

Методы и модели анализа прогнозирования экономических моделей.

Кольцов С.Н
2014

www.linis.ru

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 $y = f(x_1, x_2, x_3...)$

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

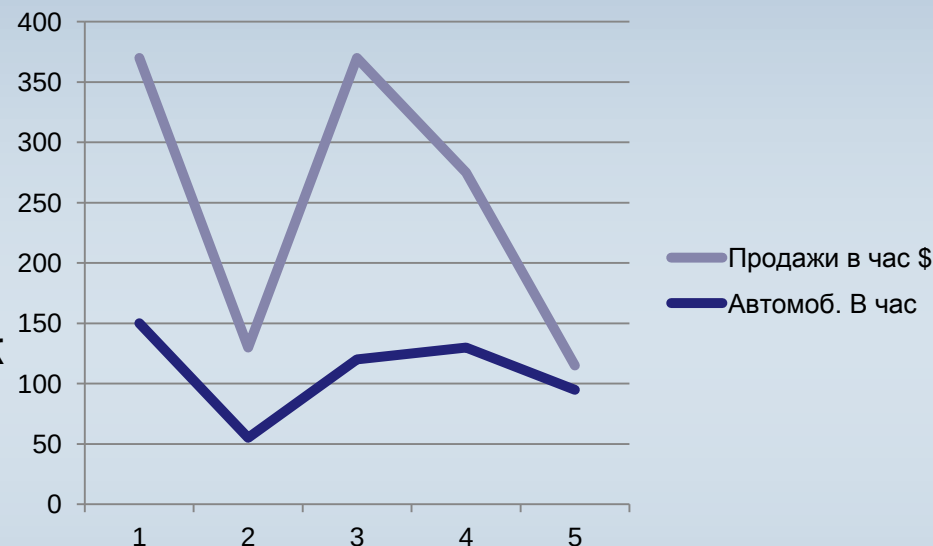
Причинно – следственная модель

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

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

где

y – реальное значение

$\hat{y}$ – аппроксим значение

Средний квадрат ошибки

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

где

y – реальное значение

$\hat{y}$ – аппроксим 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%.

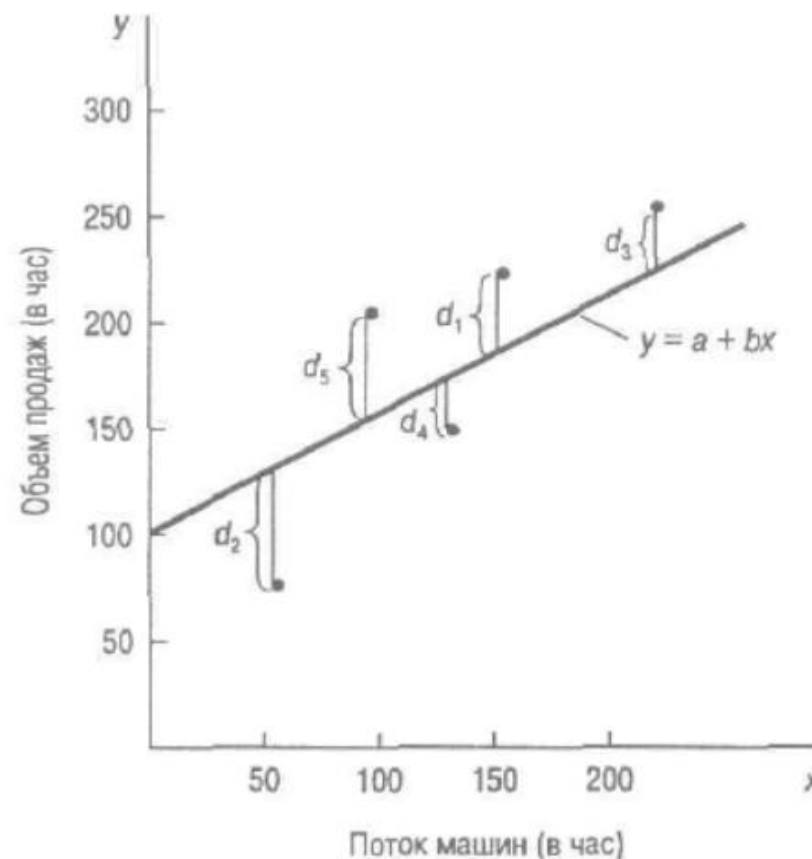
Метод наименьших квадратов

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

$$x_1, x_2, \dots, x_i, \dots, x_n;$$

$$y_1, y_2, \dots, y_i, \dots, y_n.$$

По данным такого эксперимента можно построить график зависимости $y = f(x, a, b, c, \dots)$. 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

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



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

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

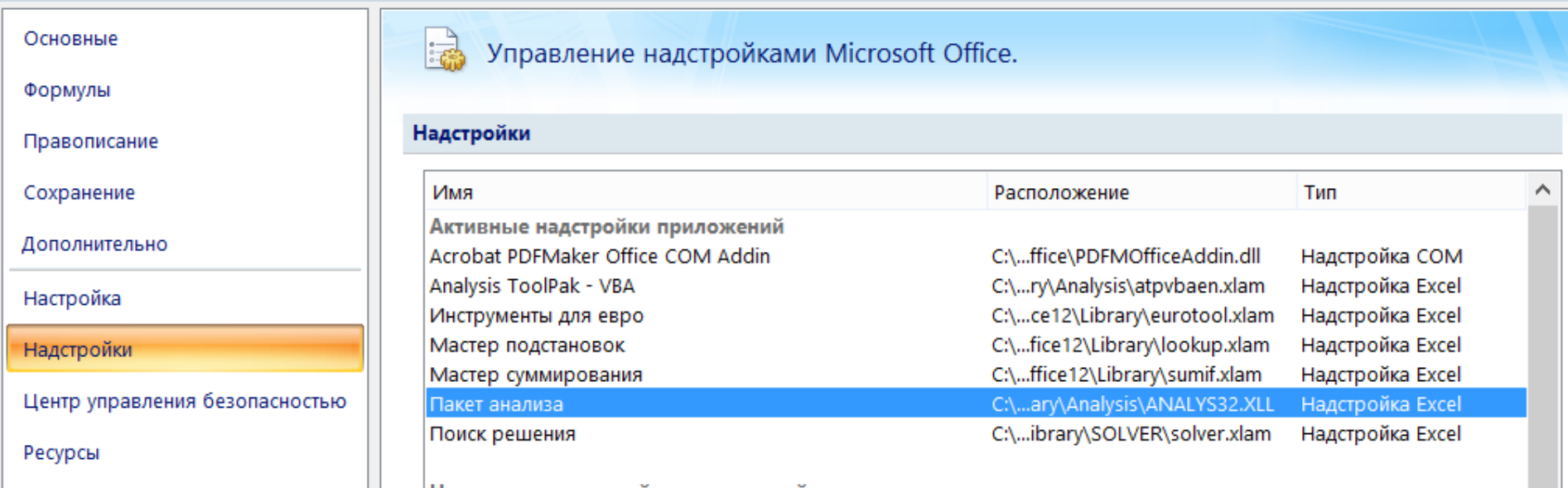


$$b = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2}$$

Получить самостоятельно

Метод наименьших квадратов в Экселе

Включим пакет анализ данных.



Управление надстройками Microsoft Office.

Настройки

Имя	Расположение	Тип
Активные надстройки приложений		
Acrobat PDFMaker Office COM Addin	C:\...ffice\PDFMOfficeAddin.dll	Надстройка COM
Analysis ToolPak - VBA	C:\...ry\Analysis\atpvbaen.xlam	Надстройка Excel
Инструменты для евро	C:\...ce12\Library\eurotool.xlam	Надстройка Excel
Мастер подстановок	C:\...fice12\Library\lookup.xlam	Надстройка Excel
Мастер суммирования	C:\...ffice12\Library\sumif.xlam	Надстройка Excel
Пакет анализа	C:\...ary\Analysis\ANALYS32.XLL	Надстройка Excel
Поиск решения	C:\...ibrary\SOLVER\solver.xlam	Надстройка Excel

Теперь нужно перейти в опцию 'Настройки' – 'Attestat' - аппроксимация зависимости

Метод наименьших квадратов в Экселе

Аппроксимация зависимостей и регрессионный анализ

Интервал абсцисс: Лист1!\$B\$2:\$B\$6

Интервал ординат: Лист1!\$C\$2:\$C\$6

Интервал вывода решения: Лист1!\$E\$2:\$E\$6

Тип модели

- ☐ Экспоненциальная функция
- ☐ Степенная функция
- ☐ Логарифмическая функция
- ☐ Эксп.-степенная функция
- ☐ Логистическая функция
- ☒ Полином
- ☐ Гипербола
- ☐ Пользовательская функция

Параметры полиномов

Минимальная степень: 0

Степень полинома: 1

Параметры пользовательской функции

Интервал параметров: Лист1!\$G\$2:\$G\$7

Ячейка формулы:

Рабочая ячейка:

Допустимое число итераций: 100

Точность: 0,0001

Метод оптимизации

- ☒ Бройдена-Флетчера-Голдфарба-Шанно
- ☐ Гаусса-Ньютона

Дополнительно

Доверительная вероятность: 0,95

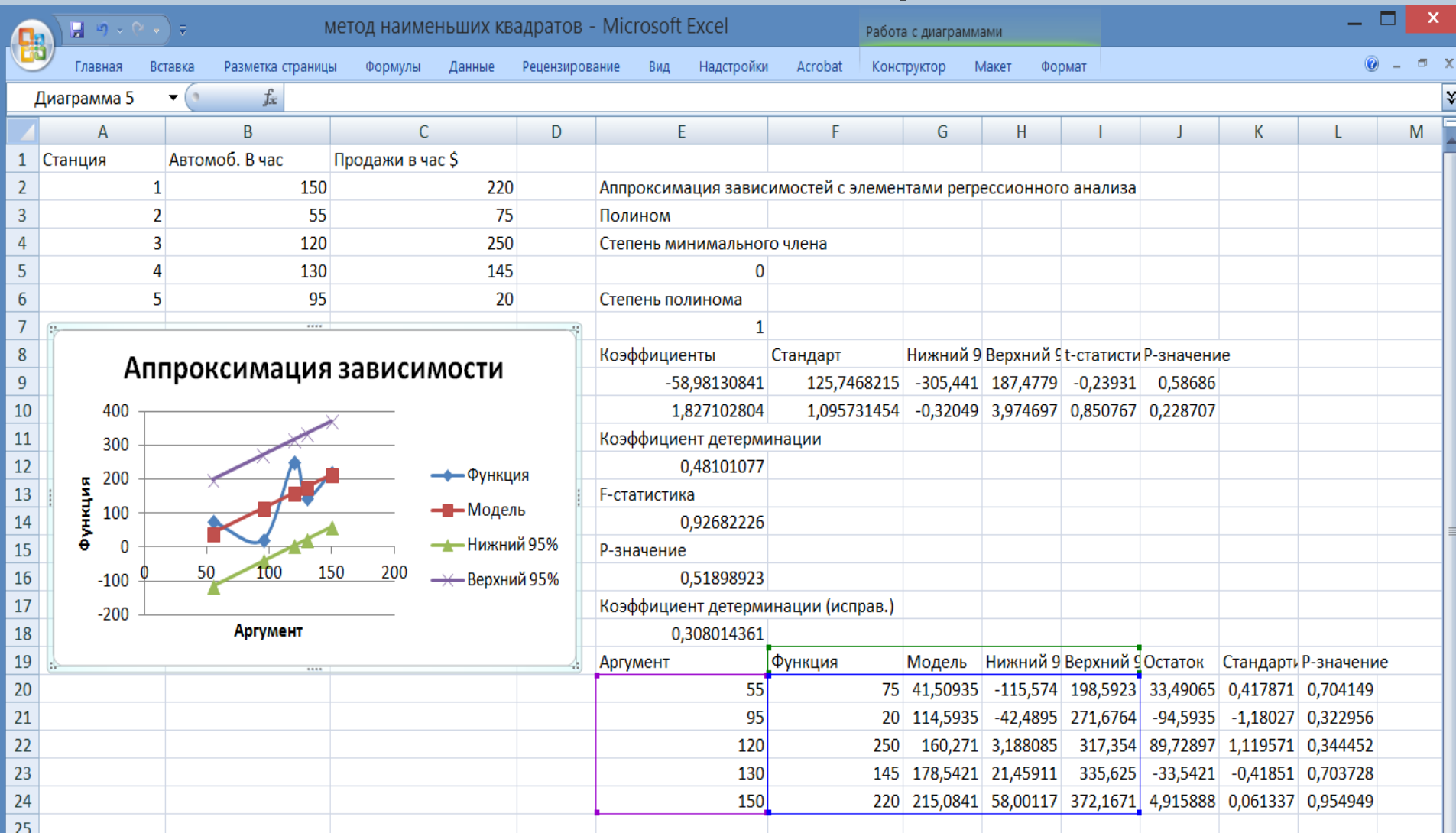
АtteStat

- Модуль XTAG - Кросстабуляция
- Модуль TRFG - Преобразования данных
- Модуль TIME - Анализ временных рядов
- Модуль SURVIVE - Анализ выживаемости
- Модуль SQC - Контроль качества
- Модуль RNZ - Рандомизация
- Модуль PS - Параметрическая статистика
- Модуль PRI - Распознавание образов
- Модуль QP - Обработка выбросов
- Модуль NP - Непараметрическая статистика
- Модуль NDC - Проверка нормальности
- Модуль MDP - Обработка пропущенных данных
- Модуль MQS - Многомерное шкалирование
- Модуль KNOW - Экспертные оценки
- Модуль JA - Информационный анализ
- Модуль FAA - Факторный анализ
- Модуль EXACT - Точные критерии
- Модуль DS - Описательная статистика
- Модуль CORA - Корреляционный анализ
- Модуль CLA - Кластерный анализ
- Модуль AV - Дисперсионный анализ
- Модуль APX - Аппроксимация зависимостей**
- О программе
- Как начать работу

Справка по APX

А Аппроксимация зависимостей

Метод наименьших квадратов в Экселе



Метод наименьших квадратов для квадратичной функции

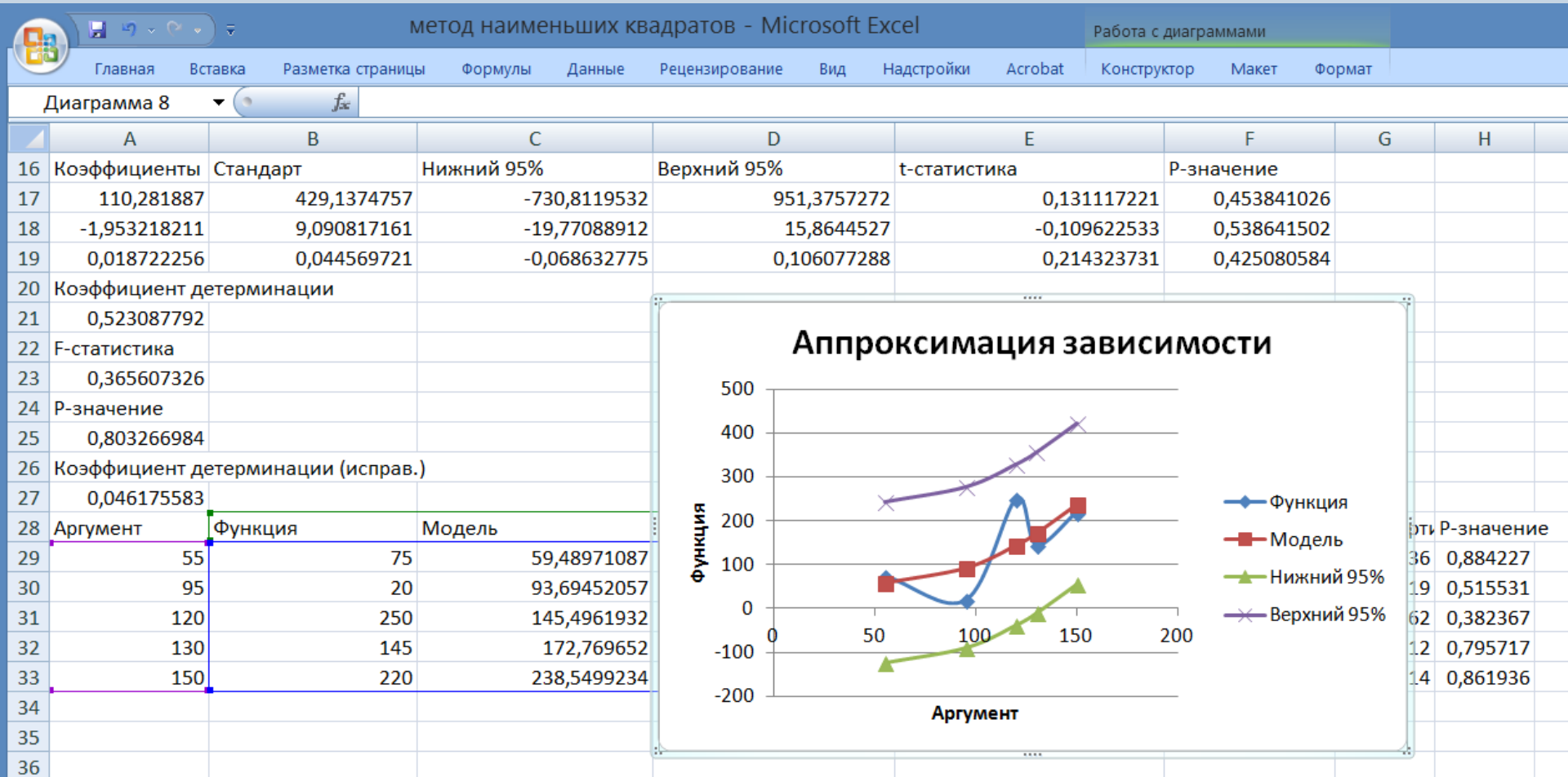
Рассмотрим аппроксимационную функцию в виде квадратичной функции, соответственно ищем квадрат разности между экспериментальными данными и нашей аналитической функцией с не известными коэффициентами.

$$\sum_{i=1}^5 (y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2 .$$

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

$$\begin{aligned} 5a_0 + (\sum x_i) a_1 + (\sum x_i^2) a_2 &= \sum y_i , \\ (\sum x_i) a_0 + (\sum x_i^2) a_1 + (\sum x_i^3) a_2 &= \sum x_i y_i , \\ (\sum x_i^2) a_0 + (\sum x_i^3) a_1 + (\sum x_i^4) a_2 &= \sum x_i^2 y_i . \end{aligned}$$

Метод наименьших квадратов для квадратичной функции в Экселе.



Метод наименьших квадратов

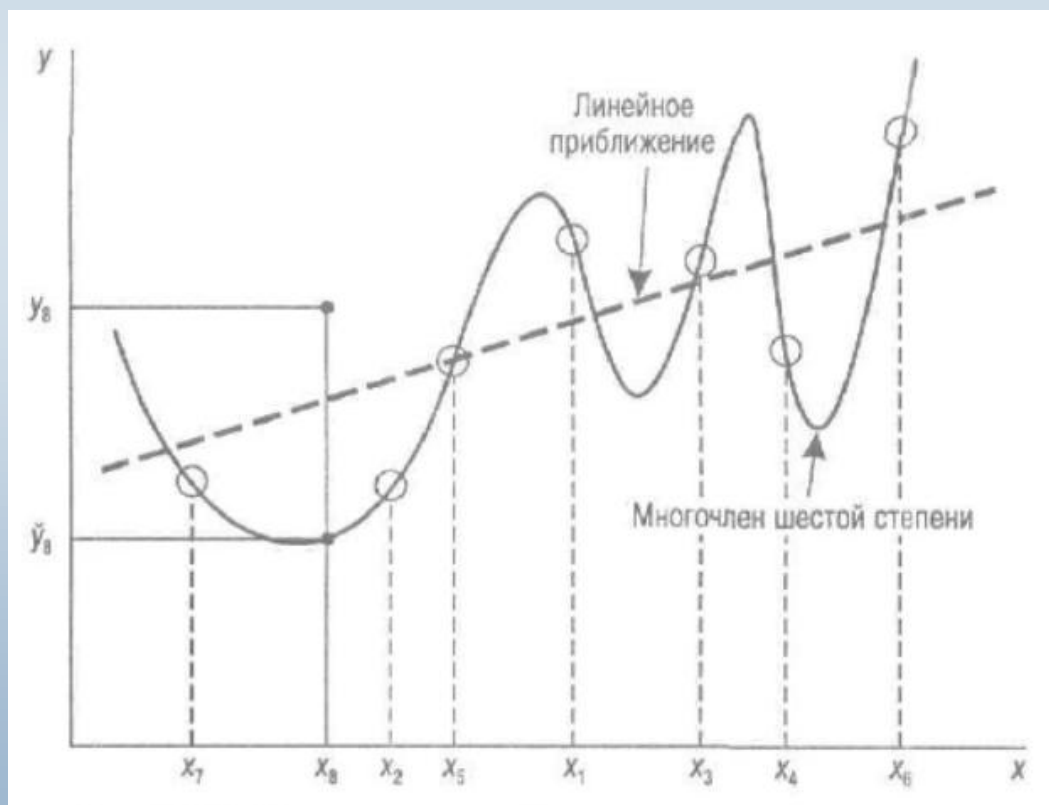
Проблема выбора степени полинома.

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 событиях 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;

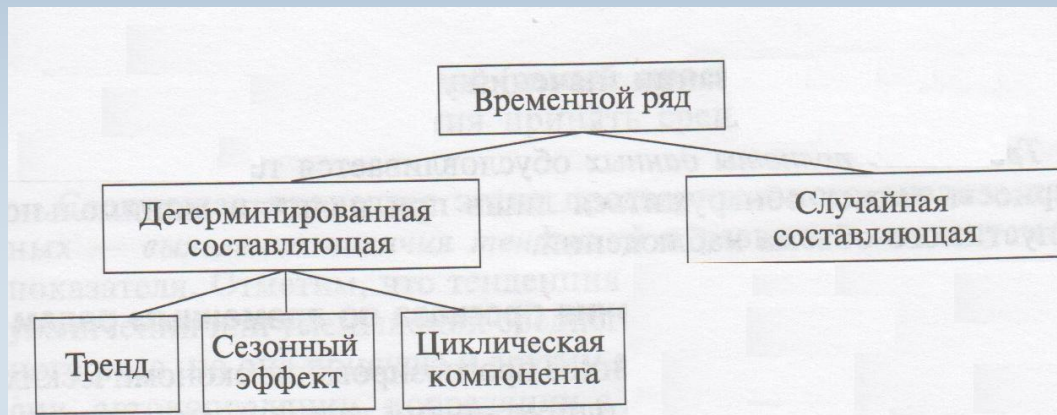
б) прогноз развития состояния.

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 (multiple 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

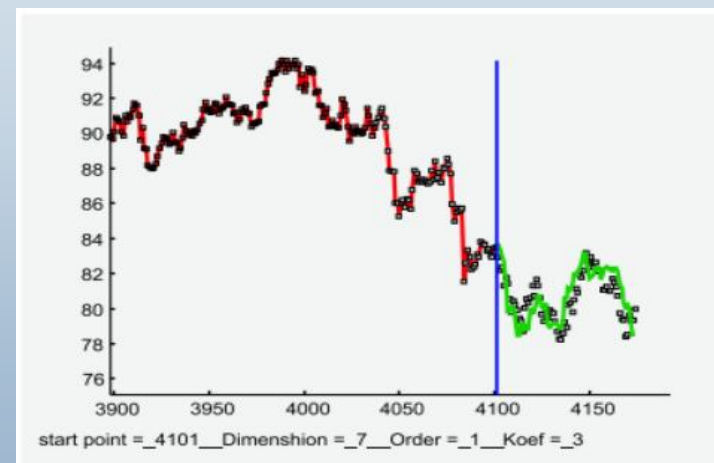
Экспоненциальное сглаживание (exponential smoothing) (Модель Брауна)

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.



Экспоненциальное сглаживание (Модель Брауна).

Модели сглаживания относятся 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.

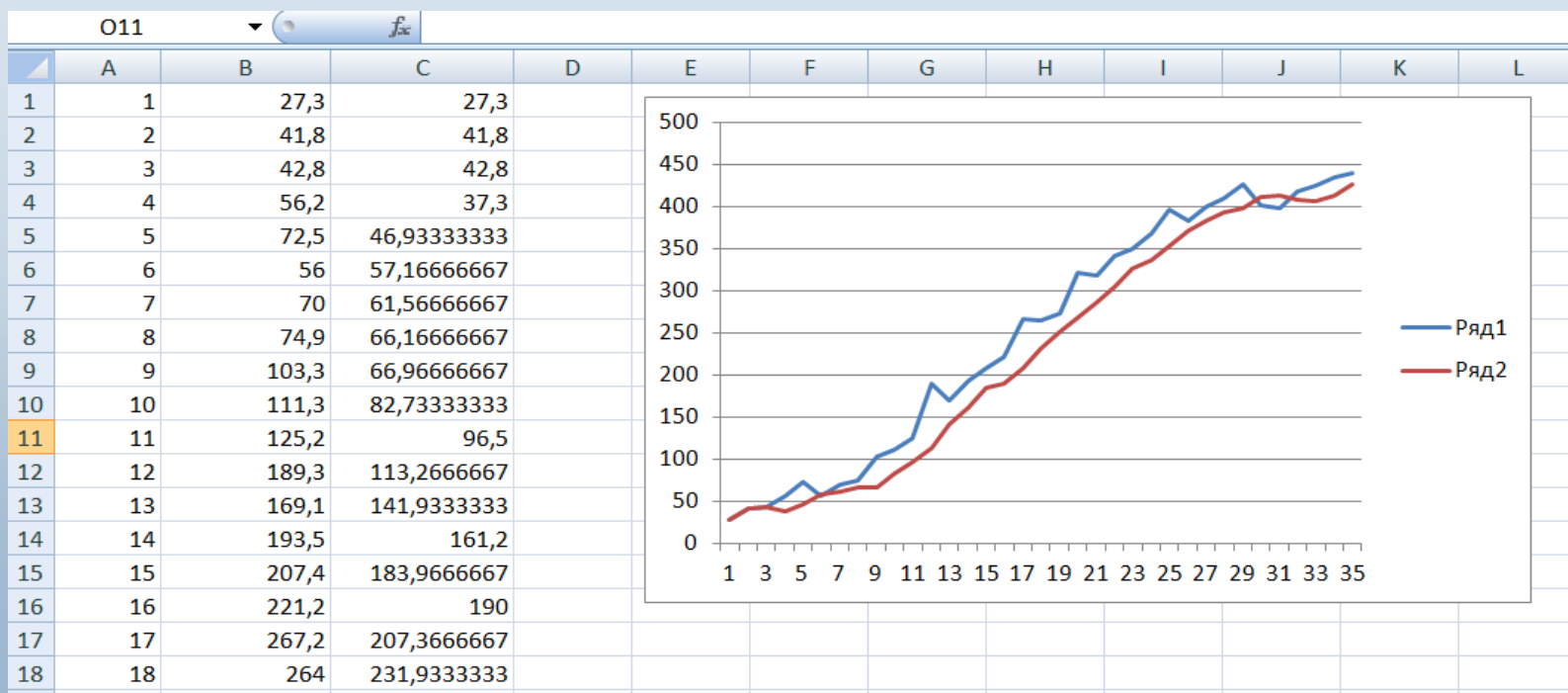
В СС - моделях сглаживание производится с помощью параметра сглаживания, который принимает значения в интервале от 0 до 1.

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.

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



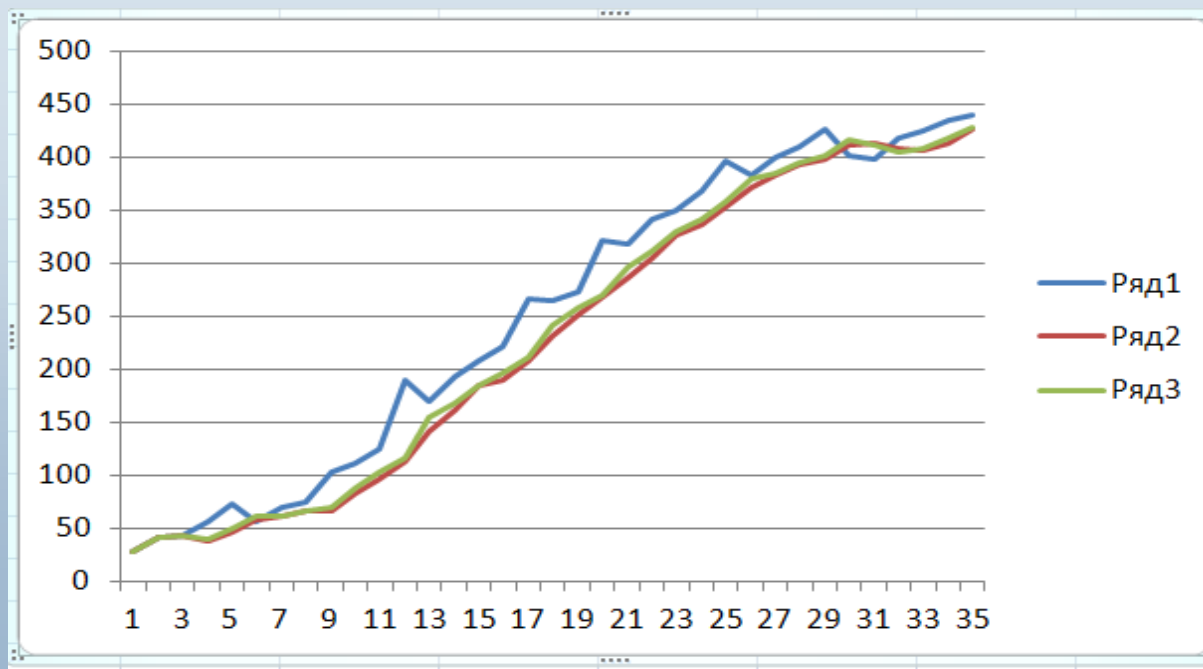
Взвешенное скользящее среднее.

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.

Например

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

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



Экспоненциальное сглаживание (Модель Брауна).

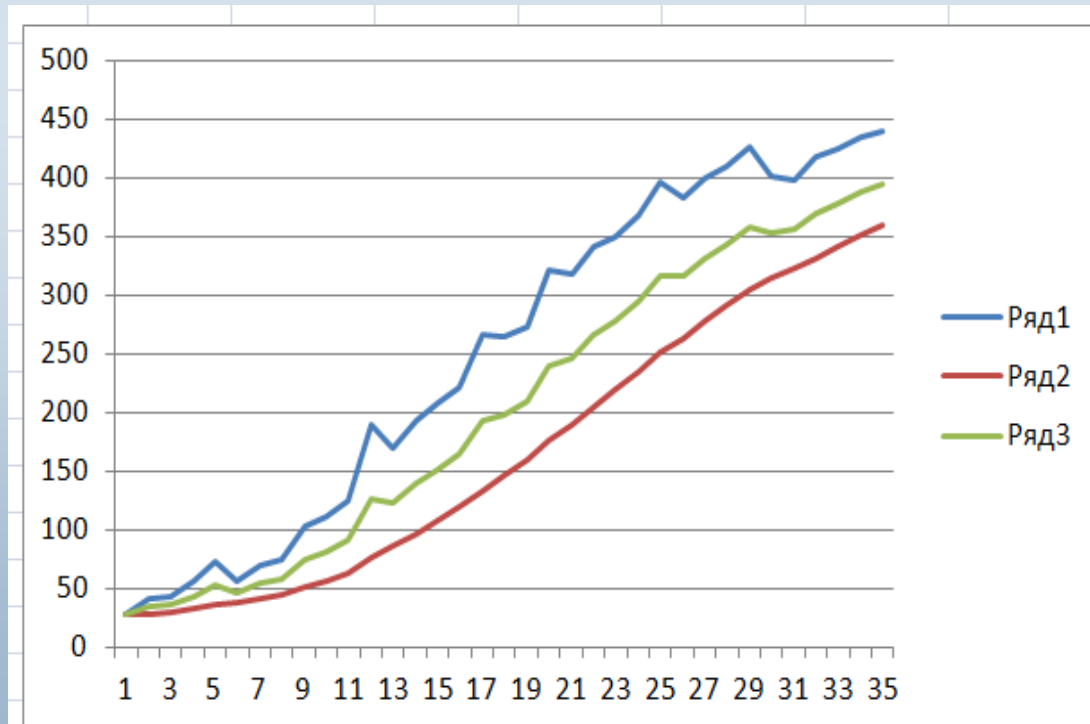
Модель экспоненциального сглаживания отличается от метода взвешенного скользящего среднего выбором коэффициента:

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

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

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 [3, p.97].

2. Когда нужно сделать краткосрочный прогноз. В таком случае наилучший результат

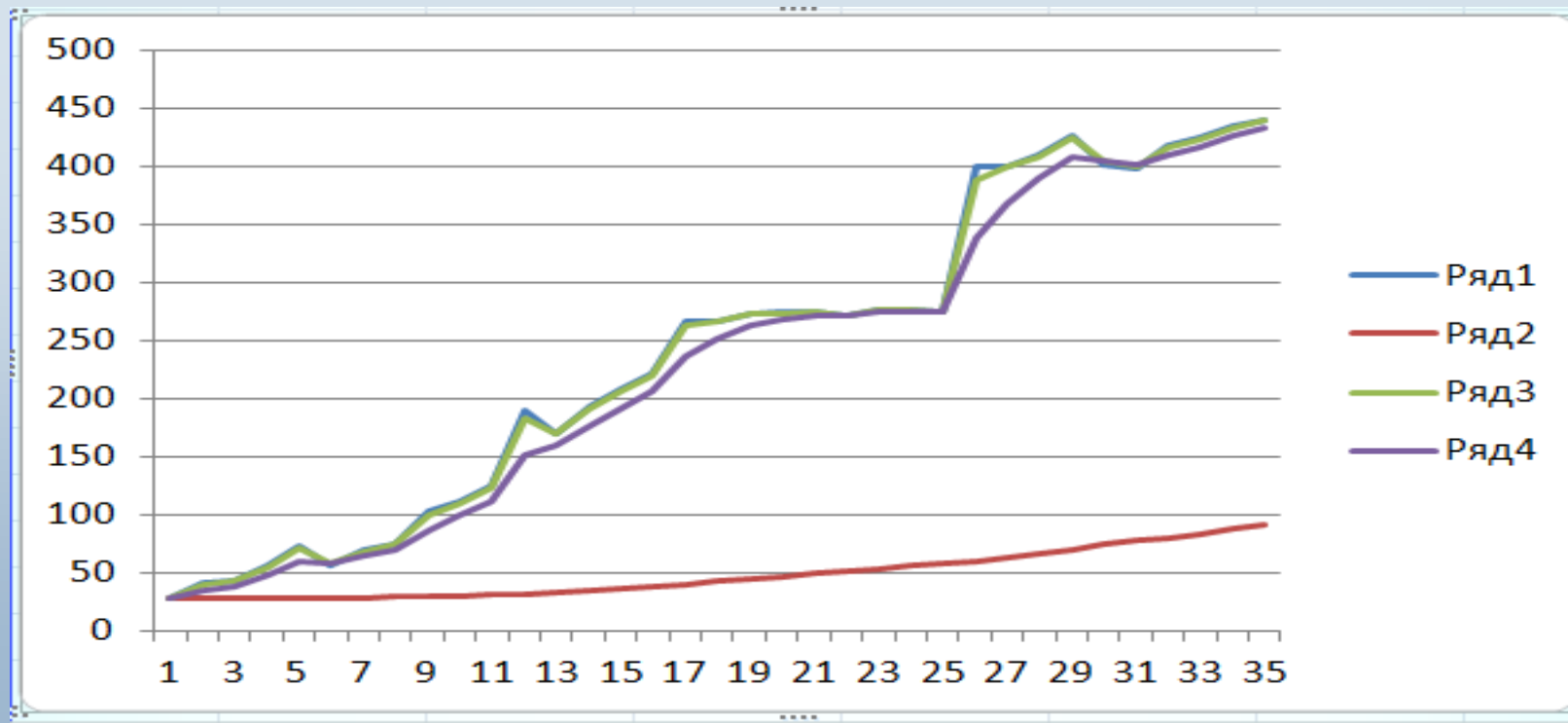
прогноза получается при задании α в пределах от 0 до 2

Экспоненциальное сглаживание (Модель Брауна).

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

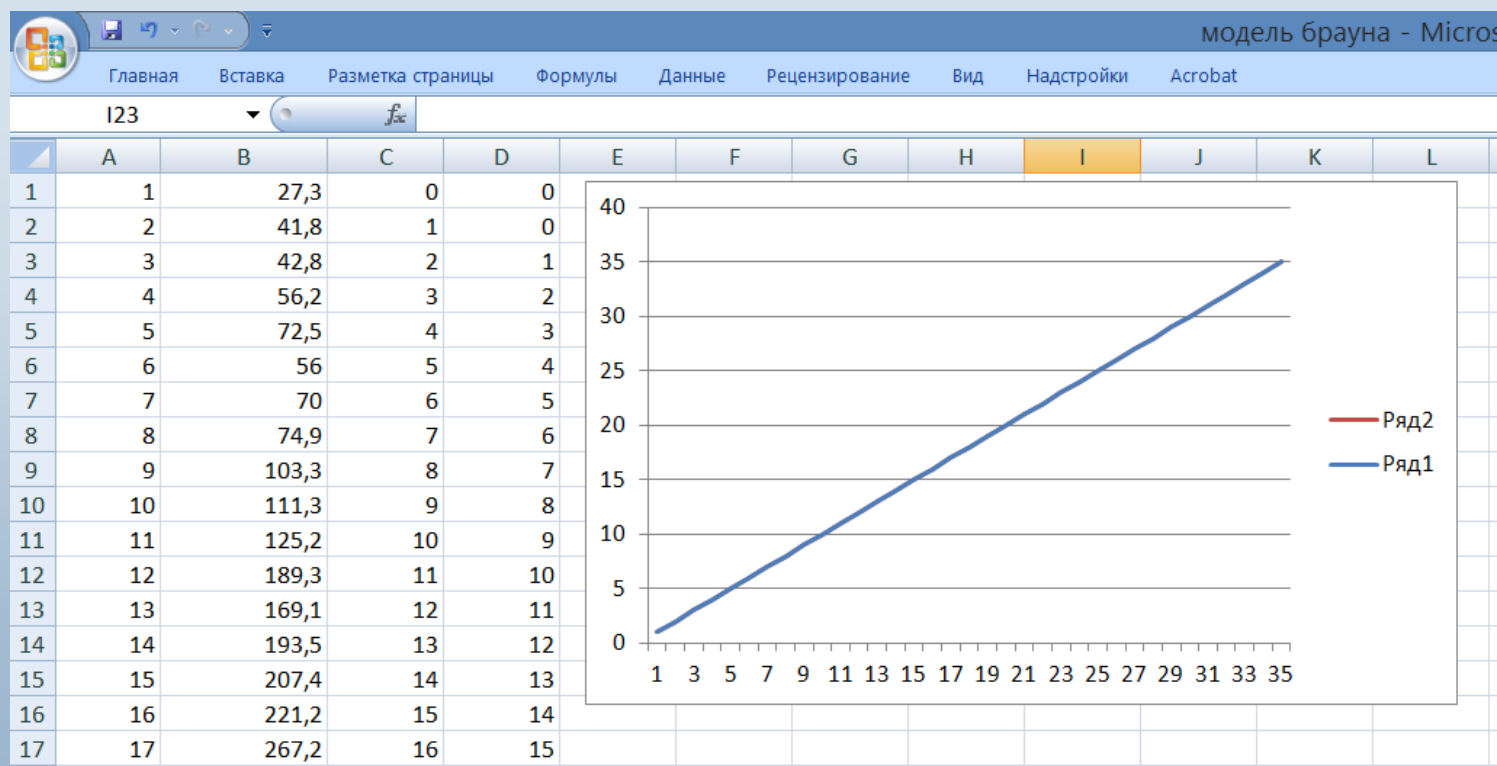
1. Реакция сглаживания на скачок.

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



1. Реакция сглаживания на постоянные изменения.

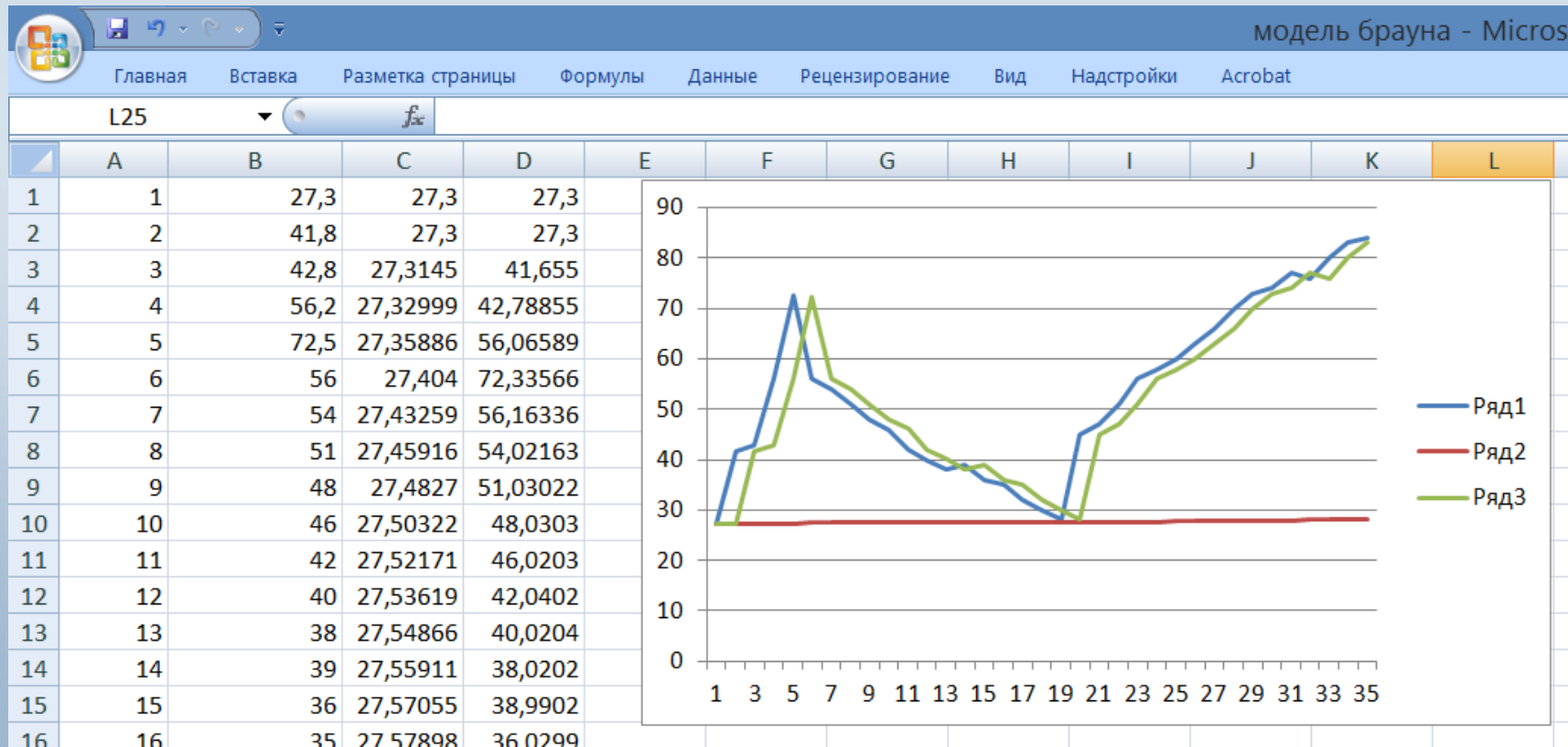
Модель Брауна плохо подходит для временных рядов, прогноз всегда занижен.



Экспоненциальное сглаживание (Модель Брауна).

1. Реакция сглаживания на колебания.

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



Модель Хольта

$$\begin{cases} \hat{Y}_{t+\tau} = a_t + \tau b_t \\ a_t = \alpha_1 Y_t + (1 - \alpha_1)(a_{t-1} + b_{t-1}) , \\ b_t = \alpha_2 (a_t - a_{t-1}) + (1 - \alpha_2)b_{t-1} \end{cases}$$

где $\hat{Y}_{t+\tau}$ - прогноз, делающийся на τ шагов вперёд, a_t – коэффициент уровня ряда, b_t – коэффициент пропорциональности, α_1, α_2 – постоянные сглаживания.

Для того чтобы получить прогноз по модели Хольта, нужно провести некоторую подготовительную работу, а именно – рассчитать значения коэффициентов a_0 и b_0 по имеющемуся ряду данных. После этого по какому-либо критерию подбираются постоянные сглаживания, в результате чего исследователь получает линейную модель, на каждом шаге адаптирующуюся к фактическим данным.

THANK YOU!

← → ↺ www.linis.hse.spb.ru/index.php/glavnaja.html



[Главная](#)

[О нас](#)

[События](#)

[Публикации](#)

[Исследования](#)

[Материалы](#)

[Ссылки](#)

[Контакты](#)

ENGLISH


Новости



Чем дышит блогосфера? Семинар ЛИНИс в Москве

11:54 27.04.2012

24 апреля команда ЛИНИс провела в НИУ-ВШЭ (Москва) семинар на тему "Чем дышит блогосфера? Методы анализа больших массивов Интернет-данных для социологических задач". Выступление прошло в рамках академического семинара по социологической теории и методологии кафедры анализа социальных институтов (рук. Инна Девятко), который на этот раз проходил совместно с Лабораторией экономико-социологических исследований (рук. В.Радзев).

[Подробнее...](#)

ЛИНИс – участник Балтийского партнерства по новым медиа



Анонсы

11.05.2012 (Пятница)

Презентация Лаборатории Интернет-Исследований

11 мая в 17-00 состоится презентация новой исследовательской площадки Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург) - Лаборатории интернет-исследований (ЛИНИс).

[Подробнее...](#)

26.09.2012 (Среда)

Новые СМИ: меняющийся медийный ландшафт

27-28 сентября 2012 года лаборатория интернет-исследований совместно с зарубежными партнерами проводит конференцию "New media: changing media landscapes". Конференция