

Статистические данные

Метод наименьших

квадратов



11 класс

Статистика — наука о сборе, измерении и анализе массовых количественных данных.

Статистические данные носят приближённый, усреднённый характер, получаются путём многократных измерений.

Виды статистики:

- **Социальная**
- **Экономическая**
- **Медицинская**
и другие



**Рассмотрим пример из
медицинской статистики:**

**Определить зависимость
бронхиально-лёгочных
заболеваний от содержания
угарного газа (оксида
углерода) в воздухе.**

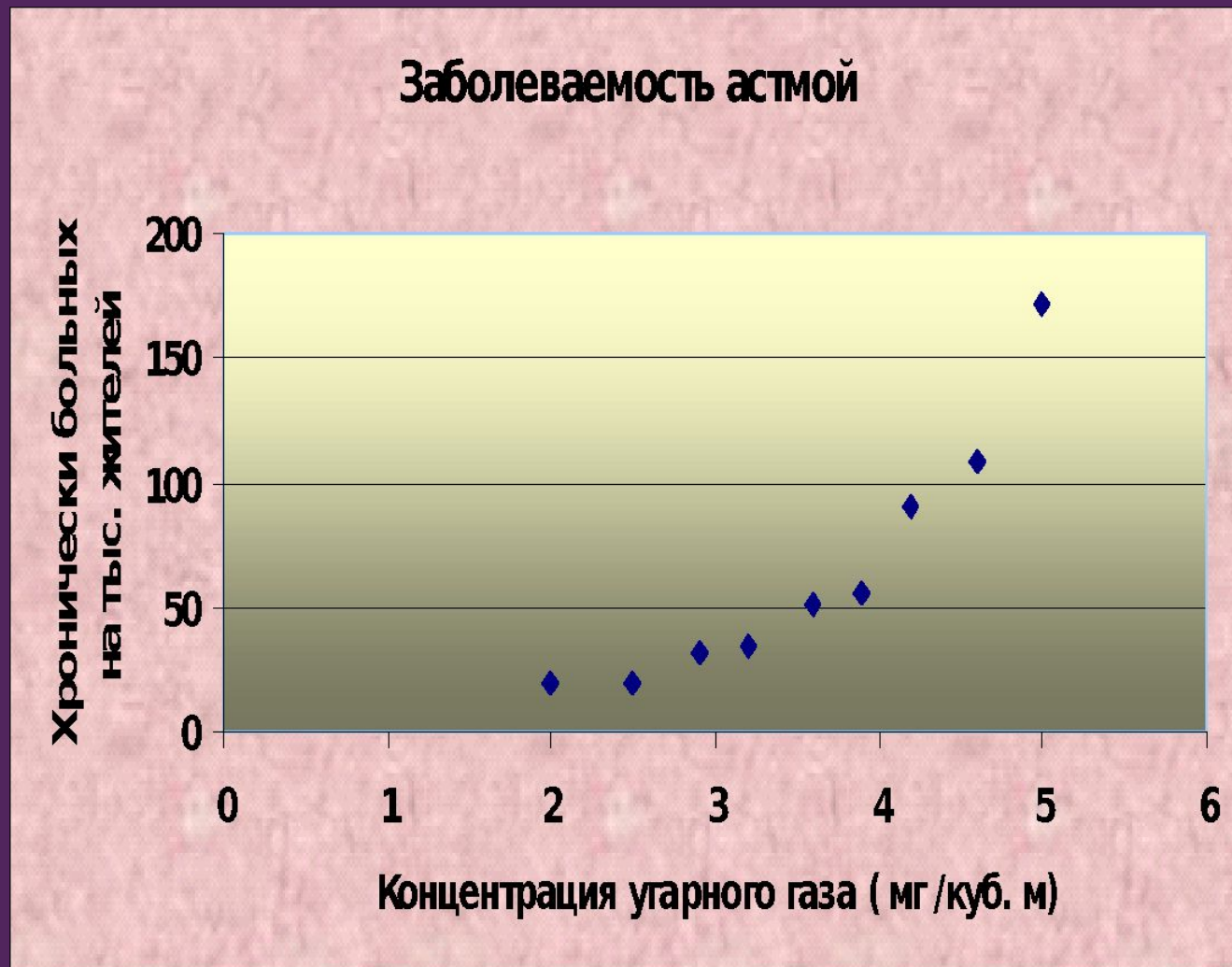


Специалисты по медицинской статистике проводят сбор данных.

Они собирают сведения из разных городов о средней концентрации угарного газа в атмосфере и о заболеваемости астмой — число хронически больных на 1000 жителей.

Полученные данные можно свести в таблицу , а также представить в виде точечной диаграммы:

$C,$ мг / куб. м	$P,$ 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



Как построить математическую модель полученных данных?

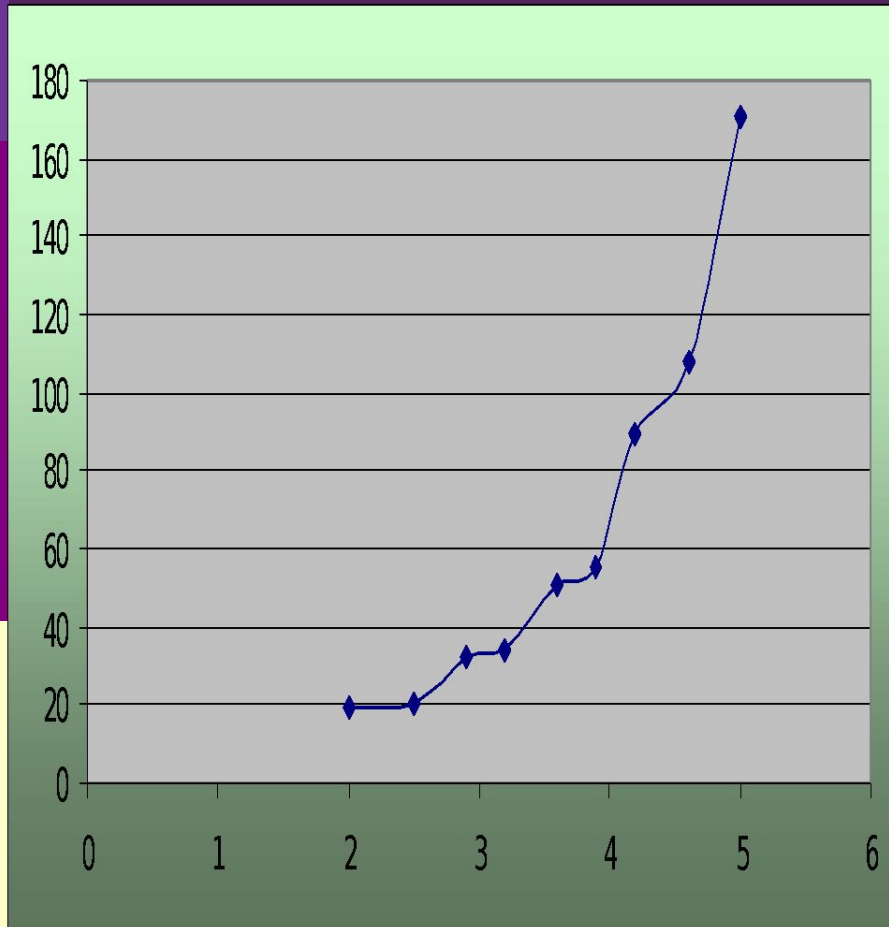
Нужно получить формулу, отражающую зависимость числа хронически больных P от концентрации угарного газа C .

То есть — функцию зависимости P от C : $P(C)$.

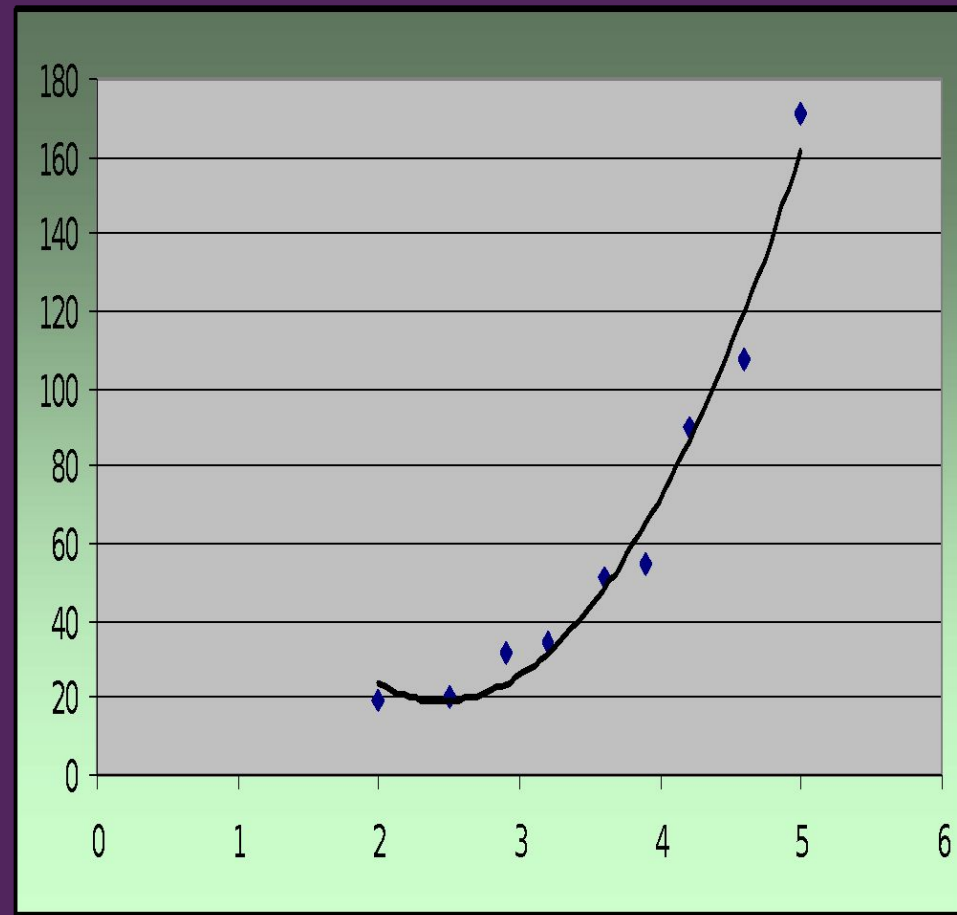
Основные требования к искомой функции:

- ▣ **Она должна быть достаточно простой для использования её в дальнейших вычислениях;**
- ▣ **График этой функции должен проходить вблизи экспериментальных точек так, чтобы отклонения от этих точек были минимальны и равномерны.**

Два варианта построения графической зависимости по данным эксперимента



а) не имеет смысла



б) Регрессионная модель

Регрессионная модель —

это функция, описывающая зависимость между количественными характеристиками сложных систем.

Вид регрессионной функции определяется путём подбора по экспериментальным данным.

Этапы получения регрессионной модели:

- 1) **подбор вида функции;**
- 2) **вычисление параметров функции.**

Наиболее часто выбор производится среди функций:

$y = ax + b$ — линейная функция;

$y = ax^2 + bx + c$ — квадратичная
функция (или полином второй степени);

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

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

$y = ax^b$ — степенная функция.

Где x — аргумент, y — значение функции,
 a, b, c — параметры функции.

Как подобрать параметры функции?

Метод наименьших квадратов (МНК) используется для вычисления параметров регрессионной модели.

Этот метод содержится в математическом арсенале электронных таблиц (в том числе и в MS Excel).

R^2 – коэффициент детерминированности

Характеристикой построенной модели является параметр R^2 – коэффициент детерминированности. Чем его значение ближе к 1, тем модель лучше.

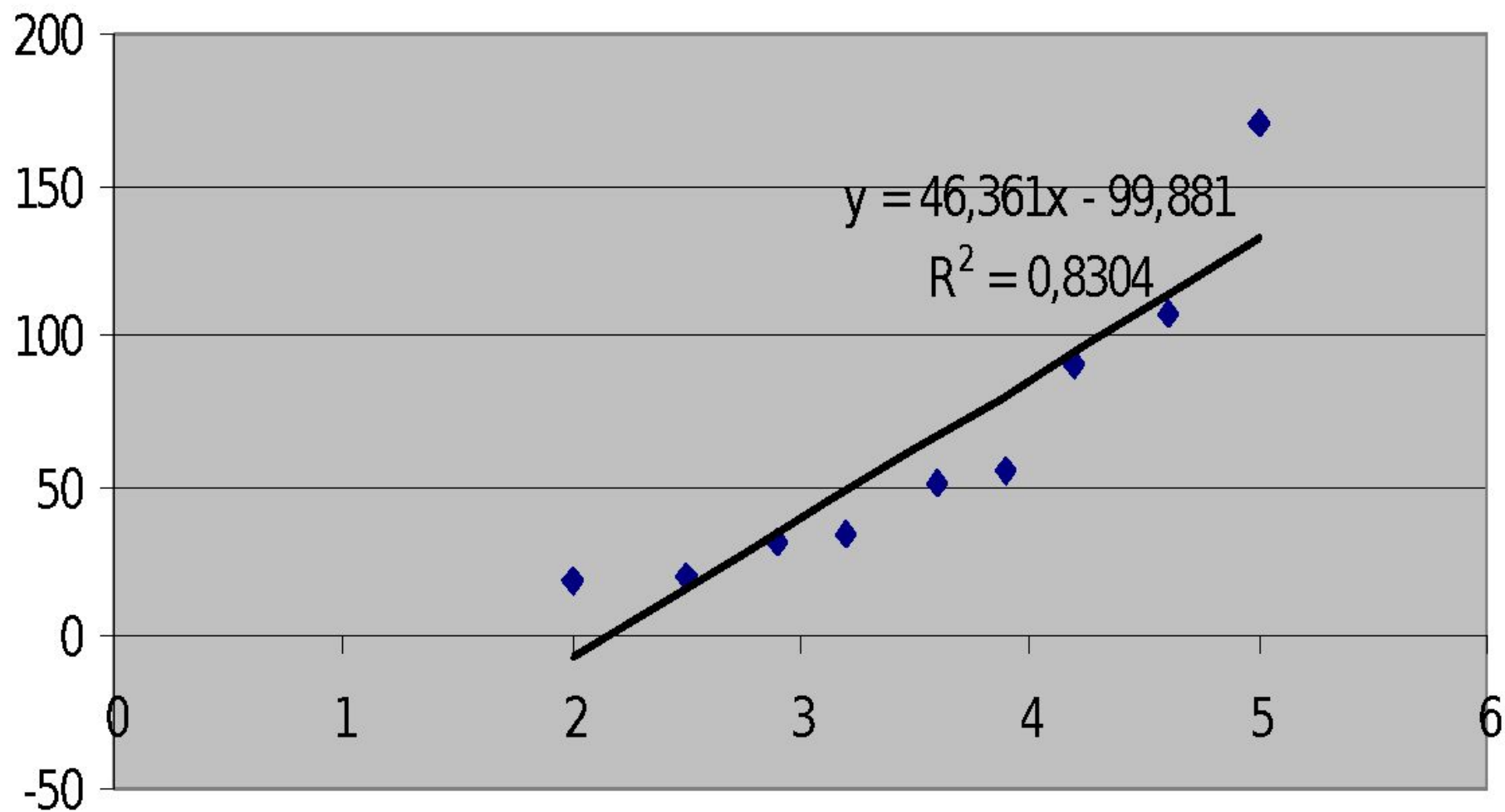
Если несколько моделей имеют близкий параметр R^2 , то пользователь выбирает из них наиболее подходящую.

Тренд:

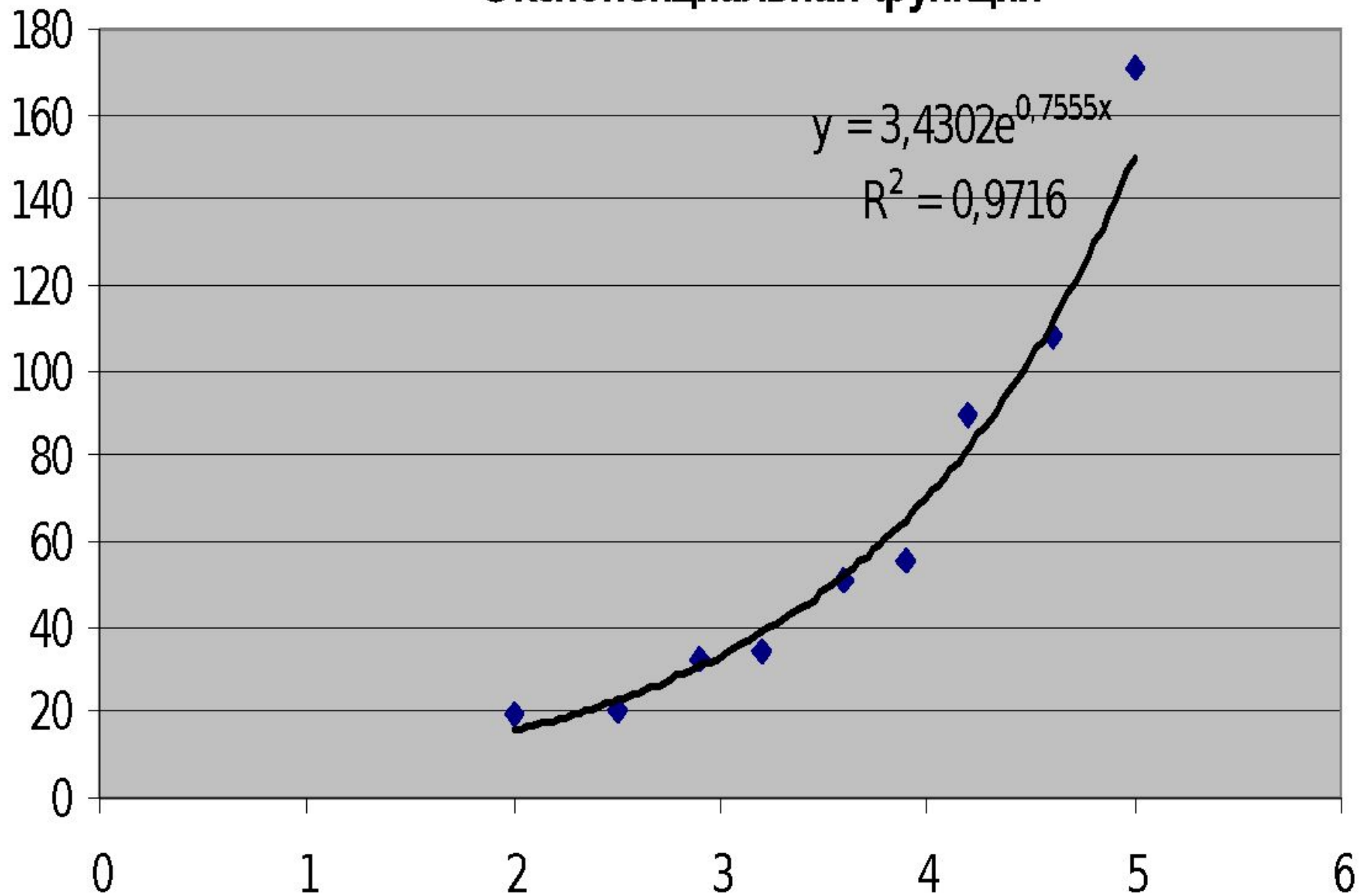
это график регрессионной модели.

Trend (англ.) – общее направление, тенденция.

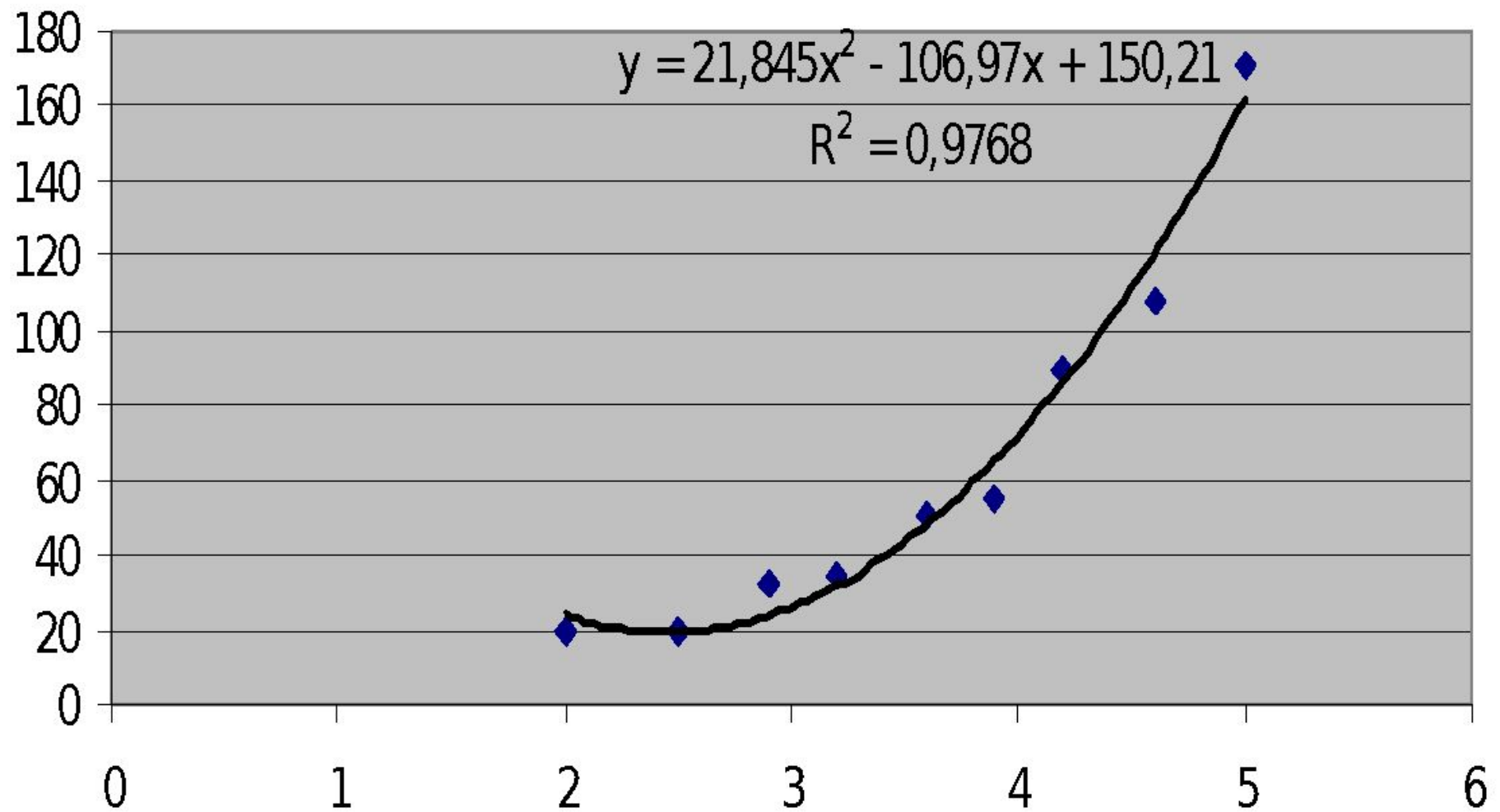
Линейная функция



Экспоненциальная функция



Квадратичная функция



Алгоритм построения регрессионной модели по МНК с построением тренда.

1. Ввести табличные данные и построить точечную диаграмму в MS Excel;
2. Щёлкнуть мышью по полю диаграммы;
3. **Макет → Линия тренда;**
4. выбрать тип функции;
5. **Дополнительные параметры линии тренда** → установить галочки на флажках «показывать уравнения на диаграмме» и «поместить на диаграмму величину достоверности R^2 », **ОК.**