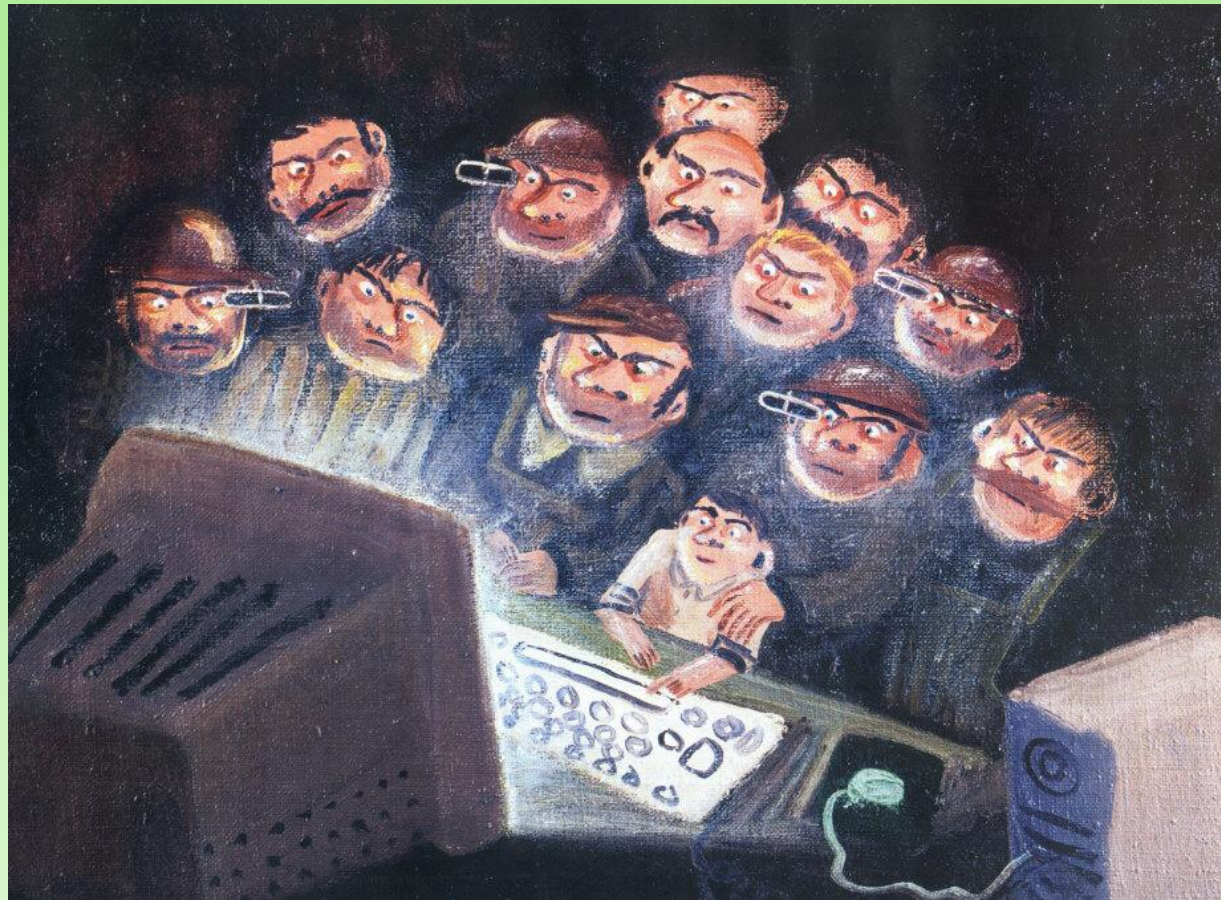


Анализ данных в системе Fitter

Алексей Померанцев

Институт химической физики РАН

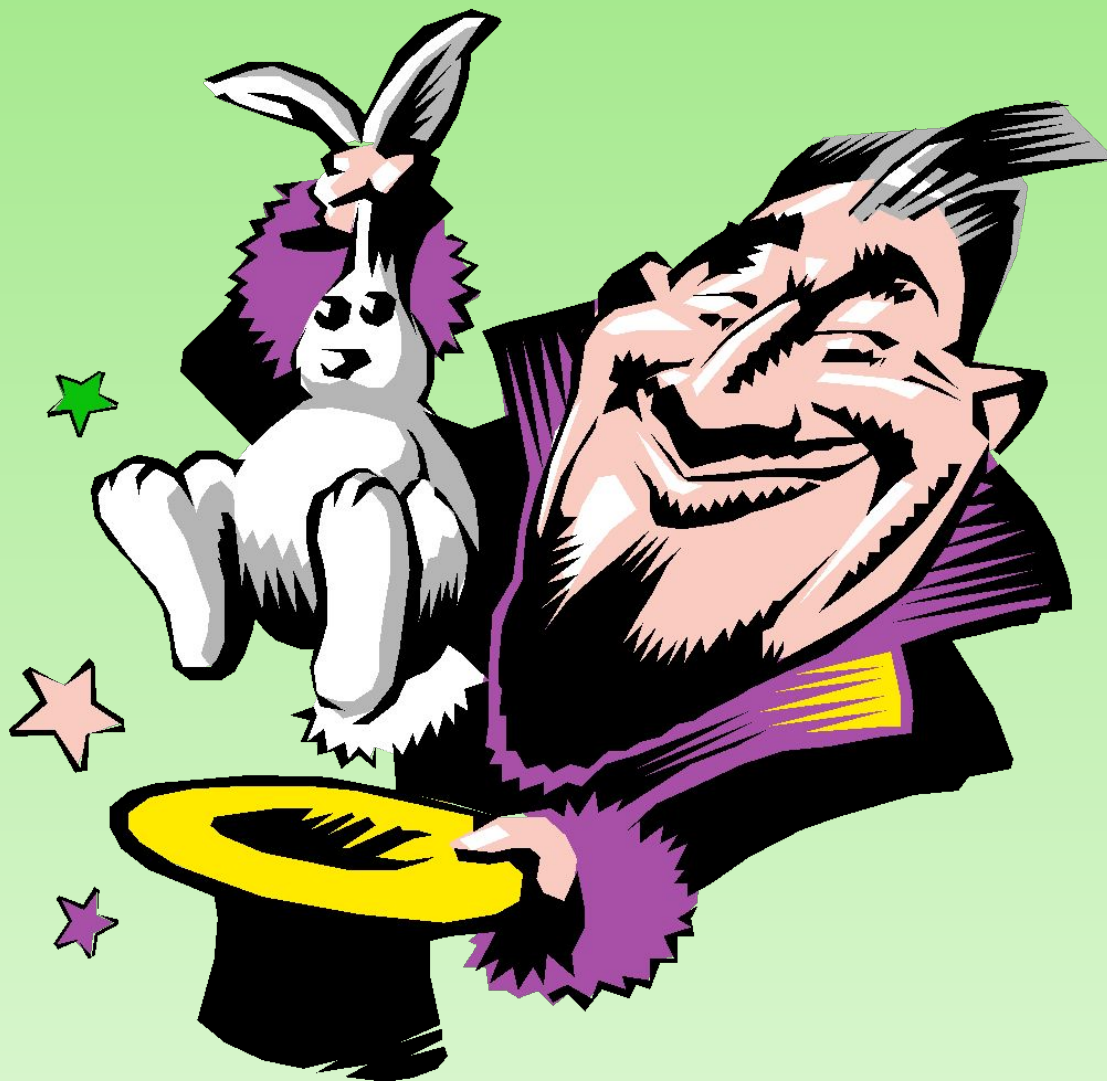


17.02.03

Содержание

1. Введение
2. Данные
3. Модель
4. Оценка параметров
5. Пример VoxVod
6. Прогнозирование
7. Тестирование модели
8. Мультиколлинеарность
9. Прогноз срока службы ПВХ изоляции
10. Выводы

1. Введение



17.02.03

Содержательно (hard) или формально (soft)?

Содержательно

Формально

Откуда

Физика, химия,

Математика, ...

Формула

Разнообразие

Трудно ?

Легко ?

украшает жизнь!

Построение

Придумать модель

Получить данные

Проблемы

Экстраполяция

Интерполяция

Fitter, Matlab, Origin, ...

Unscrambler, Simca, SIC, ...

Назначение

17.02.03

Программы

Система Fitter

Fitter - это

Фактор

Отклики

Весы

Эксперимент

Комментарии

настройка

(Add-In)

к программе

Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		BoxBod Data				Parameters							
3		x	y	w	f	a	213.8941						
4		0		0	0.00	b	0.54735						
5		1	109	1	90.11								
6		2	149	1	142.24								
7		3	149	1	172.41								
8		5	191	1	199.95								
9		7	213	1	209.17								
10		10	224	1	212.91								
11													

'BoxBOD model
 $y = a * [1 - \exp(-b * x)]$
a=?
b=?



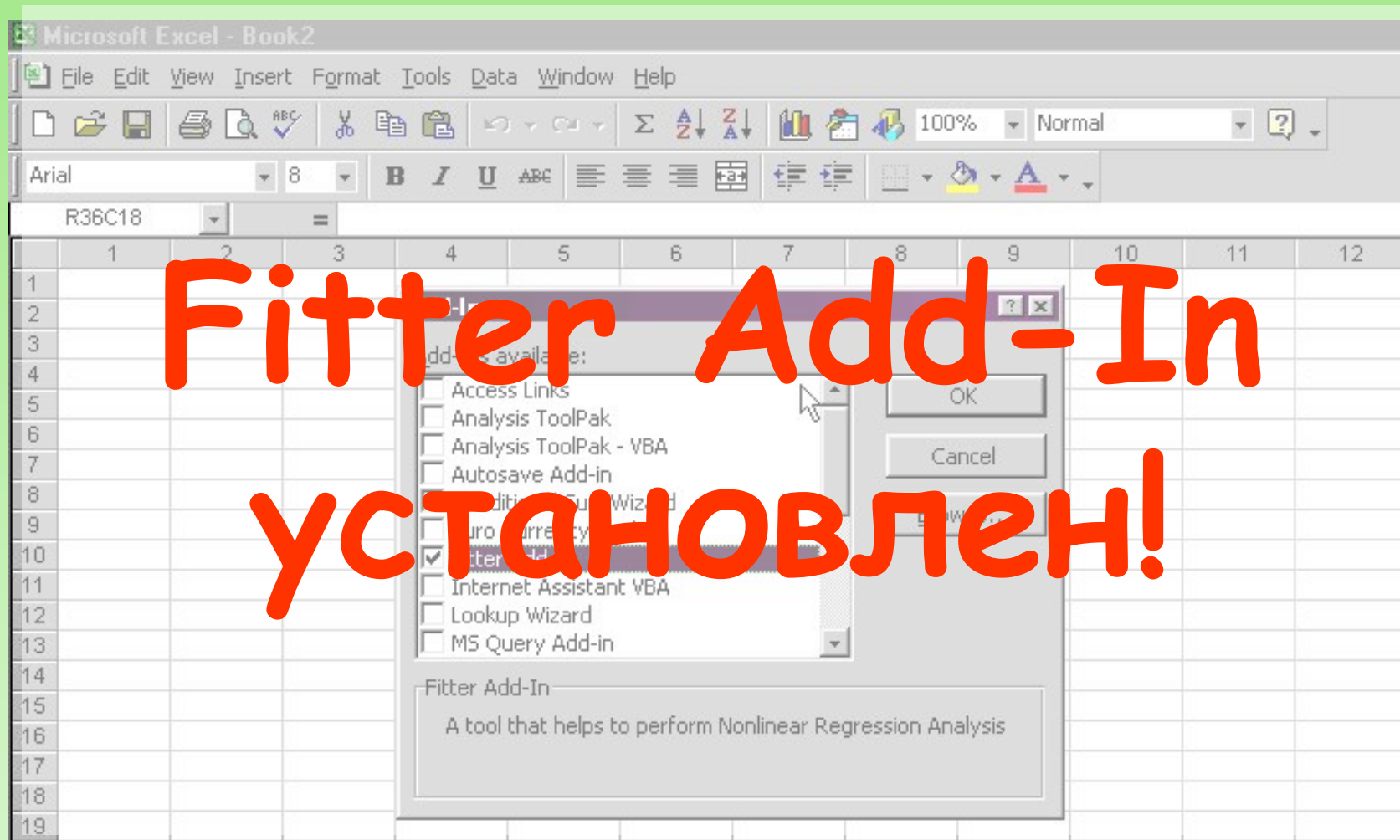
The graph shows a scatter plot of data points (blue 'x' marks) and a fitted curve (red line) on a grid. The x-axis is labeled 'x' and has values 0, 4, 8. The y-axis is labeled 'y' and has values 0, 100, 200. The data points are approximately at (0,0), (1,109), (2,149), (3,149), (5,191), (7,213), (10,224). The fitted curve starts at the origin and levels off towards a value of approximately 213.

Модель

Параметры

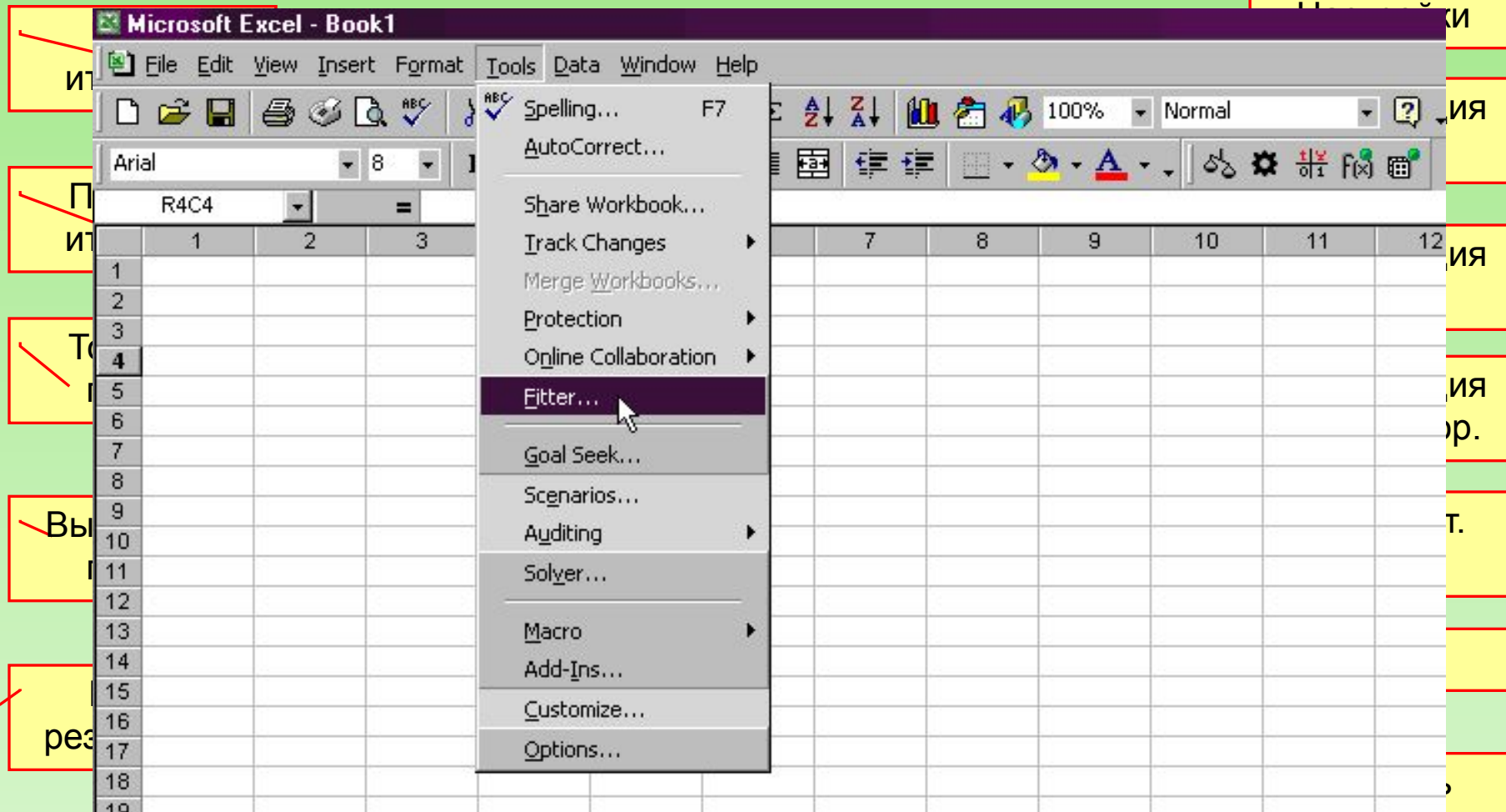
Уравнение

Установка Fitter Add-In



Главный диалог системы Fitter

Поиск



17.02.03

7

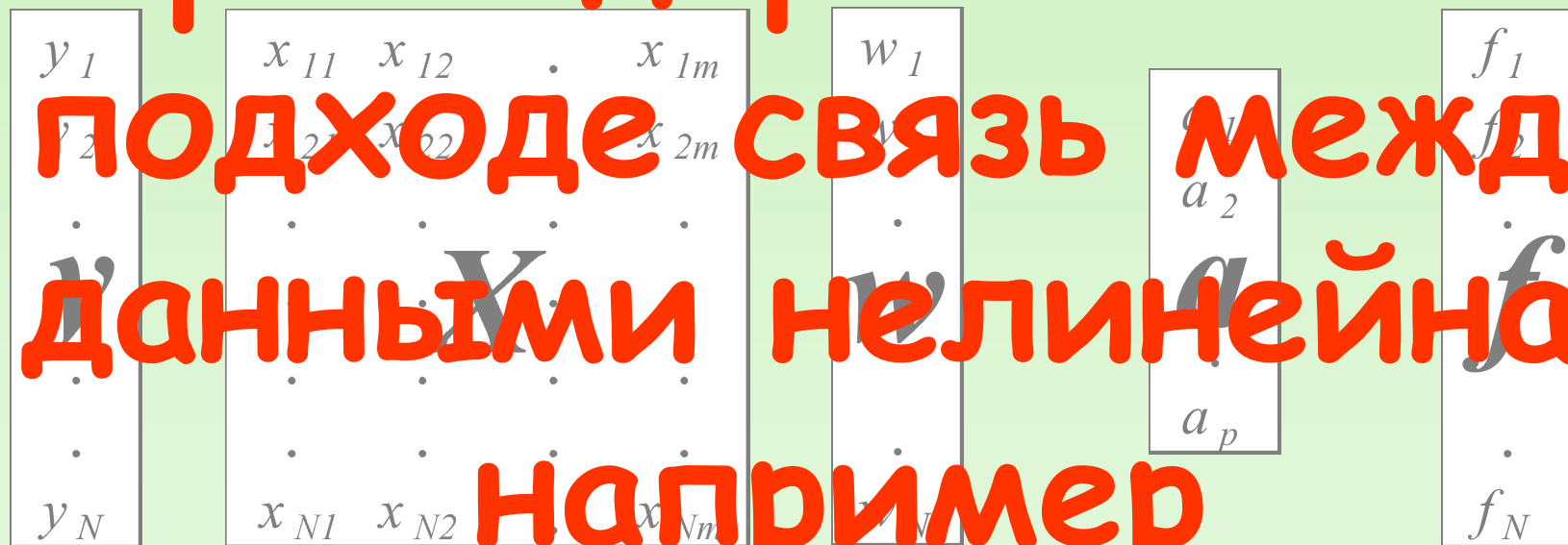
2. Данные



17.02.03

Данные и ошибки

Отклик Препдикторы Вес Параметр Модель



При содержательном подходе связь между данными нелинейная например

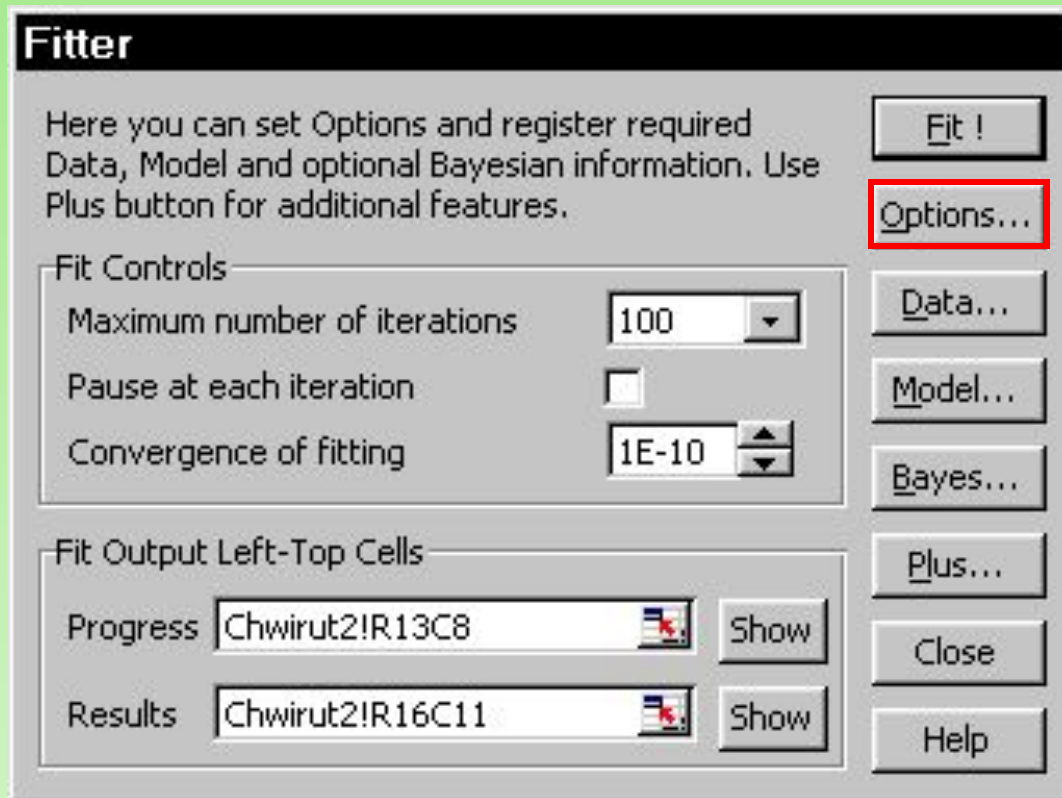
Абсолютная $y = a_1 * \exp(a_2 * x) + a_3 * \exp(a_4 * x)$ Вес и дисперсия

Относительная

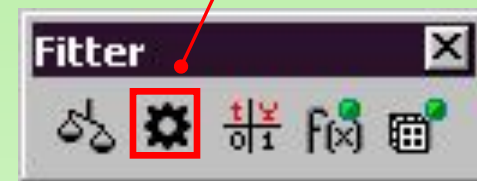
$$y_i = f_i(1 + \varepsilon_i)$$

$$w_i^2 \text{cov}(\varepsilon_i, \varepsilon_i) = \sigma_\varepsilon^2 = \text{Const}$$

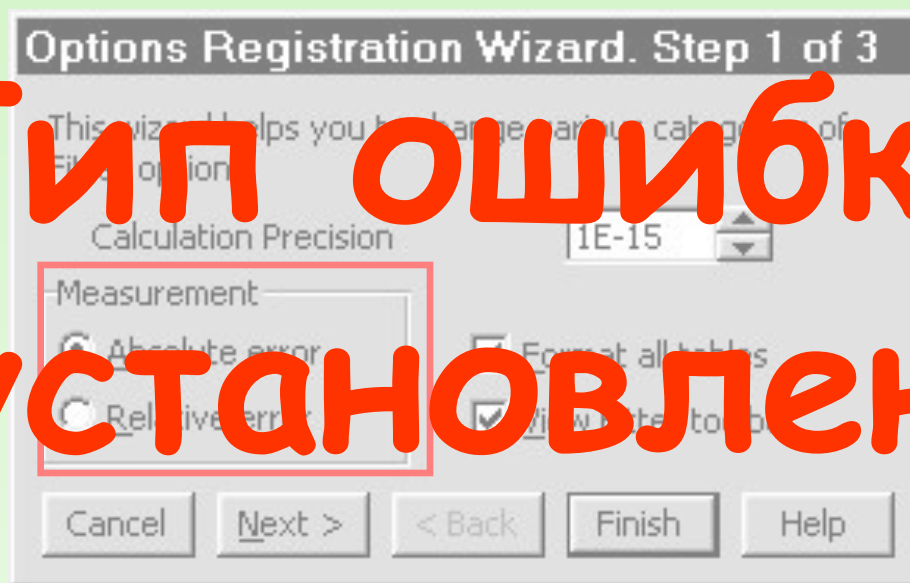
Регистрация типа ошибки



Настройки



Регистрация типа ошибки . Шаг 1



Абсолютная
ошибка

Точность
вычислений

Формат
таблиц

Панель
инструментов

**Тип ошибки
установлен!**

Относительная
ошибка

Продолжить

Закончить

Представление данных в системе Fitter

Левая и правая границы доверительных интервалов

Отклик

Модель

Промежуточная переменная

Дисперсия

Вес

Предикторы

Все данные вводятся и выводятся на страницу Excel

Повторные измерения

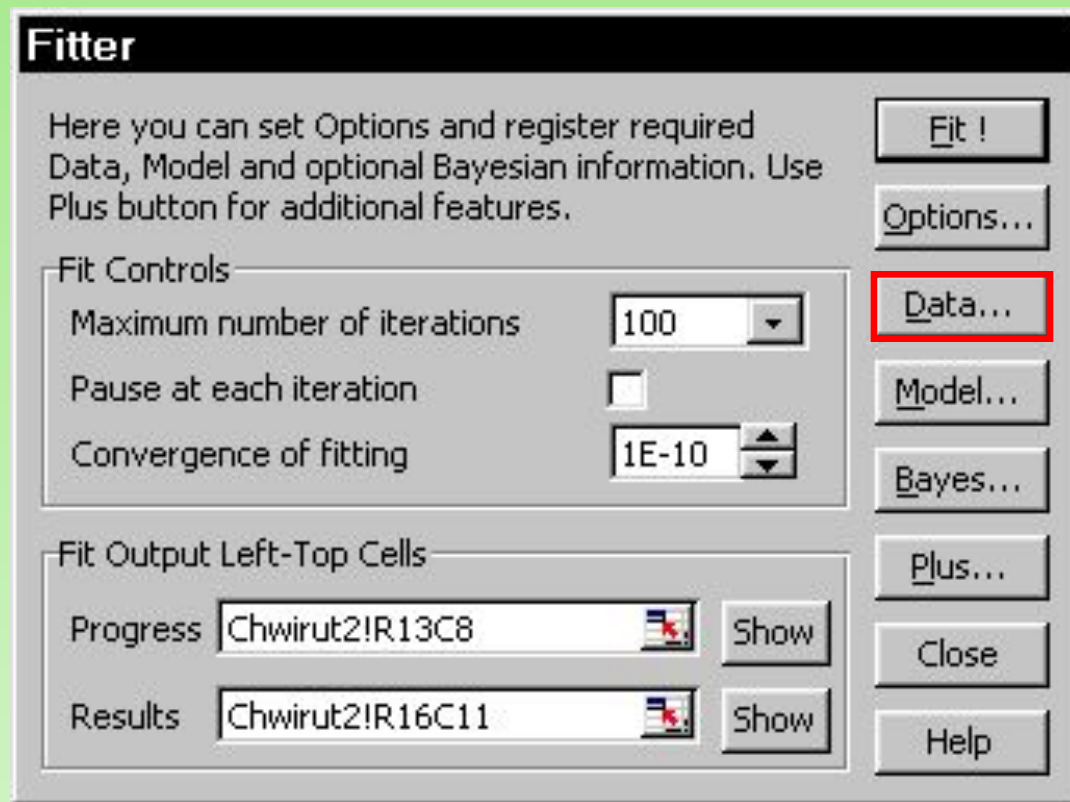
Титул

Заголовок

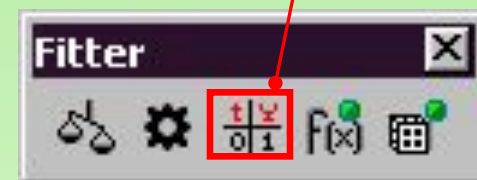
Тело

P	t	CO	W	Fit	ROOH	Left	Right	Var
0.1	0	0.000	0	0.000	0.000	0.000	0.000	
0.1	0	0.011	1	0.011	0.053	0.010	0.011	0.018
0.1	1	0.008	1	0.011	0.053	0.010	0.011	0.018
0.1	1	0.011	1	0.011	0.053	0.010	0.011	0.018
0.1	1	0.011	1	0.011	0.053	0.010	0.011	0.018
0.1	2	0.066	1	0.063	0.161	0.061	0.066	0.002
0.1	2	0.066	1	0.063	0.161	0.061	0.066	0.002
0.1	2	0.067	1	0.063	0.161	0.061	0.066	0.002
0.1	2	0.067	1	0.063	0.161	0.061	0.066	0.002
10	16	9.384	1	9.942	1.416	9.418	10.311	0.179
10	16	9.911	1	9.942	1.416	9.418	10.311	0.179
10	16	10.400	1	9.942	1.416	9.418	10.311	0.179
10	16	9.737	1	9.942	1.416	9.418	10.311	0.179

Регистрация данных



Регистрация
данных



Регистрация данных. Шаг 1

Адрес
таблицы
данных

Data Registration Wizard. Step 1 of 3

This wizard helps you to register Data.

If the selected cells do not contain data you want to register, select other range now.

You may include row containing title information. Then the next row of selection should contain Data names (predictors and response), which appear in the Model.

Range

Cancel < Back Next > Finish Help

Продолжить

Закончить

Регистрация данных. Шаг 2

Data Registration Wizard. Step 2 of 3

This screen lets you set Data layout. You can see how your Data are affected in the preview below

Treat first row as Data title text

Layout

Predictor values Intermediate values
 Response values Confidence limits
 Weight values Variance values
 Fitting values Skip this column

Data Preview

Response	Predictor	Fit	Weight
<i>y</i>	<i>x</i>	<i>f</i>	<i>w</i>
92.9	0.5	81.85575	1
81.5	0.5	81.85575	1
75.8	0.5	81.85575	1
81	0.5	81.85575	1
81.7	0.5	81.85575	1

Buttons: Cancel, < Back, Next >, Finish, Help

Первый ряд
таблицы – это
титул

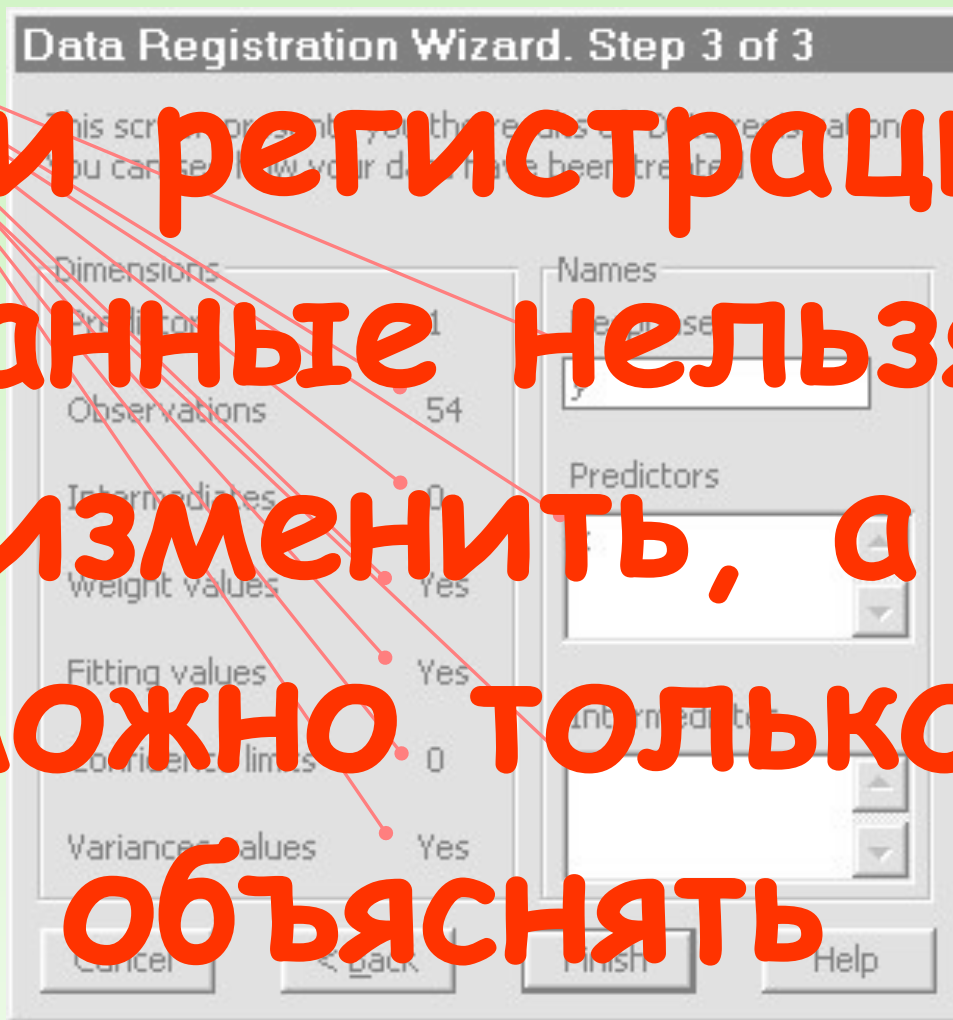
Интерпретация
столбца таблицы
данных

Вернуться
на шаг
назад

Продолжить

Закончить

Регистрация данных. Шаг 3



Результаты
регистрации
таблицы
данных

Вернуться
на шаг
назад

Закончить

**При регистрации
данные нельзя
изменить, а
можно только
объяснить**

3. Модель



$$y = f(x_1, x_2, \dots, a_1, a_2, \dots)$$

отклик предикторы параметры

известны неизвестны

данных и модели -

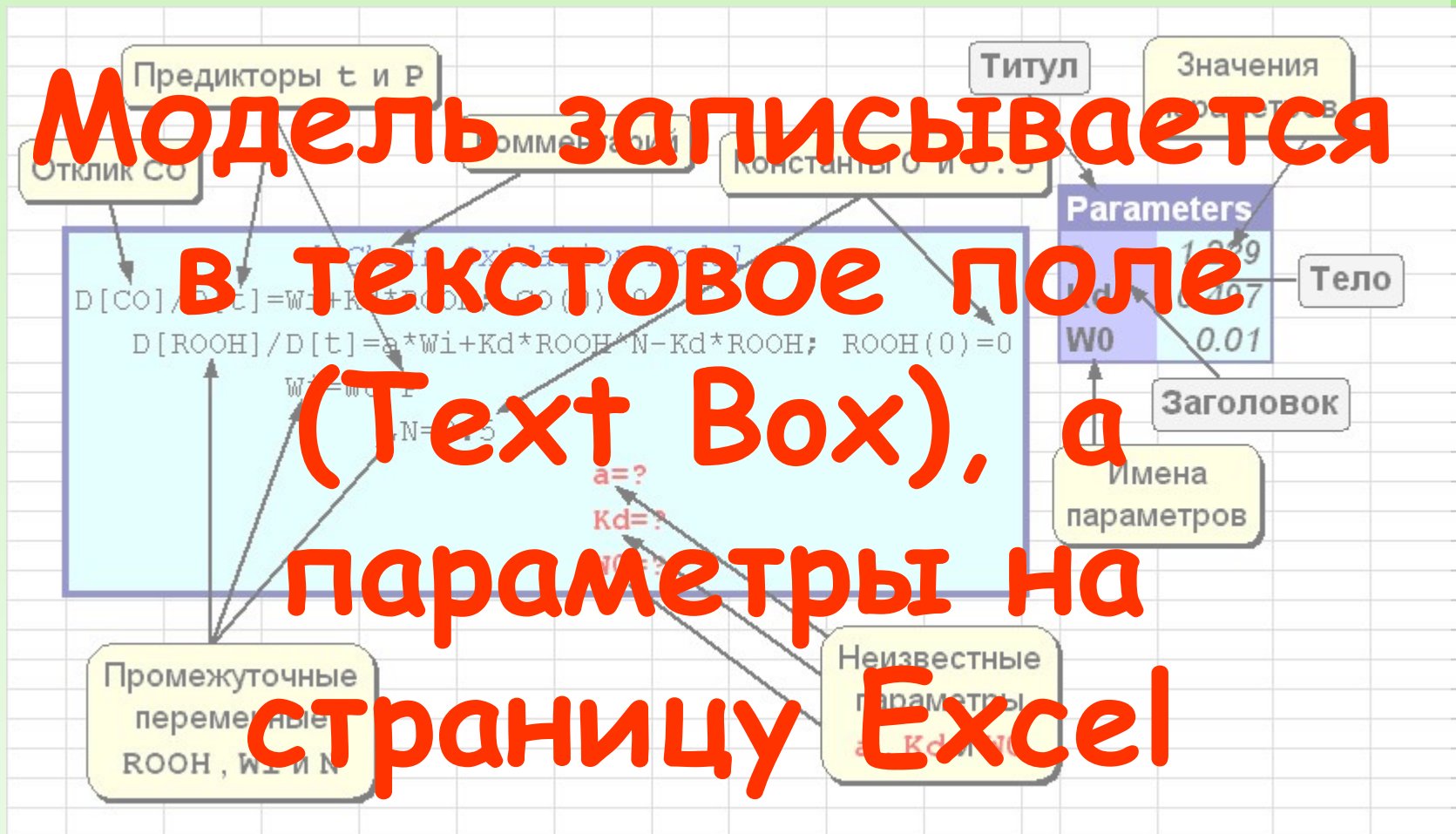


**это сумма
квадратов
отклонений**

Мера согласия

$$S(a) = \sum_{i=1}^N (y_i - f_i)^2$$

Представление модели в системе Fitter



Модель $f(x)$

Различные

Явная

Неявная

Дифф. уравне

```
' Explicit
y=a+(b-a)*e
```

a=

b=

c=

17.02.03

```
        'Цикл "увлажнение-сушка"
M=Sor*hev(t1-t)+Des*[hev(t-t1)+imp(t-t1)]
'Кинетика "увлажнения"
Sor=Sor1*hev(USESor1)+Sor2*[hev(-USESor1)+imp(-USESor1)]
'Кинетика "сушки"
Des=Des1*hev(USEDes1)+Des2*[hev(-USEDes1)+imp(-USEDes1)]
'Условие применимости асимптотик
USESor1=Sor2-Sor1
USEDes1=Des1-Des2
'константы и промежуточные величины
t3=(t-t1)*hev(t-t1)
t4=t*hev(t1-t)+t1*[hev(t-t1)+imp(t-t1)]
P2=PI*PI
P12=(PI)^(-0.5)
R=r*(M1-M0)*exp(-r*t4)
K=M1+(M0-M1)*exp(-r*t4)
V0=M0-C0
V1=M1-C0
'асимптотика десорбции при 0<t<tau
Sor1=C0*4*PI*(d*t)^0.5*[M0-C0+(M1-M0)*beta]
bet=1-exp(-R)
x=t
z=1+1/2*(a1*x+a2*x*x+a3*x*x*x)/(1+1/2*(a1*x+a2*x*x+a3*x*x*x))
a1=0.666653925002
a2=0.0121051017749
a3=0.0099225322428
b1=0.0848006232519
b2=0.0246634591223
b3=0.0017549947958
'кинетика сорбции при t1<t<tau
Sor2=K-fS1
U1=(U01/U0+U11/n1+U22/n2+U33/n3+U44/n4)*S1
n0=P2
U01=[(V0*n0*d-V1*r)*exp(-n0*d*t4)+R]/(n0*d-r)
n1=P2*9
U11=[(V0*n1*d-V1*r)*exp(-n1*d*t4)+R]/(n1*d-r)
n2=P2*25
U21=[(V0*n2*d-V1*r)*exp(-n2*d*t4)+R]/(n2*d-r)
n3=P2*49
U31=[(V0*n3*d-V1*r)*exp(-n3*d*t4)+R]/(n3*d-r)
n4=P2*81
U41=[(V0*n4*d-V1*r)*exp(-n4*d*t4)+R]/(n4*d-r)
'асимптотика десорбции при t1<t<tau
Des1=K*[1-4*PI2*(d*t3)^0.5]-8*S1
'кинетика десорбции при t1+tau<t
Des2=8*S2
S2=U02/n0+U12/n1+U22/n2+U32/n3+U42/n4
U02=(K-U01)*exp(-n0*d*t3)
U12=(K-U11)*exp(-n1*d*t3)
U22=(K-U21)*exp(-n2*d*t3)
U32=(K-U31)*exp(-n3*d*t3)
U42=(K-U41)*exp(-n4*d*t3)
'неизвестные параметры
d=?
M0=?
M1=?
C0=?
r=?
t1=?
```

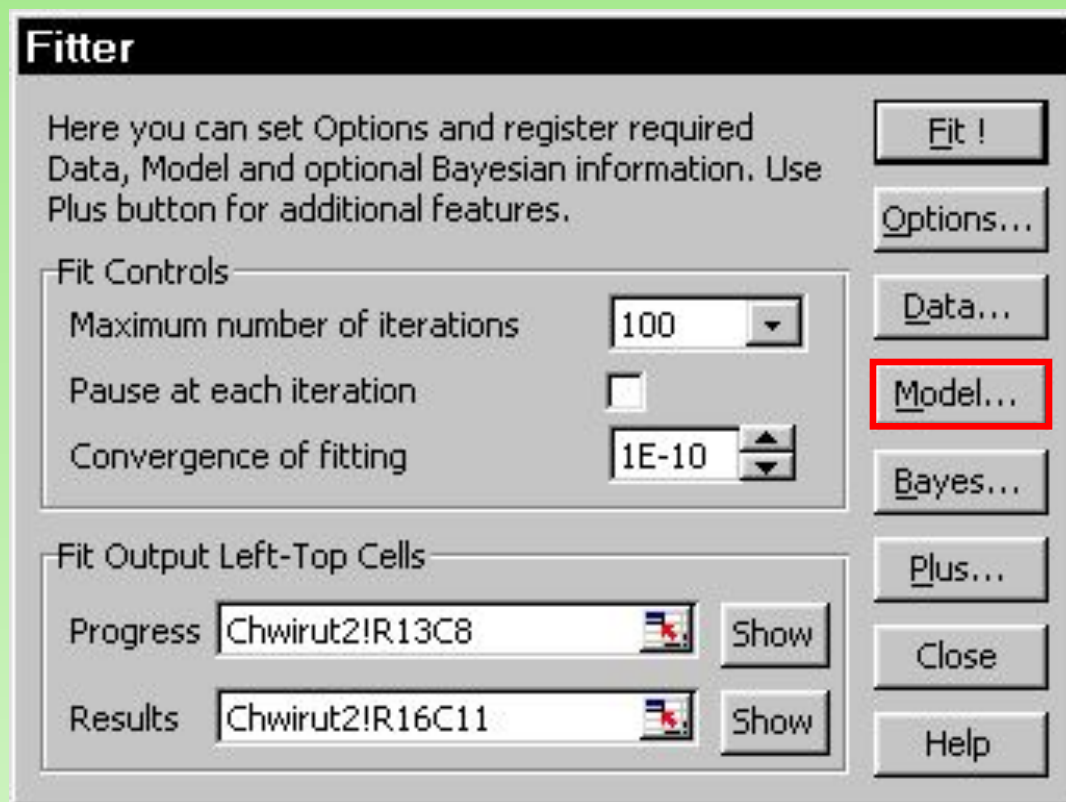
ели

b

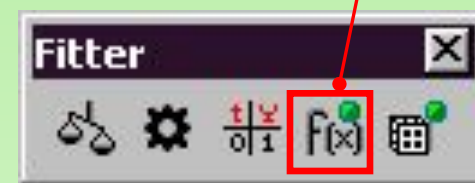
(0)=b

Очень сложная модель!

Регистрация модели



Регистрация модели



Регистрация модели. Шаг 1

Model Registration Wizard. Step 1 of 3

Model

This wizard helps you to register Model and Parameters
If the selected text box does not contain Model you want to register, select other text box from the drop-down list now.

Text Box

Parameters

Select Left-Top cell for Parameters range (names are in the left column, values are in the right column). You can leave the box blank. If you fix it, names and values will be placed here when the Model has been registered.

You may select a cell containing Parameter title text

Cell

Выбор
текстового поля
с моделью

Адрес первой
клетки
таблицы
параметров

Продолжить

Закончить

Регистрация модели. Шаг 2

~~Вид модели~~

✓ Переменные модели

~~Продолжить~~

~~Закончить~~

Вернуться на шаг назад

Model Registration Wizard. Step 2 of 3

This screen lets you preview the selected Model and check variables used in it. Select a name from the lists below and see how it appears in the Model preview.

Response:

Predictors:

Parameters:

Intermediates:

Constants:

Model preview

```
' Chwirut2
y = exp(-b1*x) / (b2+b3*x)
      b1=?
      b2=?
      b3=?
```

Cancel < Back Finish Next > Help

17.02.03

Регистрация модели. Шаг 3

Первый ряд таблицы параметров - это титул

Начальные значения параметров

Вернуться на шаг назад

Закончить

При регистрации модели нельзя изменить, а можно только объяснить

4. Оценка параметров



17.02.03

Целевая функция $Q(a)$

Целевая функция

Сумма квадратов

$$Q = \sum_{i=1}^N w_i^2 (y_i - f_i)^2 g_i^2$$

Байесовский член

$$B(a) = s_0^2 \left[N + (a - b)^t H (a - b) \right]$$

Целевая функция

$$Q(a) = \text{sum} + B(a)$$

и, может быть,

что-то еще...

Градиентный метод поиска оценок

$$Q(a) \approx Q(a_n) + b_n^t (a - a_n) + \frac{1}{2} (a - a_n)^t A (a - a_n)$$

$$a_{n+1} = a_n + A^+ b_n$$

$$A^+ A = I$$

Проблемы поиска

Исходная точка a_0

$$\det(A) \approx 0$$

Локальный минимум

**Матрица A -
ЭТО КЛЮЧ
к поиску!**



Точность оценивания (статистики)

Матрица ковариаций

$$\text{cov}(\hat{a}, \hat{a}) = C = s^2 A^{-1}$$

Средне-квадратичные отклонения

$$\text{dev}(\hat{a}_\alpha) = \sqrt{C_{\alpha\alpha}}$$

Матрица корреляции

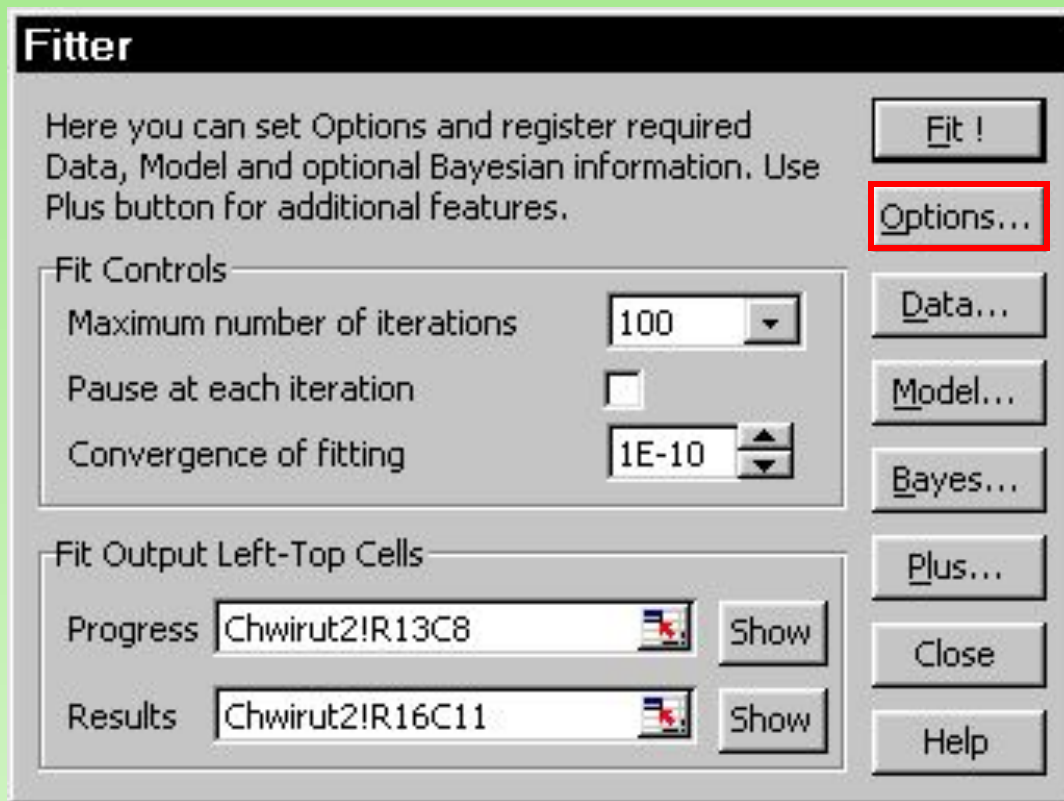
$$\text{cor}(\hat{a}_\alpha, \hat{a}_\beta) = \frac{C_{\alpha\beta}}{\sqrt{C_{\alpha\alpha} C_{\beta\beta}}}$$

F-матрица (Фишера)

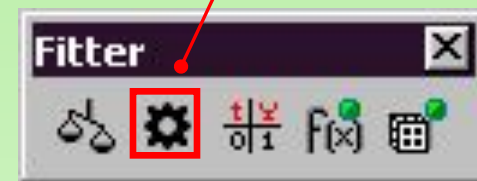
$$F = s^{-2} A = C^{-1}$$
$$s^2 = \frac{S(\hat{a})}{N_f} \quad N_f = N - p$$

Матрица A - это мера качества оценивания!

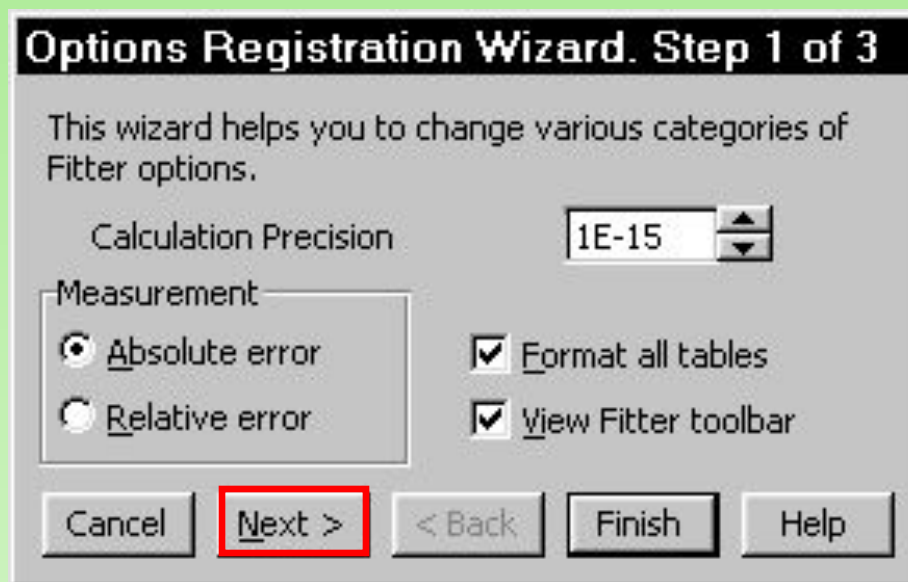
Регистрация настроек статистик



Настройки



Регистрация настроек . Шаг 1



Продолжить

Регистрация настроек статистик. Шаг 2

Общие данные

СКО

Ковариации

Корреляции

F-матрица

Итоги поиска

Options Registration Wizard. Step 2 of 3

This screen lets you select what results should be calculated and displayed. Here you can set tests significance level α and confidence probability P of prediction. If you select bootstrap method of prediction then set number of simulations N .

Statistics	Tests	Prediction
<input checked="" type="checkbox"/> General	$\alpha =$ 0.050	$P =$ 0.950
<input checked="" type="checkbox"/> Deviation	<input checked="" type="checkbox"/> Outliers	<input type="radio"/> Linearization
<input type="checkbox"/> Covariance	<input checked="" type="checkbox"/> Lack of Fit	<input checked="" type="radio"/> Bootstrap with
<input checked="" type="checkbox"/> Correlation	<input checked="" type="checkbox"/> Variance	$N =$ 1000
<input type="checkbox"/> F-matrix	<input checked="" type="checkbox"/> Series	<input type="checkbox"/> Right-side
<input checked="" type="checkbox"/> Fitting results	<input checked="" type="checkbox"/> Nonlinearity	<input type="checkbox"/> Left-side

Buttons: Cancel, Next >, < Back, Finish, All, Help

Закончить

5. Пример VoxBod (NIST)

Данные

x	y	f
1	109	90.11
2	149	142.24
3	149	142.24
5	191	199.95
7	211	209.17
10	221	212.11

Модель

$$y = a_1(1 - e^{-a_2x})$$

Посмотрим, как работает Fitter

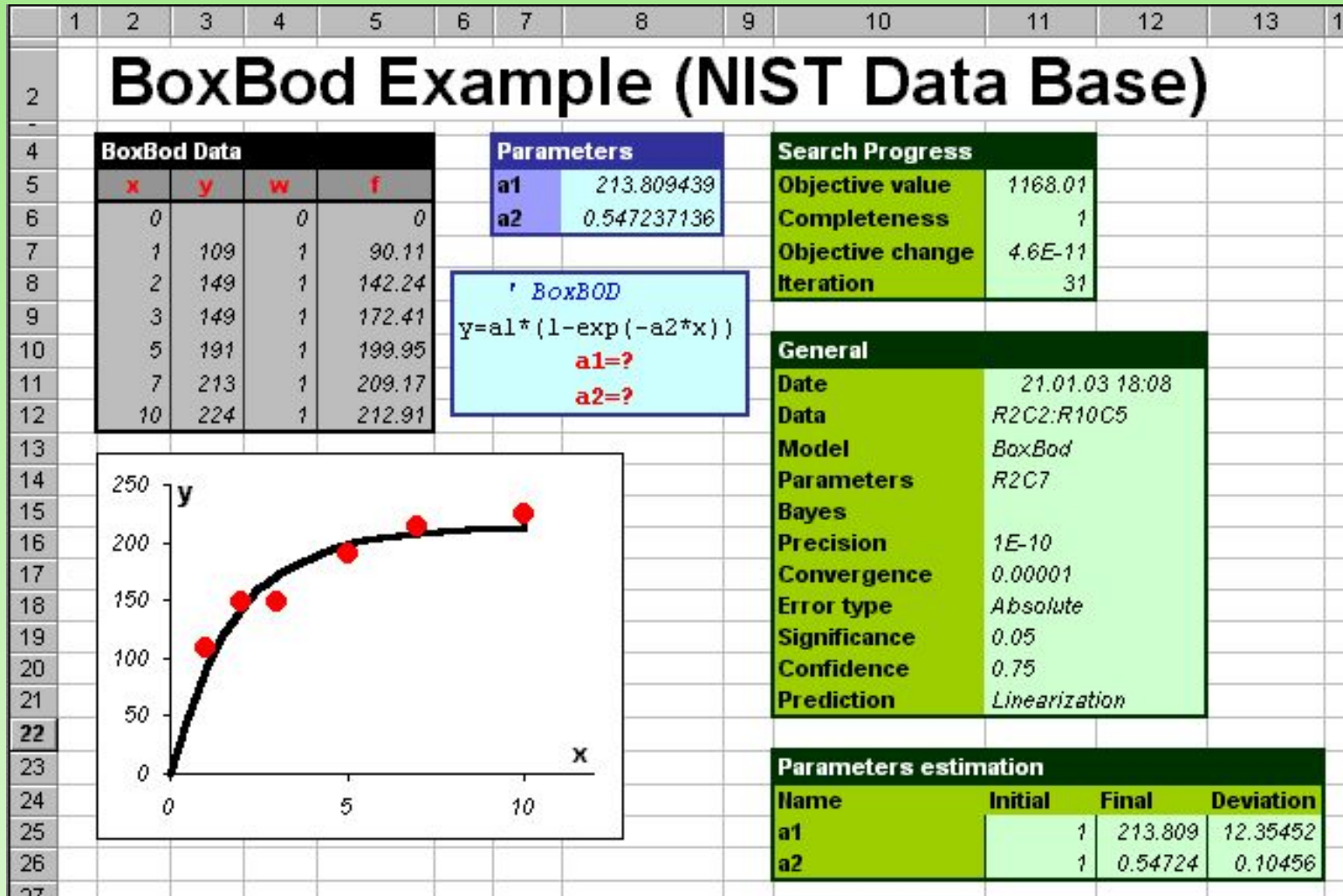
Параметры

	Оценки	СКО
a ₁	213.809	12.3545
a ₂	0.517237	0.104560

Априорная (байесовская) информация

Bayesian Information			
Name	Value	Matrix	Exclude

Переключаемся в Excel



6. Прогноз



17.02.03

Достоверный прогноз

Оценка отклика

$$\hat{y} = f(x, \hat{a})$$

прогноз

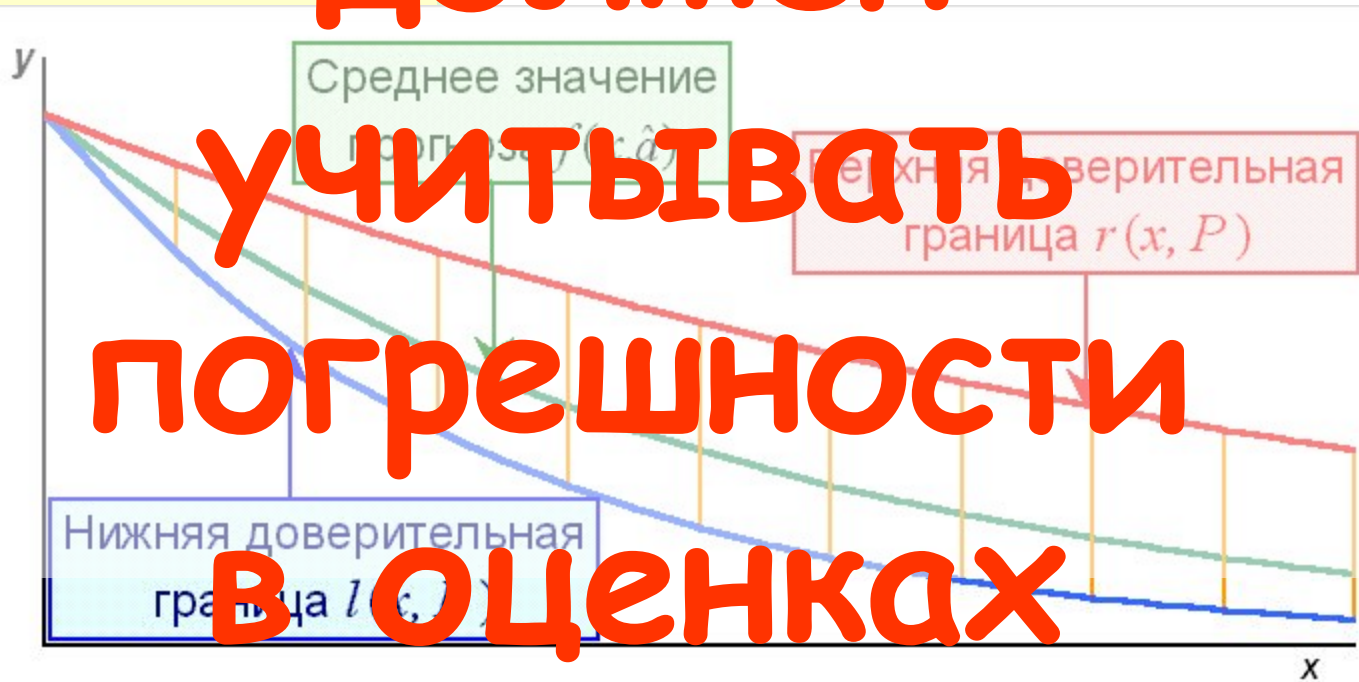
Доверительные границы

$$P\{l(x, P) < f(x, a) < r(x, P)\} \geq P$$

должен

$$r(x, P) = f(x, \hat{a}) + g(P)\sqrt{v^t C v}$$

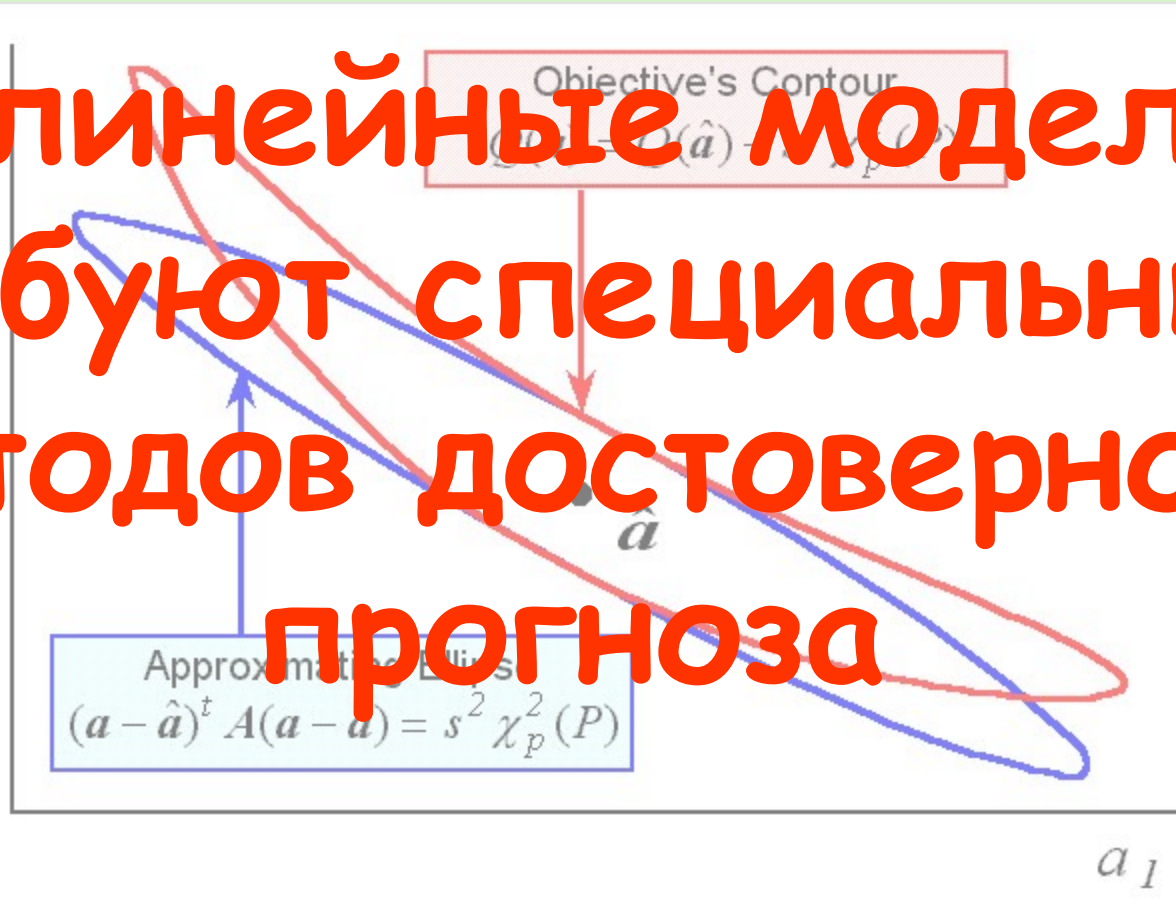
Лин



Имитационное прогнозирование

$$a^* \sim N(\hat{a}, C) \rightarrow a_1^*, \dots, a_M^* \rightarrow f_1^*, \dots, f_M^* \rightarrow r(P, x)$$

**Нелинейные модели
требуют специальных
методов достоверного
прогноза**

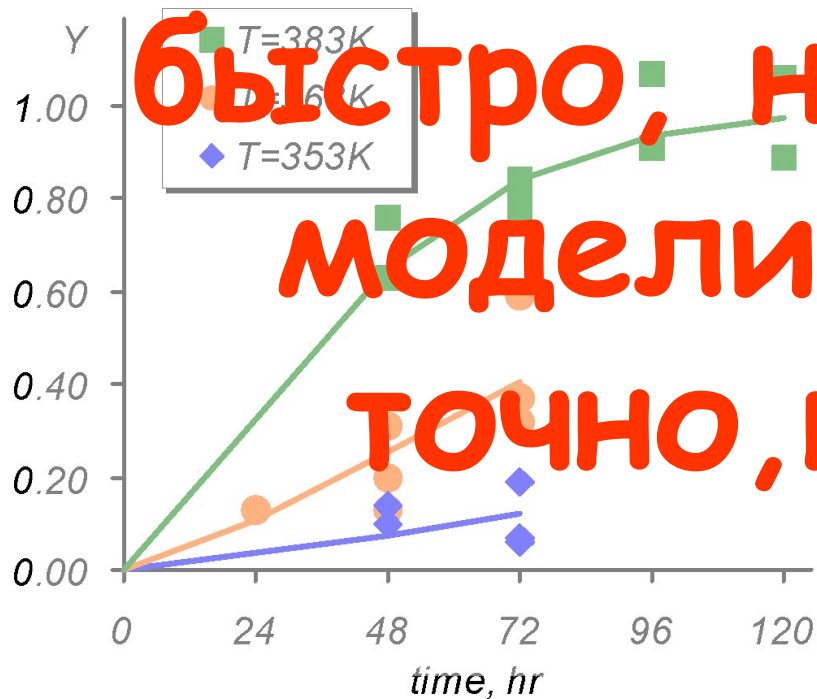


Пример разных методов прогноза

Модель старения
резины

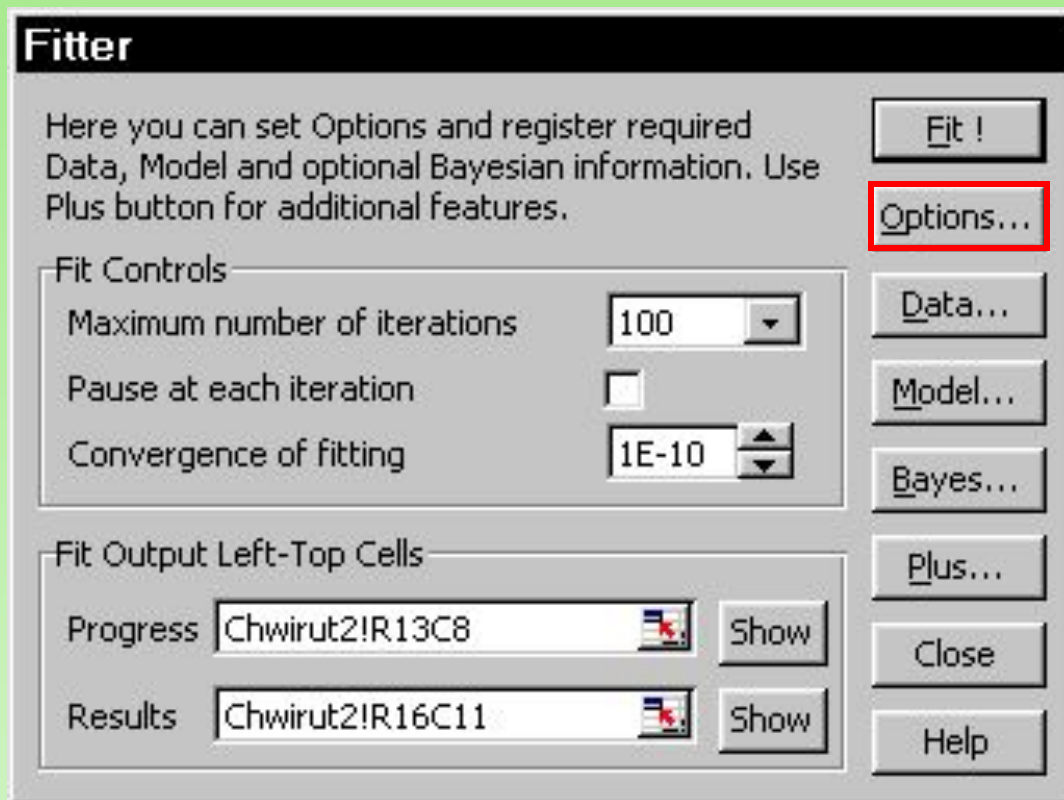
$$Y = 1 - e^{-(kt)^a}, \quad k = e^{k_0 - \frac{E}{RT}}$$

Результат ускоренных испытаний — Верхняя граница —

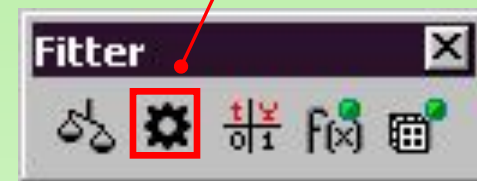


быстро, но не точно,
моделирование —
точно, но долго

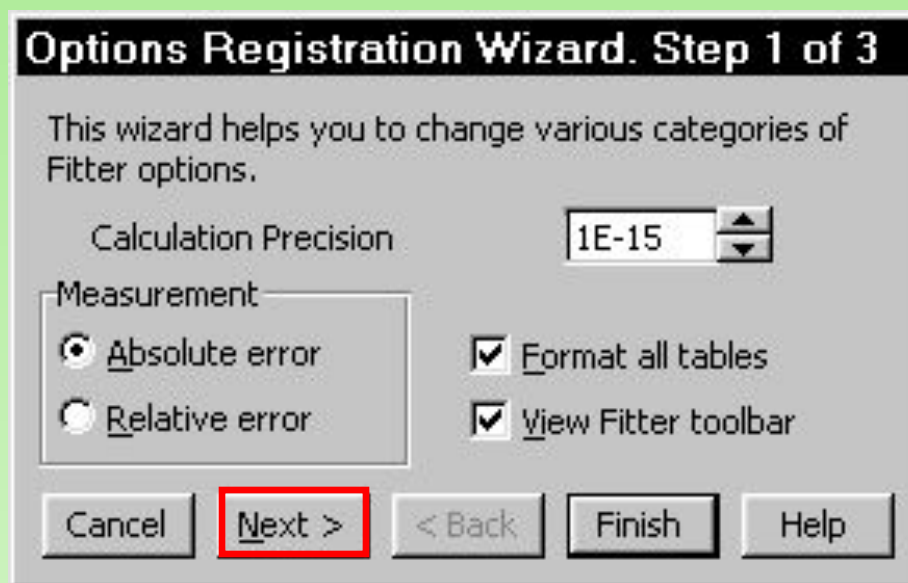
Регистрация настроек прогноза



Настройки



Регистрация настроек . Шаг 1



Продолжить

Регистрация настроек прогноза. Шаг 2

Options Registration Wizard. Step 2 of 3

This screen lets you select what results should be calculated and displayed. Here you can set tests significance level α and confidence probability P of prediction. If you select bootstrap method of prediction then set number of simulations N .

Statistics	Tests	Prediction
<input checked="" type="checkbox"/> General	$\alpha =$ 0.050	$P =$ 0.950
<input checked="" type="checkbox"/> Deviation	<input checked="" type="checkbox"/> Outliers	<input type="radio"/> Linearization
<input type="checkbox"/> Covariance	<input checked="" type="checkbox"/> Lack of Fit	<input checked="" type="radio"/> Bootstrap with
<input checked="" type="checkbox"/> Correlation	<input checked="" type="checkbox"/> Variance	$N =$ 1000
<input type="checkbox"/> F-matrix	<input checked="" type="checkbox"/> Series	<input type="checkbox"/> Right-side
<input checked="" type="checkbox"/> Fitting results	<input checked="" type="checkbox"/> Nonlinearity	<input type="checkbox"/> Left-side

Buttons: Cancel, Next >, < Back, Finish, All, Help

Достоверность прогноза

Метод -
линеаризация

Метод -
моделирование

Число повторов

Правая
граница

Левая
граница

Закончить

7. Тестирование модели



17.02.03

Проверка гипотез

Статистика ξ сравнивается с критическим значением $t(\alpha)$

**Проверка не доказывает
верности модели!
Она только показывает,
что гипотеза
противоречит или не
противоречит данным!**

Из
эксперимента

Из
теории

Тестовая статистика ξ

Уровень значимости (α)

Адекватность и гомоскедастичность

Эти тесты используют выборочные дисперсии и они не могут быть проверены без реплик!



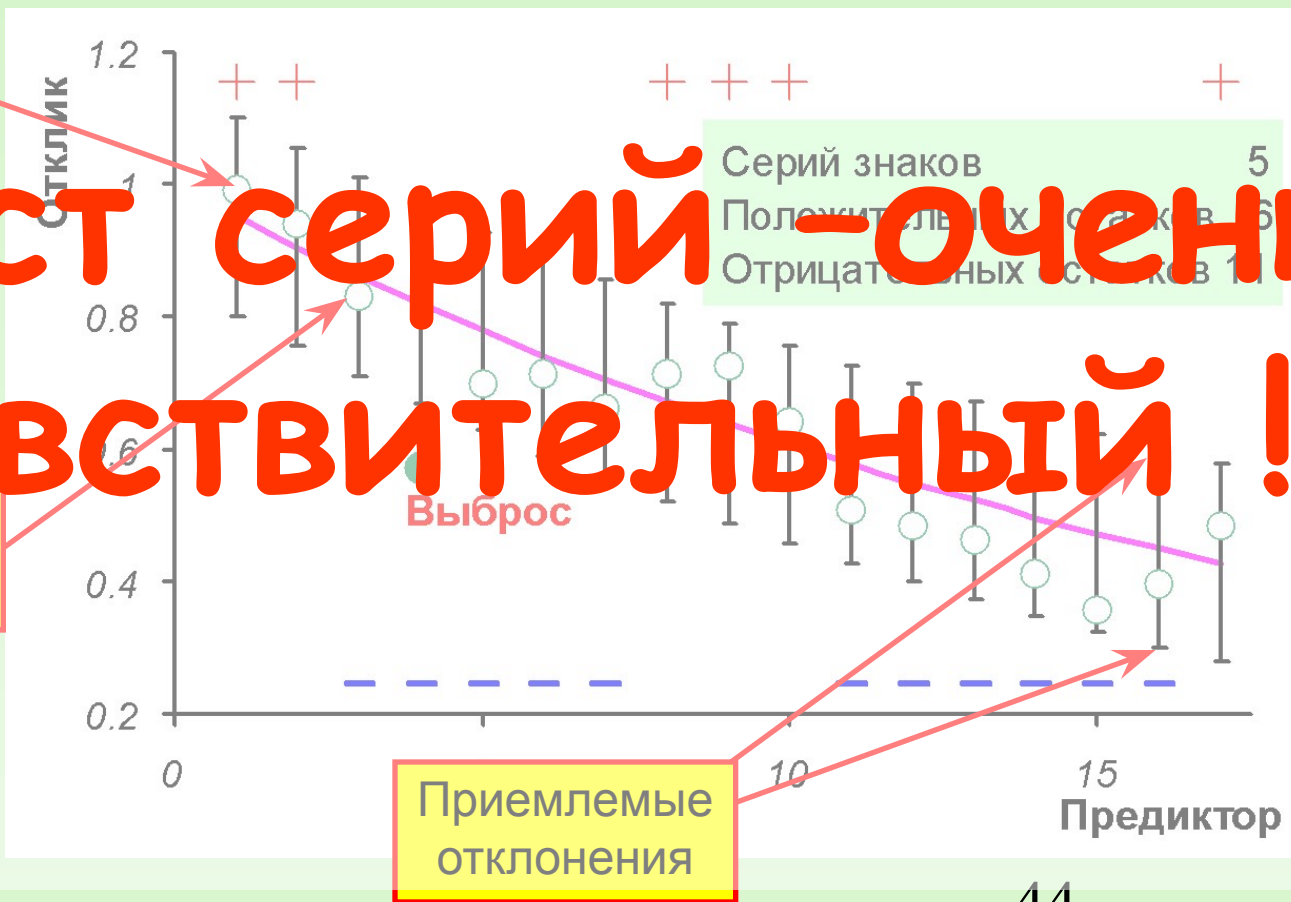
Выбросы и тест серий

Эти тесты используют только остатки, поэтому они могут быть проверены и без реплик

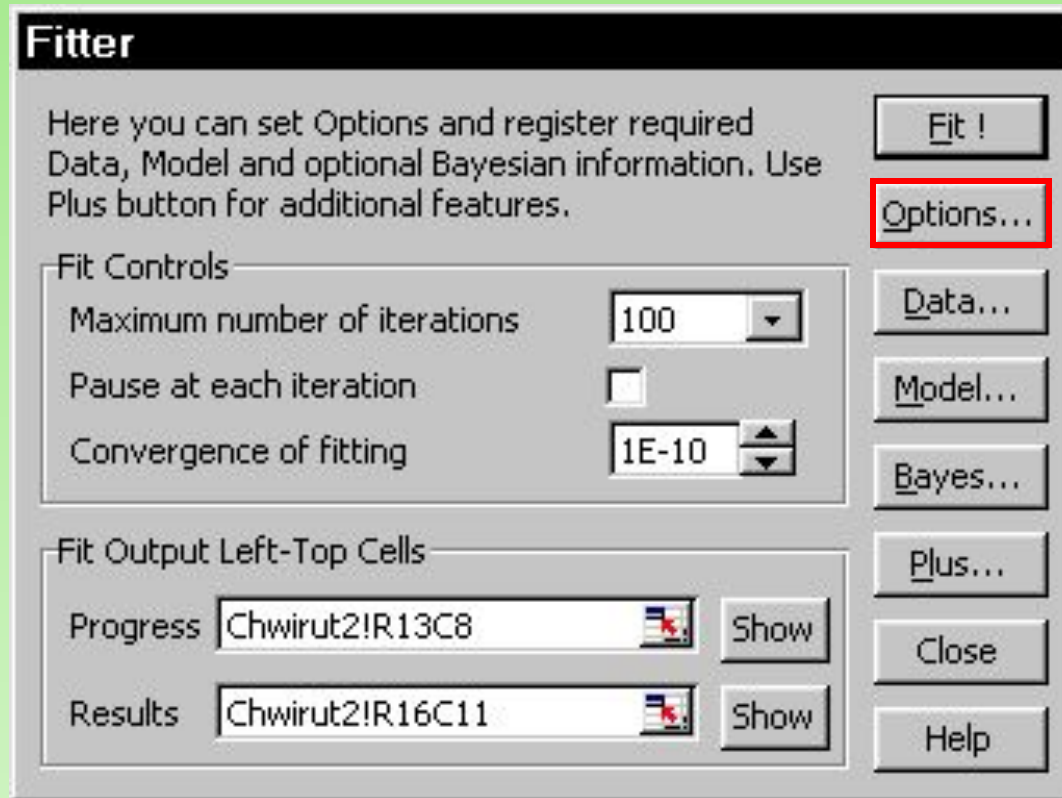
Положительные остатки

Отрицательные остатки

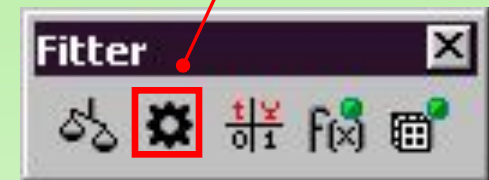
Тест серий - очень чувствительный!



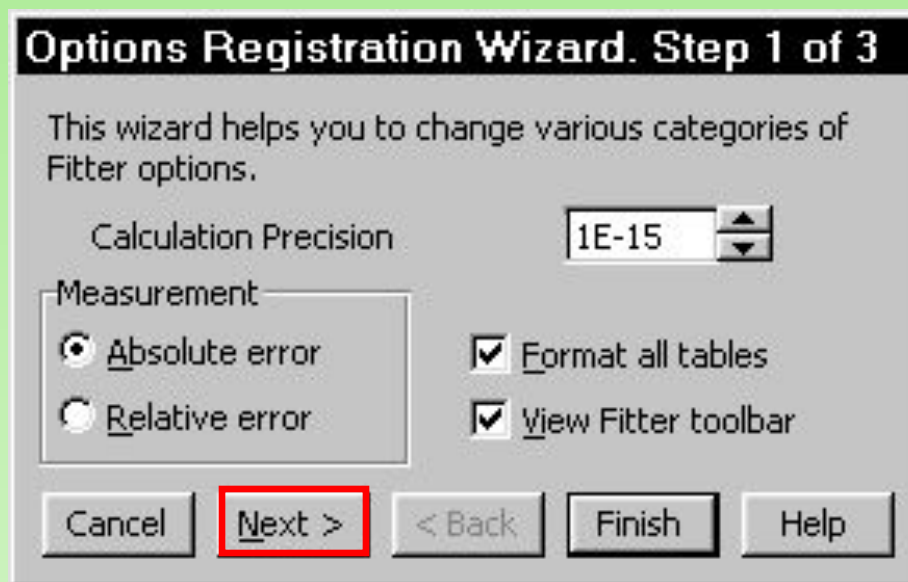
Регистрация настроек проверки гипотез



Настройки



Регистрация настроек . Шаг 1



Продолжить

Регистрация настроек гипотез. Шаг 2

Options Registration Wizard. Step 2 of 3

This screen lets you select what results should be calculated and displayed. Here you can set tests significance level α and confidence probability P of prediction. If you select bootstrap method of prediction then set number of simulations N .

Statistics	Tests	Prediction
<input checked="" type="checkbox"/> General	$\alpha =$ 0.050	P = 0.950
<input checked="" type="checkbox"/> Deviation	<input checked="" type="checkbox"/> Outliers	<input type="radio"/> Linearization
<input type="checkbox"/> Covariance	<input checked="" type="checkbox"/> Lack of Fit	<input checked="" type="radio"/> Bootstrap with
<input checked="" type="checkbox"/> Correlation	<input checked="" type="checkbox"/> Variance	N = 1000
<input type="checkbox"/> F-matrix	<input checked="" type="checkbox"/> Series	<input type="checkbox"/> Right-side
<input checked="" type="checkbox"/> Fitting results	<input checked="" type="checkbox"/> Nonlinearity	<input type="checkbox"/> Left-side

Buttons: Cancel, Next >, < Back, Finish, All, Help

Уровень
значимости

Тест выбросов

Тест
адекватности

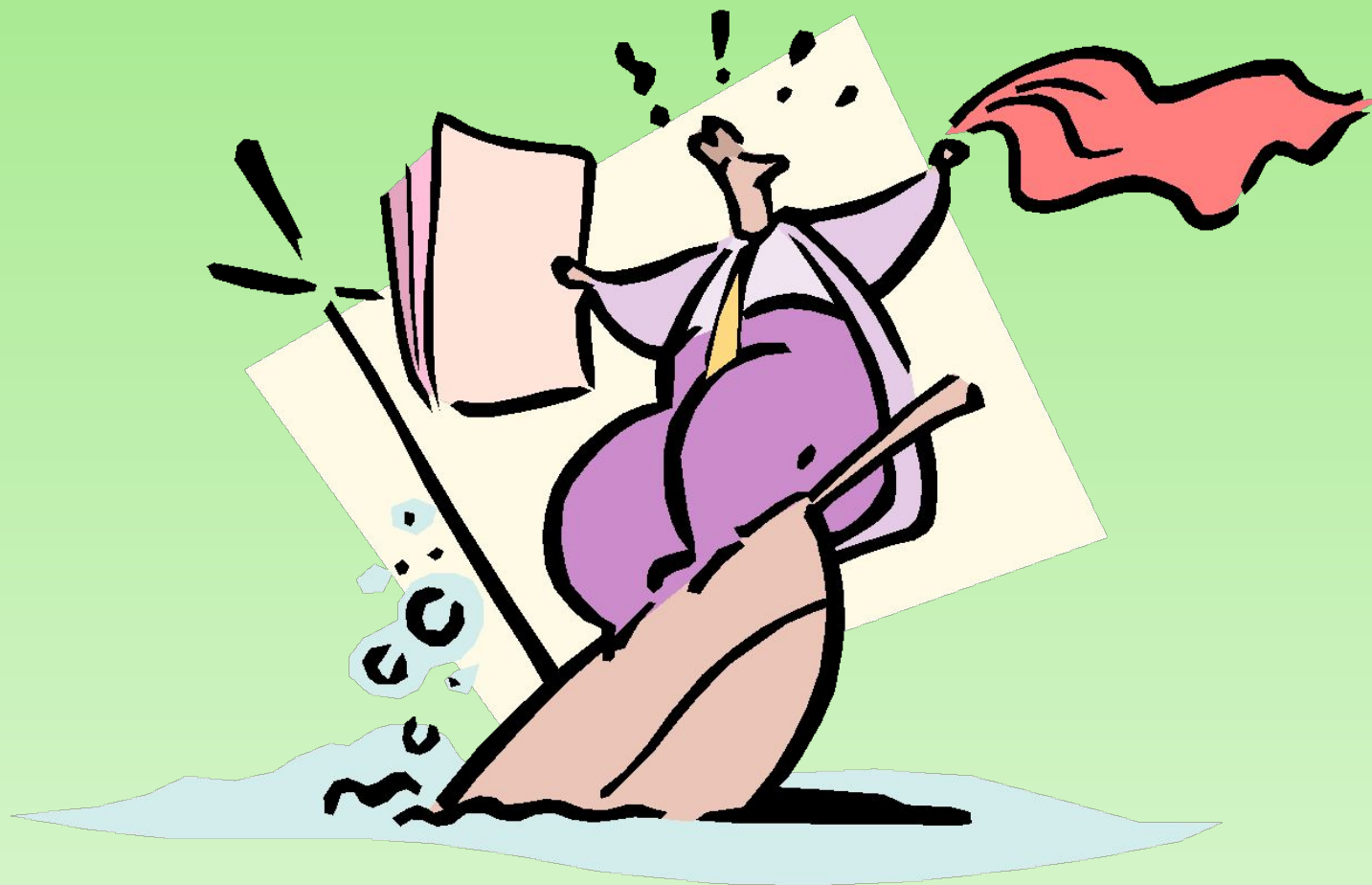
Тест дисперсий

Тест серий

Коэффициент
нелинейности

Закончить

8. Мультиколлинеарность



Что такое мультиколлинеарность

Мультиколлинеарность – это вырождение матрицы A

Разброс собственных значений:

$$N(A) = \log_{10} \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right)$$

это мера вырождения матрицы A

Целевая функция $Q(a)$



**Мультиколлинеарность
– это наш главный враг!**

$$N(A) = 7$$

Причины мультиколлинеарности

“Строгая” мультиколлинеарность

$$y = a_1 a_2 x$$

“Нестрогая” мультиколлинеарность

$$y = a_1 (1 - e^{-a_2 x}) \approx a_1 a_2 x$$



Пример: закон Аррениуса

Обычная форма записи

$$k \exp\left(-\frac{A}{RT}\right)$$

Улучшенная форма записи

$$\exp(a_1 - a_2 X)$$

Простые преобразования,

Шкалирование и центрирование

дающие БОЛЬШОЙ

Перепараметризации

ВЫИГРЫШ!

1

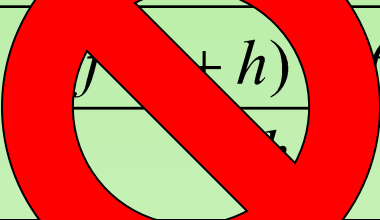
$$N(A) = 20$$

$$N(A) = 2$$

Вычисление производных и точность

$N(A)$	A^{-1}	$y=f(a,x)$	$0=f(y,a,x)$	$dy/dx=f(y,a,x)$
10	$10+2=12$	$12+0=12$	$12+2=14$	$14+2=16$
6	$6+2=8$	$8+0=8$	$8+2=10$	$10+2=12$

1) Численное вычисление разностных производных

$$\frac{\partial f(y, a+h)}{\partial a} - \frac{\partial f(y, a)}{\partial a}$$


2) Автовычисление аналитических производных

$$f = \exp(-a * t)$$

$$df/da = -t * \exp(-a * t)$$



9. Прогноз срока службы ПВХ изоляции



17.02.03

Постановка задачи

Объект

Кабельная изоляция из ПВХ

Предсказать

значения, которые

нельзя измерить

Эксперимент

Термогравиметрический метод

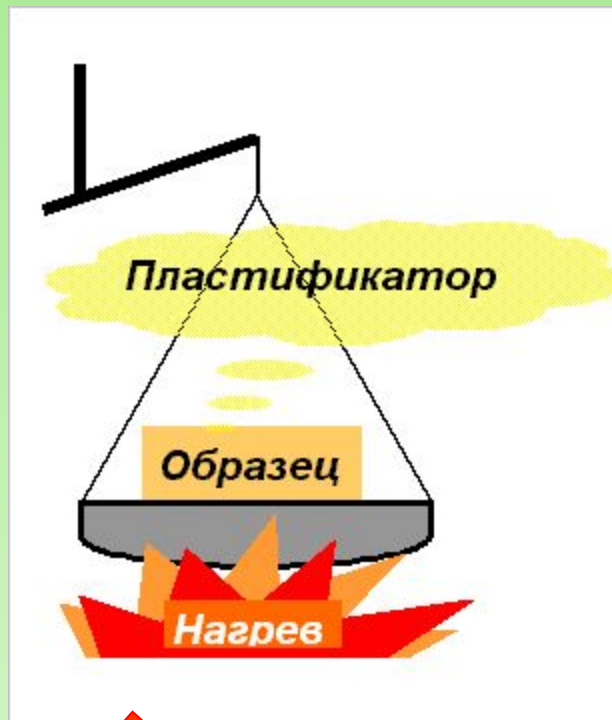
- ЭТО ЭКСТРАПОЛЯЦИЯ!

Обработка

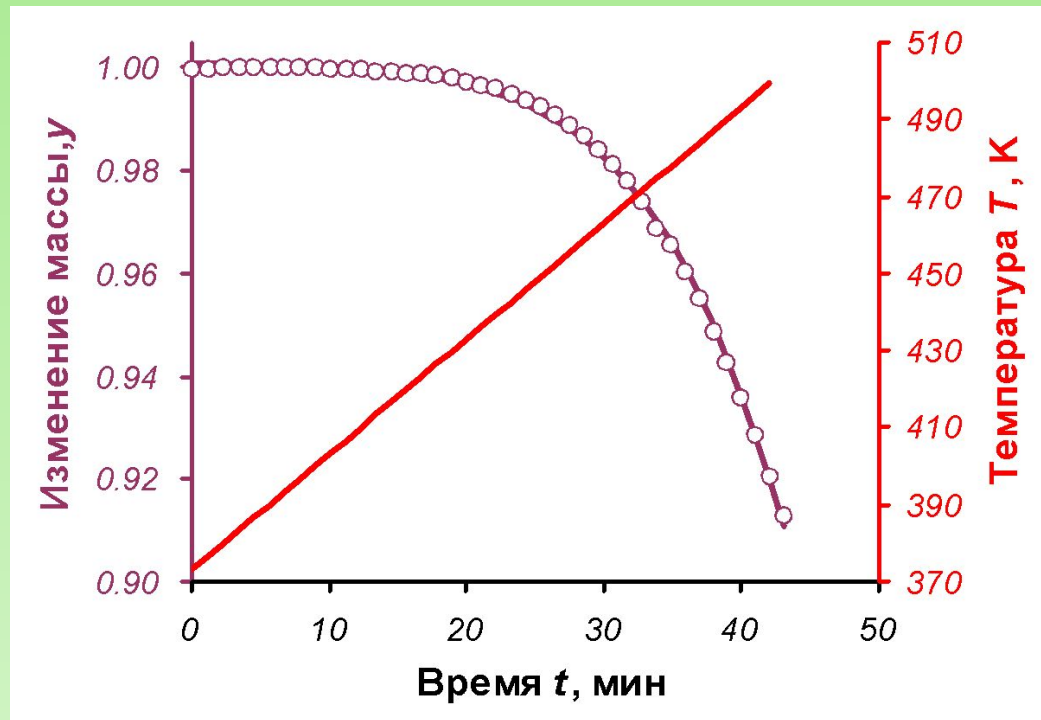
Нелинейная регрессия + Fitter

ТГА эксперимент и данные

ТГА эксперимент



ТГА данные



Это схема эксперимента, а не грешник в аду!

Переменные в примере ТГА

Измеряемые

Оцениваемые

Отклик

$$y = m/m_0$$

Изменение массы

Промежуточные

Контроль качества дифференциатора

Факторы

t

Время

C_0

Молярная концентрация

v

Скорость нагрева

T_0

Начальная температура

F

Удельная поверхность образца

Параметры

y_0

Начальное значение y

k_0

Константа скорости испарения

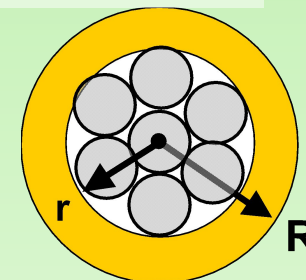
E

Энергия активации

**Нелинейная
задача малой
размерности!**

Удельная поверхность образца

$$F = \frac{S}{V} = \frac{2R}{R^2 - r^2}$$



Модель испарения пластификатора

Закон испарения

$$\frac{dy}{dt} = -kC, \quad y(0) = y_0$$

Диффузия

Изменение объема

$$C = 1 - \frac{1 - C_0}{y}$$

здесь

Закон Аррениуса

не важна!

$$\left(A e^{-\frac{E}{RT}} \right)$$

Рост температуры

$$T = T_0 + vt$$

Пример ТГА на рабочем листе Fitter

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1																
2		Data														
3		v	T0	C	F	t	y	f	Left							
4		3	373	0.3	2.4	0.00	1.000	1.001	1.001							
5		3	373	0.3	2.4	0.00	1.000	1.001	1.001							
6		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
7		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
8		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
9		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
10		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
11		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
12		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
13		3	373	0.3	2.4	0.00	1.000	1.001	1.000							
14																
15		0	293	0.4	2	0.00	1.000	0.871	0.844							
16		0	293	0.4	2	5E+06	0.871	0.871	0.844							
17		0	293	0.4	2	8E+06	0.821	0.788	0.788							
18		0	293	0.4	2	1E+07	0.781	0.746	0.746							
19		0	293	0.4	2	1E+07	0.748	0.714	0.714							
20																

$$\frac{dy}{dt} = -kC, \quad y(0) = y_0$$

$$C = 1 - \frac{1 - C_0}{1 + v \cdot t}$$

$$k = F \cdot \exp\left(k_0 - \frac{E}{RT}\right)$$

$$T = T_0 + vt$$

Fitter [X]

🔍 ⚙️ $\frac{t}{y}$ $f(x)$ 📊

'TGA Desorbption Model

D[y]/D[t]=k*[1-(1-C0)/y];y(0)=y0

k=F*exp(k0-E/R/T)

T=T0+v*t

1.98717

y0=?

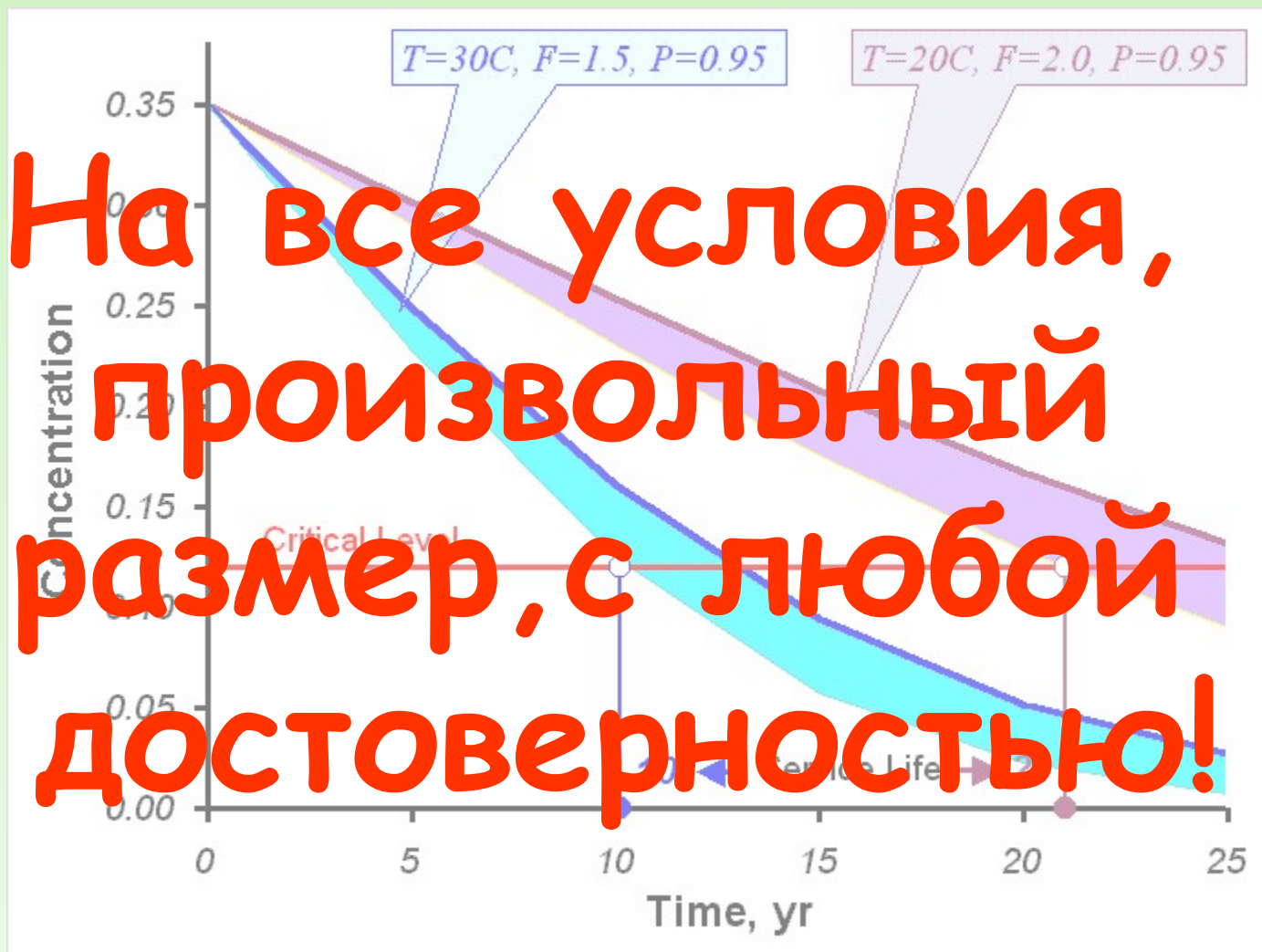
k0=?

E=?

Parameters estimation

Name	Initial	Final	Deviation
y0	1	1.00088	0.0002
k0	10	13.9964	0.24016
E	10000	18052.3	225.424

Прогноз срока службы



На все условия,
произвольный
размер, с любой
достоверностью!

10. Выводы

Формальный
подход



Загадочная природа



Содержательный
подход



Спасиб
о
за
вниман
ие!

17.02.03

61