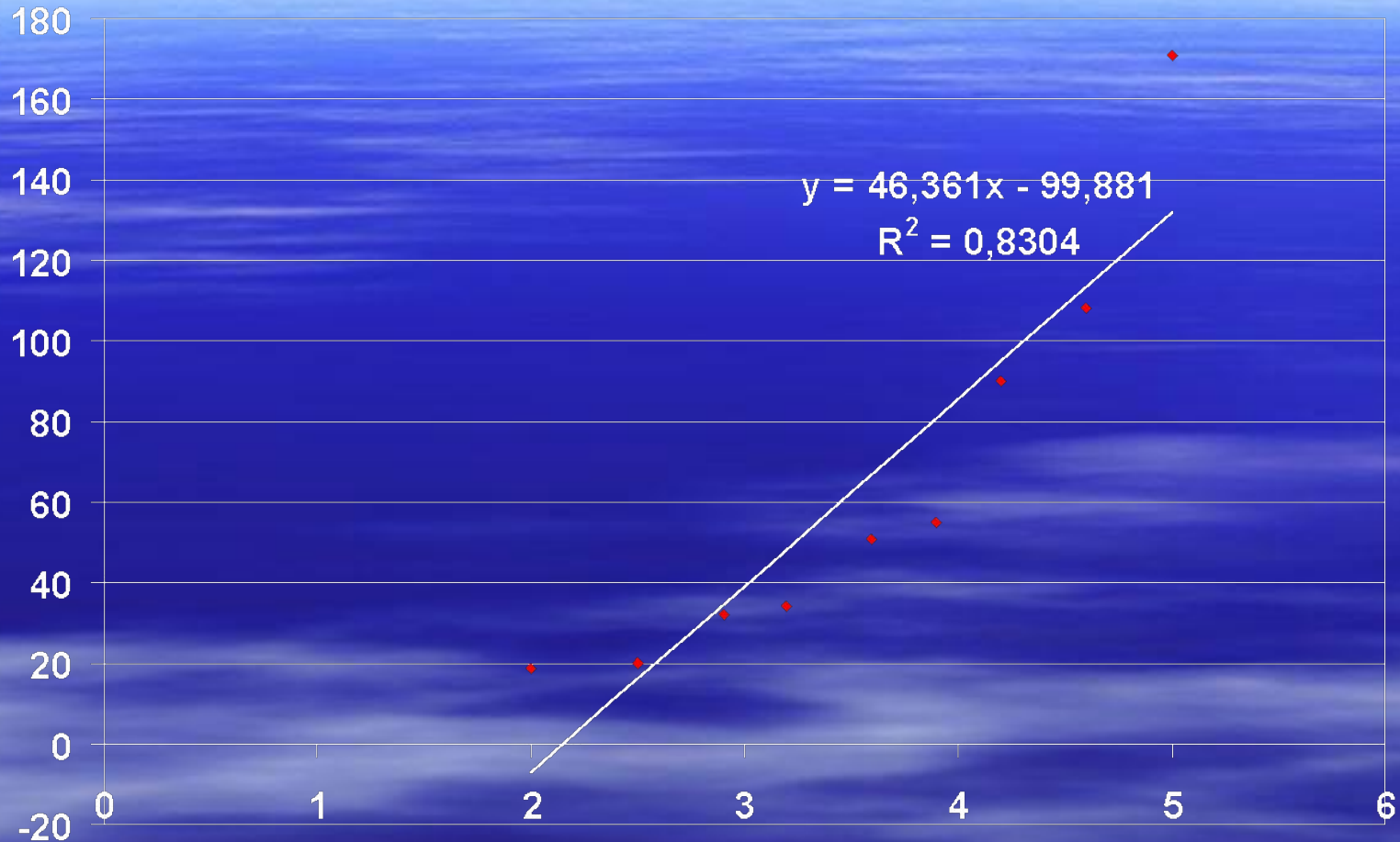


Рис.2.а

Посмотрите на график квадратичной функции.

Отметьте для себя, что коэффициент детерминации здесь лучше

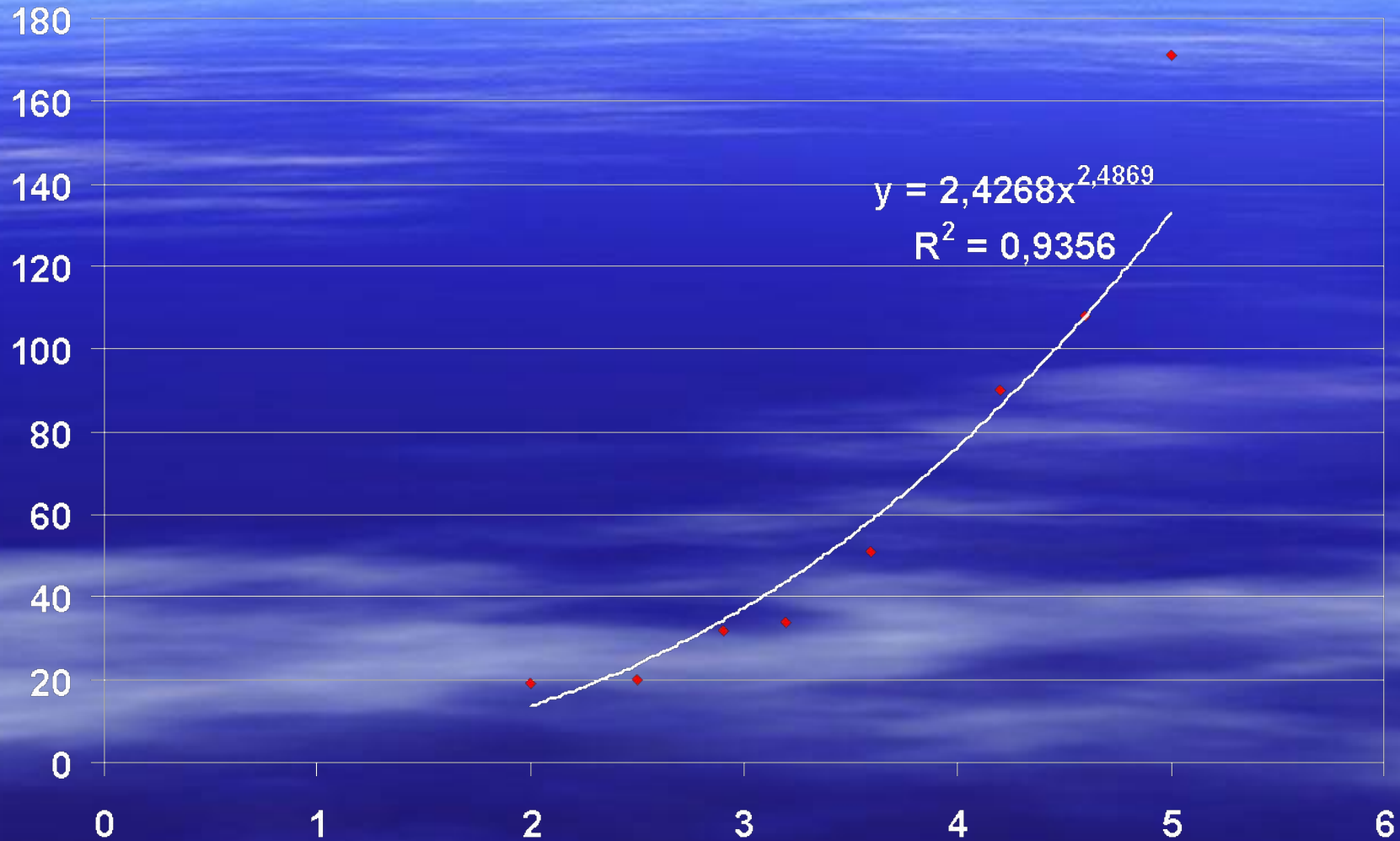


Рис.26

- Два типа графиков получены.
- График регрессионной модели называется **трендом**.
- На графиках подписаны функции графиков, обратите внимание как они выглядят и запишите их в тетрадь.
- Кроме регрессионных моделей на графиках записан и **коэффициент детерминированности  $R^2$** . Именно эта величина определяет точность подбора регрессионной модели, она должна быть в пределах от  **$(0;1)$  ( $0 < R^2 < 1$ )**. И чем  **$R^2$  ближе к 1, тем точнее регрессионная модель**.
- Определите самую удачную модель из предложенных выше. Запишите её в тетрадь.
- Повторите расчеты сами, подберите еще одну функцию («Практическую работу № 2»).

Рассчитайте сами графики  
по заданной проблеме.

Найдите лучший, по нему спрогнозируйте  
количество больных в нашем городе,  
если выделение угарного газа станет  
равным 6,2 мг/куб.м.

Завершите работу экспертов и  
заполните слайд № 11 в режиме  
редактирования,  
установите этот слайд в полноэкранном  
режиме  
подпишитесь.

***Желаю удачи!***