

Регрессионные модели с переменной структурой (фиктивные переменные)

1

1. ПОНЯТИЕ О ФИКТИВНЫХ ПЕРЕМЕННЫХ
2. ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ПРИ АНАЛИЗЕ ПРОСТРАНСТВЕННЫХ ДАННЫХ
3. ФИКТИВНЫЕ ПЕРЕМЕННЫЕ ПРИ АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

Понятие о фиктивных переменных

Dummy variables -

фиктивные переменные, искусственные, структурные, бинарные, булевые, манекены, индикаторы, переменные переключатели, дискретные переменные

$$D = \begin{cases} 0, & \text{фактор не действует} \\ 1, & \text{фактор действует} \end{cases}$$

Фиктивные переменные в регрессионных моделях

Пространственные данные

Временные ряды

Фиктивная переменная в
левой части модели

Фиктивная переменная в
правой части модели

линейная
вероятностная
модель (LPM-
модель)

Logit-модель

Tobit-модель

Модели
множественного
выбора

Регрессионные модели при
наличии у фиктивной переменной
двух альтернатив

Регрессионные модели при
наличии у качественных
переменных более двух
альтернатив

Регрессия с одной количественной
и двумя качественными
переменными

Выявление
сезонности

Выделение
временного
тренда

Моделирование
скачкообразных
структурных
сдвигов

Фиктивные переменные при анализе пространственных данных

5

- A. Регрессии с фиктивной переменной дифференциации свободного члена
- B. Регрессии с фиктивной переменной наклона
- C. Регрессии с фиктивными переменными: общий случай

ANCOVA-модель при наличии у фиктивной переменной двух альтернатив

$$\tilde{y}_i = a_0 + a_1 x_i + \alpha D_i + \varepsilon_i$$

y - заработная плата сотрудника фирмы;

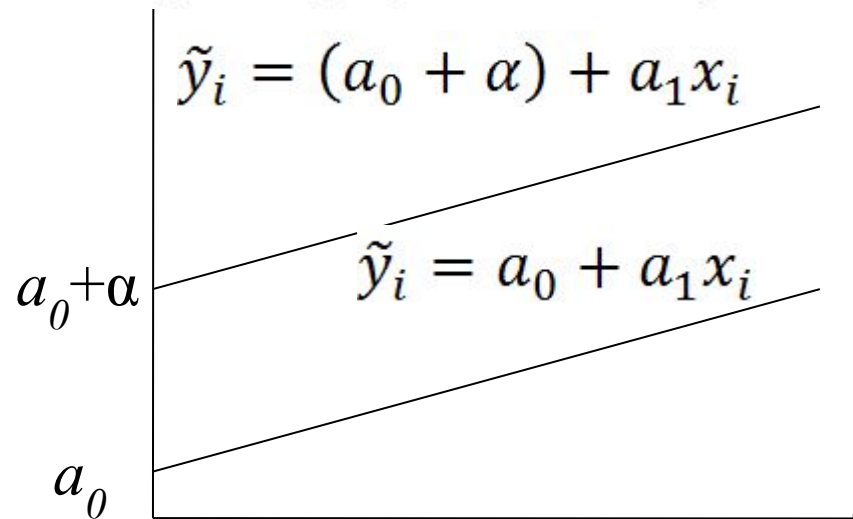
x - стаж сотрудника;

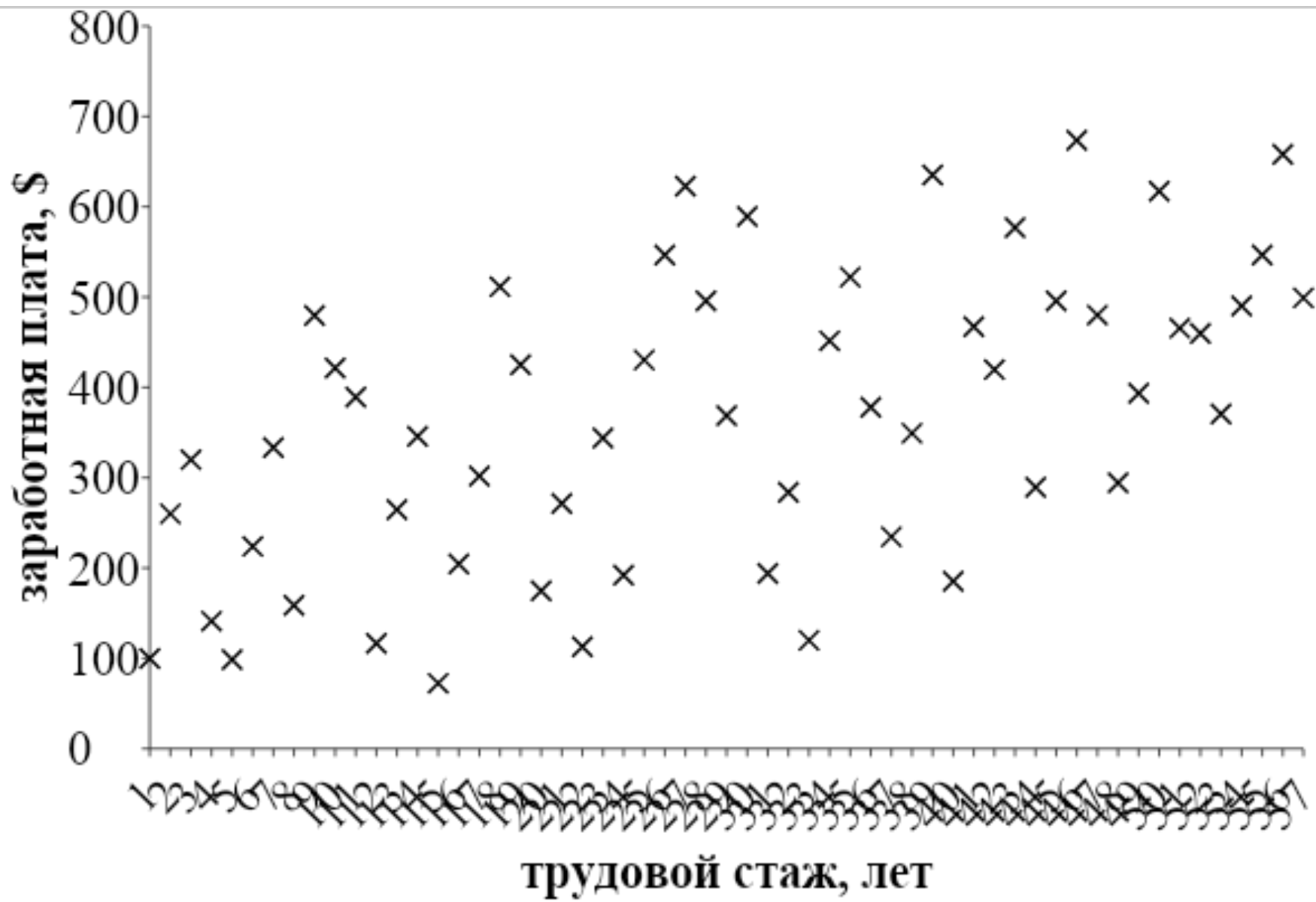
D - пол сотрудника.

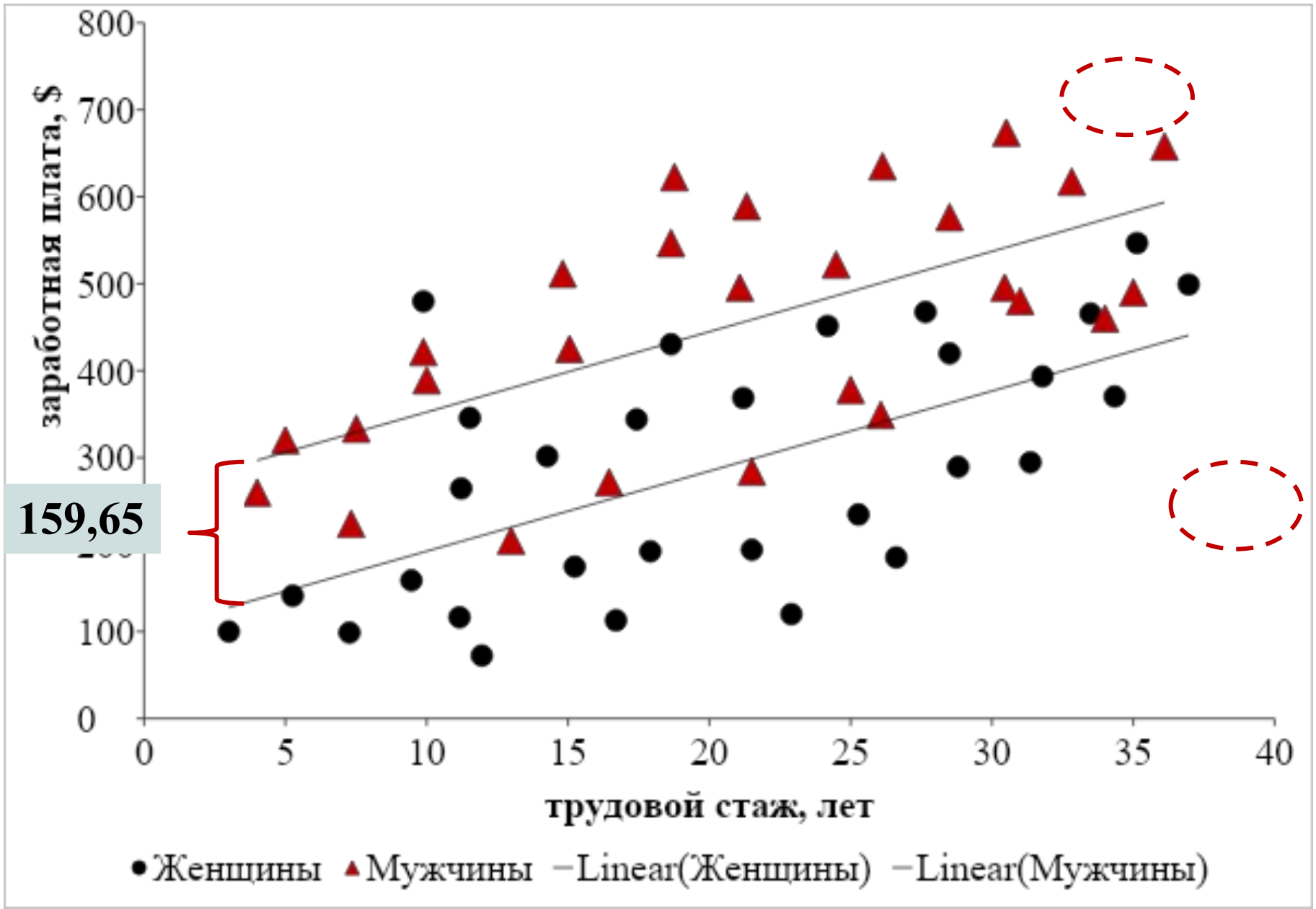
$$D = \begin{cases} 0, & \text{если сотрудник - женщина} \\ 1, & \text{если сотрудник - мужчина} \end{cases}$$

$f(y|x, D = 0) = a_0 + a_1 x_i$ для женщин

$f(y|x, D = 1) = a_0 + a_1 x_i + \alpha = (a_0 + \alpha) + a_1 x_i$ для мужчин







Оценки регрессии

9

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Свободный член	259,44	38,30	6,77	0,00
X	9,25	1,52	6,07	0,00
D	-160,05	29,20	-5,48	0,00

Модели ANCOVA при наличии у качественных переменных более двух альтернатив

$$\tilde{y}_i = a_0 + a_1 x_i + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \varepsilon_i$$

y – расходы; x – доходы домохозяйств.

$$D_1 = \begin{cases} 0, & \text{если дошкольник} \\ 1, & \text{в противном случае} \end{cases}$$

$$D_2 = \begin{cases} 0, & \text{если дошкольник или младший школьник} \\ 1, & \text{в противном случае} \end{cases}$$

Средний расход на дошкольника:

$$f(y|D_1 = 0, D_2 = 0) = a_0 + a_1 x_i$$

$$\tilde{y}_i = (a_0 + \alpha_1 + \alpha_2) + a_1 x_i$$

Средний расход на младшего школьника:

$$f(y|D_1 = 1, D_2 = 0) = (a_0 + \alpha_1) + a_1 x_i$$

$$\tilde{y}_i = (a_0 + \alpha_1) + a_1 x_i$$

Средний расход на старшего школьника:

$$f(y|D_1 = 1, D_2 = 1) = (a_0 + \alpha_1 + \alpha_2) + a_1 x_i$$

$$\tilde{y}_i = a_0 + a_1 x_i$$

Регрессия с одной количественной и двумя качественными переменными

y - заработная плата сотрудников фирмы, x - стаж работы,

D_1 - наличие высшего образования, D_2 - пол сотрудника

$$D_1 = \begin{cases} 0, & \text{если сотрудник - женщина} \\ 1, & \text{если сотрудник - мужчина} \end{cases}$$
$$D_2 = \begin{cases} 0, & \text{если нет высшего образования} \\ 1, & \text{в противном случае} \end{cases}$$
$$\tilde{y}_i = a_0 + a_1 x_i + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \varepsilon_i$$

Средняя заработная плата женщины без высшего образования:

$$f(y|D_1 = 0, D_2 = 0) = a_0 + a_1 x_i$$

Средняя заработная плата женщины с высшим образованием:

$$f(y|D_1 = 0, D_2 = 1) = (a_0 + \alpha_2) + a_1 x_i$$

Средняя заработная плата мужчины без высшего образования:

$$f(y|D_1 = 1, D_2 = 0) = (a_0 + \alpha_1) + a_1 x_i$$

Средняя заработная плата мужчины с высшим образованием:

$$f(y|D_1 = 1, D_2 = 1) = (a_0 + \alpha_1 + \alpha_2) + a_1 x_i$$

Фиктивные переменные при анализе пространственных данных

12

- A. Регрессии с фиктивной переменной дифференциации свободного члена
- B. Регрессии с фиктивной переменной наклона
- C. Регрессии с фиктивными переменными: общий случай

Дифференциальный коэффициент наклона прямой (фиктивная переменная наклона)

$$\tilde{y}_i = a_0 + a_1 x_i + \beta D_i x_i + \varepsilon_i$$

$$D_i x_i = \begin{cases} x_i & \text{если } D_i = 1 \\ 0 & \text{в остальных случаях} \end{cases}$$

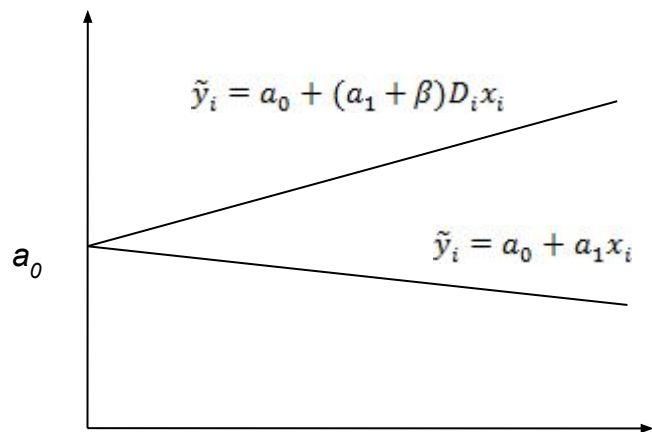
фиктивная переменная наклона

базовая категория

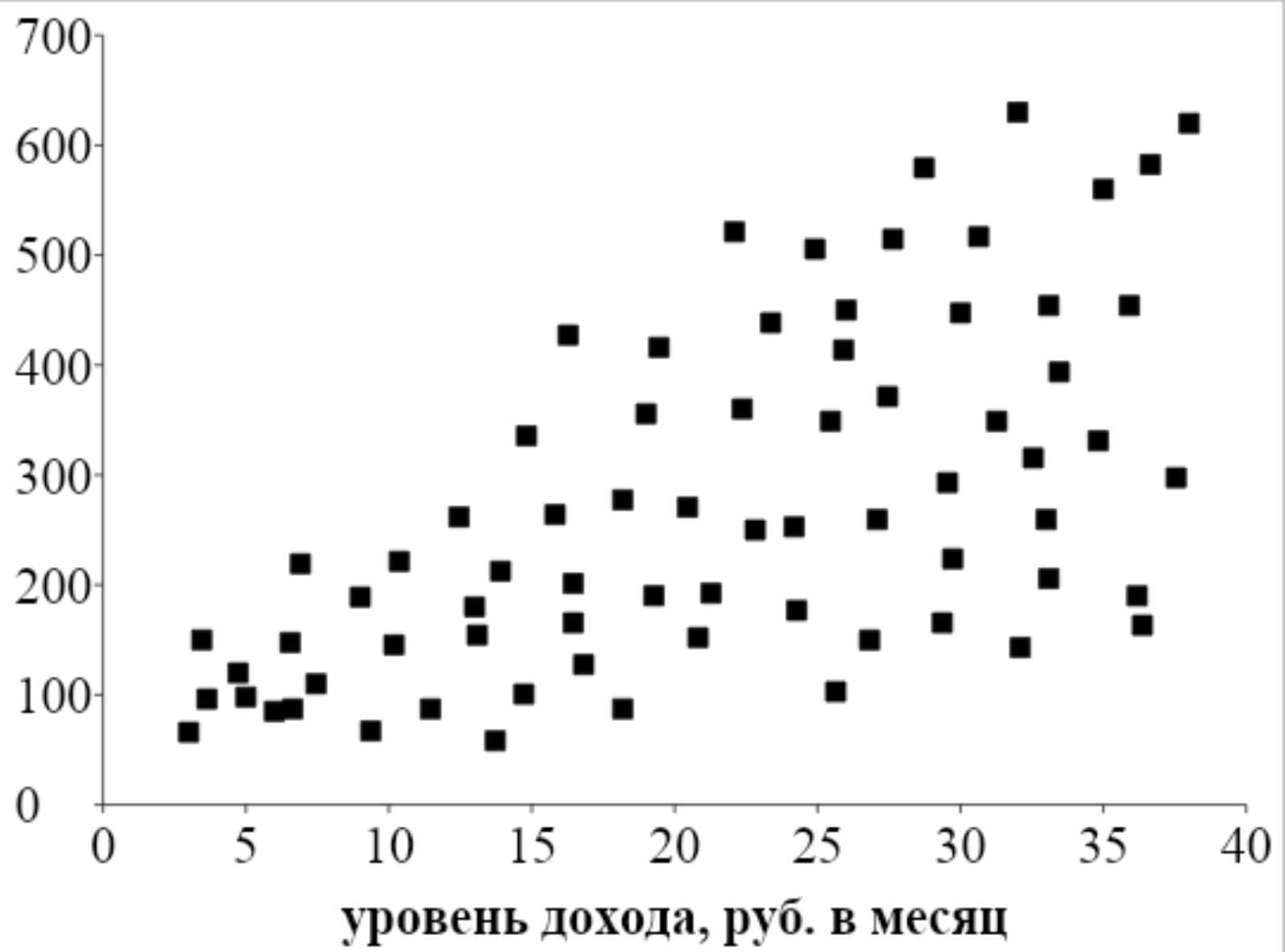
$$f(y|x, D = 0) = a_0 + a_1 x_i$$

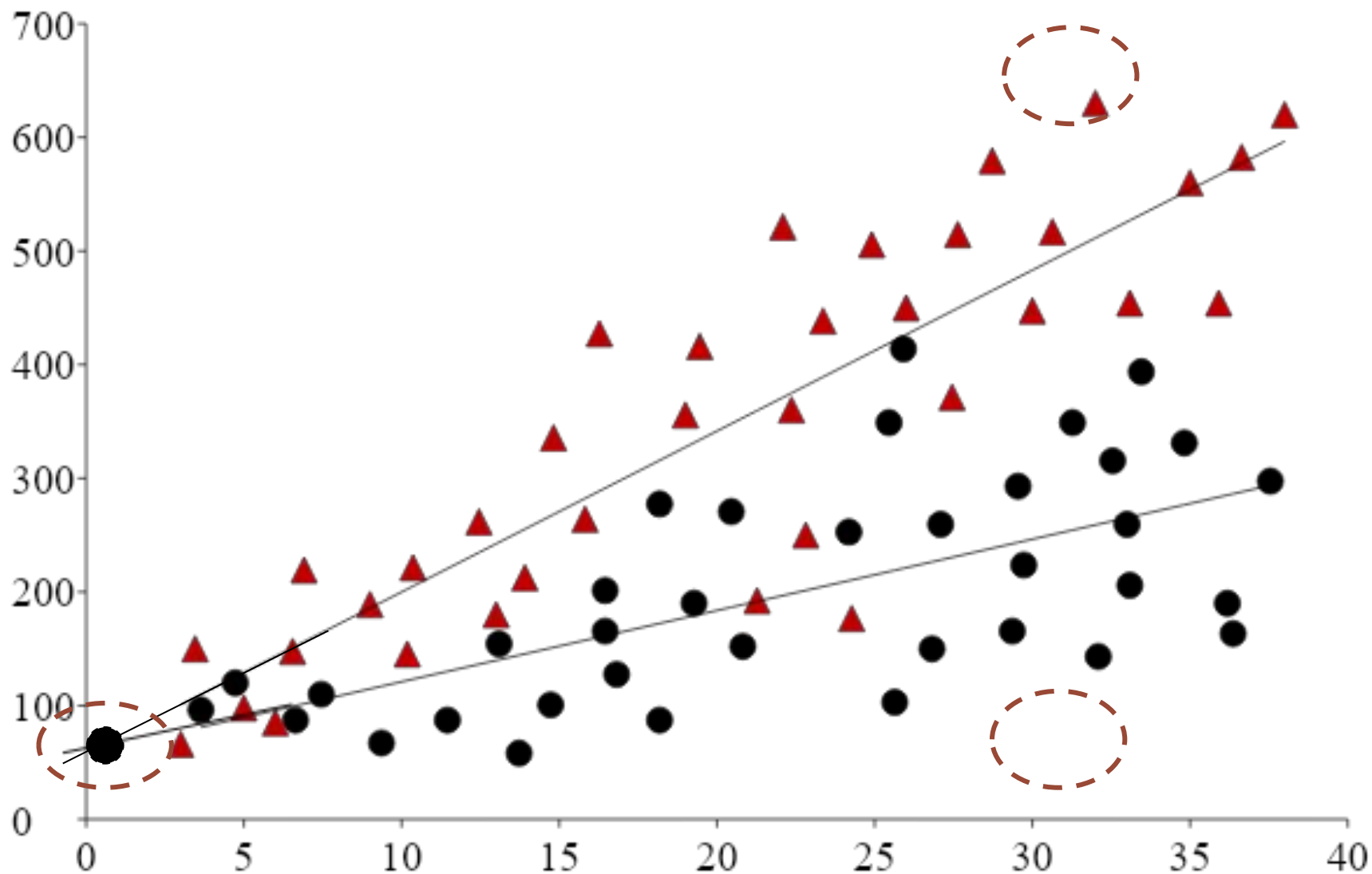
альтернативная категория

$$f(y|x, D = 1) = a_0 + (a_1 + \beta) D_i x_i$$



расходы на одежду, руб. в месяц

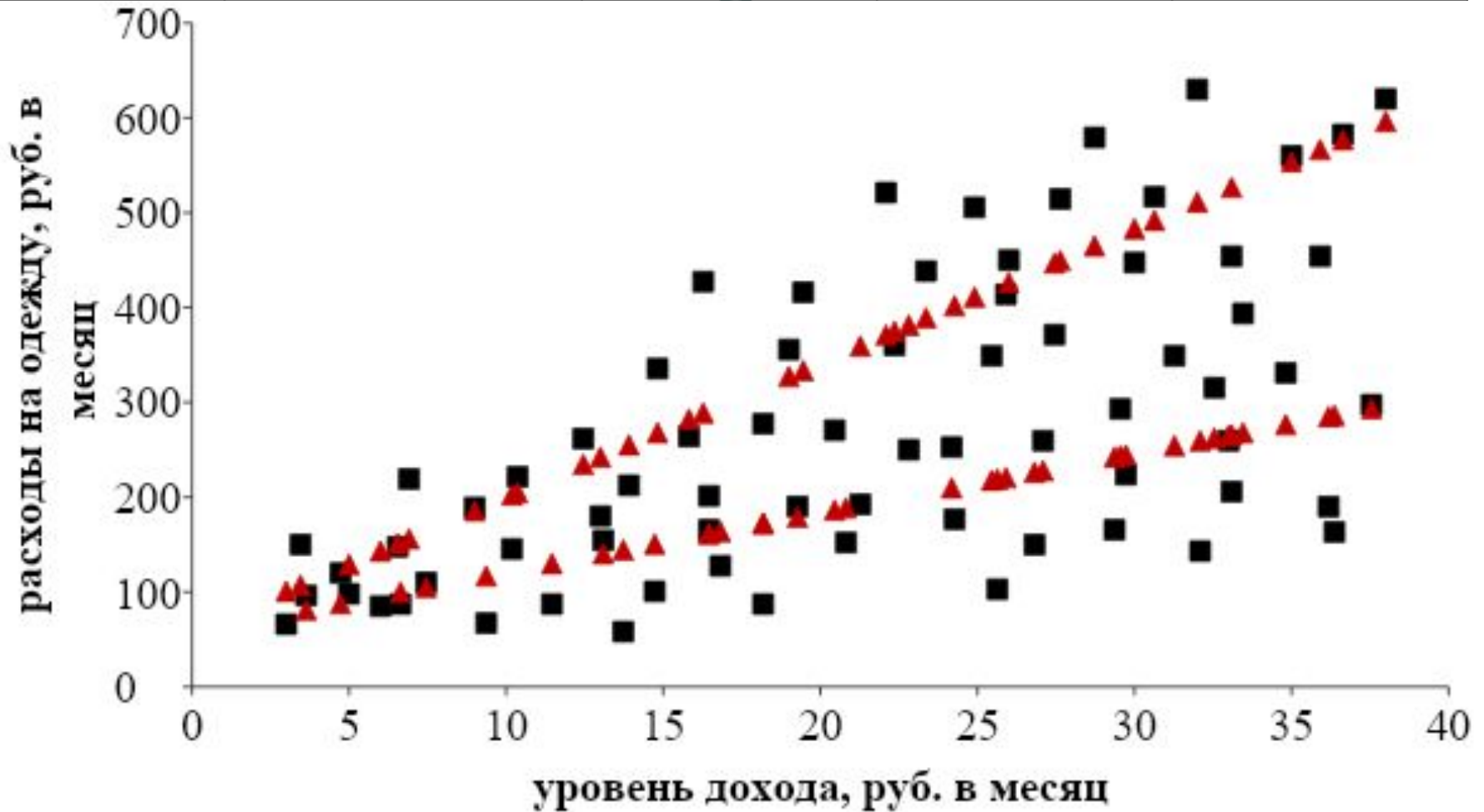




● Мужчины ▲ Женщины - Linear(Мужчины) - Linear(Женщины)

Оценки регрессии

16



Фиктивные переменные при анализе пространственных данных

17

- A. Регрессии с фиктивной переменной дифференциации свободного члена
- B. Регрессии с фиктивной переменной наклона
- C. Регрессии с фиктивными переменными: общий случай

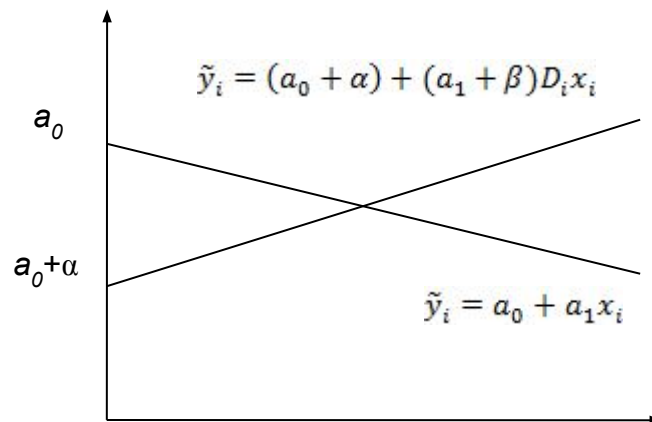
Общий случай

18

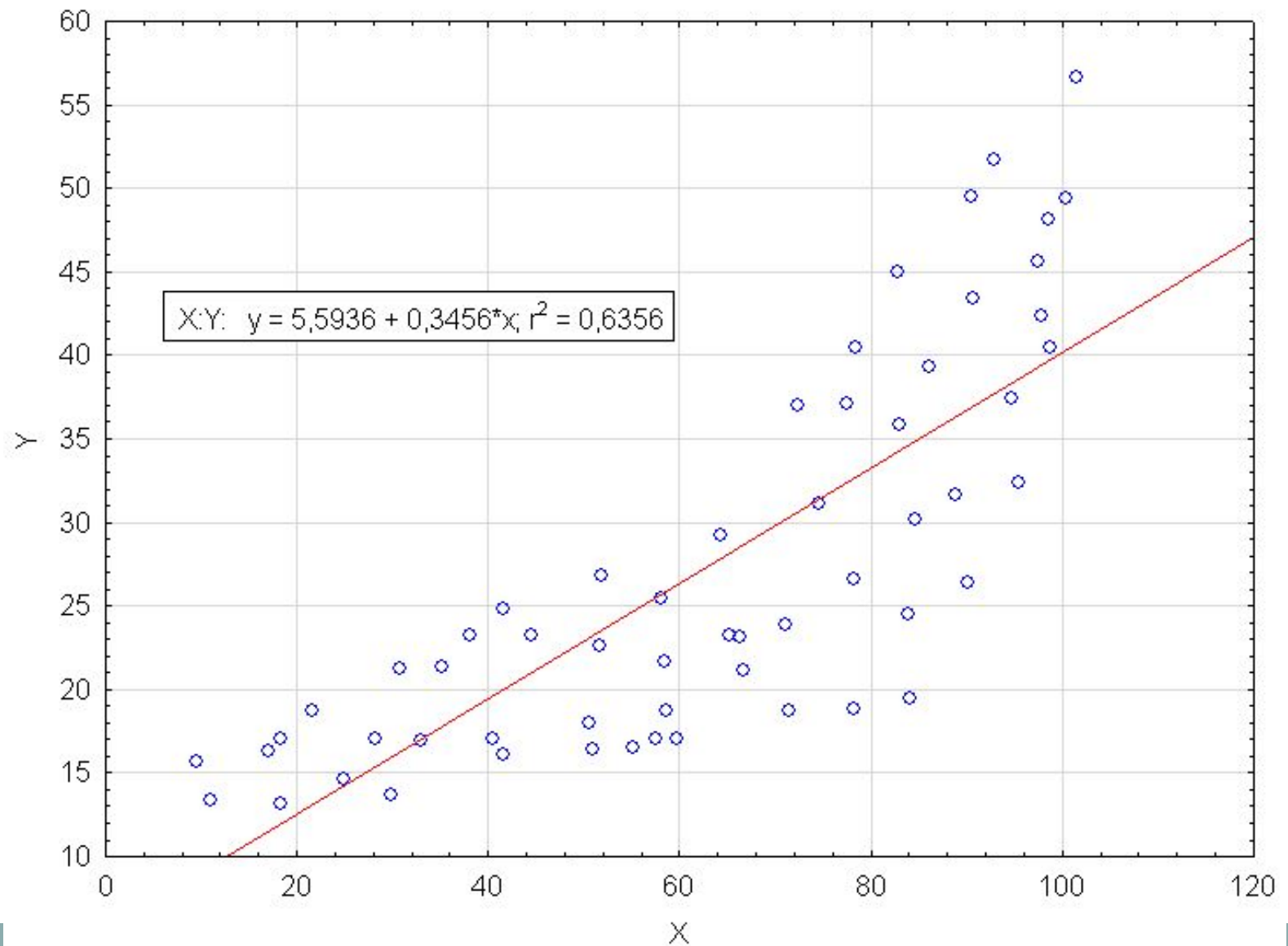
$$\tilde{y}_i = a_0 + a_1x_i + \alpha D_i + \beta D_i x_i + \varepsilon_i$$

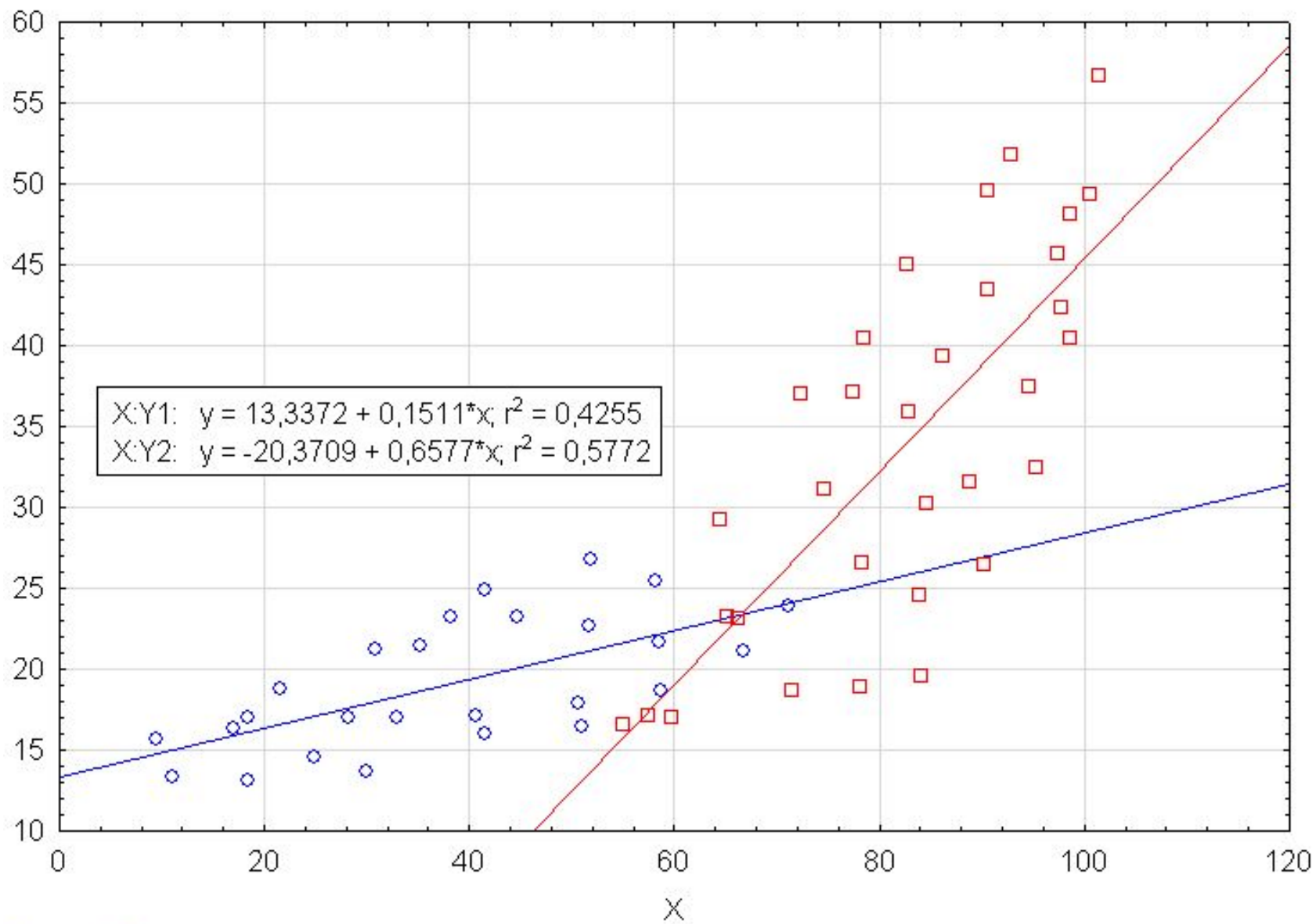
D – фиктивная переменная дифференциации

Dx - фиктивная переменная наклона



Пример структурного сдвига





Y1 Y2

Тест (критерий) Г. Чоу

1 этап: $S_{ост}^{к-л} = S_{ост}^1 + S_{ост}^2$

2 этап: $\Delta S_{ост} = S_{ост}^3 - S_{ост}^{к-л}$

Условные обозначения для алгоритма теста Чоу

Периоды	Число наблюдений в совокупности	Остаточная сумма квадратов	Число параметров в уравнении	Число степеней свободы остаточной дисперсии
Первое уравнение	T_1	$S_{ост}^1$	m_1	$T_1 - m_1$
Второе уравнение	T_2	$S_{ост}^2$	m_2	$T_2 - m_2$
Объединенное уравнение	T	$S_{ост}^3$	m_3	$T - m_3 = (T_1 + T_2) - m_3$

3 этап: $F_{факт} = \frac{Q_{\Delta S}}{Q_{к-л}} = \frac{\Delta S_{ост}}{m_1 + m_2 - m_3} / \frac{S_{ост}^{к-л}}{T - m_1 - m_2}$

Регрессия по все
совокупности :

Эффект	Сумма квадр.	сс	Средн. квадр.	F	р-знач.
Регресс.	4849,39	1,00	4849,39	97,66	0,00
Остатки	2780,71	56,00	49,66		
Итого	7630,10				

$$F_{\phi} = \frac{(2780,71 - 227,45 - 1737,25)}{(227,45 + 1737,25)} \bigg/ \frac{(1+1)}{(58-1-1)} = 11,63 > 3,16$$

Регрессия по
первой подвыборке
совокупности :

Эффект	Сумма квадр.	сс	Средн. квадр.	F	р-знач.
Регресс.	168,45	1,00	168,45	17,77	0,000
Остатки	227,45	24,00	9,48		
Итого	395,90				


Регрессия по
второй подвыборке
совокупности :

Эффект	Сумма квадр.	сс	Средн. квадр.	F	р-знач.
Регресс.	2371,32	1,00	2371,32	40,95	0,00
Остатки	1737,25	30,00	57,91		
Итого	4108,56				


Фиктивные переменные при анализе временных рядов

**В регрессионных моделях с временными рядами
используется :**


24



Переменные-
индикаторы
принадлежности
наблюдения к
определенному
периоду



Сезонные
переменные



Линейный
временной
тренд

Единственный временной тренд

Нечетное число членов

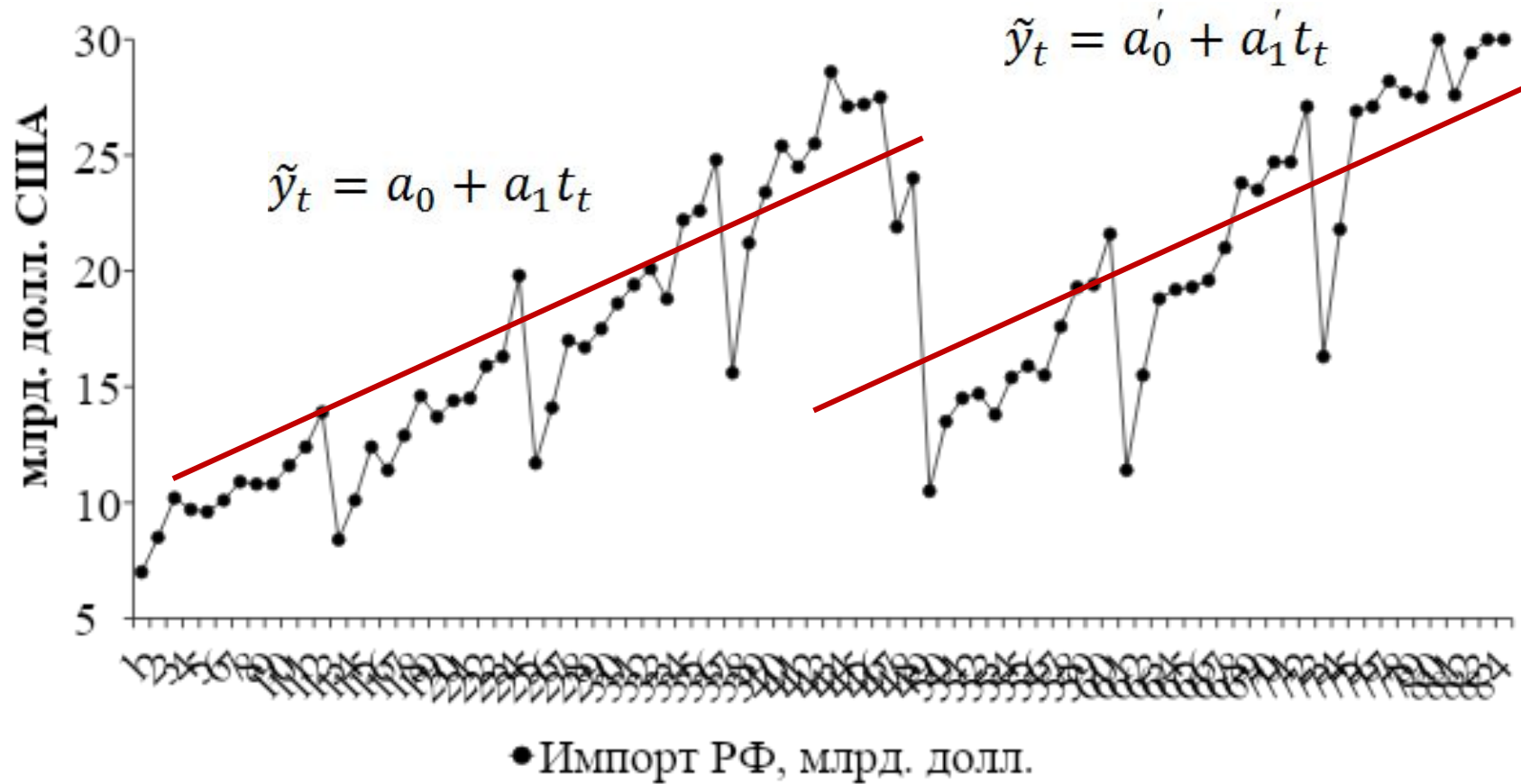
Дата	y	t	t'
1996г.	2,3	1	-4
1997г.	3,8	2	-3
1998г.	6,5	3	-2
1999г.	7,4	4	-1
2000г.	10,2	5	0
2001г.	10,5	6	1
2002г.	12,1	7	2
2003г.	13,2	8	3
2004г.	13,6	9	4

Четное число членов

Дата	y	t	t'
1996г.	2,3	1	-9
1997г.	3,8	2	-7
1998г.	6,5	3	-5
1999г.	7,4	4	-3
2000г.	10,2	5	-1
2001г.	10,5	6	1
2002г.	12,1	7	3
2003г.	13,2	8	5
2004г.	13,6	9	7
2005г.	14,5	10	9

Модель «краха» или «прорыва»

26



Оценки регрессии

27

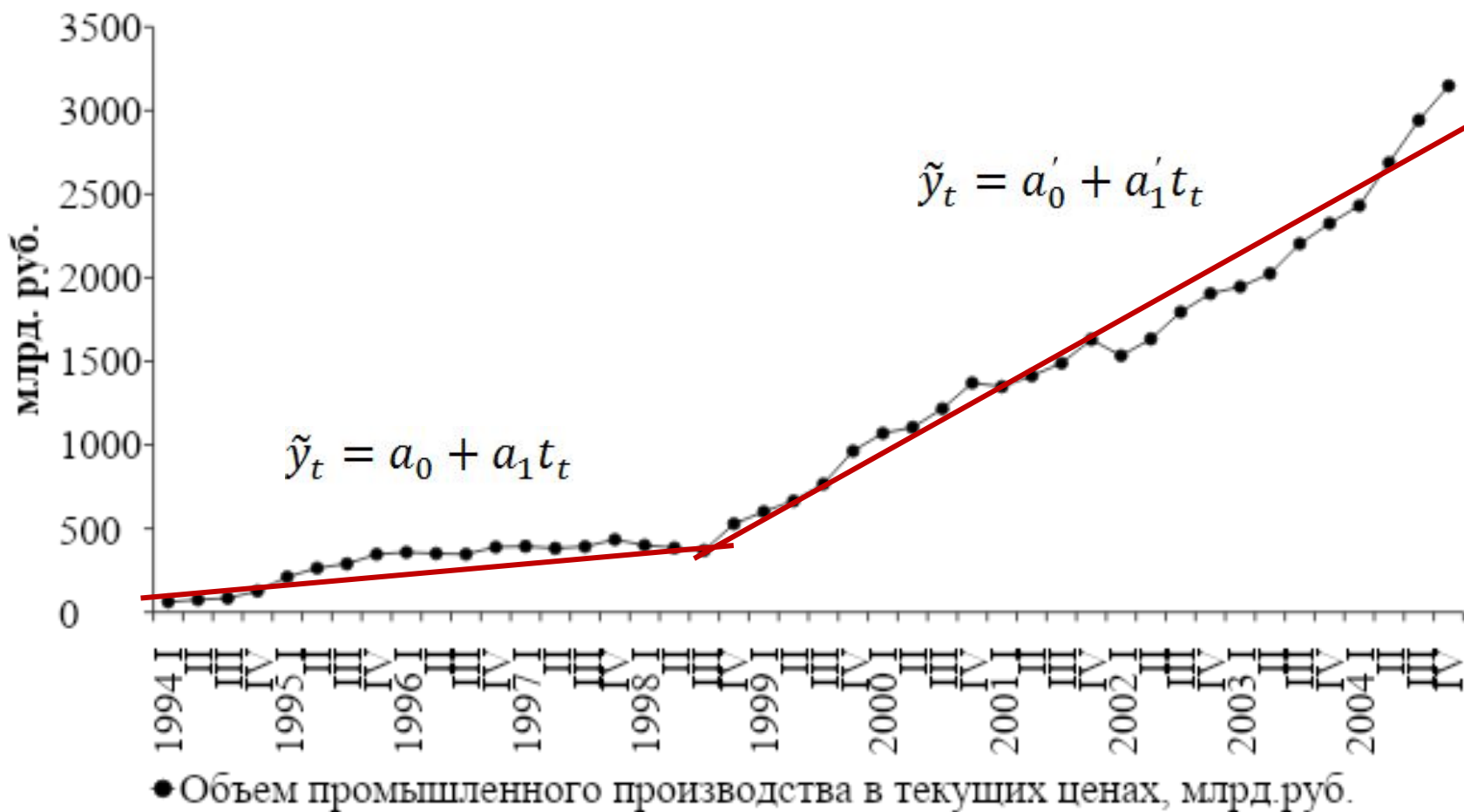
$$\tilde{y}_t = a_0 + a_1 t_t + (a'_0 - a_0) D_{1t} + \varepsilon_t$$

$$D_{1t} = \begin{cases} 1, & \text{если } t > t' \\ 0, & \text{если } t \leq t' \end{cases}$$

	Искомые параметры	Стандартная ошибка искомых параметров	$t(81)$ -статистика Стьюдента	p -уровень значимости
Свободный член	6,340	0,676	9,374	0,000
t	0,428	0,023	18,469	0,000
D_1	-13,320	1,133	-11,755	0,000

Модель «изменения роста (падения)»

28



Оценки регрессии

29

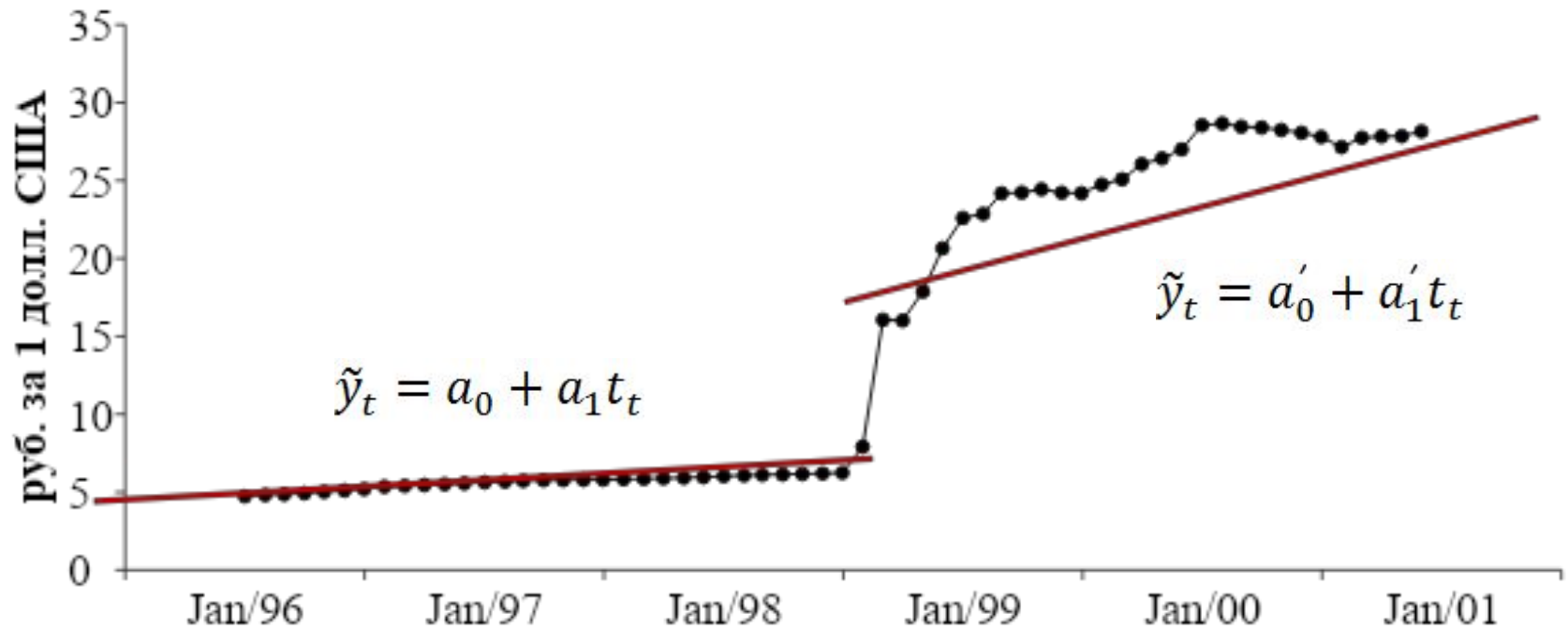
$$\tilde{y}_t = a_0 + a_1 t_t + (a'_1 - a_1) D_{2t} + \varepsilon_t$$

$$D_{2t} = \begin{cases} t - t', & \text{если } t > t' \\ 0, & \text{если } t \leq t' \end{cases}$$

	Искомые параметры	Стандартная ошибка искомых параметров	$t(41)$ -статистика Стьюдента	p -уровень значимости
Свободный член	127,791	50,065	2,552	0,015
t	15,713	3,685	4,264	0,000
D_2	77,579	5,553	13,972	0,000

Смешанная модель

30



● Официальный курс доллара на конец периода, рублей за 1 доллар США

Оценки регрессии

31

$$\tilde{y}_t = a_0 + a_1 t_t + (a'_0 - a_0)D_{1t} + (a'_1 - a_1)D_{2t} + \varepsilon_t$$

	Искомые параметры	Стандартная ошибка искомых параметров	$t(56)$ -статистика Стьюдента	p -уровень значимости
Свободный член	4,776	0,465	10,270	0,000
t	0,056	0,025	2,275	0,027
D_1	12,927	0,668	19,365	0,000
D_2	0,333	0,039	8,565	0,000

Сезонные фиктивные переменные: дифференциация свободного члена (линейный тренд)

Общий вид модели: $\tilde{y}_t = a_0 + a_1 t_t + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \varepsilon_i$

долговременная
компонента

сезонная
компонента

$$D_{1t} = \begin{cases} 0, \\ 1, \end{cases}$$

если рассматривается II квартал
в противном случае

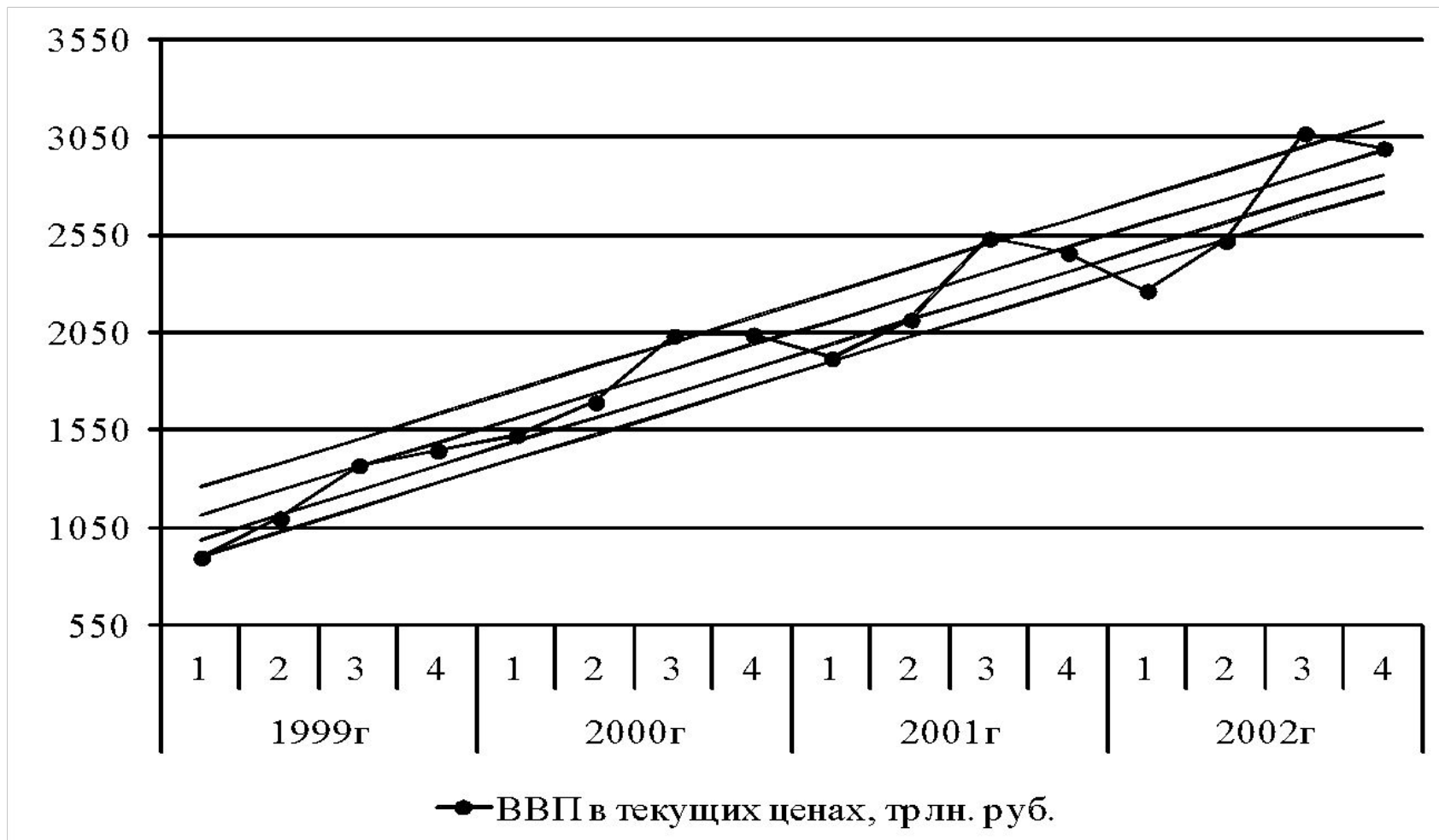
$$D_{2t} = \begin{cases} 0, \\ 1, \end{cases}$$

если рассматривается III квартал
в противном случае

$$D_{3t} = \begin{cases} 0, \\ 1, \end{cases}$$

если рассматривается IV квартал
в противном случае

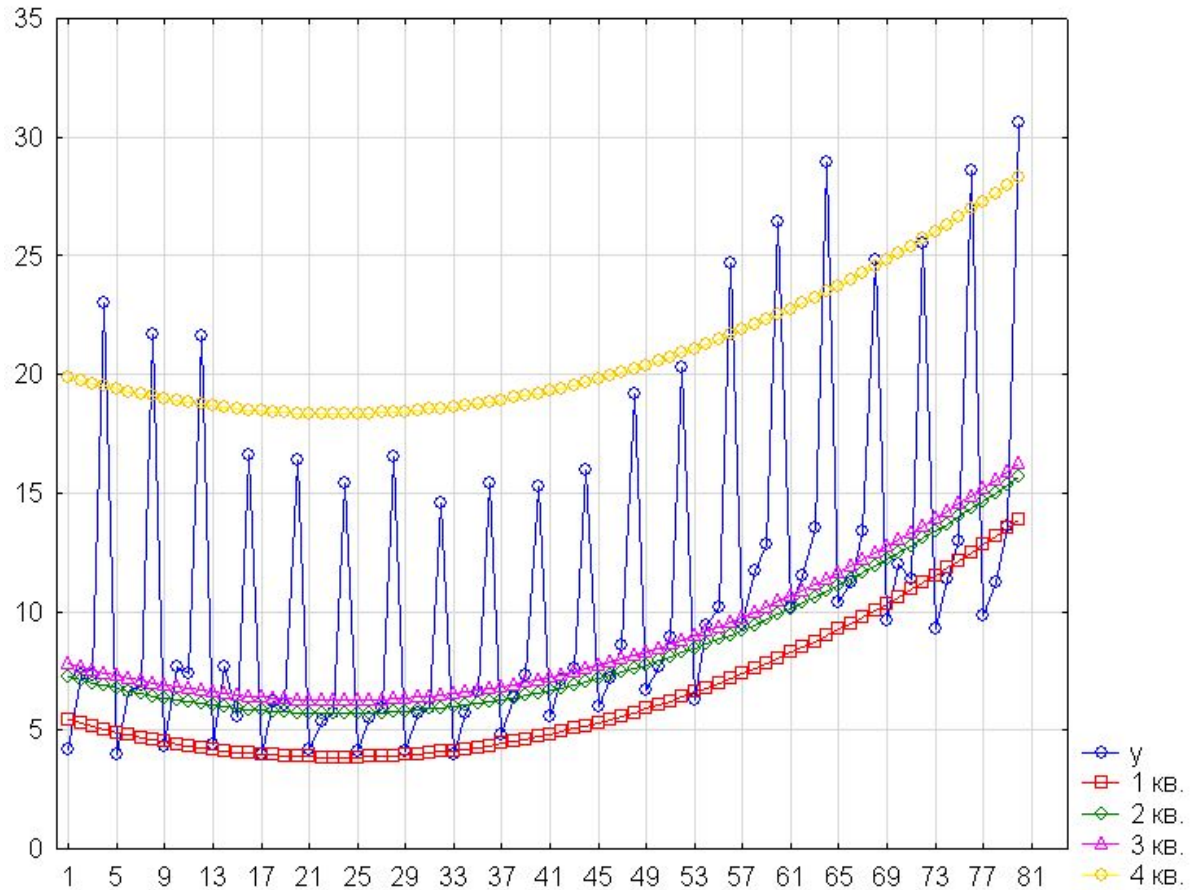
Моделирование сезонных колебаний ряда ВВП Российской Федерации в текущих ценах с помощью фиктивных переменных



Сезонные фиктивные переменные: дифференциация свободного члена (не линейный тренд)

$$\tilde{y}_t = a_0 + a_1 t_t + a_1 t_t^2 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \varepsilon_i$$

Динамика ввода в
действие жилых домов,
млн. кв. м



Сезонные фиктивные переменные: наклон тренде

$$\tilde{y}_t = a_0 + a_1 t_t + \alpha_1 D_{1t} t_t + \alpha_2 D_{2t} t_t + \alpha_3 D_{3t} t_t + \varepsilon_i$$

$$D_{1t} t_t = \begin{cases} t_t, & \text{если рассматривается II квартал} \\ 0, & \text{в противном случае} \end{cases}$$

$$D_{2t} t_t = \begin{cases} t_t, & \text{если рассматривается III квартал} \\ 0, & \text{в противном случае} \end{cases}$$

$$D_{3t} t_t = \begin{cases} t_t, & \text{если рассматривается IV квартал} \\ 0, & \text{в противном случае} \end{cases}$$

Динамика инвестиций в основной капитал России, млрд. руб.

36

