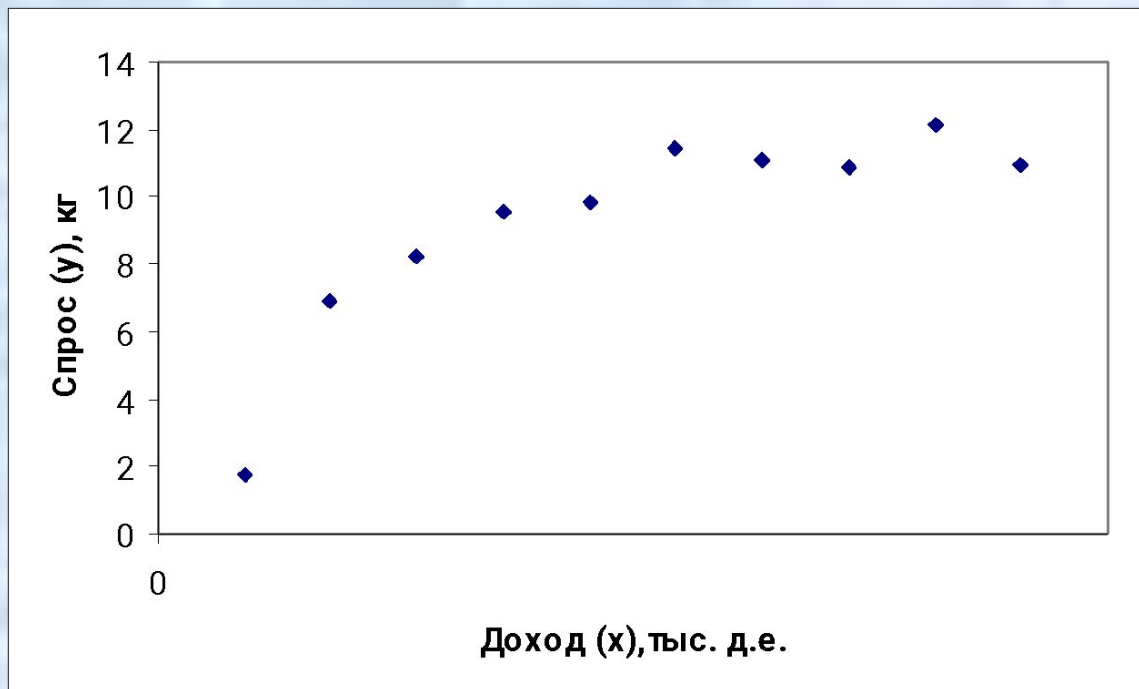


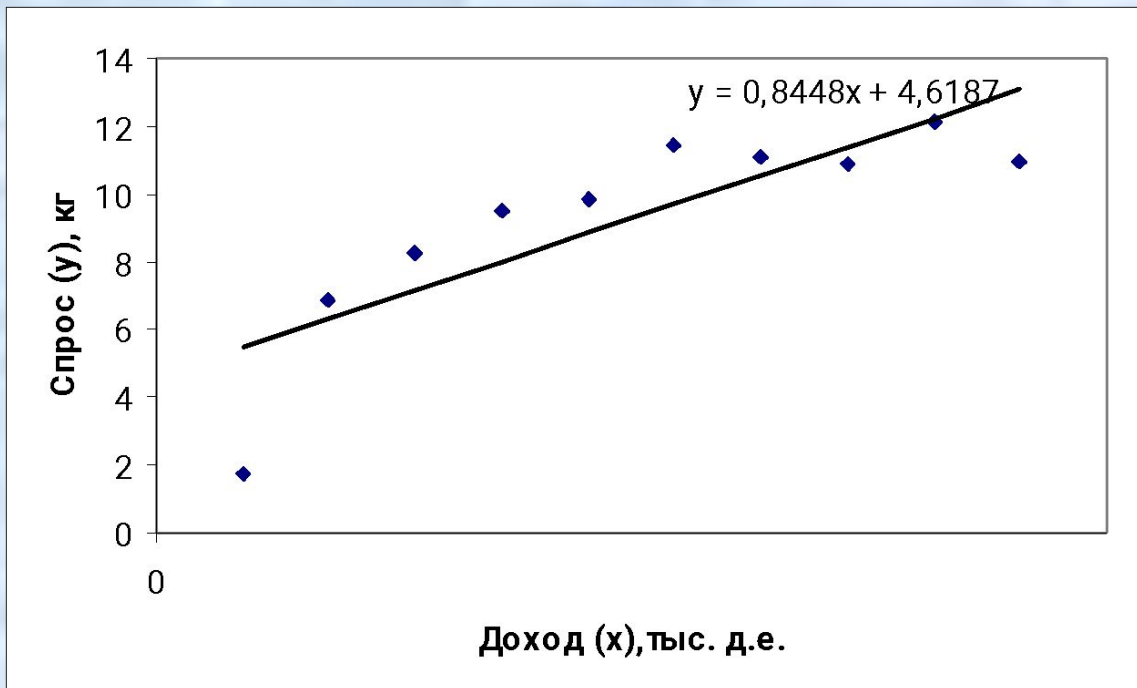
# Модели нелинейной регрессии

№ домохозяйства	Среднедушевой доход домохозяйства, тыс. д.е.	Объем спроса, кг в месяц
1	1	1,71
2	2	6,88
3	3	8,25
4	4	9,52
5	5	9,81
6	6	11,43
7	7	11,09
8	8	10,87
9	9	12,15
10	10	10,94



x	y
1	1,71
2	6,88
3	8,25
4	9,52
5	9,81
6	11,43
7	11,09
8	10,87
9	12,15
10	10,94

**Зависимость нелинейная!**



x	y
1	1,71
2	6,88
3	8,25
4	9,52
5	9,81
6	11,43
7	11,09
8	10,87
9	12,15
10	10,94

Попытка провести прямую

# 1) Логарифмическая модель

$$y = a \cdot \ln x + b + \varepsilon$$

Для оценки такой зависимости создаем столбец с  $\ln(x)$

	A	B	C
1	x	ln(x)	y
2	1	0	1,7
3	2	0,7	6,9
4	3	1,1	8,3
5	4	1,4	9,5
6	5	1,6	9,8
7	6	1,8	11
8	7	1,9	11
9	8	2,1	11
10	9	2,2	12
11	10	2,3	11

# 1) Логарифмическая модель

$$y = a \cdot \ln x + b + \varepsilon$$

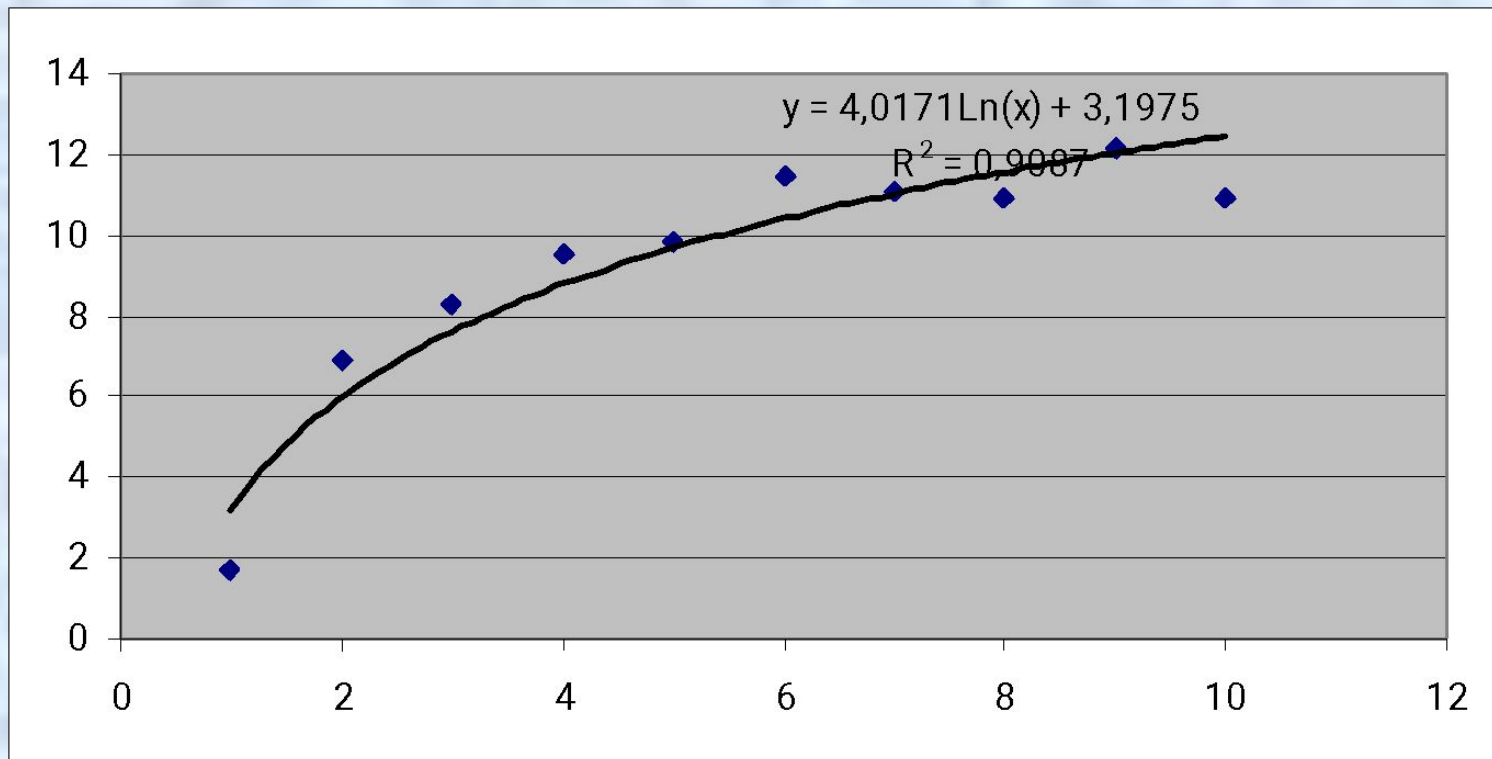
Используя сервис Анализ данных построим модель линейной регрессии, используя в качестве зависимой переменной  $y$ , а в качестве независимой  $\ln(x)$ .

	Кoeffициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересеч	3,197464	0,748798	4,270129	0,002724
$\ln(x)$	4,017062	0,450314	8,920585	1,98E-05

$$Y = 4.017 \ln(x) + 3.197$$

# 1) Логарифмическая модель

$$y = a \cdot \ln x + b + \varepsilon$$



# 1) Логарифмическая модель

$$y = a \cdot \ln x + b + \varepsilon$$

Интерпретация коэффициента  $a$ : при увеличении  $x$  на 1%  $y$  увеличится на  $a/100$  единиц.

$$Y = 4.017 \ln(x) + 3.197$$

При увеличении дохода на 1% спрос на товар увеличится на 0,0417 единиц.

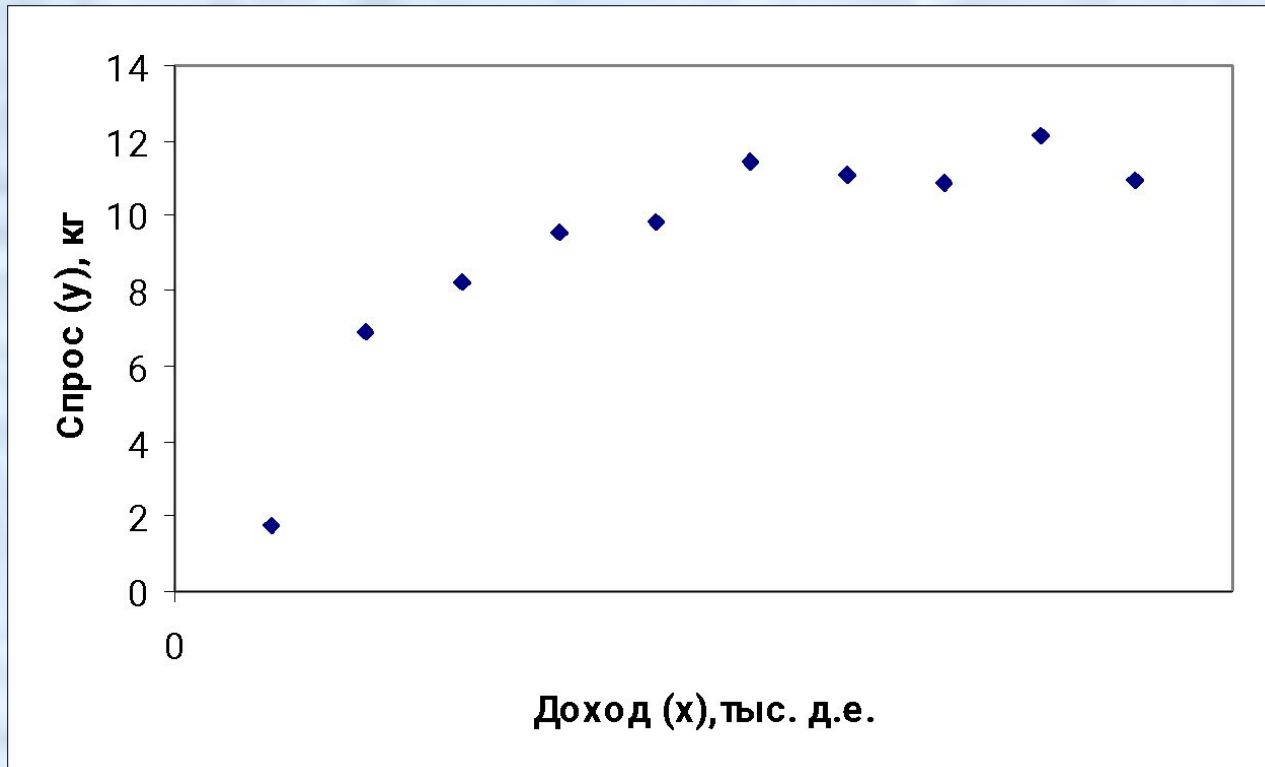
# 1) Логарифмическая модель

Также как в линейной модели рассчитывается средняя относительная ошибка аппроксимации

$$Y=4.017\ln(x)+3.197$$

	A	B	C	D	E	F
1	X	ln(x)	y	прогноз	e	Относительная ошибка
2	1	0	1,7	3,197	-1,5	86,99%
3	2	0,7	6,9	5,982	0,9	13,05%
4	3	1,1	8,3	7,611	0,6	7,75%
5	4	1,4	9,5	8,766	0,8	7,92%
6	5	1,6	9,8	9,663	0,1	1,50%
7	6	1,8	11	10,395	1	9,05%
8	7	1,9	11	11,014	0,1	0,68%
9	8	2,1	11	11,551	-0,7	6,26%
10	9	2,2	12	12,024	0,1	1,04%
11	10	2,3	11	12,447	-1,5	13,78%
12			Средняя относительная ошибка			14,80%





x	y
1	1,71
2	6,88
3	8,25
4	9,52
5	9,81
6	11,43
7	11,09
8	10,87
9	12,15
10	10,94

2) Попробуем провести гиперболу  
наилучшим образом.

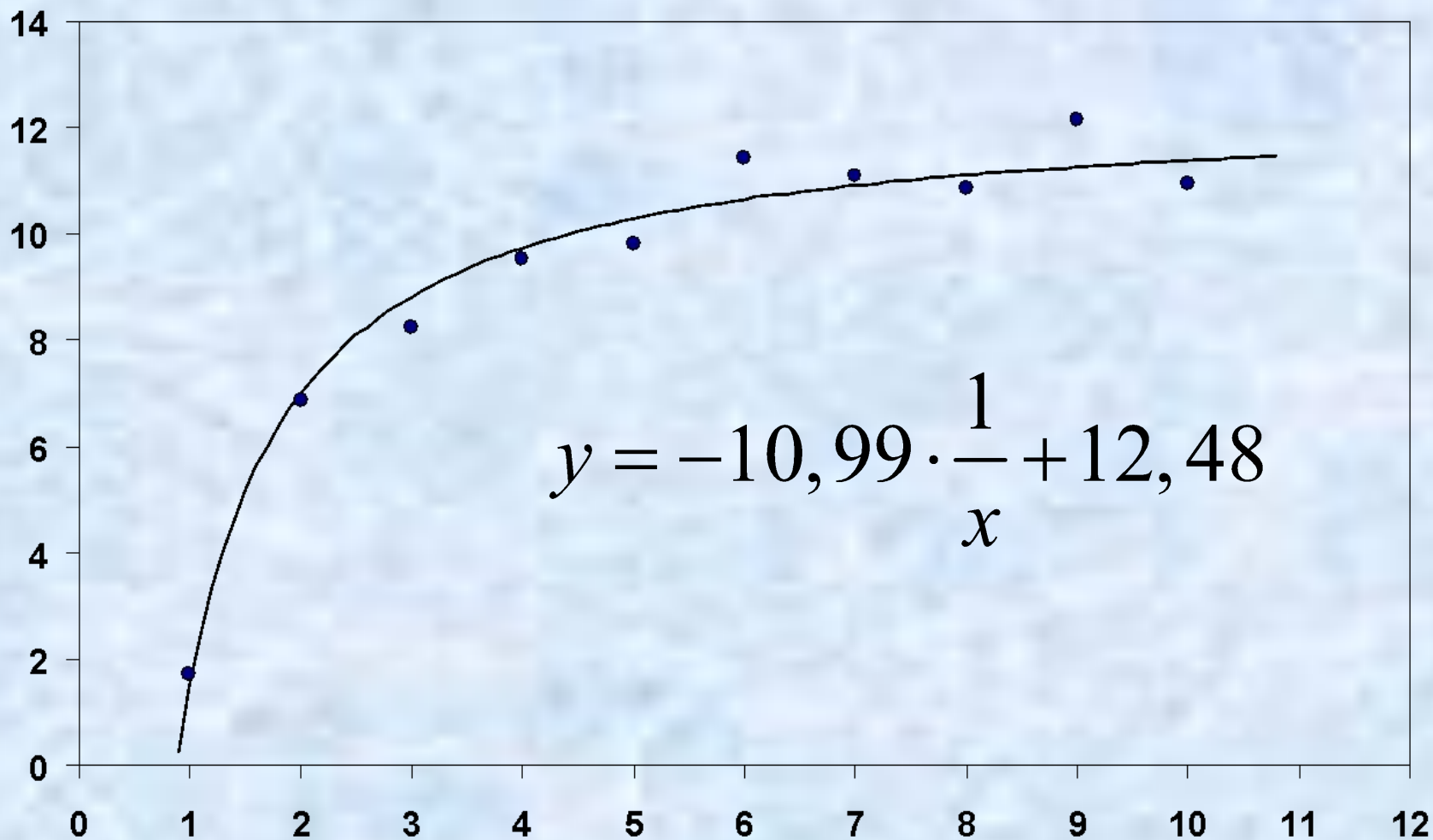
$$y = a \frac{1}{x} + b$$

	A	B	C
1	x	1/x	y
2	1	1	1,7
3	2	0,5	6,9
4	3	0,3	8,3
5	4	0,3	9,5
6	5	0,2	9,8
7	6	0,2	11
8	7	0,1	11
9	8	0,1	11
10	9	0,1	12
11	10	0,1	11

Сначала рассчитаем столбик  $1/x$


	<i>Козффициенты</i>	<i>Стандар тная ошибка</i>	<i>t- статис тика</i>	<i>p- Значение</i>
Y-пересеч	12,48354	0,255751	48,81128	3,43E-11
1/x	-10,9887	0,649657	-16,9145	1,51E-07

$$y = -10,99 \cdot \frac{1}{x} + 12,48$$



С ростом дохода объем потребления товара стремится к 12.48 ед.

# Вычисляем ошибку аппроксимации

$$y = -10,99 \cdot \frac{1}{x} + 12,48$$


x	1/x	y	прогноз y	e	Относительная ошибка
1	1	1,7	1,495	0,215	12,58%
2	0,5	6,9	6,989	-0,109	1,59%
3	0,3	8,3	8,821	-0,571	6,92%
4	0,3	9,5	9,736	-0,216	2,27%
5	0,2	9,8	10,29	-0,476	4,85%
6	0,2	11	10,65	0,778	6,81%
7	0,1	11	10,91	0,176	1,59%
8	0,1	11	11,11	-0,24	2,21%
9	0,1	12	11,26	0,887	7,30%
10	0,1	11	11,38	-0,445	4,06%
Средняя ошибка аппроксимации					5,02%

### 3) Степенная модель

$$y = bx^a$$

Интерпретация коэффициента  $a$  – эластичность зависимой переменной по объясняющей переменной  $a$  показывает, на сколько процентов возрастает  $y$  при возрастании  $x$  на 1%.

# Степенная модель

$$y = bx^a$$

Сводится к линейной модели логарифмированием

$$\ln y = \ln b + a \ln x$$

# Степенная модель

Создаем столбцы с логарифмами

	A	B	C	D
1	x	ln(x)	y	ln(y)
2	1	0	1,71	0,536
3	2	0,7	6,88	1,929
4	3	1,1	8,25	2,110
5	4	1,4	9,52	2,253
6	5	1,6	9,81	2,283
7	6	1,8	11,4	2,436
8	7	1,9	11,1	2,406
9	8	2,1	10,9	2,386
10	9	2,2	12,2	2,497
11	10	2,3	10,9	2,392



Используя сервис Анализ данных построим модель линейной регрессии, используя в качестве зависимой переменной  $\ln(y)$ , а в качестве независимой  $\ln(x)$ .

	<i>Козффиц иенты</i>	<i>Стандартн ая ошибка</i>	<i>t-статистика</i>	<i>p- Значение</i>
Y-пересеч	1,063664	0,22029814	4,828294976	0,001307
$\ln(x)$	0,701353	0,13248337	5,293894844	0,000734

$$\ln(Y)=0.701\ln(x)+1.063$$

Используя сервис Анализ данных построим модель линейной регрессии, используя в качестве зависимой переменной  $\ln(y)$ , а в качестве независимой  $\ln(x)$ .

	Кoeffициенты	Стандартная ошибка	t-статистика	p-Значение
Y-пересеч	1,063664	0,22029814	4,828294976	0,001307
$\ln(x)$	0,701353	0,13248337	5,293894844	0,000734

$$\ln(Y)=0.701\ln(x)+1.063 \quad \ln y = a \ln x + \ln b$$

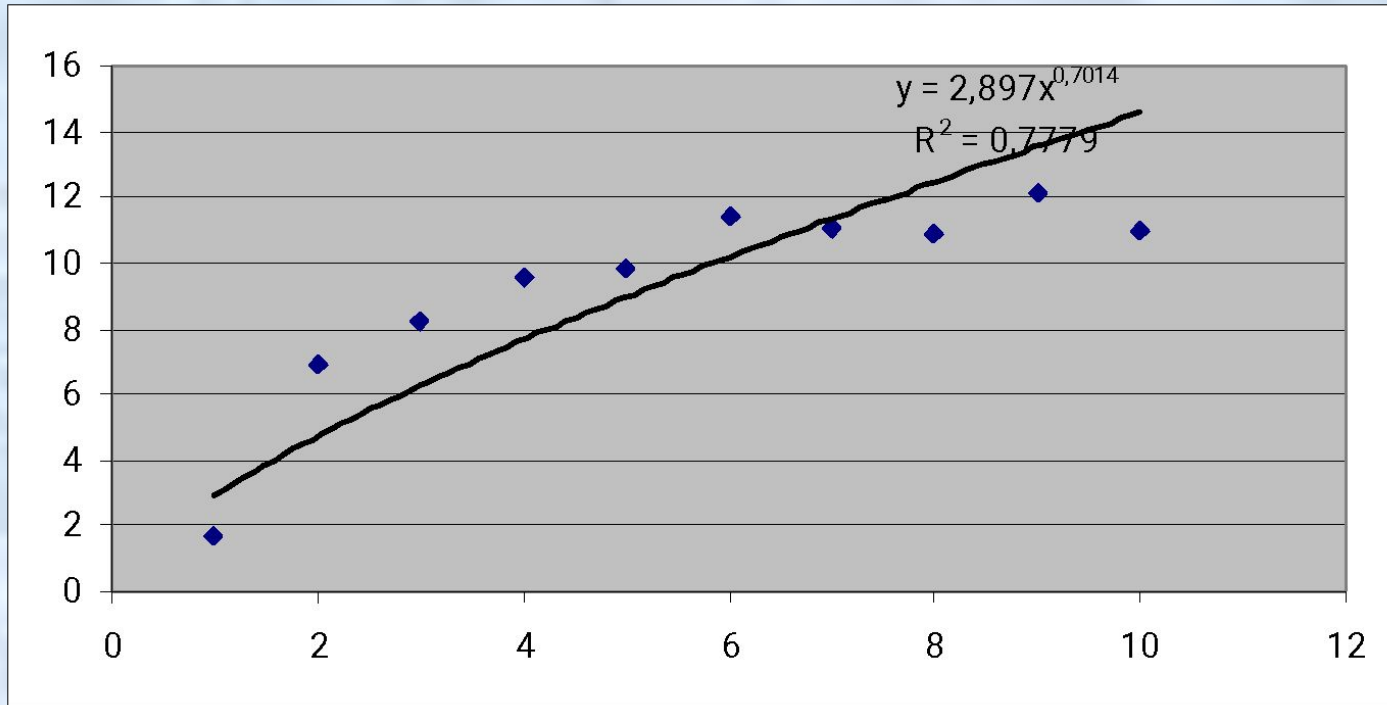
$$\ln b = 1.063 \quad b = \exp(1.063) = 2.9$$

Используя сервис Анализ данных построим модель линейной регрессии, используя в качестве зависимой переменной  $\ln(y)$ , а в качестве независимой  $\ln(x)$ .

	Коэффициенты	Стандартная ошибка	t-статистика	p-значение
Y-пересеч	1,063664	0,22029814	4,828294976	0,001307
$\ln(x)$	0,701353	0,13248337	5,293894844	0,000734

$$\ln(Y) = 0.701 \ln(x) + 1.063 \quad b = \exp(1.063) = 2.9$$

$$y = bx^a \quad y = 2.9x^{0.701}$$



$$y = 2.9x^{0.701}$$

Также как в линейной модели рассчитывается средняя относительная ошибка аппроксимации

$$y = 2.9x^{0.701}$$

A	B	C	D	E	F	G	
				↓		Относит ельная ошибка	
x	ln(x)	y	ln(y)	прогноз	e		
1	0	1,71	0,536	2,897	-1,187	69,41%	
2	0,7	6,88	1,929	4,711	2,169	31,53%	
3	1,1	8,25	2,110	6,260	1,990	24,12%	
4	1,4	9,52	2,253	7,659	1,861	19,54%	
5	1,6	9,81	2,283	8,957	0,853	8,69%	
6	1,8	11,4	2,436	10,179	1,251	10,95%	
7	1,9	11,1	2,406	11,341	-0,251	2,26%	
8	2,1	10,9	2,386	12,455	-1,585	14,58%	
9	2,2	12,2	2,497	13,527	-1,377	11,33%	
10	2,3	10,9	2,392	14,565	-3,625	33,13%	
		Средняя относительная ошибка					22,56%

$y = -10,99 \cdot \frac{1}{x} + 12,48$  - наилучшая функция спроса

в зависимости от дохода.

- 1) Выполнить прогноз потребления продукта домохозяйством с доходом 4 тыс.д.е.
- 2) Имеется ли уровень насыщения для данного продукта? Если да, найти его.
- 2) Найти предельную склонность к потреблению продукта.
- 3) Найти эластичность спроса по доходу при доходе 1000 д.е. и 10000 д.е.

# Модели парной нелинейной регрессии

Существует 2 типа нелинейных моделей:

1. модели, сводящиеся к линейным;
2. модели, не сводящиеся к линейным.

# 1 тип моделей

## 1) Гиперболическая модель

$$y = a \cdot \frac{1}{x} + b + \varepsilon$$

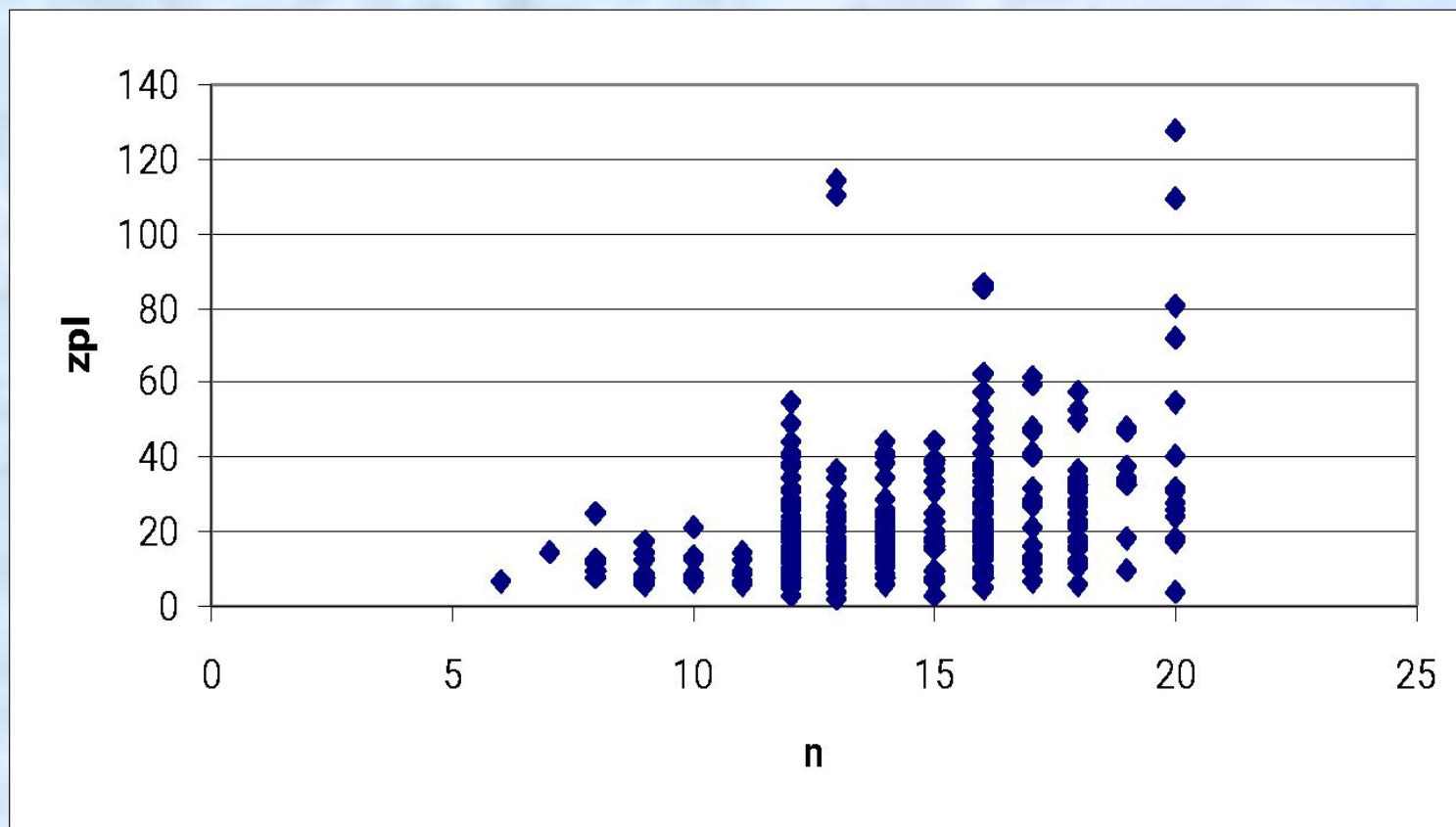


# 1 тип моделей

## 3) Экспоненциальная модель

$$y = b \cdot e^{ax}$$

## Пример применения экспоненциальной модели для моделирования оплаты труда



Данные 2002 г. о часовой заработной плате (\$ США) и уровне образования (лет) по 540 респондентам из национального опроса в США.

12 лет – средняя школа

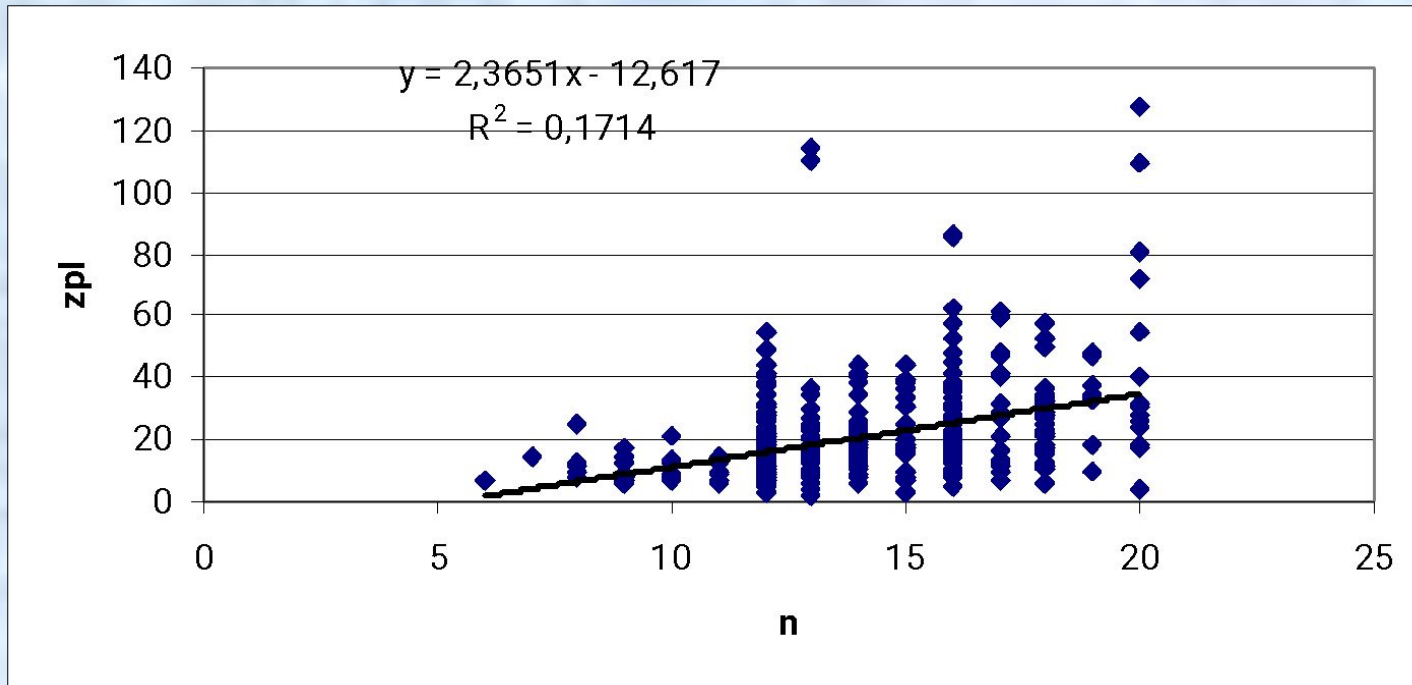
13-16 лет – колледж (бакалавриат)

17-18 лет – университет ( магистратура)

19-20 лет - PhD

# ПРИМЕР УРАВНЕНИЯ ПАРНОЙ РЕГРЕССИИ

$$Z_{pl} = -12,617 + 2,3651N$$



**Увеличению уровня образования на один год приведет в среднем к увеличению почасовой заработной платы на \$2.37**

## Пример применения экспоненциальной модели для моделирования оплаты труда

$$Zpl = be^{aN}$$

$$\ln(Zpl) = \ln b + aN$$

---

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	1,424862	0,125762	11,32982	7,81E-27
n	0,100131	0,009042	11,07396	8,12E-26

---

## Пример применения экспоненциальной модели для моделирования оплаты труда

$$Z_{pl} = be^{aN}$$

$$\ln(Z_{pl}) = \ln b + aN$$

---

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	1,424862	0,125762	11,32982	7,81E-27
n	0,100131	0,009042	11,07396	8,12E-26

---

$$\ln(Z_{pl}) = 1,42 + 0,1N$$

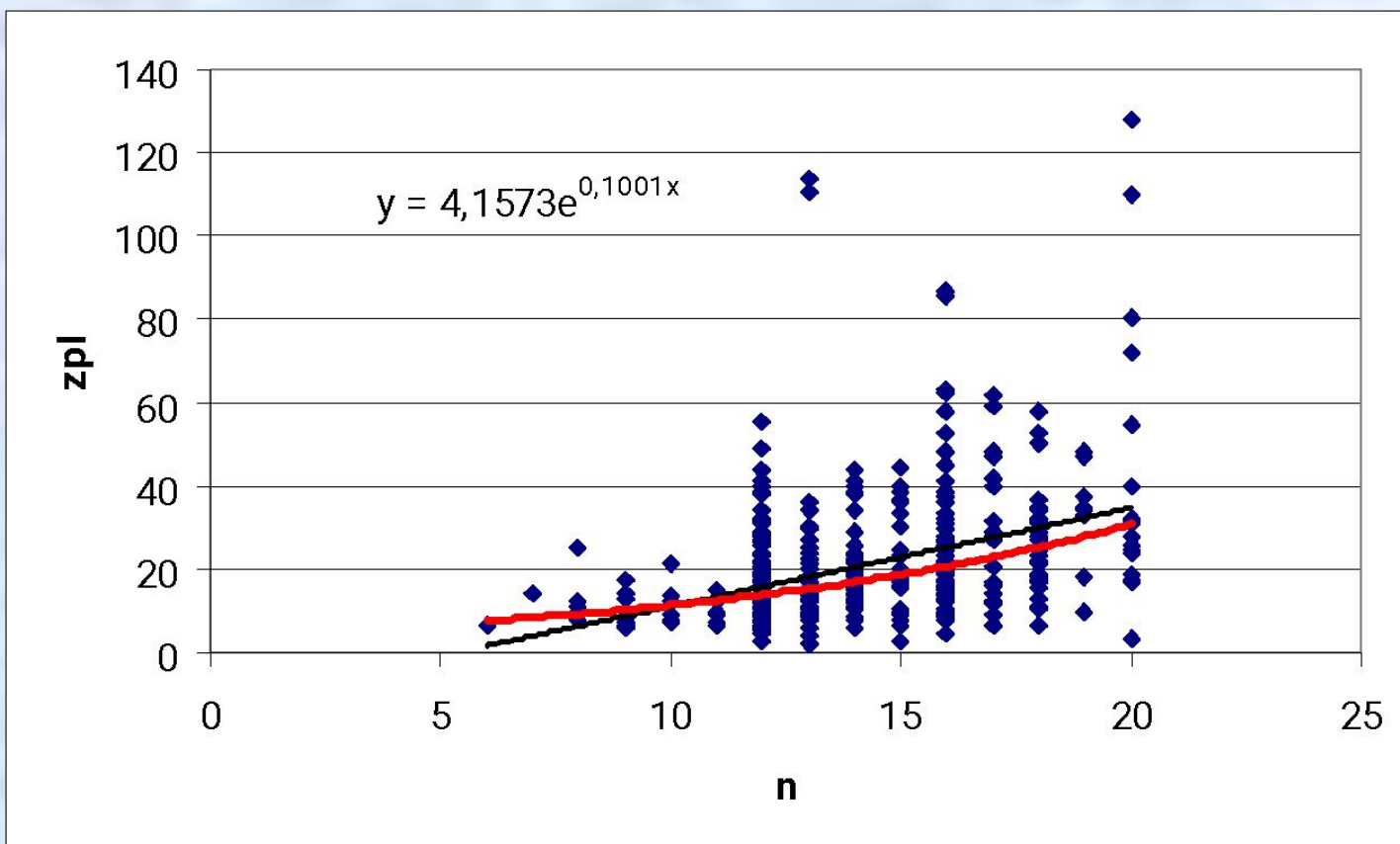
$$Z_{pl} = e^{1,42+0,1N}$$

$$Z_{pl} = 4,14e^{0,1N}$$

# Пример применения экспоненциальной модели для моделирования оплаты труда

$$Z_{pl} = 4,14e^{0,1N}$$

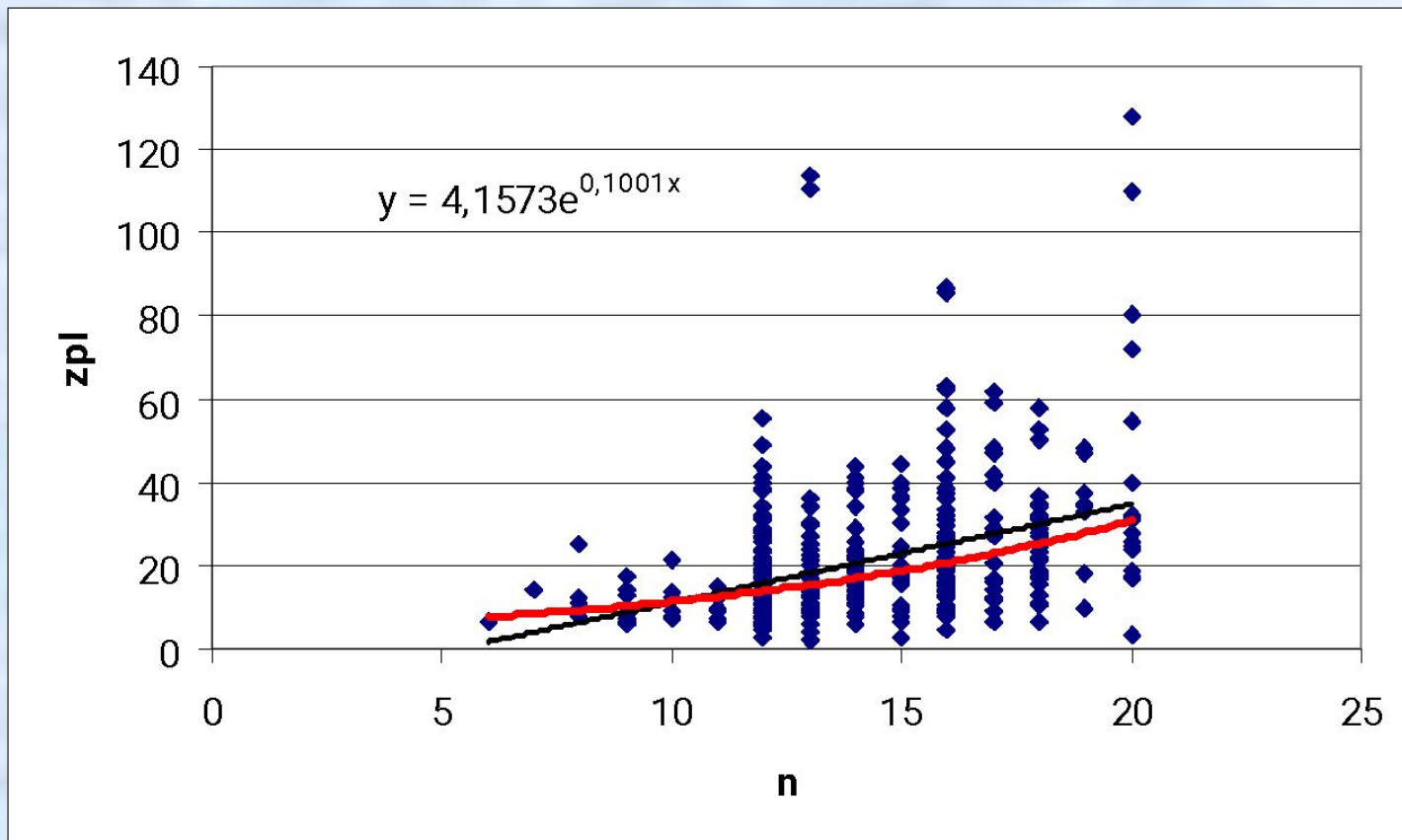
Каждый дополнительный год обучения приводит к росту заработка на 10%



# Пример применения экспоненциальной модели для моделирования оплаты труда

Преимущества экспоненциальной модели:

- 1) Она не предсказывает отрицательного заработка индивидам с низким образовательным уровнем
- 2) Она показывает возрастание прироста зарплат в расчете на 1 дополнительный год обучения при повышении образовательного уровня.

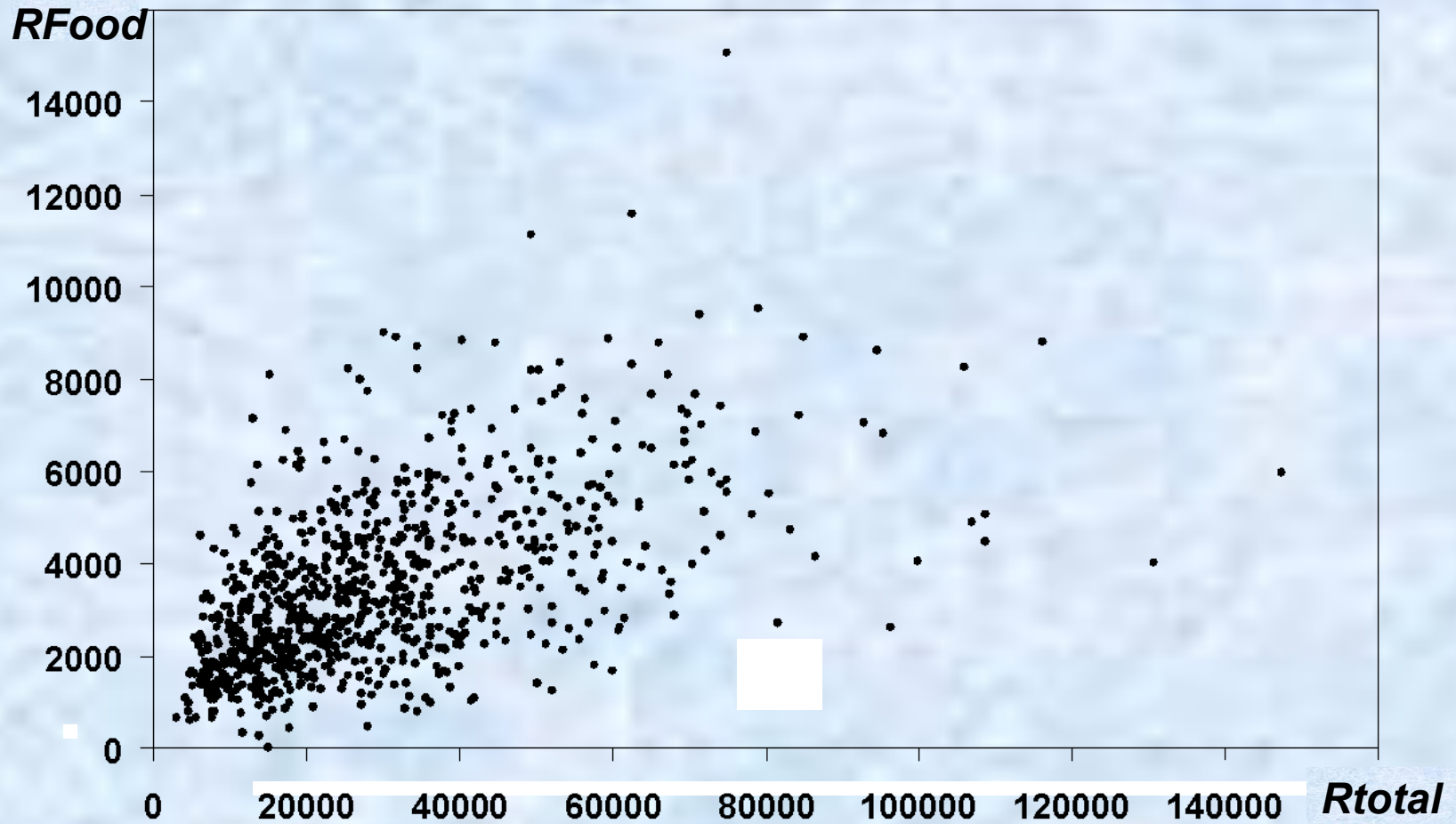


## 4) Степенная модель

$$y = bx^a$$



# Пример. Линейная и степенная модель



Расходы на продукты питания и общие расходы в 1995 (обе - в долларах) по данным 869 домохозяйств США

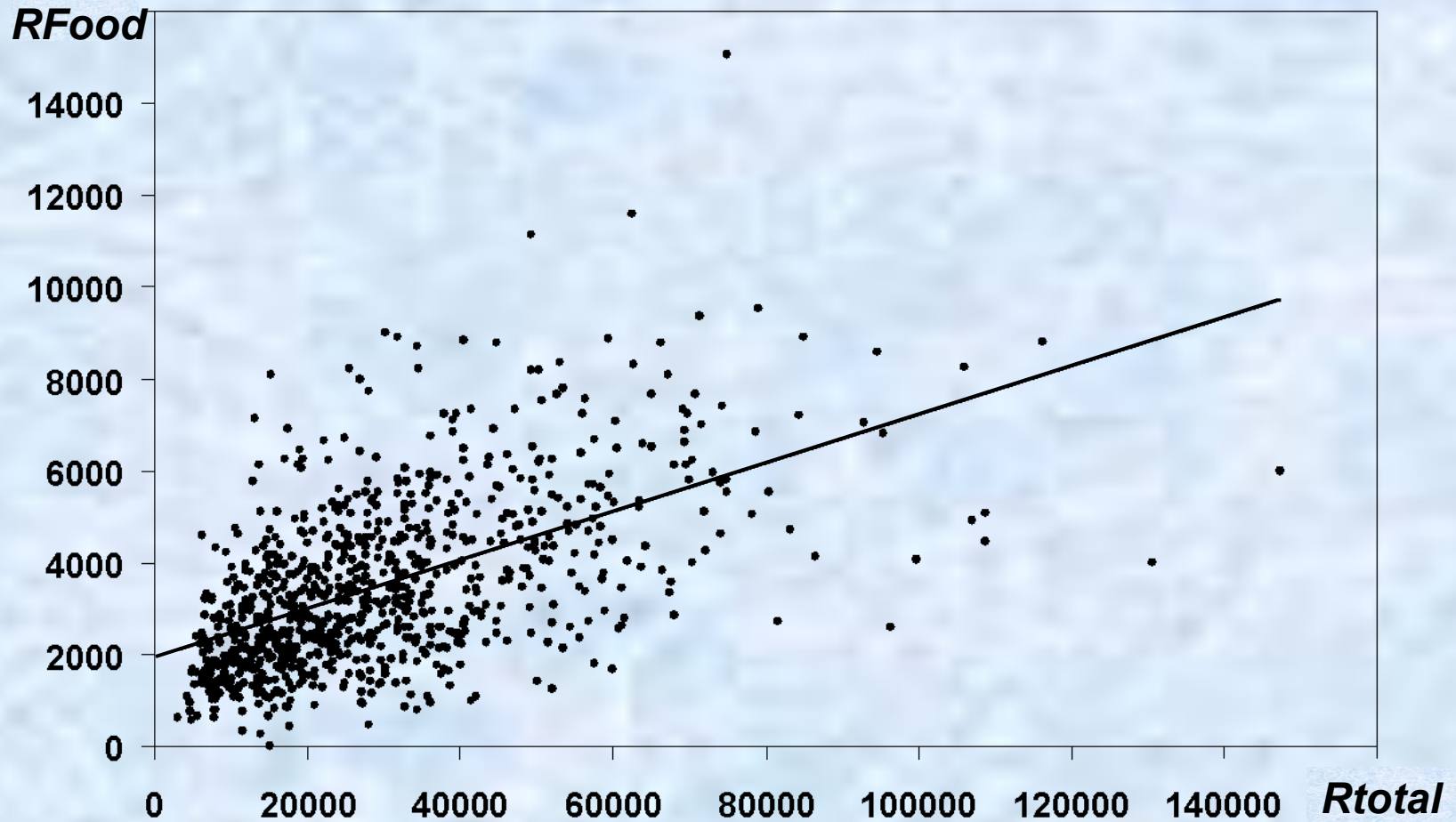
# Линейная модель

Number of obs = 869  
F( 1, 867) = 381.47  
Prob > F = 0.0000  
R-squared = 0.3055  
Adj R-squared = 0.3047  
Root MSE = 1549.5

Rtotal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
RFood	.0528427	.0027055	19.531	0.000	.0475325	.0581529
_cons	1916.143	96.54591	19.847	0.000	1726.652	2105.634

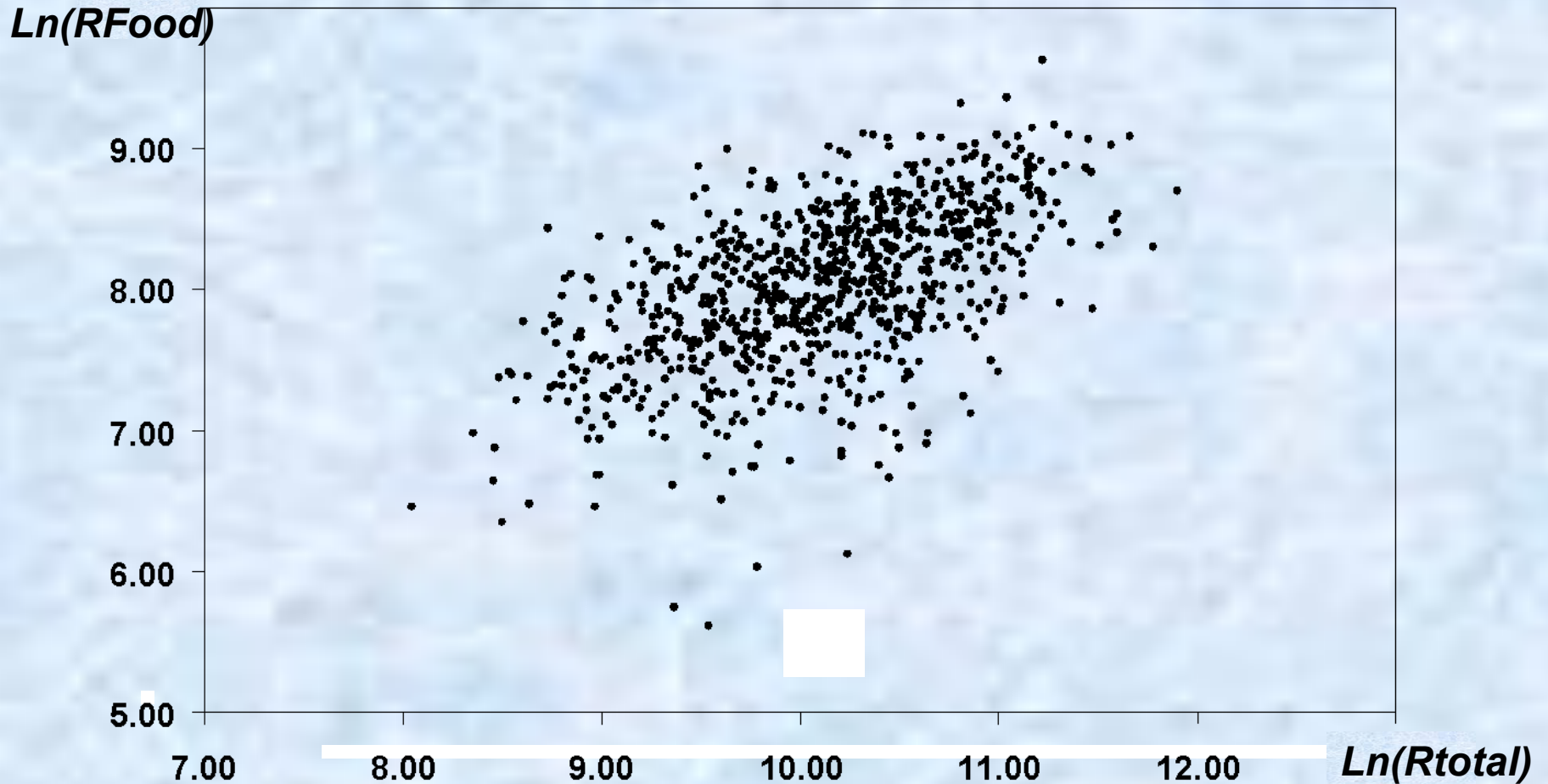
**Коэффициенты представляются разумными, хотя предельный эффект несколько занижен, а константа-завышена.**

# Линейная модель



Несоответствие коэффициентов хорошо видно на графике

# Степенная модель



**Между логарифмически преобразованными переменными линейная зависимость кажется более адекватной**

## Степенная модель

Number of obs = 868  
F( 1, 866) = 396.06  
Prob > F = 0.0000  
R-squared = 0.3138  
Adj R-squared = 0.3130  
Root MSE = .46167

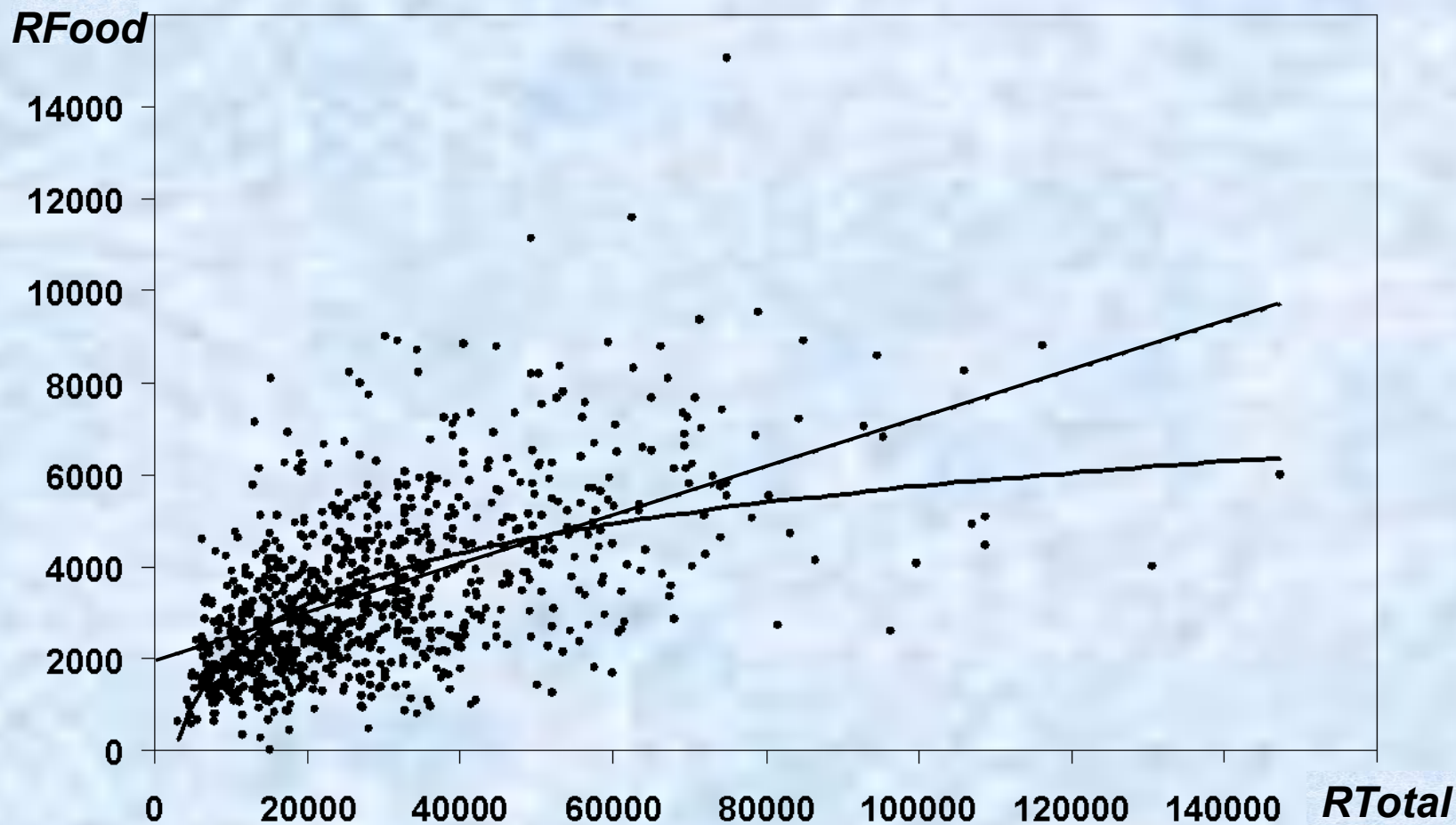
LnRtotal	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
LnRFood	.4800417	.0241212	19.901	0.000	.4326988	.5273846
_cons	3.166271	.244297	12.961	0.000	2.686787	3.645754

Модель высокозначима. Коэффициент эластичности расходов на товары питания по совокупным расходам положителен и меньше единицы, как и полагается для нормального товара первой необходимости

$$\ln R_{Food} = 3.17 + 0.48 \ln R_{total} \Rightarrow R_{Food} = 23.8 \text{EXP}^{0.48 R_{total}}$$

Константа не имеет хорошей интерпретации.  $e^{3.16} = 23.8$ , то есть просто некий масштабный множитель

## Степенная модель



Сопоставление линейной и степенной регрессии на исходном графике четко делает выбор в пользу последней. Хотя различие не кажется особенно сильным, но степенная модель лучше объясняет данные при малых значениях  $R_{total}$ , более обоснована с теоретической точки зрения (постоянная эластичность) и гетероскедастичность меньше выражена

## Полиномиальная модель

$$y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k + \varepsilon$$

Появляются возможность исследования зависимостей, для которых существенно наличие максимумов и минимумов

$$y = a_0 + a_1x + a_2x^2 + \varepsilon \quad \text{квадратичная модель}$$

## Полиномиальная модель

$$y = a_0 + a_1x + a_2x^2 + \varepsilon$$

$$x_1 = x \quad x_2 = x^2$$

$$y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$$

- модель множественной регрессии.

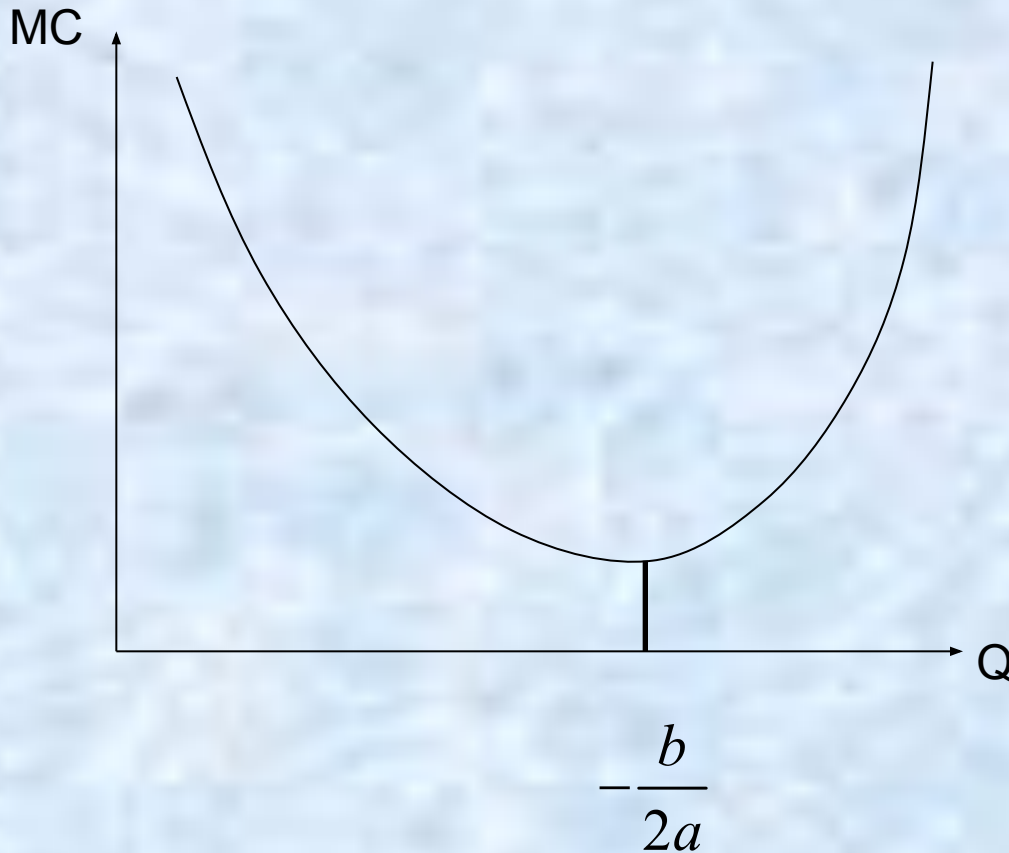


# Полиномиальная модель

## Примеры

1) Пусть  $Q$  – объем выпуска продукции,  $MC$  – предельные издержки производства.

$$MC = aQ^2 + bQ + c + \xi \quad a > 0, b < 0$$

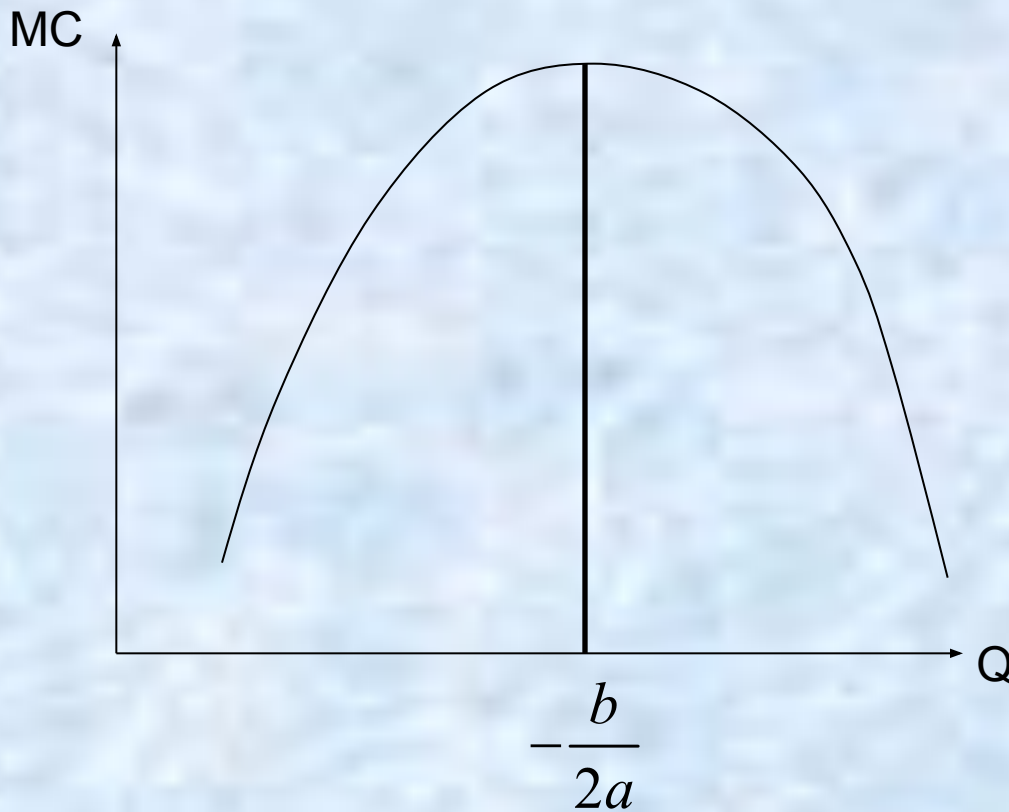


# Полиномиальная модель

## Примеры

2)  $x$  – возраст работника физического труда,  $y$  – заработная плата

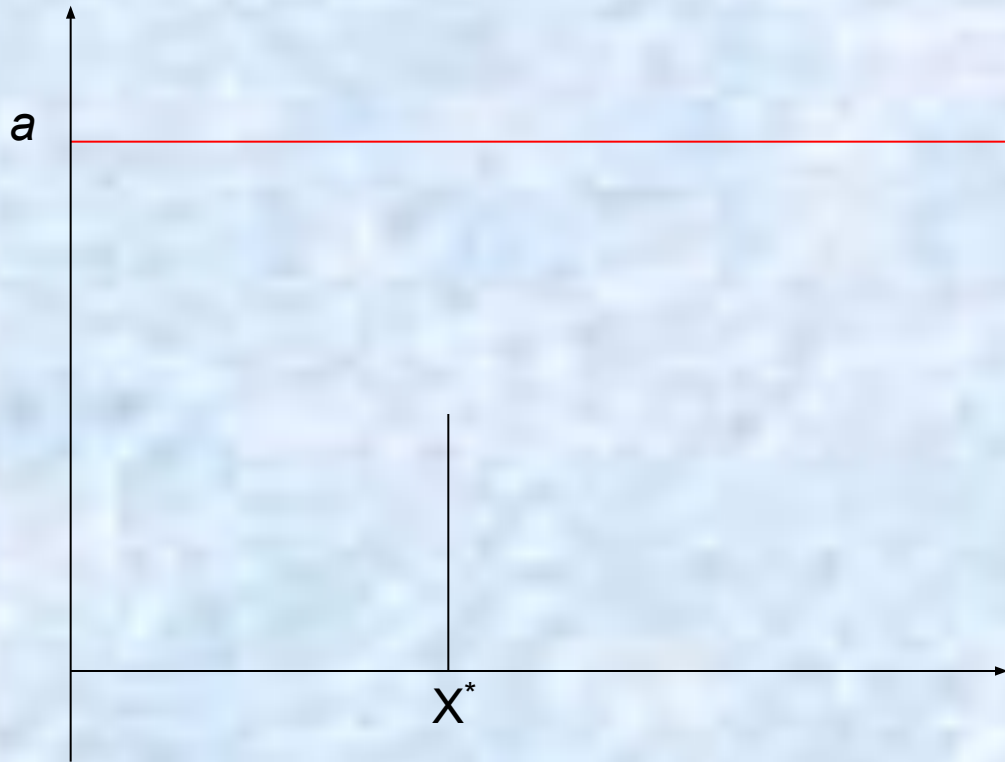
$$y = ax^2 + bx + c + \xi \quad a < 0, b < 0$$



## 2 тип модели (модели, не сводящиеся к линейным)

Например, Логистическая модель

$$y = \frac{a}{1 + be^{-cx}} + \varepsilon$$



## 2 тип моделей (модели, не сводящиеся к линейным)

Для оценки коэффициентов таких моделей используется МНК:

$$y = \frac{a}{1 + be^{-cx}} + \varepsilon$$

$$S(a, b, c) = \sum_{i=1}^n \left( y_i - \frac{a}{1 + be^{-cx_i}} \right)^2$$

$$\min_{a, b, c} S(a, b, c)$$

$$\begin{cases} S'_a = 0 \\ S'_b = 0 \\ S'_c = 0 \end{cases}$$

Задача решается численными методами.  
В Excel через поиск решения