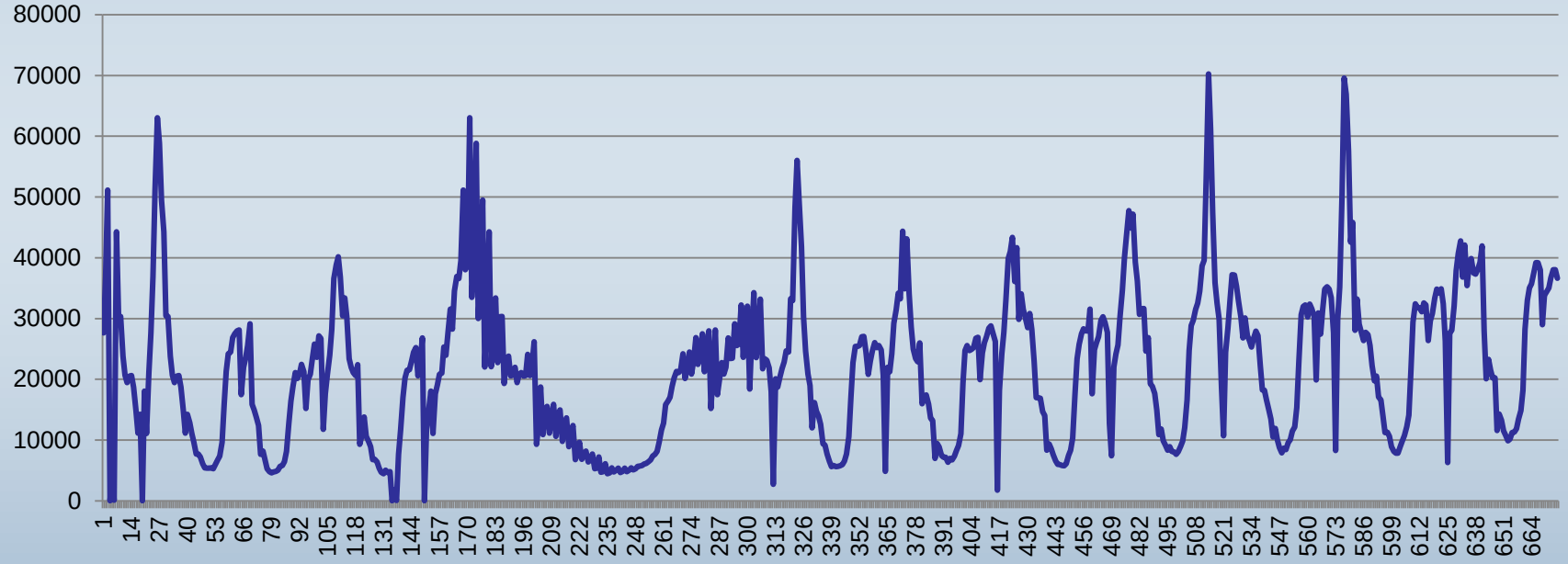
# Методы и модели анализа прогнозирования.




Кольцов С.Н

# Основные понятия и определения

Два метода прогнозирования: 1. Количественное прогнозирование.
2. Качественное прогнозирование (опрос экспертов).

**Сложности прогнозирования:**

1. **Очень сложно сделать правдивый прогноз, особенно если он касается будущего.**
2. **Предсказать не сложно, сложно предсказать правильно☺**
3. **Числа, если их правильно преподнести, могут сказать о чем угодно.**

Количественное прогнозирование: 1. Каузальные модели (причинно – следственные).
2. Временные модели (например, модель Брауна).

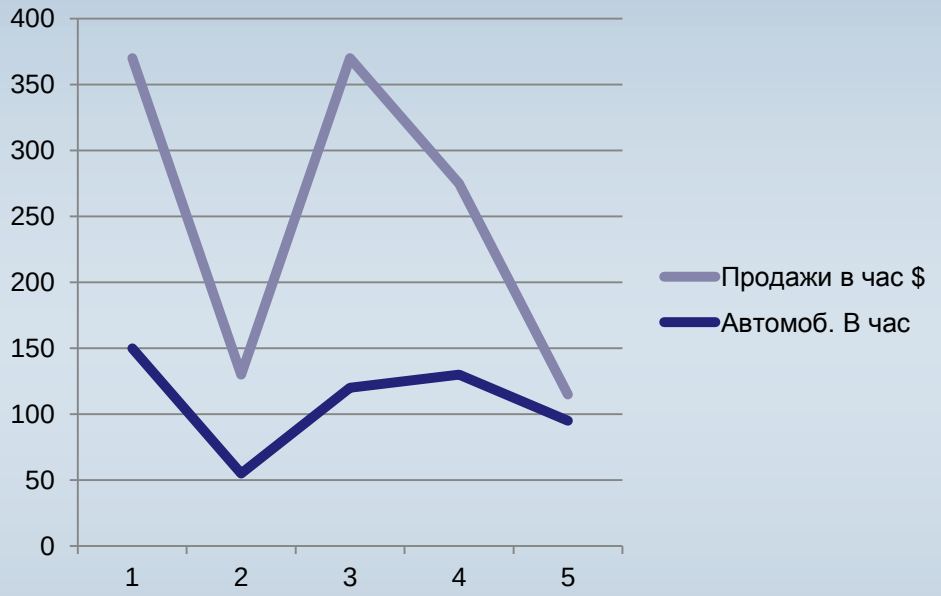
**Математическая постановка каузальной задачи** $y = f(x1, x2, x3...)$

Например: **y** – спрос на детскую еду, **x** – число детей, 1,2,3. – нумерация месяцев (лет)

или: **y** – спрос на медицинское оборудование, **x** – число больниц или поликлиник

# Причинно – следственная модель

Причинно - следственные модели используются в том случае, когда независимые переменные (**x**) известны заранее или их спрогнозировать проще чем зависимую переменную **y**. Например, можно спрогнозировать объем продаж на следующий месяц как функцию от производства за прошлый месяц.
Таким образом для выбора П-С модели нужно что бы выполнялись два условия.




1. Должна существовать связь (за прошлый месяц) между независимой переменной и зависимой переменной.

2. Значения независимой переменной (**x**) должны быть известны на период, на который мы собираемся делать прогноз.

Формирование математической связи между зависимой и независимой переменными основано на применение метода наименьших квадратов.

# Оценка прогнозирования

Для оценки точности модели можно относительную ошибку аппроксимации и средний квадрат ошибки.

Относительная ошибка

$$error = y - \hat{y},$$

*где*

$y$ – *реальноез значение*

$\hat{y}$ – *апроксим значение*

Средний квадрат ошибки

$$error = \frac{(y - \hat{y})^2}{y^2},$$

*где*

$y$ – *реальное значение*

$\hat{y}$ – *апроксим значение*

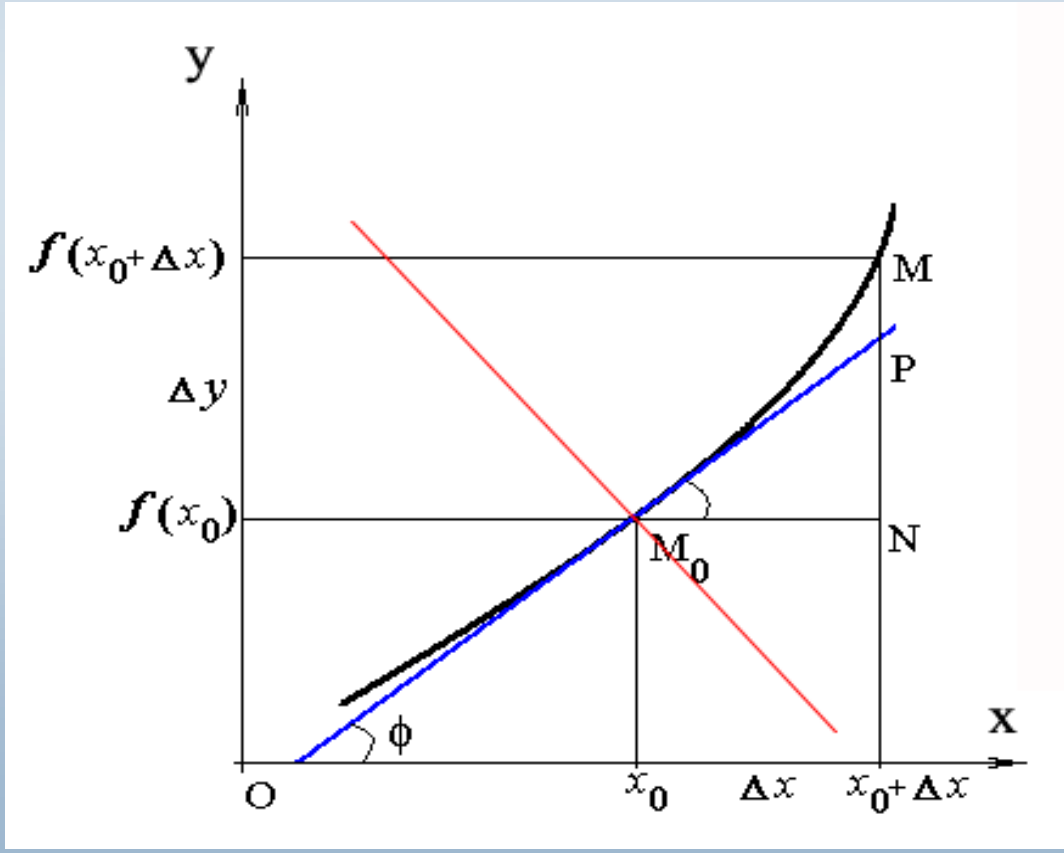
Считается, что точность модели хорошая, если среднее значение относительной погрешности не превышает 5% , удовлетворительная, если среднее значение относительной погрешности не превышает 15%, и неудовлетворительная, если среднее значение относительной погрешности больше 15%.

# Определение и геометрическая интерпретация производной

Пусть *x* - *независимая переменная, y = f(x)*

Значение производной $f'(x_0)$ равняется угловому коэффициенту касательной к графику функции $y = f(x)$ в точке $M_0(x_0, f(x_0))$; $f'(x_0) = \operatorname{tg} \Phi$, где Φ - угол наклона касательной к оси 0X

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$



# Таблица производных

| | |
|---|---|
| $C' = 0$ | $(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$ |
| $x' = 1$ | $(\arccos x)' = -\frac{1}{\sqrt{1-x^2}}$ |
| $(x^n)' = n \cdot x^{n-1}$ | $(\operatorname{arctg} x)' = \frac{1}{1+x^2}$ |
| $(\sqrt{x})' = \frac{1}{2\sqrt{x}}$ | $(\operatorname{arcctg} x)' = -\frac{1}{1+x^2}$ |
| $(e^x)' = e^x$ | $(\operatorname{sh} x)' = \operatorname{ch} x$ |
| $(a^x)' = a^x \ln a$ | $(\operatorname{ch} x)' = \operatorname{sh} x$ |
| $(\ln x)' = \frac{1}{x}$ | $(\operatorname{th} x)' = \frac{1}{\operatorname{ch}^2 x}$ |
| $(\log_a x)' = \frac{1}{x \ln a}$ | $(\operatorname{cth} x)' = -\frac{1}{\operatorname{sh}^2 x}$ |
| $(\sin x)' = \cos x$ | $(\operatorname{arcsh} x)' = \frac{1}{\sqrt{x^2+1}}$ |
| $(\cos x)' = -\sin x$ | $(\operatorname{arcch} x)' = \frac{1}{\sqrt{x^2-1}}$ |
| $(\operatorname{tg} x)' = \frac{1}{\cos^2 x}$ | $(\operatorname{arcth} x)' = \frac{1}{1-x^2}$ |
| $(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}$ | $(\operatorname{arccth} x)' = \frac{1}{1-x^2}$ |

| Производные простых функций ($x$ – независимая переменная) | Производные сложных функций ($u = u(x)$ – любая дифференцируемая функция) |
|---|---|
| 1. $(x^2)' = 2x$ | 1. $(u^2)' = 2u \cdot u'$ |
| 2. $(\sqrt{x})' = \frac{1}{2\sqrt{x}}$ | 2. $(\sqrt{u})' = \frac{1}{2\sqrt{u}} \cdot u'$ |
| 3. $\left(\frac{1}{x}\right)' = -\frac{1}{x^2}$ | 3. $\left(\frac{1}{u}\right)' = -\frac{1}{u^2} \cdot u'$ |
| 4. $(e^x)' = e^x$ | 4. $(e^u)' = e^u \cdot u'$ |
| 5. $(\ln x)' = \frac{1}{x}$ | 5. $(\ln u)' = \frac{1}{u} \cdot u'$ |
| 6. $(\sin x)' = \cos x$ | 6. $(\sin u)' = \cos u \cdot u'$ |
| (……) | (……) |

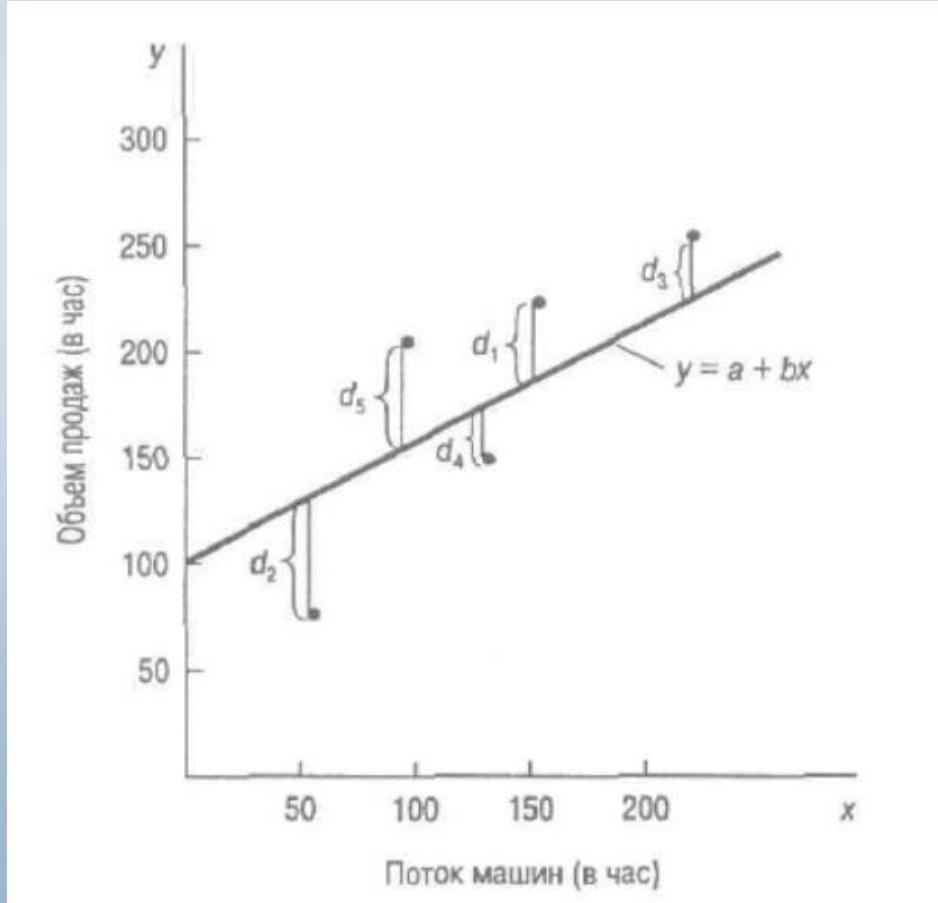
# Метод наименьших квадратов

Если некоторая величина зависит от другой величины , то эту зависимость можно исследовать, измеряя y при различных значениях x . В результате измерений получается ряд значений:

$x_1, x_2, ..., x_i, , ... , x_n;$

$y_1, y_2, ..., y_i, , ... , y_n.$

По данным такого эксперимента можно построить график зависимости $y = f(x, a, b, c....)$. Полученная кривая дает возможность судить о виде функции $f(x)$. Однако постоянные коэффициенты, которые входят в эту функцию, остаются неизвестными.



# Метод наименьших квадратов для линейной функции

Аппроксимация наших данных, например линейной функцией можно выразить следующим образом:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

**Суть метода НКв заключается в том что можно выбрать такие коэффициенты a,b, так что бы минимизировать квадрат разности между аналитической функцией и экспериментальными данными.**

Что бы минимизировать квадрат разности нужно найти частные производные по коэффициентам a, b.

$$\sum_{i=1}^{n} -2(y_i - (a + bx_i)) = 0$$

$$\sum_{i=1}^{n} -2x_i(y_i - (a + bx_i)) = 0$$

# Метод наименьших квадратов для линейной функции

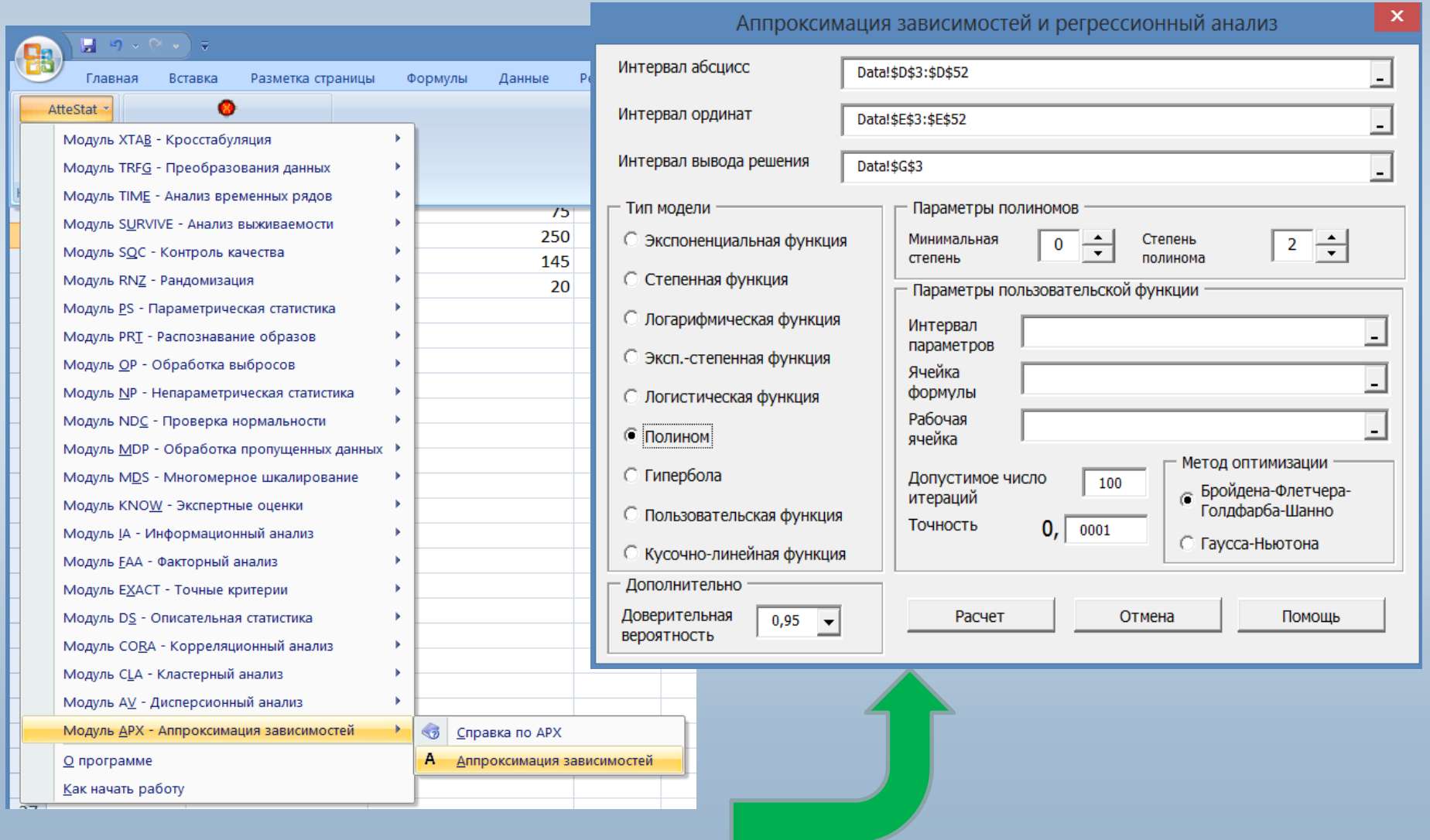
$$\sum_{i=1}^{n} -2(y_i - (a + bx_i)) = 0$$

$$\sum_{i=1}^{n} -2x_i(y_i - (a + bx_i)) = 0$$

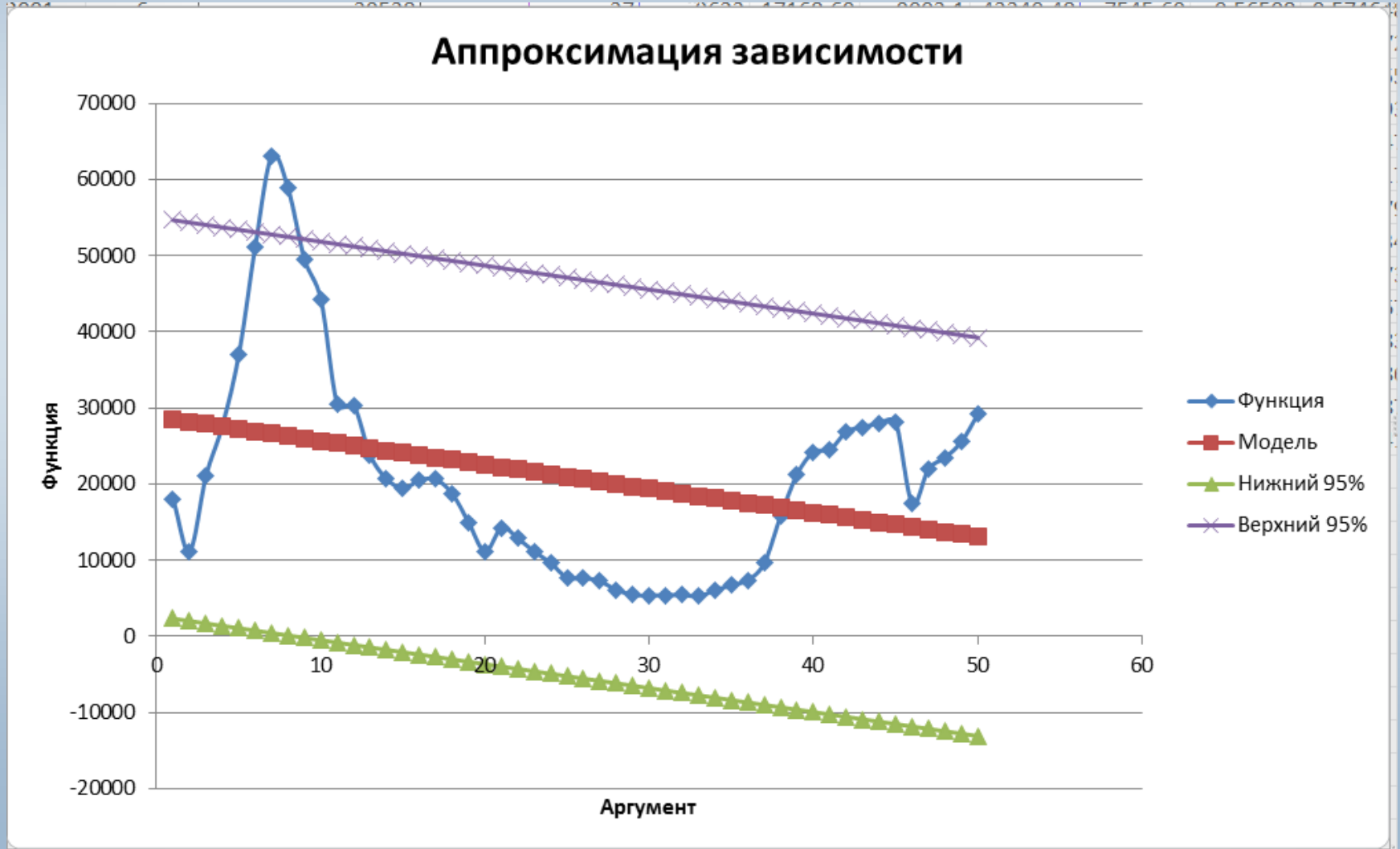
$$a = \frac{1}{n}\sum_{i=1}^{n} y_i - b\frac{1}{n}\sum_{i=1}^{n} x_i$$

$$b = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$$

# Метод наименьших квадратов в Экселе



# Метод наименьших квадратов в Экселе, линейная функция



# Метод наименьших квадратов для полимальной функции

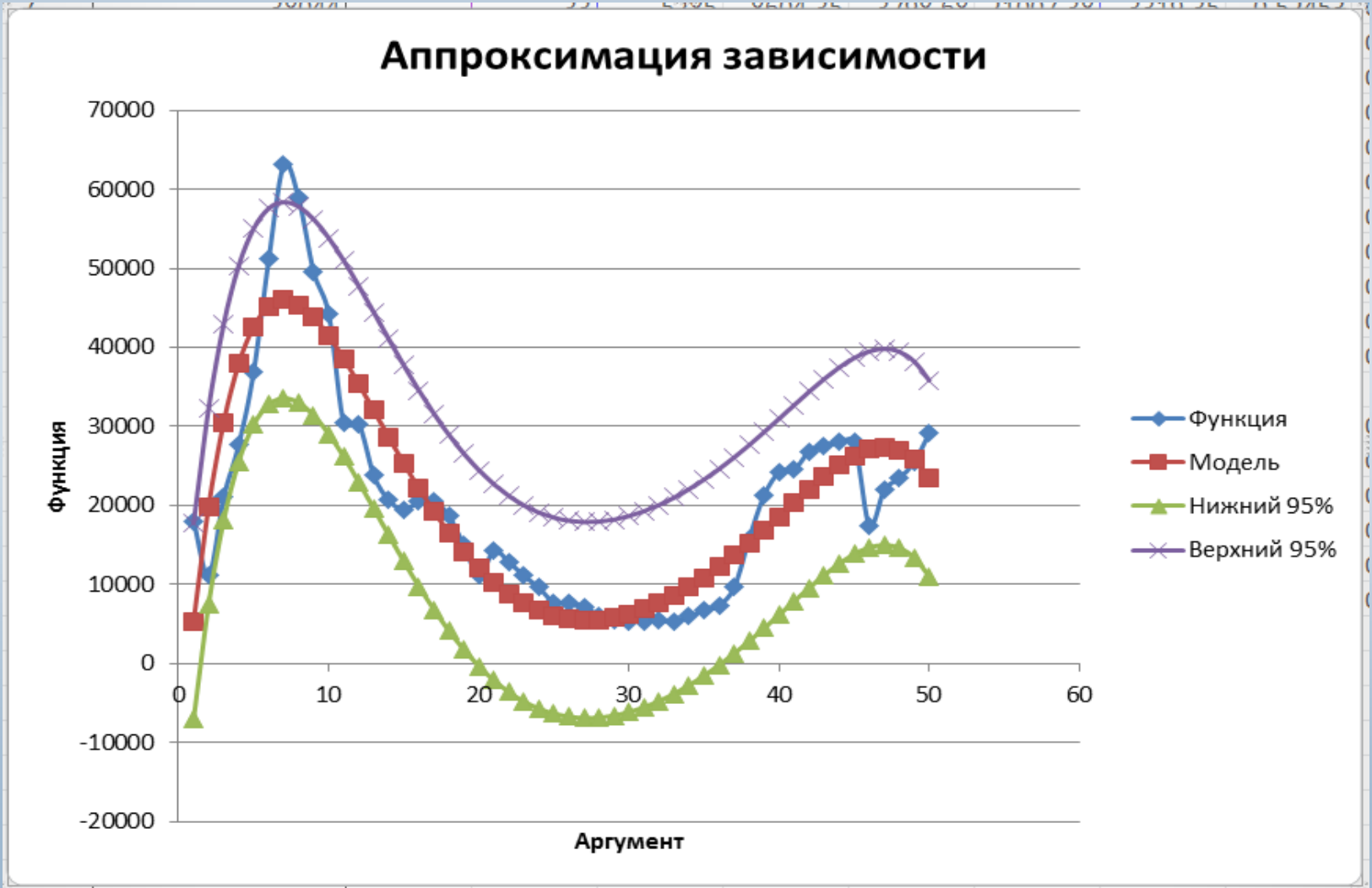
Рассмотрим аппроксимационную функцию в виде квадратичной функции, соответственно ищем квадрат разности между экспериментальными данными и нашей аналитической функцией с не известными коэффициентами.

$$\sum_{i=1}^{5}(y_i-(a_0+a_1x_i+a_2x_i^2))^2 .$$

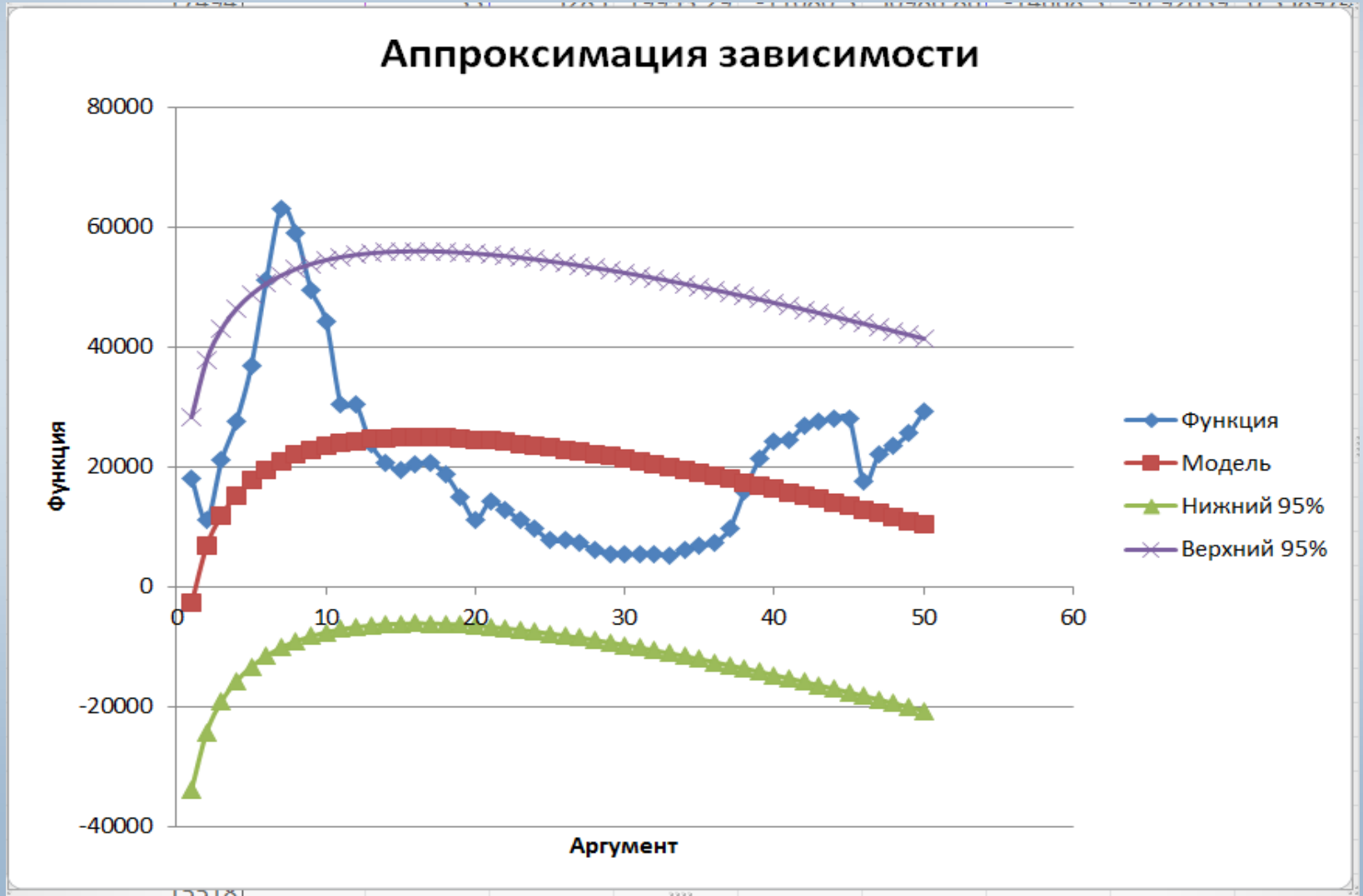
Приравняем нулю три производные по трем параметрам, соответственно получим три набора уравнений:

$$5a_0+(\sum x_i)a_1+(\sum x_i^2)a_2=\sum y_i ,$$
$$(\sum x_i)a_0+(\sum x_i^2)a_1+(\sum x_i^3)a_2=\sum x_iy_i ,$$
$$(\sum x_i^2)a_0+(\sum x_i^3)a_1+(\sum x_i^4)a_2=\sum x_i^2y_i .$$

# Метод наименьших квадратов в Экселе, полином 6 степени



# Метод наименьших квадратов для логарифмической функции.



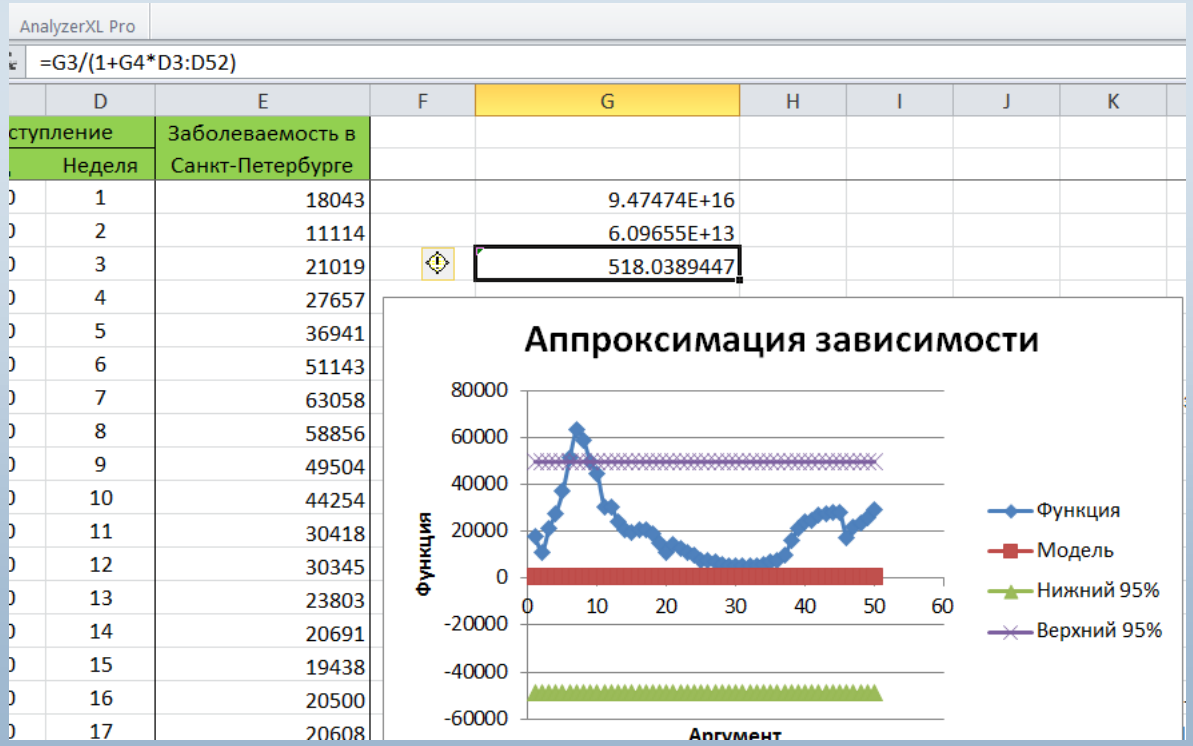
# Пример пользовательской функции

Пусть мы хотим аппроксимировать данные функцией следующего вида:

$$y(t) = \theta_1 / (1 + \theta_2 t)$$

В ходе аппроксимации методом наименьших квадратов ищутся параметры: $\theta_1, \theta_2$

Нам нужно задать начальные данные для этих параметров (в отдельных ячейках), и задать вид функции (также в отдельной ячейке), например



# Оценка точности метода наименьших квадратов

Дисперсия функции вычисляется следующим образом:

$$\sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

где y (с чертой) является средним значением, и рассчитывается следующим образом:

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

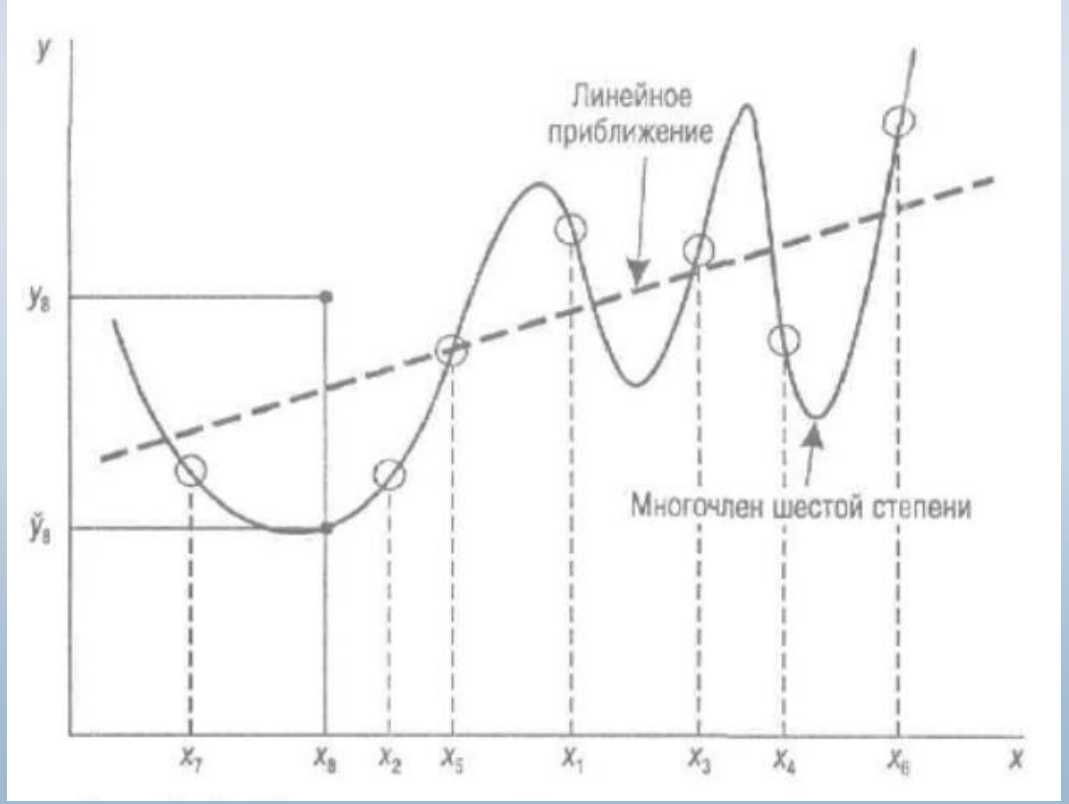
# Метод наименьших квадратов Проблема выбора степени полинома.

Собственно выбирая большую степень полинома мы будем увеличивать точность аппроксимации, то есть будет уменьшаться сумма квадратов.

Однако это приводит к тому что между точками будет возникать большие флуктуации.

Для того что бы выбрать ‘наилучшее приближение’ можно использовать следующее правило.

Нужно сравнивать среднее значения квадратов ошибок, то, есть для каждого приближения берется величина квадрата ошибки и делится на число точек минус число параметров. Чем меньше эта величина тем лучше.



# Модель временных рядов.

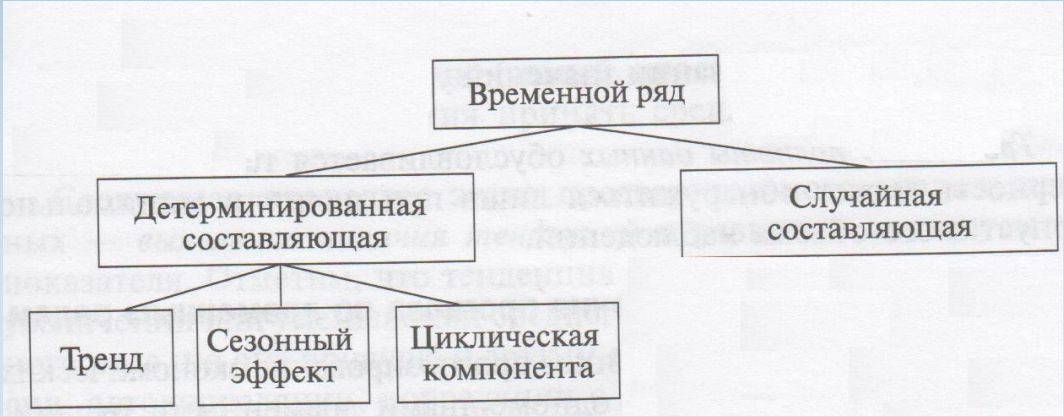
В статистике под **временным рядом** понимаются последовательно измеренные через некоторые (зачастую равные) промежутки времени данные. **Анализ временных рядов** объединяет методы изучения временных рядов, как пытающиеся понять природу точек данных, так и пытающиеся построить прогноз. **Прогнозирование временных рядов** заключается в построении модели для предсказания будущих событий основываясь на известных событий прошлого, предсказания будущих данных до того как они будут измерены. На пример — предсказание цены открытия биржи основываясь на предыдущей её деятельности.

Какие задачи здесь возникают?

1. **Физика солнца**: а) скрытые периодичности; б) прогноз активности.
**Электрокардиограммы** (ЭКГ): а) природа наблюдающихся аритмий;
б) прогноз развития состояния.
**Экономические ряды**: а) задача сегментации; б) задача прогноза.
**Химическая кинетика**: а) анализ динамики; б) построение модели.

# Модель временных рядов.

**Тренд** (тенденция) - устойчивая закономерность, наблюдаемую в течении длительного времени. Например, демографическая характеристика или рост потребления.




**Сезонная компонента** - функция, которая характеризует сезонные колебания. Как правило это колебания носят периодический или почти периодический характер, например загруженность дорог, пик продаж товаров.

**Циклическая компонента** - функция, описывающая длительные периоды (более года) спада или подъема. примером таких колебаний являются волны Кондратьева, демографические ямы.

**Случайная компонента** - шум, который отражает случайное действие многочисленных факторов.

# Обзор моделей прогнозирования на основе временных рядов.

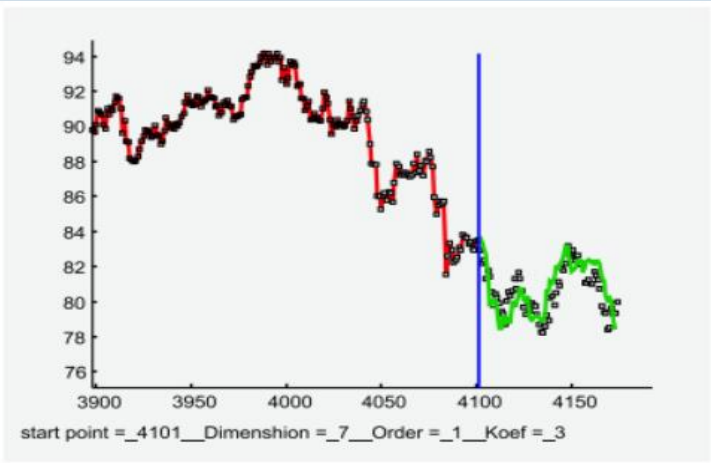
**1. Регрессионные модели прогнозирования:**
•Простая линейная регрессия (linear regression)
•Множественная регрессия (multilple regression)
•Нелинейная регрессия (nonlinear regression)

**2. Авторегрессионные модели прогнозирования:**
ARIMAX (autoregression integrated moving average extended)
GARCH (generalized autoregressive conditional heteroskedasticity),
ARDLM (autoregression distributed lag model)

**3. Модели экспоненциального сглаживания (ES)**:
Взвешенное скользящее среднее
Экспоненциальное сглаживание (exponential smoothing) (Модель Брауна)
Модель Хольта
Модель Хольта-Винтерса

# Модель временных рядов.

4. Модель по выборке максимального подобия (MMSP)
5. Модель на нейронных сетях (ANN)
6. Модель на цепях Маркова (Markov chains)
7. Модель на классификационно-регрессионных деревьях (CART)
8. Модель на основе генетического алгоритма (GA)
9. Модель на опорных векторах (SVM)
10. Модель на основе передаточных функций (TF)
11. Модель на нечеткой логике (FL)
12. Модель сингулярного спектрального анализа (SSA)
13. Модель локальной аппроксимации (LA)
14. Модель фрактальных временных рядов
15. Модель на основе Вейвлет - преобразования
16. Модель на основе Фурье-преобразования.




Прогноз Индекса Доу–Джонса

# Экспоненциальное сглаживание (Модель Брауна).

Модели сглаживания относится к адаптивным моделям прогнозирования, способным изменять свою структуру и параметры, приспосабливаясь к изменению условий. Все адаптивные модели делятся на два класса:

1. Модели скользящего среднего (СС-модели) (различные вариации)
2. Авторегрессии (АР-модели). – это мы не рассматриваем

Согласно схеме скользящего среднего оценкой текущего уровня (наблюдения) является взвешенное среднее всех предшествующих уровней, причем вес (множитель), который отражает информационную ценность наблюдения, тем больше, чем ближе оно находится к текущему уровню. Такие модели хорошо отражают тенденцию, но не позволяют отражать колебания.

В СС - моделях сглаживание производятся с помощью параметра сглаживания, который принимает значения в интервале от 0 до 1.

# Скользящее среднее.

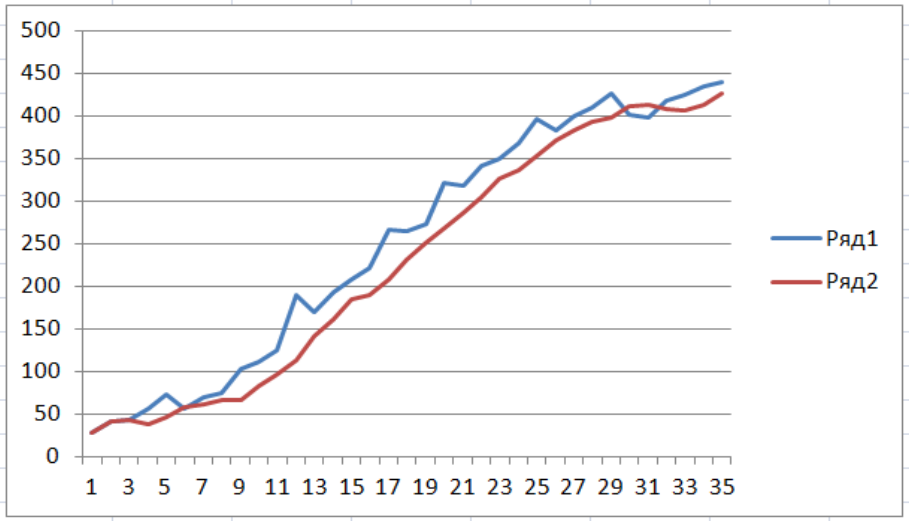
Метод скользящего среднего заключается в том что предсказанное значение вычисляется как среднее арифметическое по трем предыдущим данным:

Этот метод годится для медленно меняющихся кривых, с небольшой шумовой компонентой.

$$\hat{y}_{t+1} = \frac{1}{n}(y_t + y_{t-1} + \ldots + y_{t-n+1}).$$

O11

| | A | B | C |
|---|---|---|---|
| 1 | 1 | 27,3 | 27,3 |
| 2 | 2 | 41,8 | 41,8 |
| 3 | 3 | 42,8 | 42,8 |
| 4 | 4 | 56,2 | 37,3 |
| 5 | 5 | 72,5 | 46,93333333 |
| 6 | 6 | 56 | 57,16666667 |
| 7 | 7 | 70 | 61,56666667 |
| 8 | 8 | 74,9 | 66,16666667 |
| 9 | 9 | 103,3 | 66,96666667 |
| 10 | 10 | 111,3 | 82,73333333 |
| 11 | 11 | 125,2 | 96,5 |
| 12 | 12 | 189,3 | 113,2666667 |
| 13 | 13 | 169,1 | 141,9333333 |
| 14 | 14 | 193,5 | 161,2 |
| 15 | 15 | 207,4 | 183,9666667 |
| 16 | 16 | 221,2 | 190 |
| 17 | 17 | 267,2 | 207,3666667 |
| 18 | 18 | 264 | 231,9333333 |



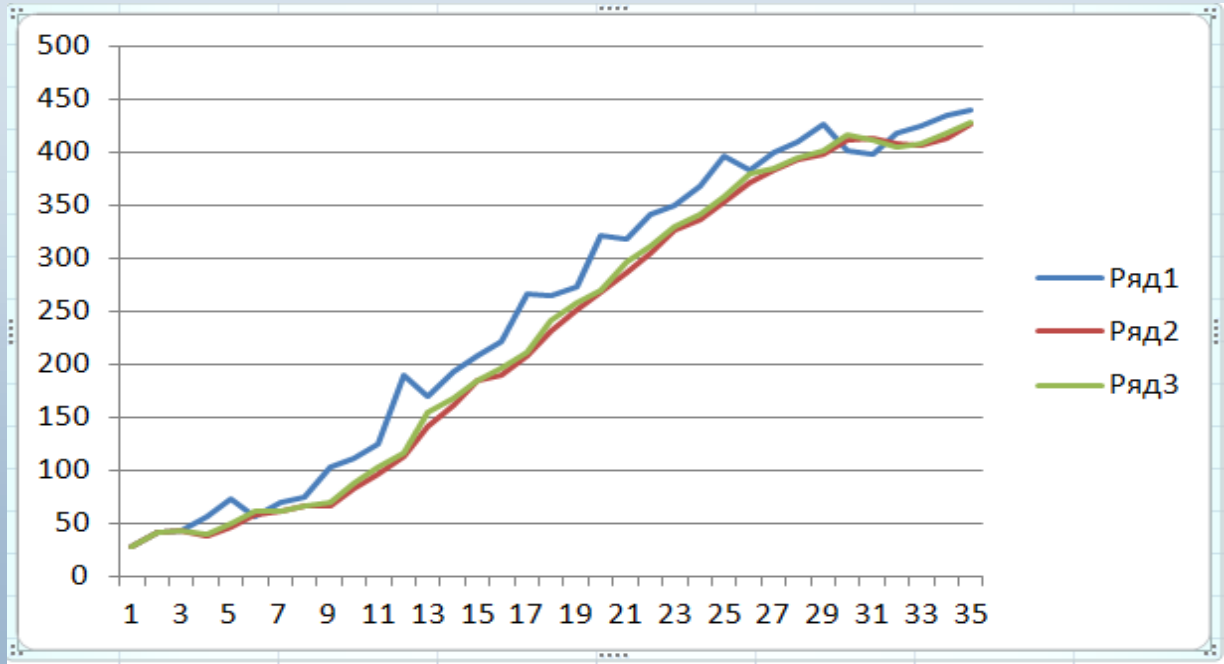
# Взвешенное скользящее среднее.

Метод взвешенного скользящего среднего заключается в том что предсказанное значение вычисляется как среднее арифметическое по трем предыдущим данным, при этом каждое из трех величин умножается на веса. Сумма весов должна быть равна единице.

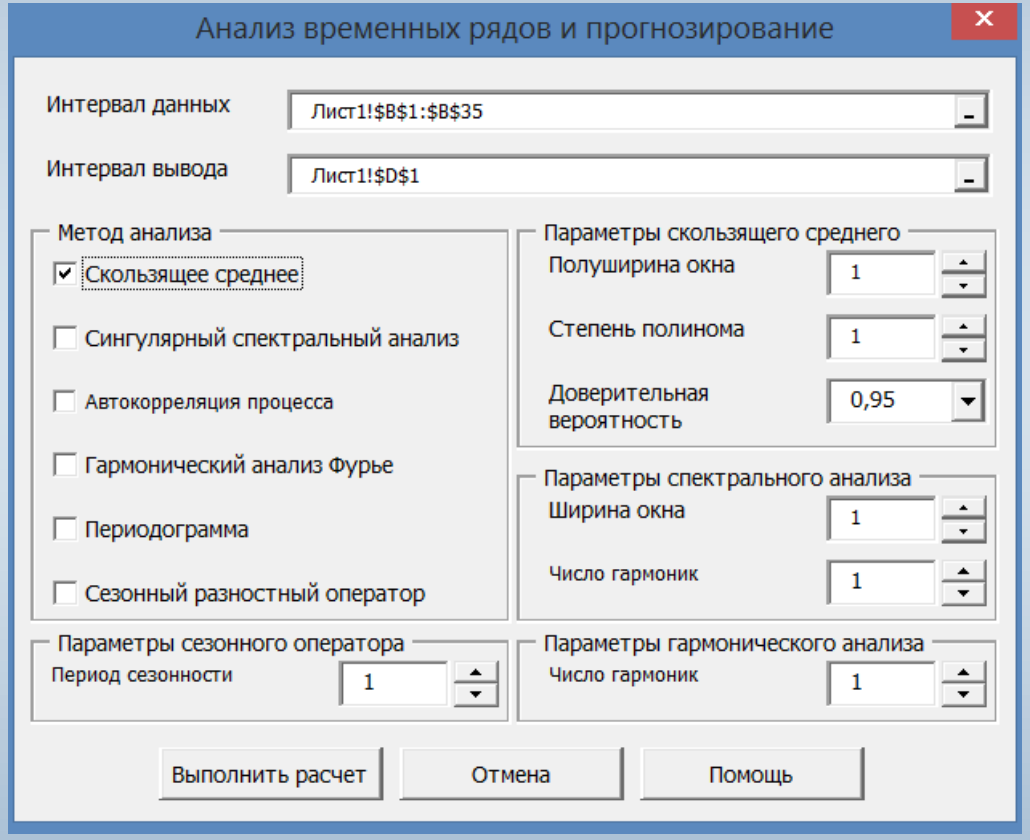
$$\hat{y}_7 = \alpha_1 y_6 + \alpha_2 y_5 + \alpha_3 y_4,$$

Например

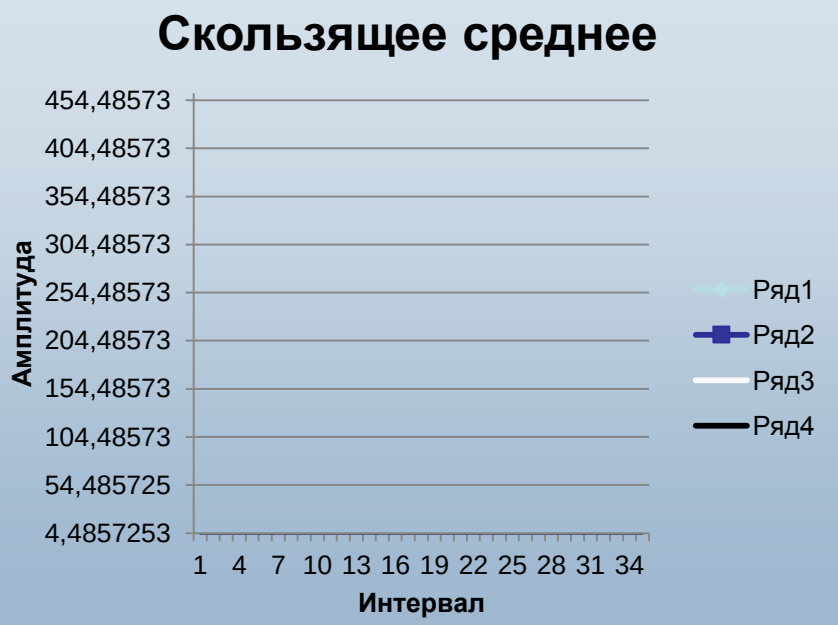
$$\hat{y}_7 = \frac{3}{6} y_6 + \frac{2}{6} y_5 + \frac{1}{6} y_4 .$$



# Взвешенное скользящее среднее в AtteStat




Увеличение степени полинома улучшает качество апроксимации



# Экспоненциальное сглаживание (Модель Брауна).

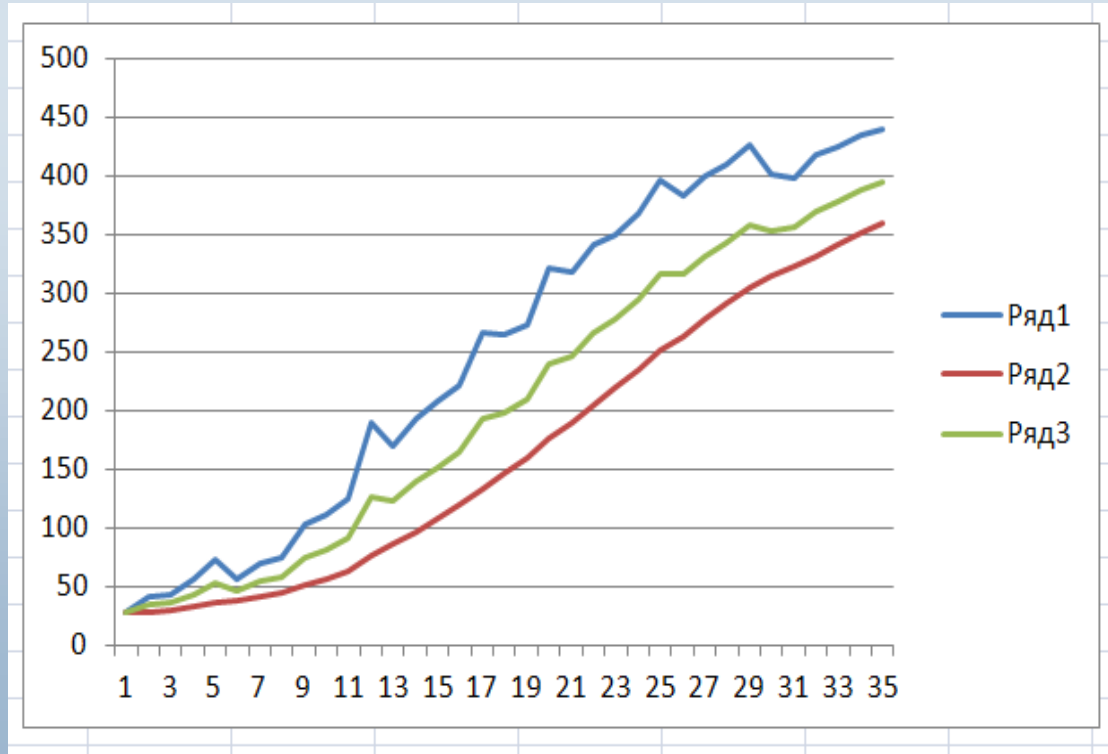
Модель экспоненциального сглаживания отличается от метода взвешенного скользящего среднего выбором коэффициента:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha\left(Y_t - \hat{Y}_t\right)$$

где $\hat{Y}_t$ предсказанное значение за прошлую дату

Идея метода заключается в том, что прогнозное значение определяется через предыдущее спрогнозированное значение, но скорректированное на величину отклонения факта от прогноза

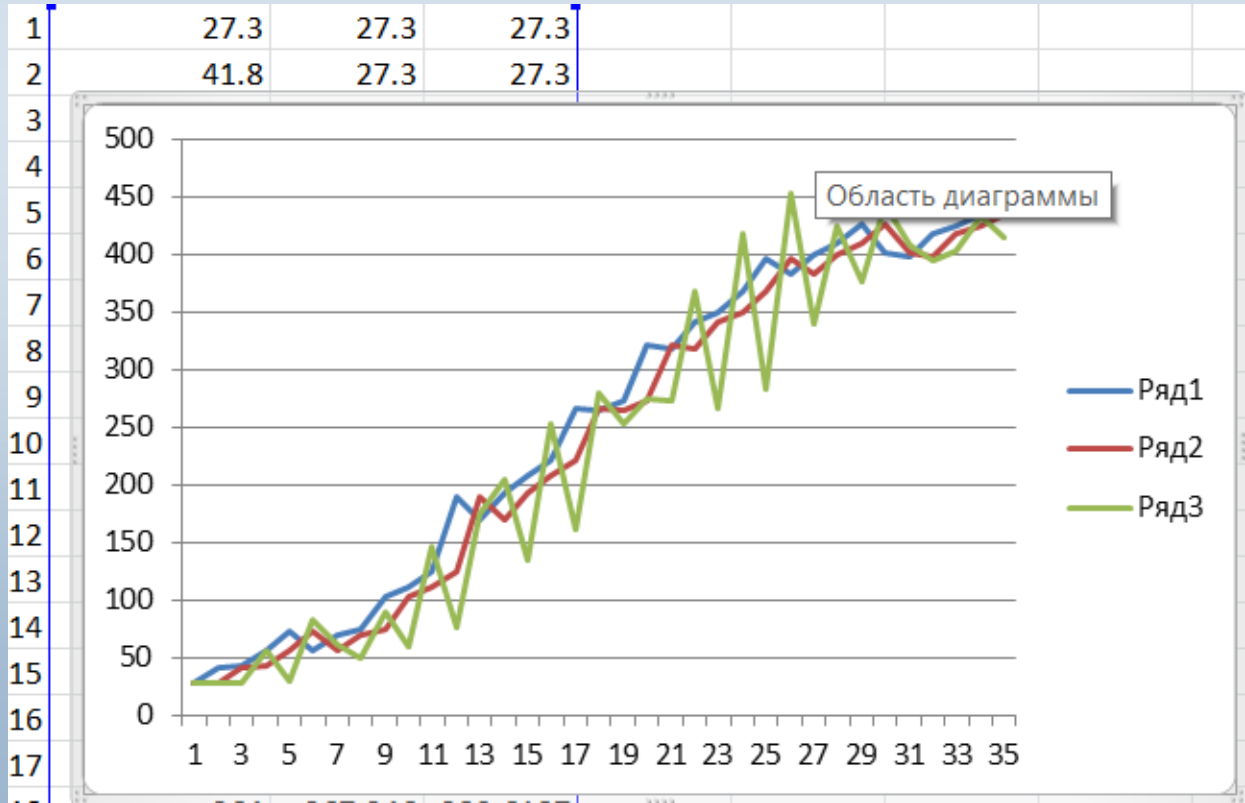
$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$$



# Экспоненциальное сглаживание (Модель Брауна).

Вообще же модель Брауна может применяться в двух случаях:

1. Когда нужно сгладить имеющийся ряд данных для выявления какой-либо тенденции(обычно в случае со стационарными процессами). Тогда обычно исследователь задаёт значение *α в пределах от 0 до 1.*

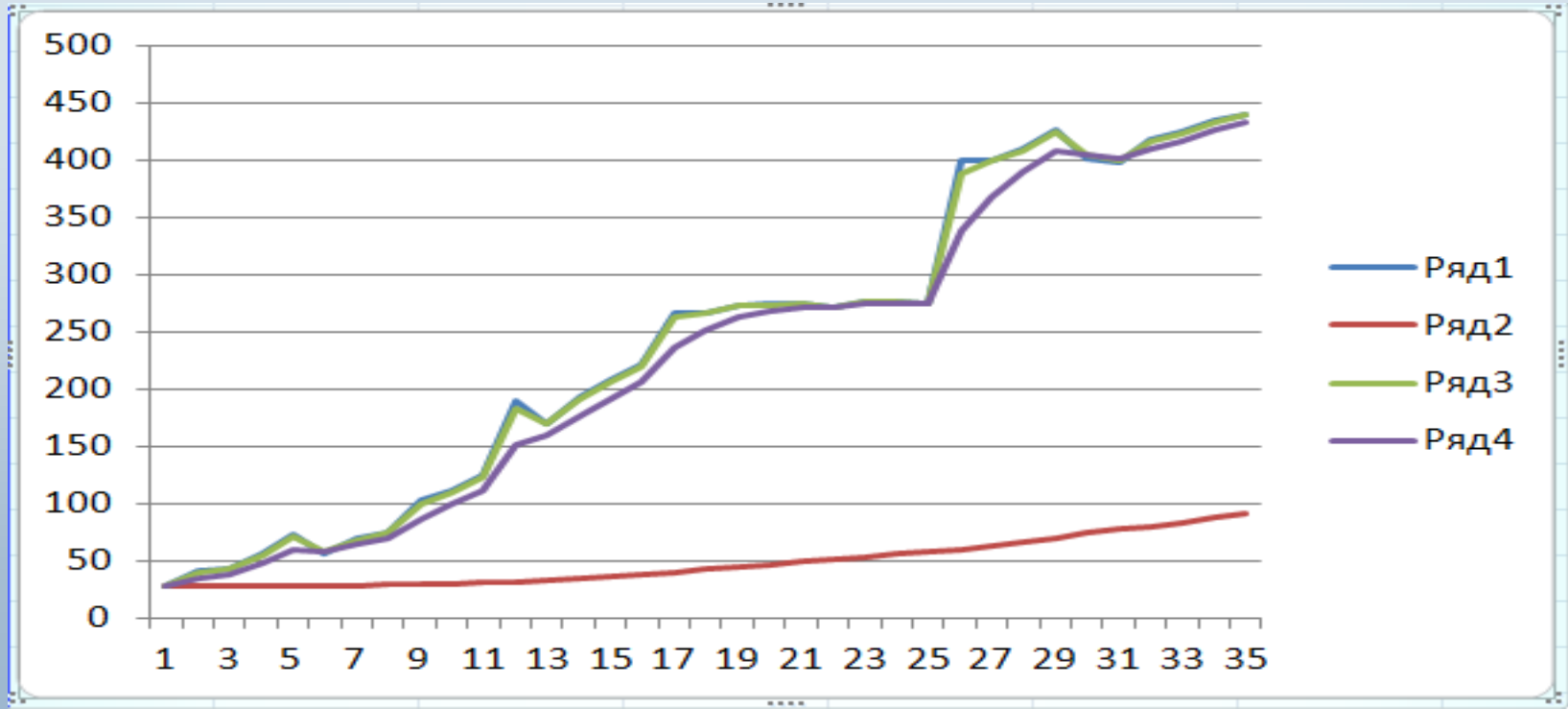


# Экспоненциальное сглаживание (Модель Брауна).

Очевидно, что величина α сильно влияет на результат сглаживания. Рассмотрим как метод сглаживания влияет на три ситуации.

1. Реакция сглаживания на скачок.

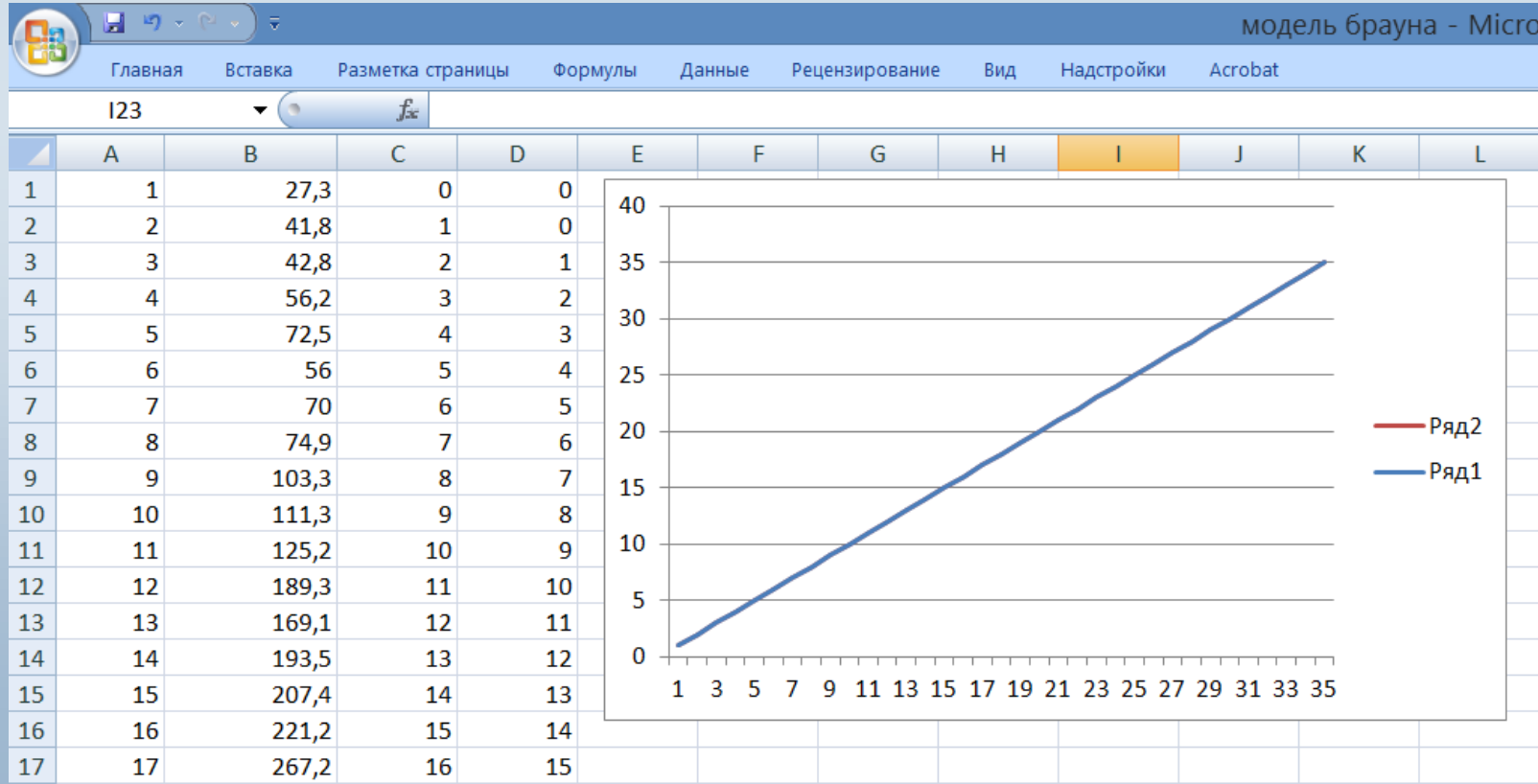
Чем меньше значение α тем хуже результат (ряд 2).



# Экспоненциальное сглаживание (Модель Брауна).

1. Реакция сглаживания на постоянные изменения.

Модель Брауна плохо годится для временных рядов, прогноз всегда занижен.

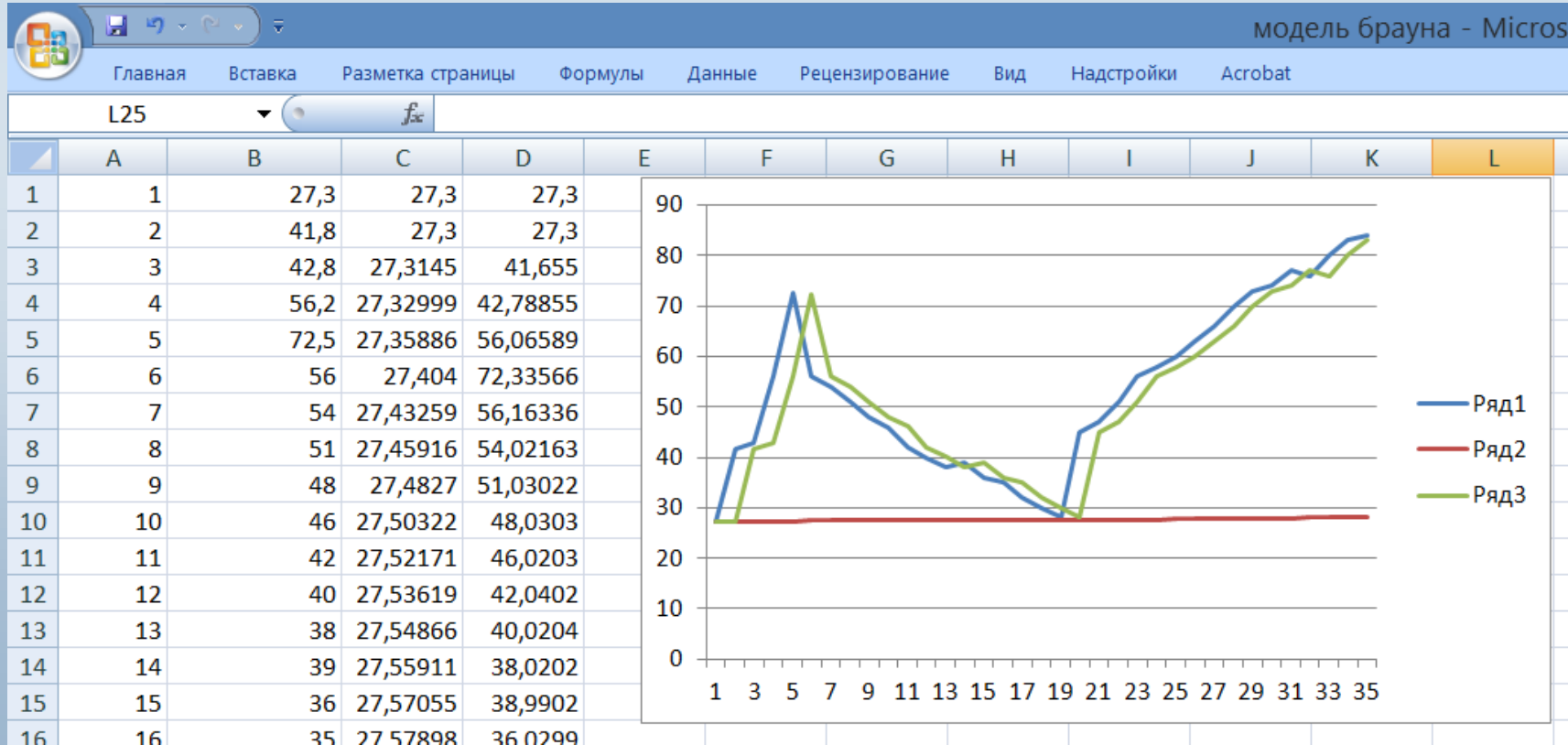



| | A | B | C | D |
|---|---|---|---|---|
| 1 | 1 | 27,3 | 0 | 0 |
| 2 | 2 | 41,8 | 1 | 0 |
| 3 | 3 | 42,8 | 2 | 1 |
| 4 | 4 | 56,2 | 3 | 2 |
| 5 | 5 | 72,5 | 4 | 3 |
| 6 | 6 | 56 | 5 | 4 |
| 7 | 7 | 70 | 6 | 5 |
| 8 | 8 | 74,9 | 7 | 6 |
| 9 | 9 | 103,3 | 8 | 7 |
| 10 | 10 | 111,3 | 9 | 8 |
| 11 | 11 | 125,2 | 10 | 9 |
| 12 | 12 | 189,3 | 11 | 10 |
| 13 | 13 | 169,1 | 12 | 11 |
| 14 | 14 | 193,5 | 13 | 12 |
| 15 | 15 | 207,4 | 14 | 13 |
| 16 | 16 | 221,2 | 15 | 14 |
| 17 | 17 | 267,2 | 16 | 15 |

# Экспоненциальное сглаживание (Модель Брауна).

1. Реакция сглаживания на колебания.

Маленькое α почти прямая линия, большое α лучше, но дает запаздывание




| | A | B | C | D |
|---|---|---|---|---|
| 1 | 1 | 27,3 | 27,3 | 27,3 |
| 2 | 2 | 41,8 | 27,3 | 27,3 |
| 3 | 3 | 42,8 | 27,3145 | 41,655 |
| 4 | 4 | 56,2 | 27,32999 | 42,78855 |
| 5 | 5 | 72,5 | 27,35886 | 56,06589 |
| 6 | 6 | 56 | 27,404 | 72,33566 |
| 7 | 7 | 54 | 27,43259 | 56,16336 |
| 8 | 8 | 51 | 27,45916 | 54,02163 |
| 9 | 9 | 48 | 27,4827 | 51,03022 |
| 10 | 10 | 46 | 27,50322 | 48,0303 |
| 11 | 11 | 42 | 27,52171 | 46,0203 |
| 12 | 12 | 40 | 27,53619 | 42,0402 |
| 13 | 13 | 38 | 27,54866 | 40,0204 |
| 14 | 14 | 39 | 27,55911 | 38,0202 |
| 15 | 15 | 36 | 27,57055 | 38,9902 |
| 16 | 16 | 35 | 27,57898 | 36,0299 |

# Выбор адекватной модели прогнозирования

Какую из описанных моделей следует применять для прогнозирования значения временного ряда? Есть несколько эмпирических правил:

1. **Оценить величину остаточной ошибки с помощью квадратов разностей.** (то есть оценить величину дисперсии). Если модель идеально аппроксимирует значения временного ряда в предыдущие моменты времени, стандартная ошибка оценки равна нулю. С другой стороны, если модель плохо аппроксимирует значения временного ряда в предыдущие моменты времени, стандартная ошибка оценки велика. Таким образом, анализируя адекватность нескольких моделей, можно выбрать модель, имеющую минимальную стандартную ошибку оценки.

Основным недостатком такого подхода является преувеличение ошибок при прогнозировании отдельных значений. Иначе говоря, любая большая разность между величинами $Y_i$ и $\hat{Y}_i$ при вычислении суммы квадратов ошибок SSE возводится в квадрат, т.е. увеличивается.

# Выбор адекватной модели прогнозирования

2. **Оценить величину остаточной ошибки с помощью абсолютных разностей.**

$$MAD = \frac{\sum_{i=1}^{n} |Y_i - \hat{Y}_i|}{n}$$

При анализе конкретных моделей величина MAD представляет собой среднее значение модулей разностей между фактическим и предсказанными значениями временного ряда. Если модель идеально аппроксимирует значения временного ряда в предыдущие моменты времени, среднее абсолютное отклонение равно нулю. С другой стороны, если модель плохо аппроксимирует такие значения временного ряда, среднее абсолютное отклонение велико. Таким образом, анализируя адекватность нескольких моделей, можно выбрать модель, имеющую минимальное среднее абсолютное отклонение.

# Выбор адекватной модели прогнозирования

3. **Руководствоваться принципом экономии**

Если анализ стандартных ошибок оценок и средних абсолютных отклонений не позволяет определить оптимальную модель, можно воспользоваться четвертым методом, основанным на принципе экономии. Этот принцип утверждает, что из нескольких равноправных моделей следует выбирать простейшую.