ЭКОНОМЕТРИКА
ЛИТЕРАТУРА К КУРСУ:

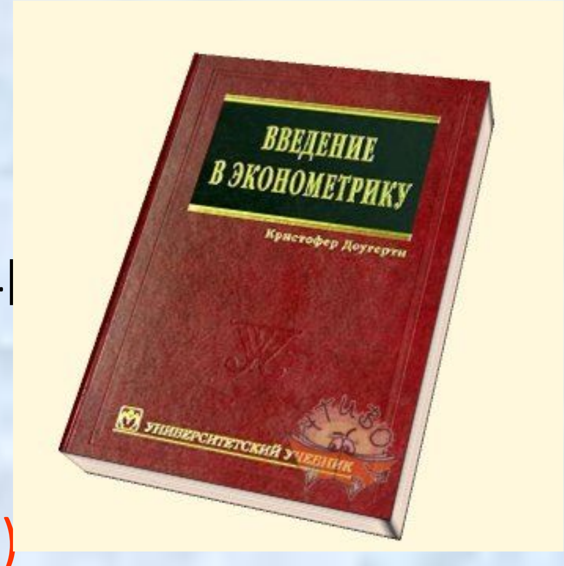
<u>Основная</u>

- К.Доугерти

«Введение в эконометрику» , М.: Инфра-

(пер. под ред. Замкова)

(желательно последнее издание 2007г!!!)



ЧТО ТАКОЕ **ЭКОНОМЕТРИКА**

Эконометрика

1. Формулирует экономические модели, основываясь на экономической теории (микро и макроэкономике)
2. Оценивает неизвестные параметры модели на базе реальных статистических данных
3. Использует построенные модели для объяснения поведения исследуемых экономических показателей, прогнозирования, а также для осмысленного проведения экономической политики.

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

1. Вычисляем средние значения $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

СРЗНАЧ(диапазон данных)

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

2. Вычисляем дисперсии

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

ДИСПР(диапазон данных)

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

3. Вычисляем ковариацию

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

КОВАР(диапазон данных x, диапазон данных y)

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

4. Вычисляем корреляцию

$$\mathrm{cor}(x, y) = \frac{\mathrm{cov}(x, y)}{\sqrt{s_x^2 \cdot s_y^2}}$$

КОРРЕЛ(диапазон данных x, диапазон данных y)

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

4. Анализируем коэффициент корреляции

Коэффициент корреляции принимает значения от -1 до 1.

Значения близкие к 1 – есть тесная прямая связь между x и y

Значения близкие к -1 – есть тесная обратная связь между x и y

Значения близкие к 0 – связь между x и y отсутствует

ПАРНАЯ РЕГРЕССИЯ

Коэффициент корреляции близок к 1

Коэффициент корреляции близок к -1

Коэффициент корреляции близок к 0

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

y – зависимая (объясняемая) переменная
x – независимая (объясняющая) переменная

6. Если коэффициент корреляции не близок к 0 строим модель парной линейной регрессии

$$y = ax + b + \varepsilon$$

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

$$y = ax + b + \varepsilon$$

Предположим, что необходимо получить функцию спроса на некоторый товар в зависимости от дохода.

Проводится опрос домохозяйств.

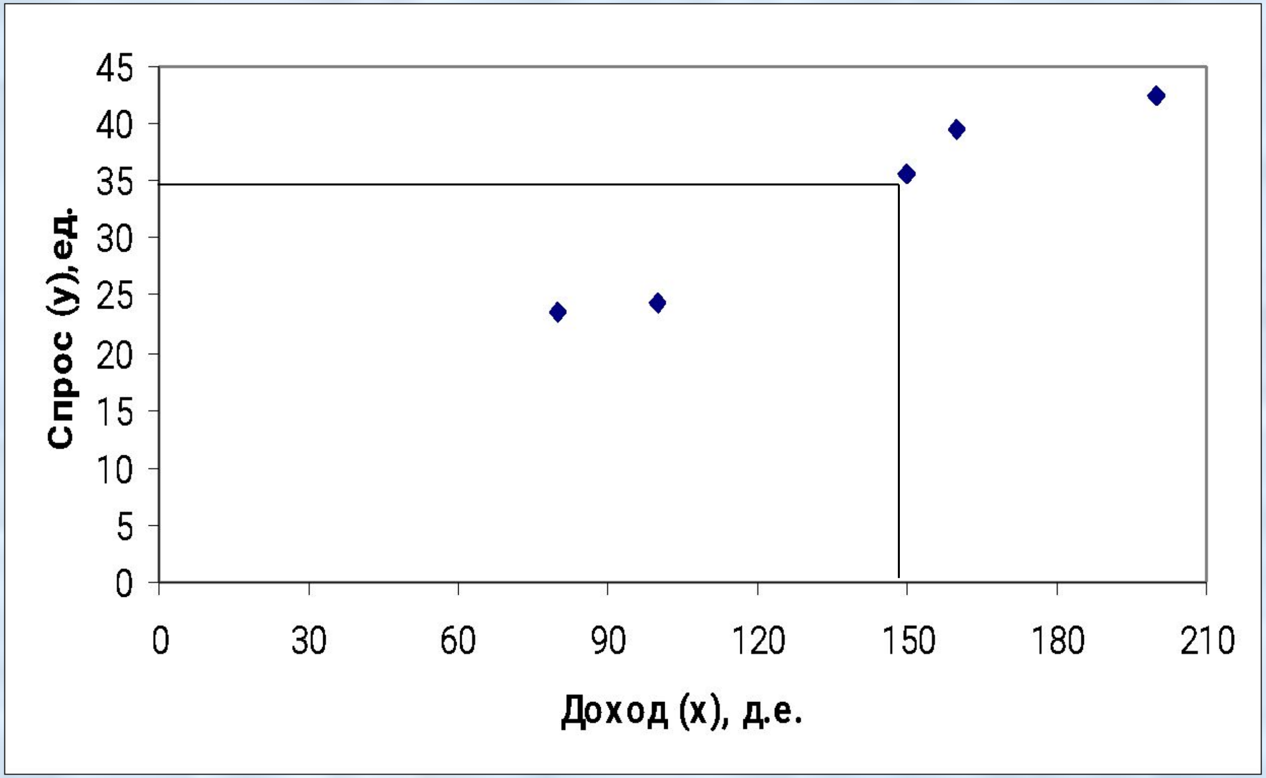
1. Среднедушевой доход домохозяйства?
2. Сколько единиц товара приобрело домохозяйство за месяц?

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

№ домохозяйства	Среднедушевой доход домохозяйства, д.е.	Объем спроса, ед.
1	100	24
2	200	42
3	150	35
4	80	24
5	160	39

МОДЕЛЬ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

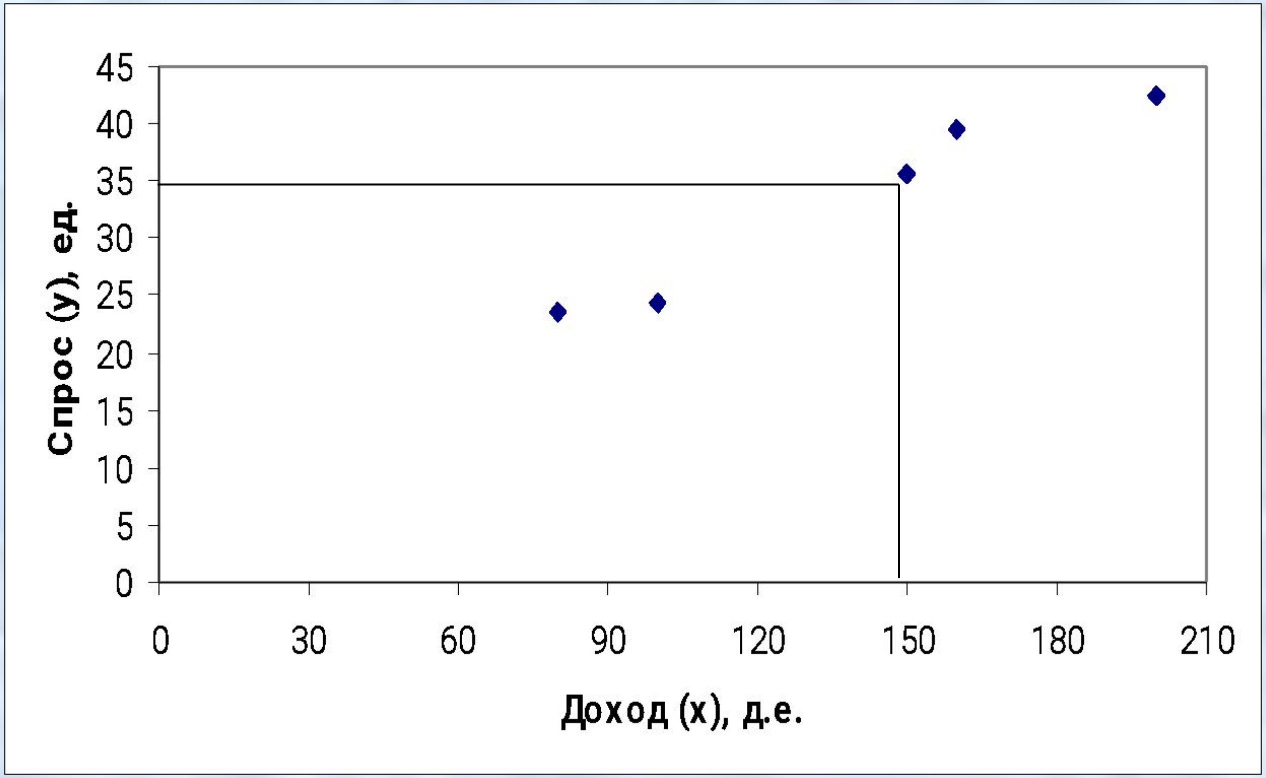
Нанесем точки на график



x	y
100	24
200	42
150	35
80	24
160	39

Метод наименьших квадратов

Нанесем точки на график

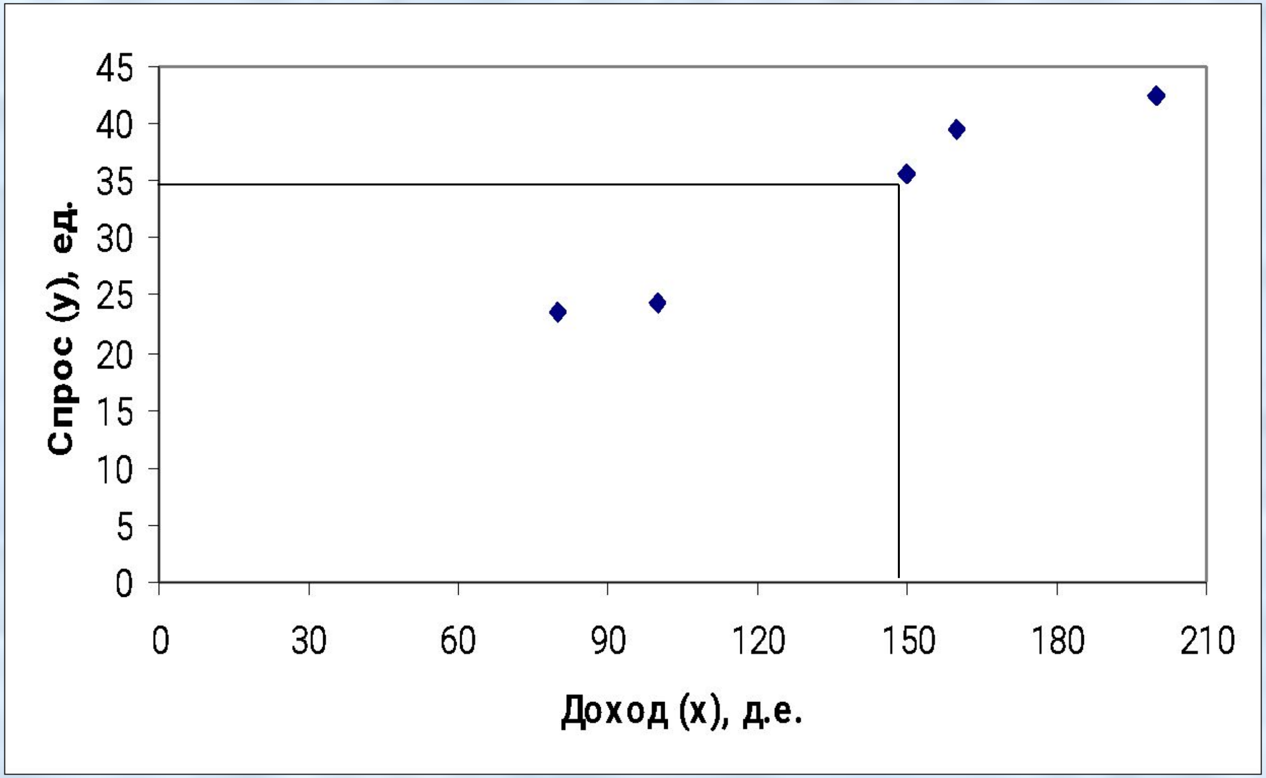


x	y
100	24
200	42
150	35
80	24
160	39

Точки разбросаны вокруг некоторой прямой!
Как ее найти?

Метод наименьших квадратов

Нанесем точки на график

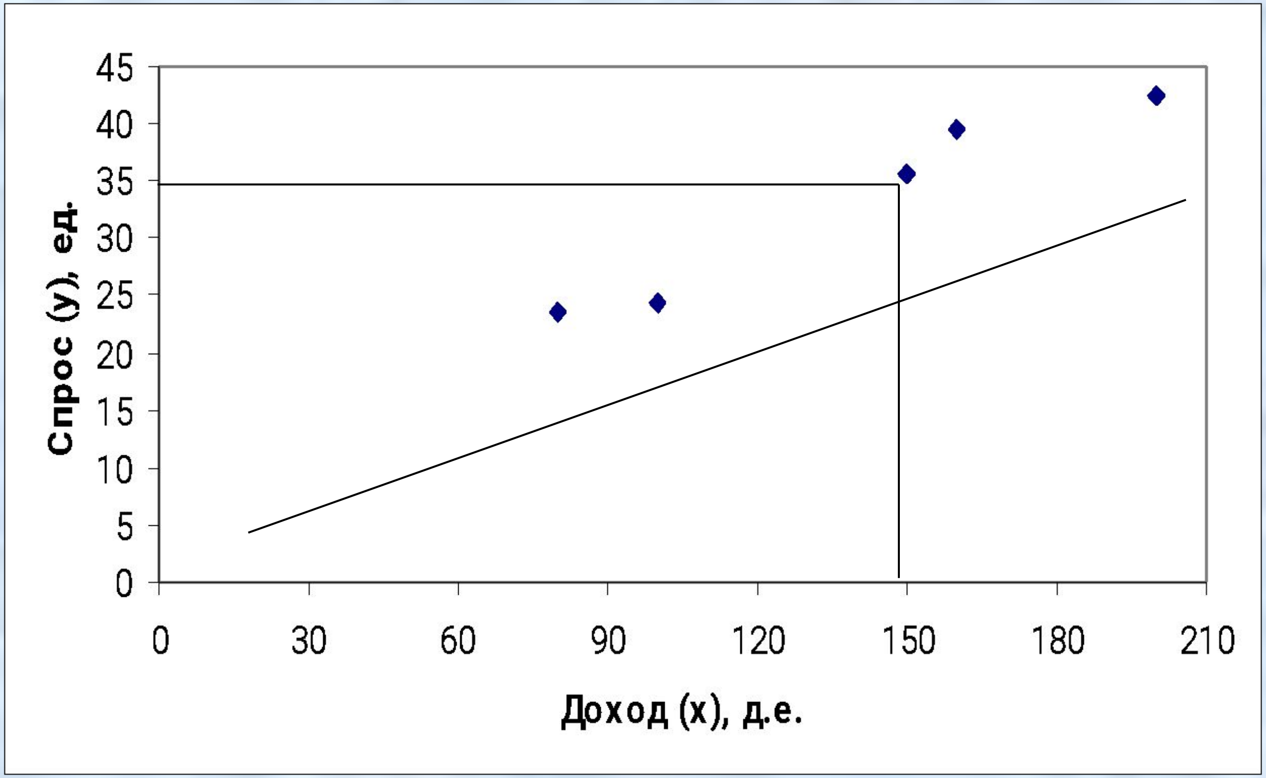


x	y
100	24
200	42
150	35
80	24
160	39

Расстояние от каждой точки до прямой должно быть как можно меньше!

Метод наименьших квадратов

Нанесем точки на график

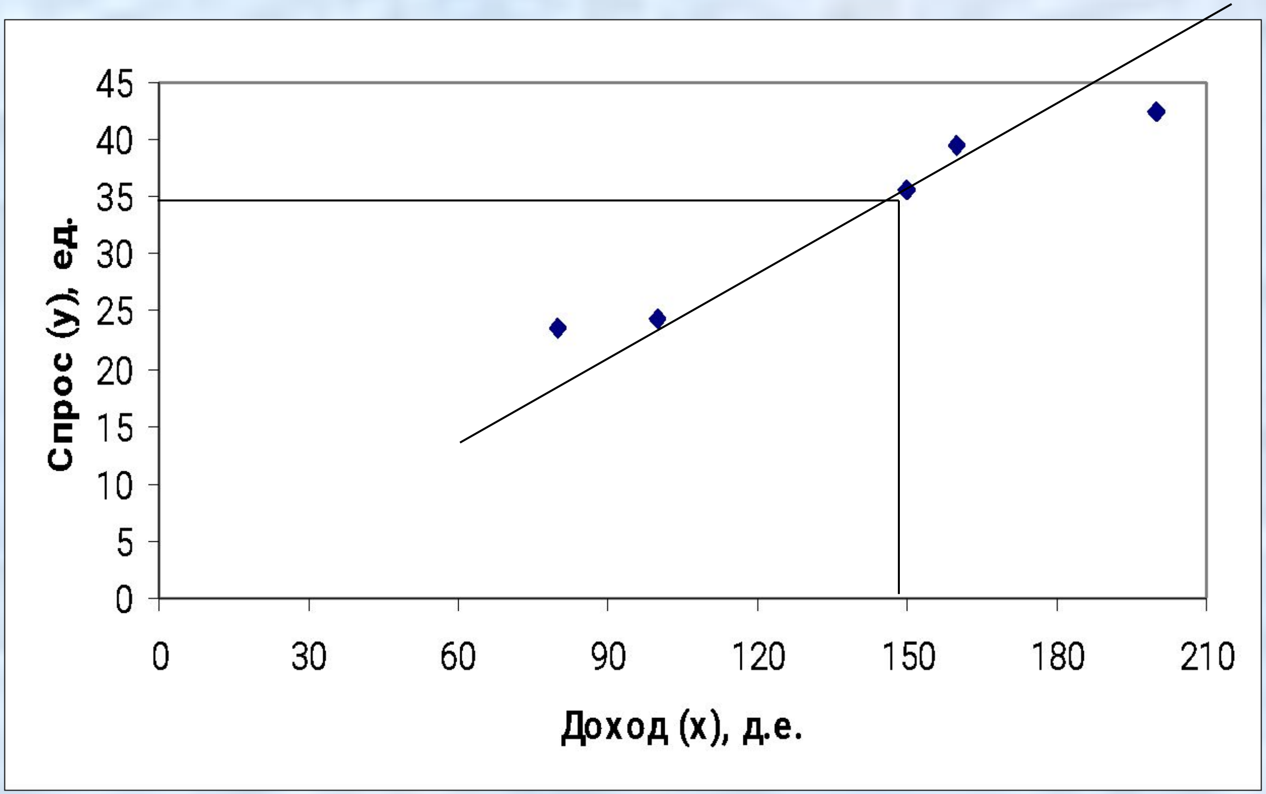


x	y
100	24
200	42
150	35
80	24
160	39

Плохая прямая!

Метод наименьших квадратов

Нанесем точки на график

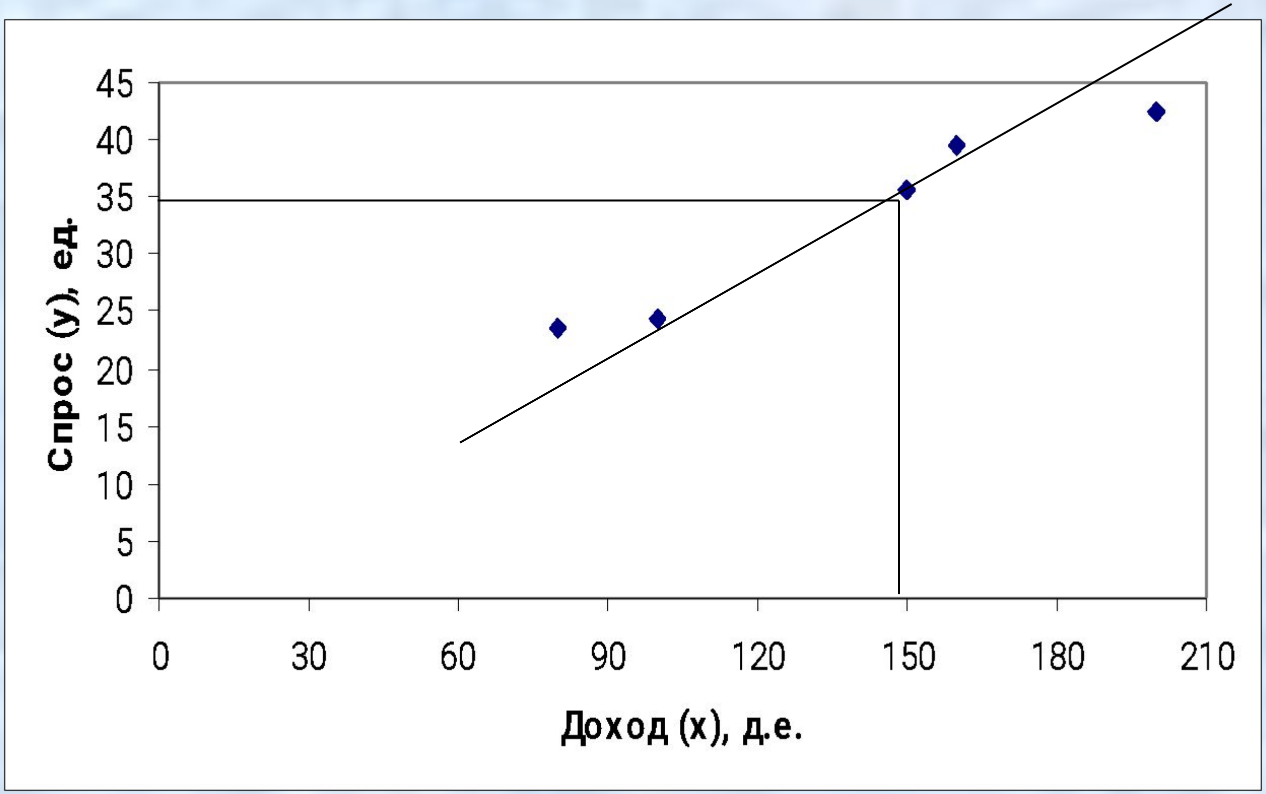


x	y
100	24
200	42
150	35
80	24
160	39

Хорошая прямая! Но может быть есть еще лучше?

Метод наименьших квадратов

Нанесем точки на график



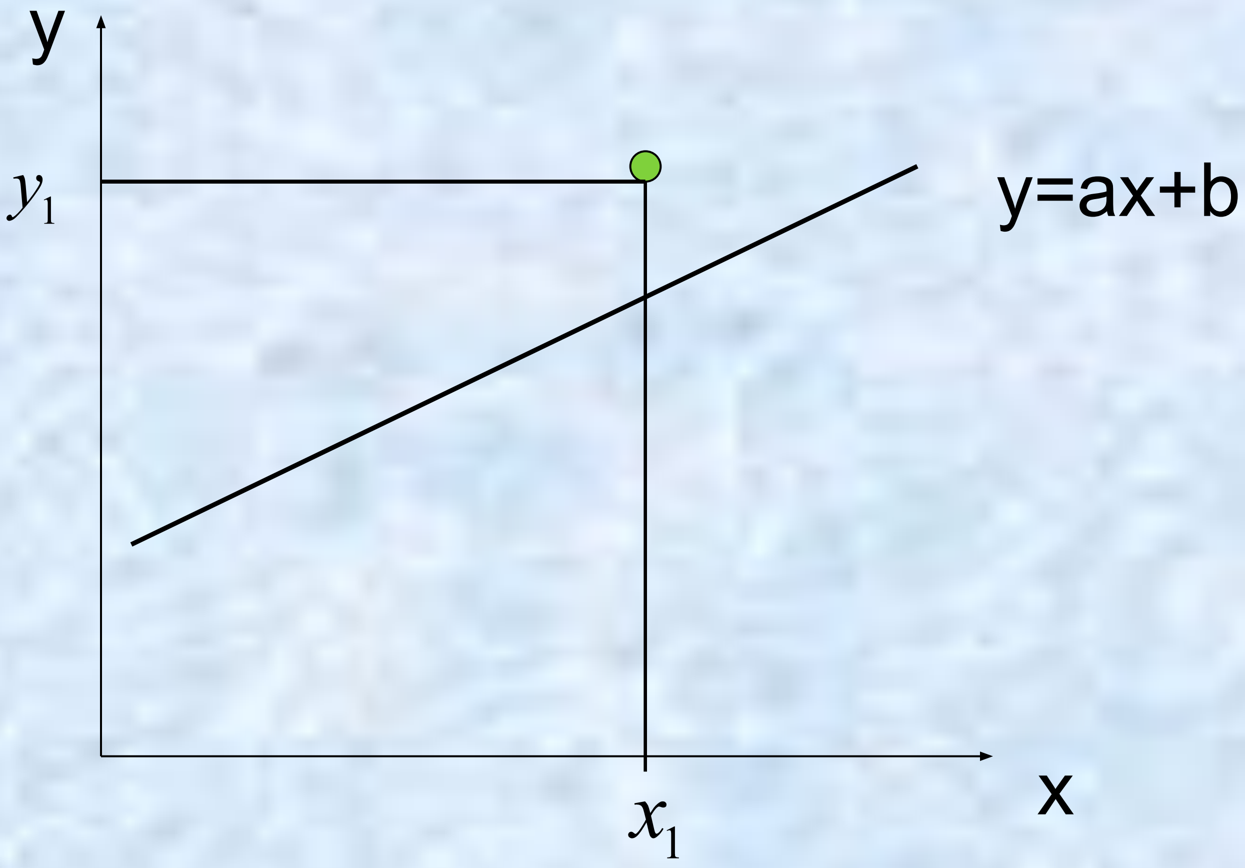
x	y
100	24
200	42
150	35
80	24
160	39

Уравнение прямой в общем виде y=ax+b. Надо найти наиболее подходящие a и b.

Обозначим

x_1 доход 1-го домохозяйства

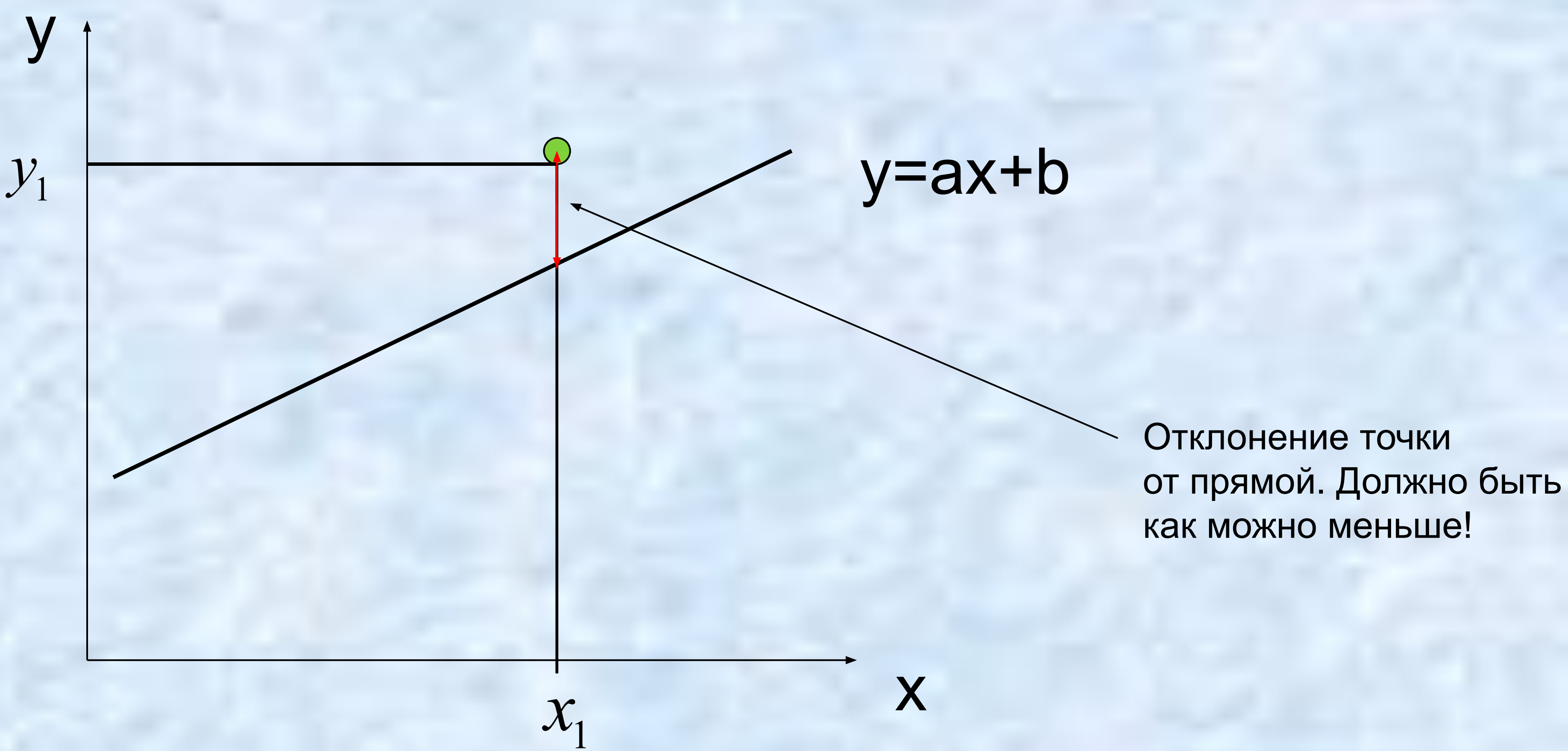
y_1 спрос 1-го домохозяйства на продукт



Обозначим

x_1 доход 1-го домохозяйства

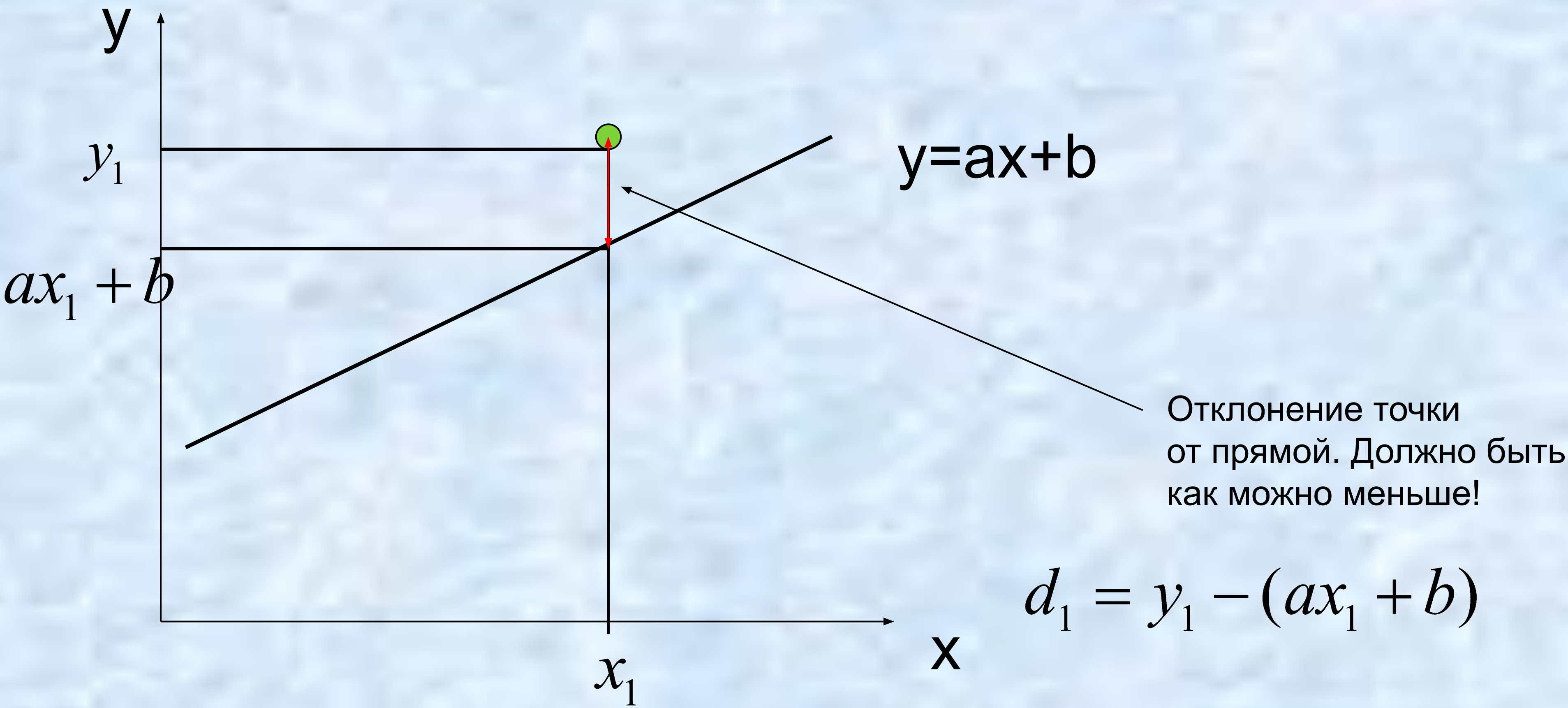
y_1 спрос 1-го домохозяйства на продукт



Обозначим

x_1 доход 1-го домохозяйства

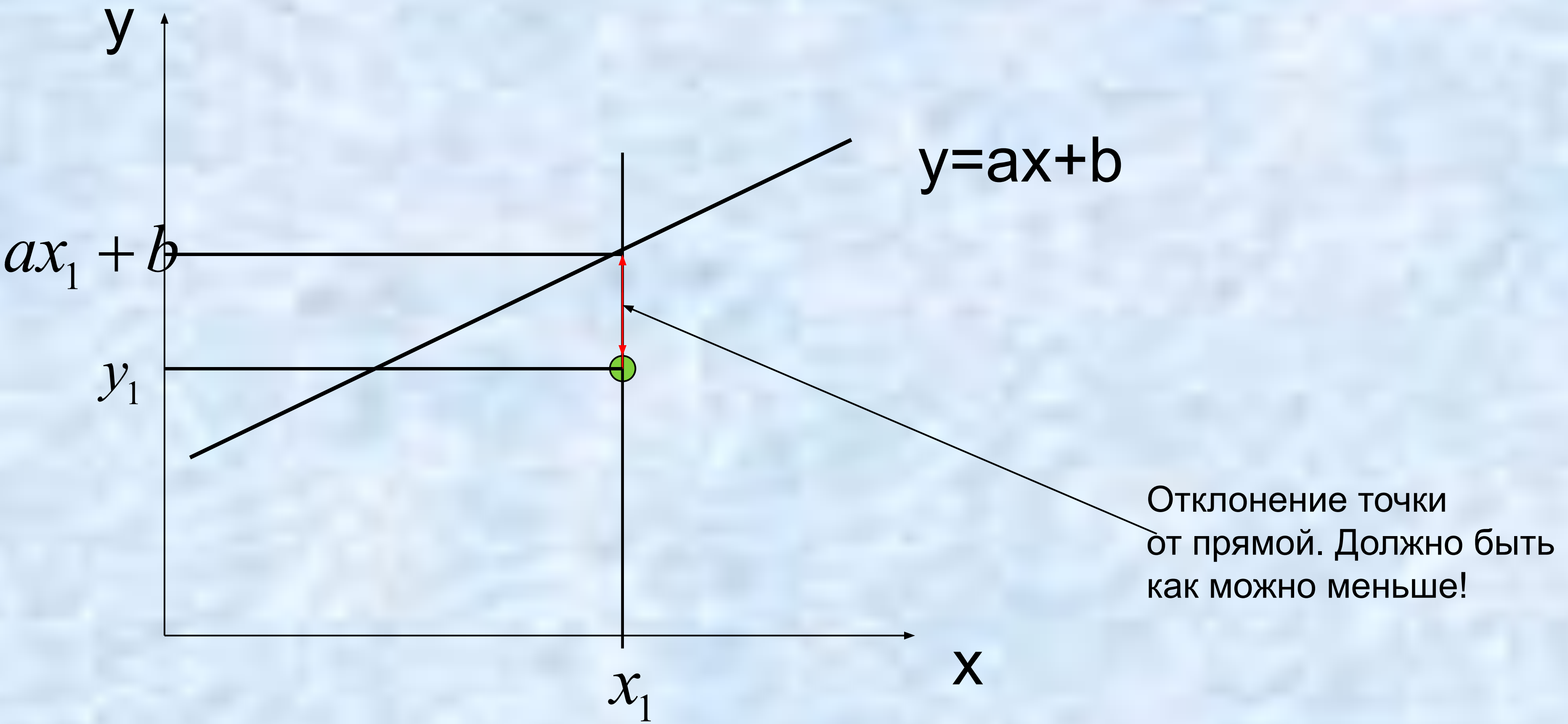
y_1 спрос 1-го домохозяйства на продукт



$$d_1 = y_1 - (ax_1 + b)$$

А если точка лежит ниже прямой?

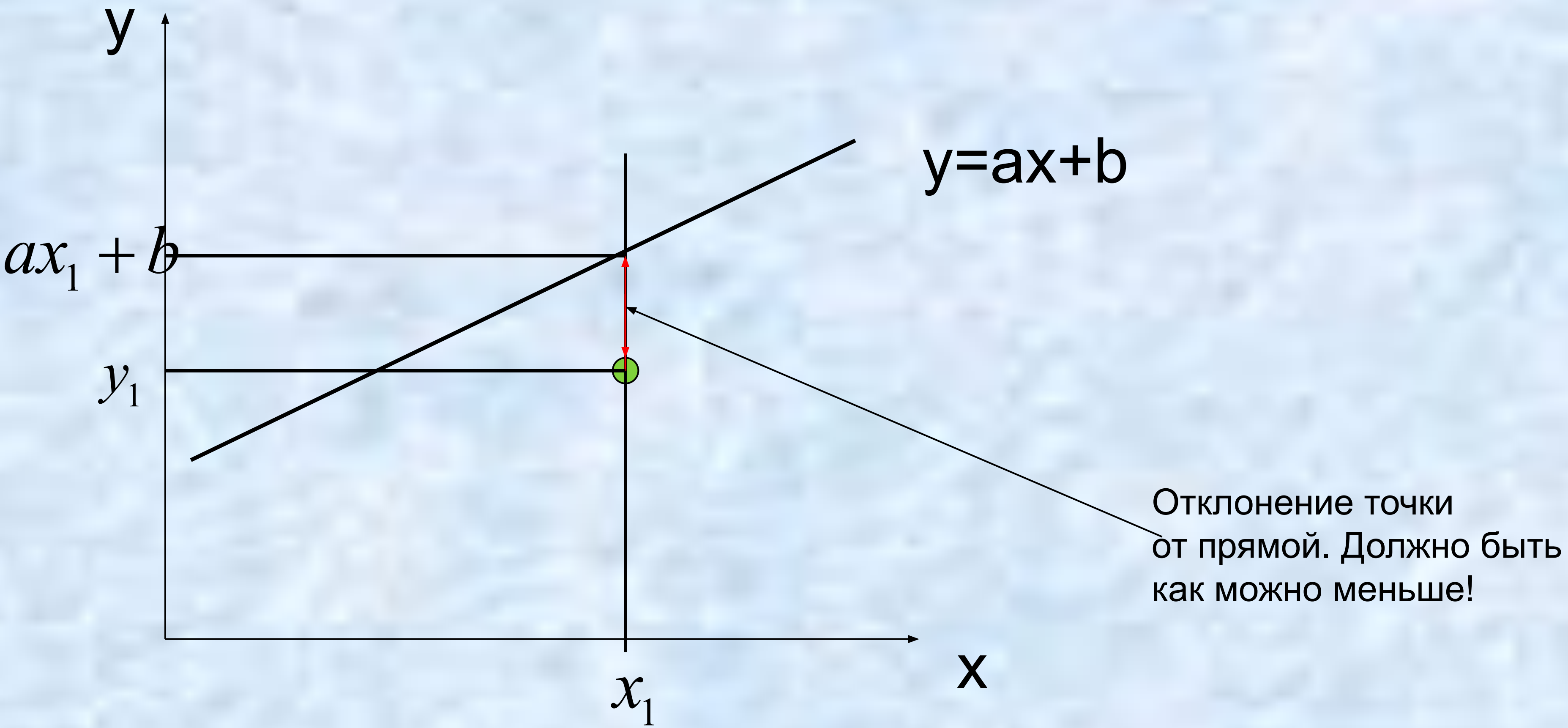
Тогда отклонение $$d_1 = (ax_1 + b) - y_1$$



Как учесть сразу оба случая?

Квадрат отклонения $d_1^2 = (y_1 - (ax_1 + b))^2$

должен быть как можно меньше.



Квадрат отклонения до второй точки тоже должен быть как можно меньше.

$$d_2^2 = (y_2 - (ax_2 + b))^2 \to \min$$

Квадрат отклонения до второй точки тоже должен быть как можно меньше.

$$d_2^2 = (y_2 - (ax_2 + b))^2 \to \min$$

И для третьей точки

$$d_3^2 = (y_3 - (ax_3 + b))^2 \to \min$$

Предположим, что у нас n точек.
Тогда и для последней точки

$$d_n^2 = (y_n - (ax_n + b))^2 \to \min$$

Как учесть все точки сразу?

$$d_1^2 + d_2^2 + d_3^2 + \ldots + d_n^2 \to \min$$

Сумма квадратов расстояний от точек до прямой должна быть как можно меньше.

Как учесть все точки сразу?

$$d_1^2 + d_2^2 + d_3^2 + \dots + d_n^2 \to \min$$

Сумма квадратов расстояний от точек до прямой должна быть как можно меньше.

$$d_1^2 + d_2^2 + d_3^2 + \dots + d_n^2 = \sum_{i=1}^{n} d_i^2$$

обозначение

Как учесть все точки сразу?

$$\sum_{i=1}^{n} d_i^2 \to \min$$

$$\sum_{i=1}^{n} (y_i - (ax_i + b))^2 \to \min$$

$$S(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

Получили функцию двух переменных, для которой надо найти минимум, т.е. надо исследовать на экстремум.

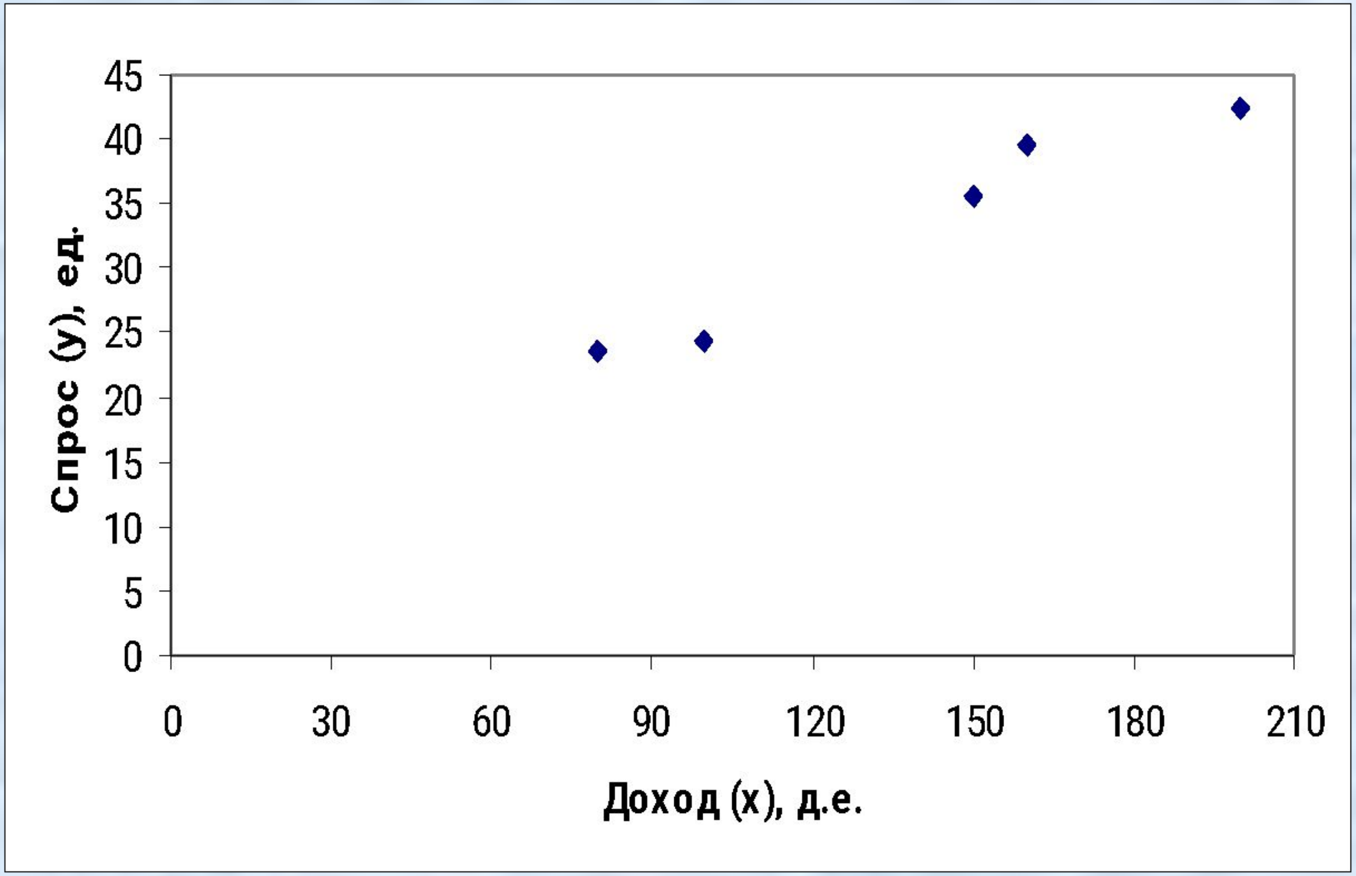
$$S(a,b) = \sum_{i=1}^{n} (y_i - a x_i - b)^2$$

x_i и y_i это просто числа, нам известные

$$a = \frac{\operatorname{cov}(x, y)}{s_x^2}$$

$$b = \overline{y} - a\overline{x}$$

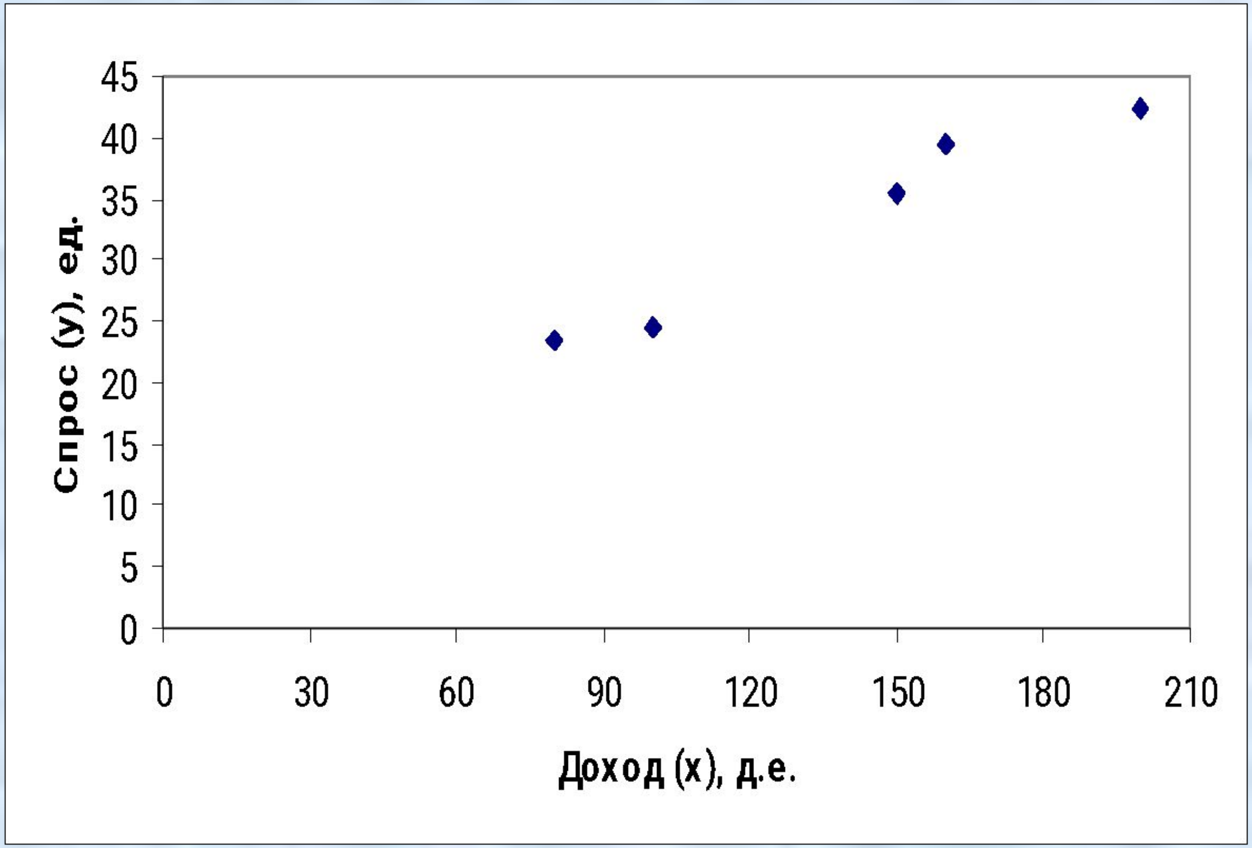
Вернемся к примеру



x	y
100	24
200	42
150	35
80	24
160	39

Надо найти $\overline{x}, \overline{y}, s_x^2, \operatorname{cov}(x, y)$

Вернемся к примеру



x	y
100	24
200	42
150	35
80	24
160	39

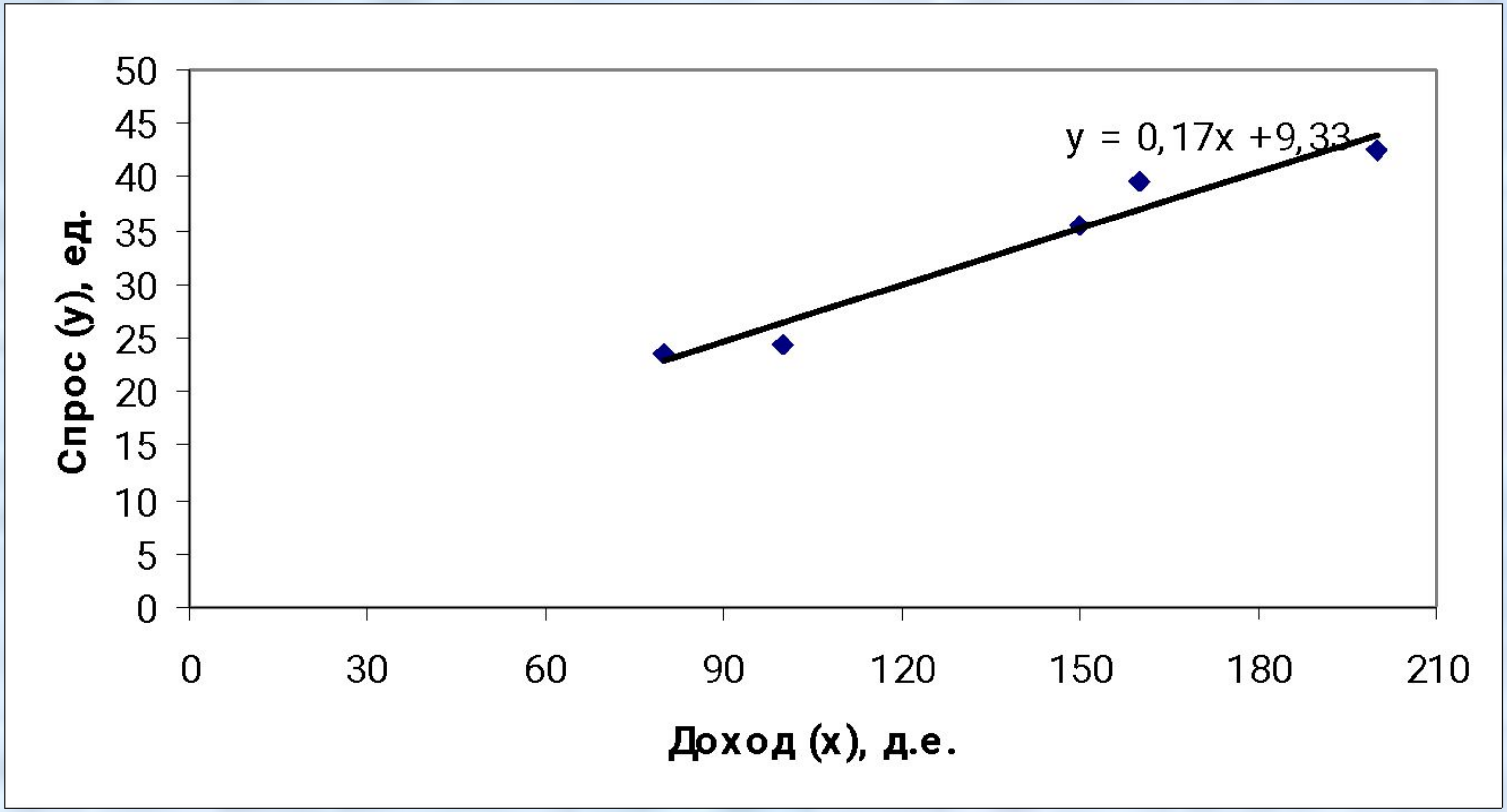
$$\bar{x} = 138,\ \bar{y} \approx 33,\ s_x^2 = 1856, \mathrm{cov}(x, y) = 315.6$$

$$\bar{x} = 138,\ \bar{y} \approx 33,\ s_x^2 = 1856, \operatorname{cov}(x, y) = 315.6$$

$$a = \frac{315.6}{1856} = 0.17 \qquad b = \bar{y} - a\bar{x}$$

a=0,17,   b=9,33

y=0,17x+9,33 - уравнение прямой, которая проходит ближе всего к точкам.



y=0,17x+9,33 - функция спроса в зависимости от дохода.

Интерпретация коэффициента a: при увеличении x на 1 ед. y увеличится на a единиц.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	9,334052	3,296116	2,831833	0,06609
Переменная X 1	0,170043	0,0228	7,458124	0,004991

y=0,17x+9,33 - функция спроса в зависимости от дохода.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	9,334052	3,296116	2,831833	0,06609
Переменная X 1	0,170043	0,0228	7,458124	0,004991

y=0,17x+9,33 - функция спроса в зависимости от дохода.
С ростом дохода на 1 ден.ед. спрос на товар растет на 0,17 ед.

Как оценить качество построенной модели?

Построим прогноз по модели по формуле $\hat{y} = 0,17x+9,33$

	A	B	C
1	x	y	прогноз y=0,17x+9,3
2	100	24	26,338
3	200	42	43,343
4	150	35	34,841
5	80	24	22,938
6	160	39	36,541

Как оценить качество построенной модели?

Вычисляем остатки $e = y - \hat{y}$

	A	B	C	D
1	x	y	прогноз y=0,17x+9,3	e
2	100	24	26,338	-2,338
3	200	42	43,343	-1,343
4	150	35	34,841	0,1595
5	80	24	22,938	1,0625
6	160	39	36,541	2,4591

Как оценить качество построенной модели?

Находим относительную ошибку аппроксимации

$$A = \frac{|y - \hat{y}|}{y}$$

	A	B	C	D	E
1	x	y	прогноз y=0,17x+9,3	e	Относитель ная ошибка
2	100	24	26,338	-2,338	9,74%
3	200	42	43,343	-1,343	3,20%
4	150	35	34,841	0,1595	0,46%
5	80	24	22,938	1,0625	4,43%
6	160	39	36,541	2,4591	6,31%

Процентный формат

Как оценить качество построенной модели?

Находим среднюю относительную ошибку аппроксимации

	A	B	C	D	E
1	x	y	прогноз y=0,17x+9,3	e	Относительная ошибка
2	100	24	26,338	-2,338	9,74%
3	200	42	43,343	-1,343	3,20%
4	150	35	34,841	0,1595	0,46%
5	80	24	22,938	1,0625	4,43%
6	160	39	36,541	2,4591	6,31%
7			Средняя относительная ошибка		4,83%

среднее по столбцу

В среднем прогноз отличается от наблюдаемого значения на 4,83%

Как оценить качество построенной модели?

Еще один показатель качества – коэффициент детерминации
Для его вычисления вычисляем сумму квадратов остатков ESS (Error Sum of Squares)

	A	B	C	D	E	F
1	x	y	прогноз y=0,17x+9,3	e	Относительная ошибка	e^2
2	100	24	26,338	-2,338	9,74%	5,468
3	200	42	43,343	-1,343	3,20%	1,803
4	150	35	34,841	0,1595	0,46%	0,025
5	80	24	22,938	1,0625	4,43%	1,129
6	160	39	36,541	2,4591	6,31%	6,047
7			Средняя относительная ошибка		4,83%	14,472

Сумма по столбцу

Как оценить качество построенной модели?

коэффициент детерминации

$$R^2 = 1 - \frac{ESS}{n \cdot s_y^2}$$

Как оценить качество построенной модели?

коэффициент детерминации

$$R^2 = 1 - \frac{ESS}{n \cdot s_y^2}$$

показывает долю вариации зависимой переменной, объясненную регрессией. Изменяется от 0 до 1

Чем ближе этот показатель к 1, тем лучше качество регрессии

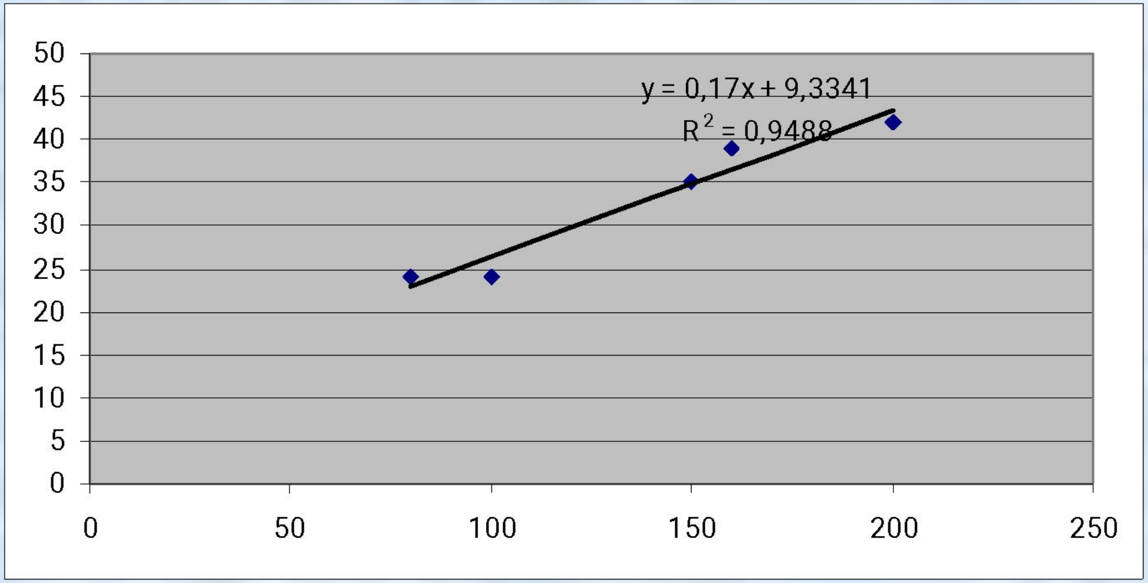
Как оценить качество построенной модели?

R2=	0,949	

94,9% вариации спроса на продукт объясняется доходом и остальные 5,1% прочими факторами, не включенными в модель

Как оценить качество построенной модели?

R2=	0,949



Проверка значимости коэффициентов модели регрессии

Построено уравнение $\hat{y} = ax + b$

Даже если в реальности у не зависит от х, уравнение можно построить. Но пользоваться им для прогноза нельзя.
В связи с этим проверяют значимость коэффициента *а*, т.е. насколько существенно *а* отличается от 0. Если коэффициент незначим, то переменная у не зависит от переменной х и моделью нельзя пользоваться

Проверка значимости коэффициентов модели регрессии

Построено уравнение $\hat{y} = ax + b$

Даже если в реальности y не зависит от x, уравнение можно построить.
Но пользоваться им для прогноза нельзя.
В связи с этим проверяют значимость коэффициента *a*, т.е. насколько существенно *a* отличается от 0. Если коэффициент незначим, то переменная y не зависит от переменной x и моделью нельзя пользоваться

Для проверки значимости коэффициента a рассчитывается величина

$$T_a = \frac{a}{s_a}$$

где, s_a - стандартная ошибка коэффициента a. Рассчитывается по специальным формулам

Проверка значимости коэффициентов модели регрессии

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересеч	9,334052	3,296116	2,831833	0,06609
Переменн	0,170043	0,0228	7,458124	0,004991

Проверка значимости коэффициентов модели регрессии

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересеч	9,334052	3,296116	2,831833	0,06609
X	0,170043	0,0228	7,458124	0,004991

На основе t-статистики рассчитывают Р-значение

Р-значение - это вероятность того, что переменная х не значима. При Р-значении меньще 0,05 обычно считают, что соответствующая переменная значима, т.е. у зависит от х

В этом примере переменная х значима, т.е. влияет на переменную у

y=0,17x+9,33 - функция спроса в зависимости от дохода.

1) *Выполнить прогноз потребления продукта домохозяйством с доходом 200 д.е.*

2) *Найти среднюю эластичность спроса по доходу*

$$\overline{E} = a\frac{\overline{x}}{\overline{y}}$$

y=0,17x+9,33 - функция спроса в зависимости от дохода.

1) Выполнить прогноз потребления продукта домохозяйством с доходом 200 д.е.

2) Найти среднюю эластичность спроса по доходу

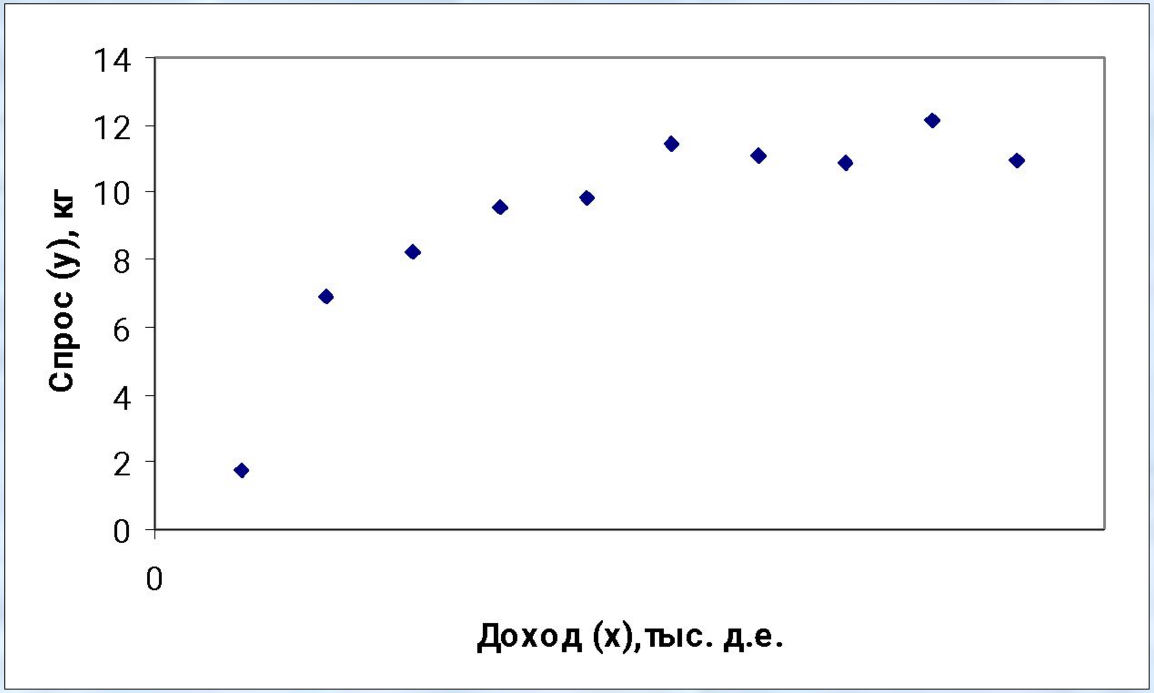
$$\bar{E} = a\frac{\bar{x}}{\bar{y}}$$

Среднее по x	Среднее по y
138,0	32,8

Эластичность	0,715

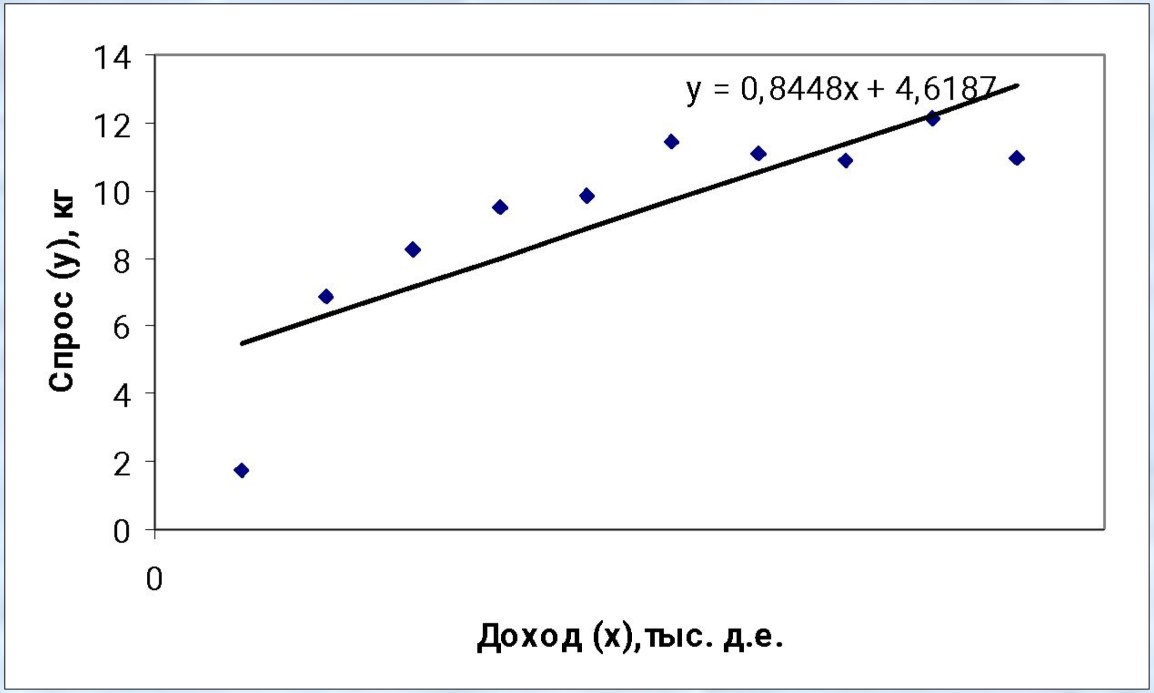
Модели парной нелинейной регрессии

№ 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

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

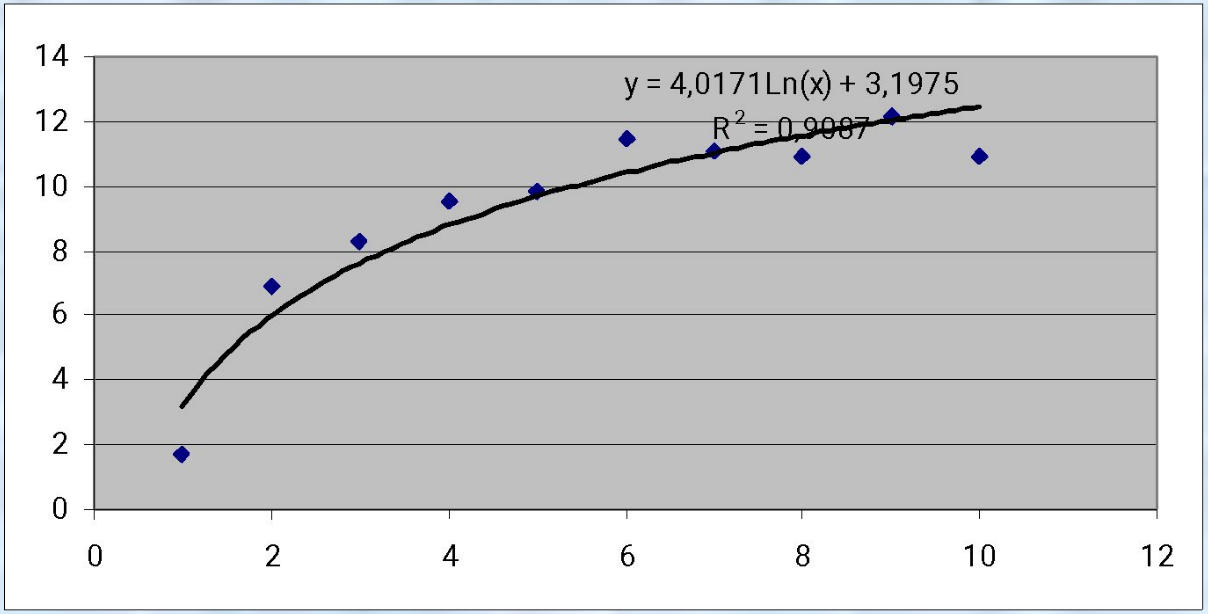
Используя сервис Анализ данных построим модель линейной регрессии, используя в качестве зависимой переменной y, а в качестве независимой ln(x).

	Коэффициенты	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=4.017ln(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=4.017ln(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=4.017ln(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%

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).

	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

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

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)=0.701ln(x)+1.063

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

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

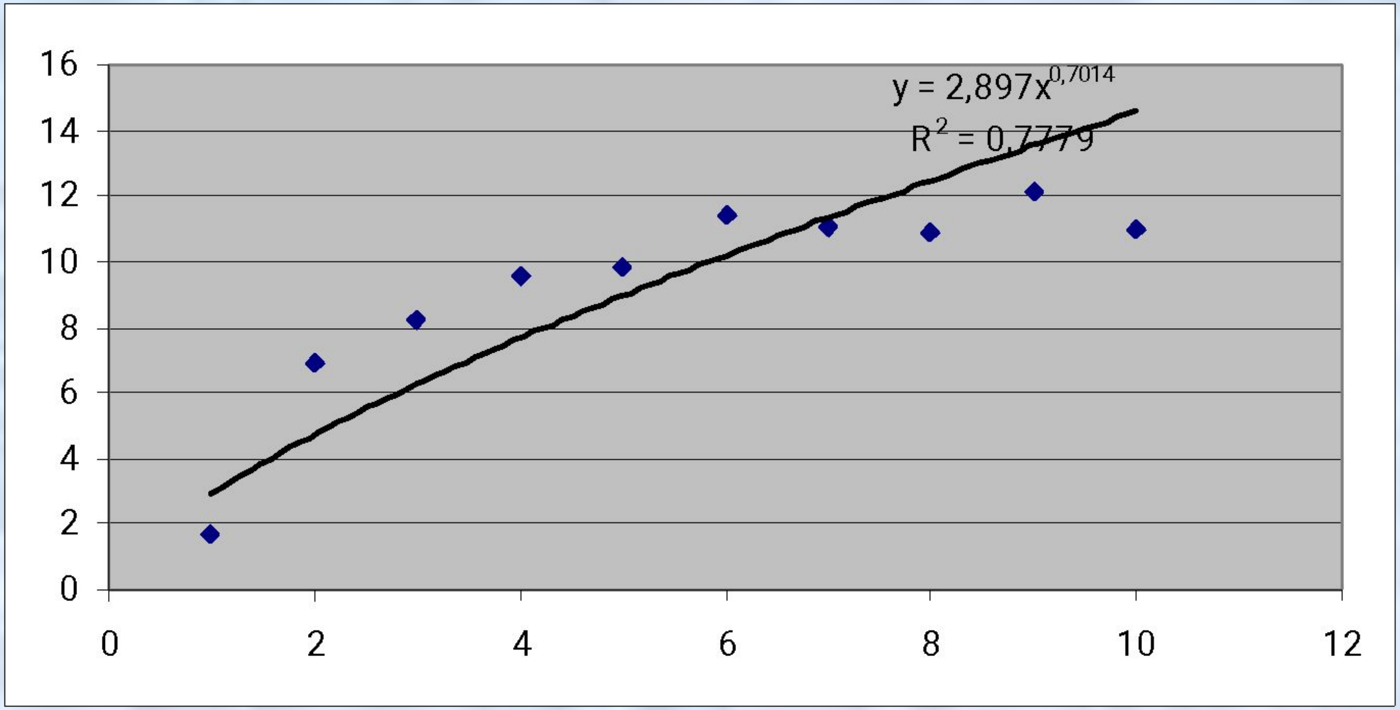
Используя сервис Анализ данных построим модель линейной регрессии, используя в качестве зависимой переменной ln(y), а в качестве независимой ln(x).

	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)=0.701ln(x)+1.063

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

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



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

Также как в линейной модели рассчитывается средняя относительная ошибка аппроксимации

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

A	B	C	D	E	F	G
x	ln(x)	y	ln(y)	прогноз	e	Относительная ошибка
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%