

# **ЭКОНОМЕТРИКА**

## **Временные ряды**

# Элементы временного ряда

- $\{y_1, y_2, \dots, y_n\} = \{y_t | t = \overline{1, \dots, n}\}$

Регулярные (систематические) компоненты ВР:

- $U_t$  – тренд (основная тенденция ряда)
- $V_t$  – сезонная компонента
- $C_t$  – циклическая компонента

Случайная (нерегулярная) компонента ВР:

- $\varepsilon_t$  – остаточная компонента

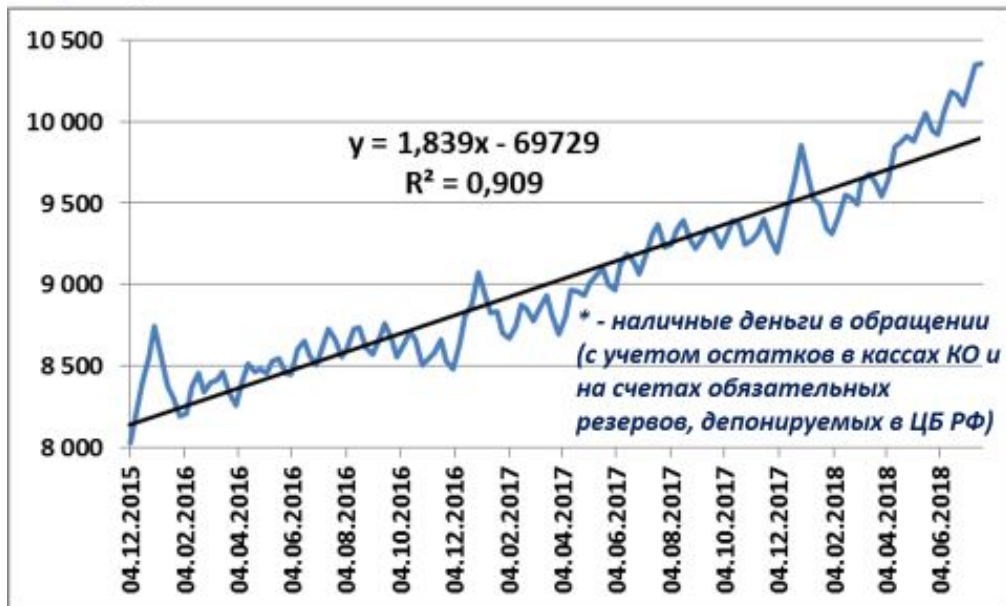
Временной ряд с аддитивной структурой:

$$y_t = U_t + V_t + C_t + \varepsilon_t$$

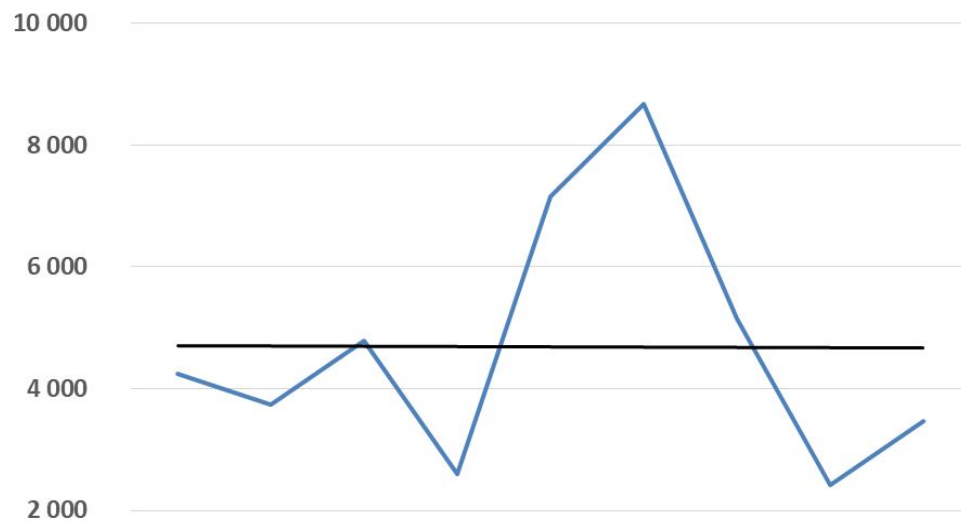
# Примеры ВР:

- Денежная база – ВР с четко выраженной тенденцией роста, сезонными и циклическими колебаниями
- Объем долговых ценных бумаг – отсутствие тенденции во временном ряду

Денежная база, млрд. руб.\*  
(по данным ЦБ РФ, с 04.12.2015 по 20.07.2018)



Объем долговых ценных бумаг в собственности российских банков, переданных по сделкам РЕПО с Банком России с 01.12.2017 по 01.08.2018), млн.руб.



# Предварительный анализ ВР

- **Выявление и устранение аномалий**
- **Аномальным считается уровень ряда, не отвечающий потенциальным возможностям исследуемой системы (процесса)**
  - аномалии (ошибки) I рода – технические ошибки, сбои в программе и т.д. (подлежат устранению)
  - ошибки II рода – неустранимые
- **Проверка гипотезы о наличии (отсутствии) тренда (тенденции) в целом**
- **Предварительный анализ проводится как графическим так и аналитическим**

# Выявление аномалий: метод Ирвина

- Значения статистики Ирвина:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, t = 1, \dots, n, \text{ где}$$

$$\sigma_y = \sqrt{\frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n - 1}}, \bar{y} = \frac{\sum_{t=1}^n y_t}{n}$$

- Сравнение с критическим значением ( $\alpha=0,05$ ):

$n$	2	3	10	20	30	50	100
$\lambda_{\text{крит.}}$	2,8	2,3	1,5	1,3	1,2	1,1	1,0

- Определение аномальных уровней

# Пример: динамика цен на целлюлозу

Дата	Цена, долл. США/т	$\lambda_t$
13.03.01	638,36	
20.03.01	632,16	0,2243
27.03.01	624,90	0,2626
03.04.01	618,40	0,2351
10.04.01	601,00	0,6295
17.04.01	586,50	0,5245
24.04.01	580,40	0,2207
01.05.01	578,00	0,0868
08.05.01	572,00	0,2171
15.05.01	560,20	0,4269



- Для  $n = 10$  и  $\alpha = 0,05$  критическое значение  $\lambda = 1,5$
- Согласно критерию Ирвина на уровне значимости  $\alpha = 0,05$  во временном ряду аномалий нет
- Проверим ВР на наличие

# Проверка равенства дисперсий

- ВР разбивается на две части
- Проверяется статистическая гипотеза о равенстве дисперсий:  $F = \frac{\sigma_i^2}{\sigma_j^2}$ , где  $\sigma_i > \sigma_j$ ;  $i, j = 1, 2$

	Переменная 1	Переменная 2
Среднее	622,964	575,42
Дисперсия	207,14968	99,392
Наблюдения	5	5
Df	4	4
<b>F</b>	<b>2,084168545</b>	
<i>P(F&lt;=f) одностороннее</i>	0,2472144	
<i>F критическое одностороннее</i>	<b>6,388233942</b>	

- Значение статистики Фишера  $F = 2,084 < 6,388$ , поэтому нет оснований отвергнуть гипотезу о равенстве дисперсий

# Проверка разности средних уровней ВР

- Значение статистики Стьюдента:

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ где } \sigma = \sqrt{\frac{(n_1 - 1) \cdot \sigma_1^2 + (n_2 - 1) \cdot \sigma_2^2}{n - 2}}$$

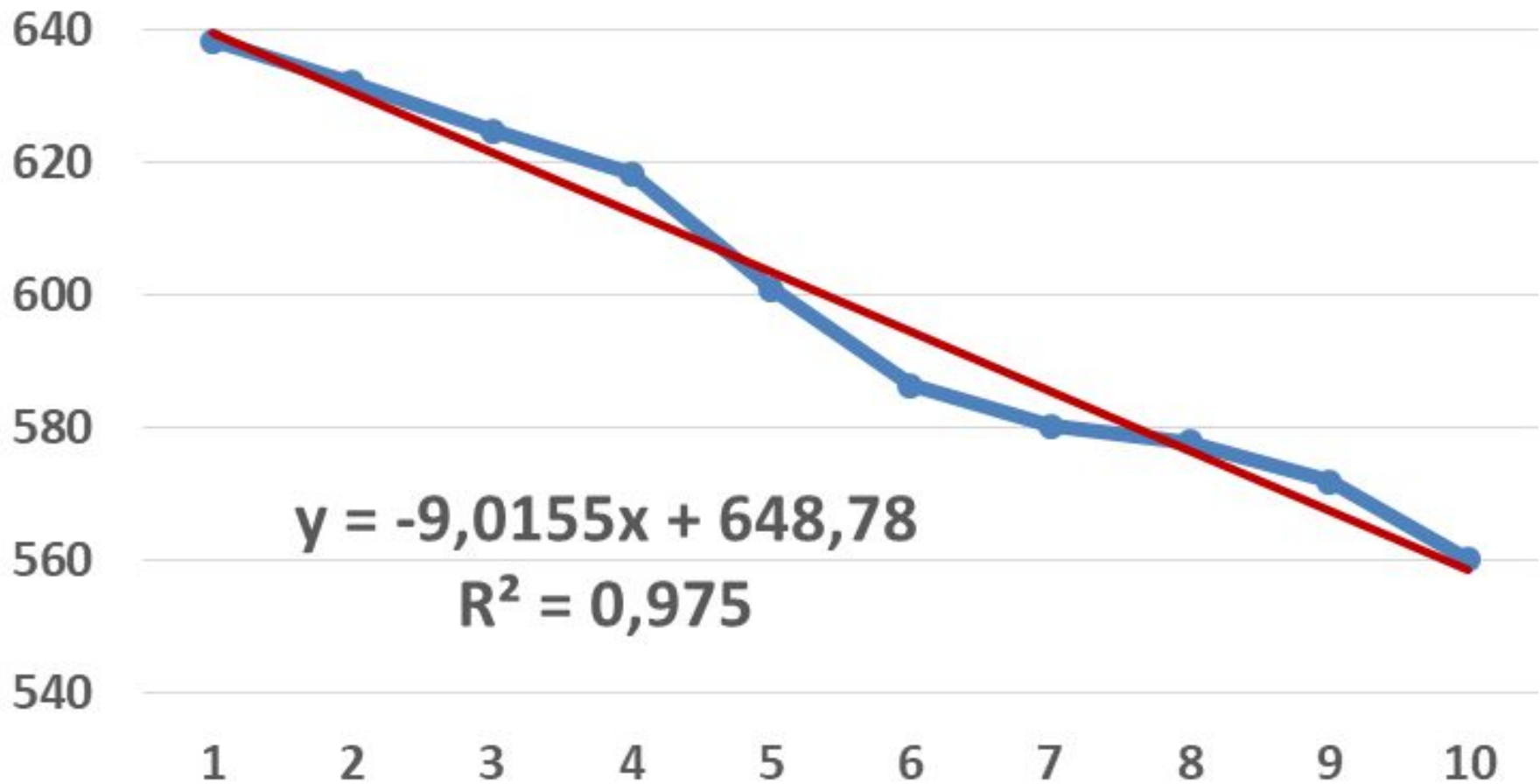
	Переменная 1	Переменная 2
Среднее	622,964	575,42
Дисперсия	207,14968	99,392
Наблюдения	5	5
Объединенная дисперсия	153,27084	
Гипотетическая разность средних	0	
Df	8	
<b>t-статистика</b>	<b>6,072058581</b>	
P(T<=t) одностороннее	0,000149268	
t критическое одностороннее	1,85954832	
P(T<=t) двухстороннее	0,000298537	
t критическое двухстороннее	<b>2,306005626</b>	

- Гипотеза о равенстве средних отвергается, т.к.  $t = 6,072 > 2,306$
- Есть тенденция в среднем (тренд)



# Линейный тренд ВР

Динамика цен на целлюлозу, долл. США/т



# Прогнозирование

- Расчет  $\hat{y}_{t+L}$ , где  $L$  – период упреждения
- Стандартная ошибка показателя:

$$S = \sqrt{\frac{\sum_{t=1}^n \varepsilon_t^2}{n - k}}, k \text{ – число параметров}$$

- Доверительный интервал прогноза (для линейного тренда):

$$\hat{y}_{n+L} \pm t_{\alpha} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{3 \cdot (n + 2L - 1)^2}{n(n^2 - 1)}}$$

## В нашем примере:

- $\hat{y}_{11} = 549,61; S = 4,633$
- Прогнозный интервал:  $549,61 \pm 7,84$ , т.е. от 541,77 до 557,45 долл. США/т
- Фактическое значение показателя в прогнозируемом периоде составило 545,20 долл. США/т, что попадает в доверительный интервал.
- Таким образом, полученный прогноз может быть верифицирован

# Метод характеристик прироста

- Предварительное сглаживание кривой методом простой скользящей средней

- Первые средние приросты:

$$\bar{u}_t = \frac{y_{t+1} - y_{t-1}}{2}, t = 2, \dots, n - 1$$

- Вторые средние приросты:

$$\bar{u}_t^{(2)} = \frac{\bar{u}_{t+1} - \bar{u}_{t-1}}{2}, t = 3, \dots, n - 2$$

- Вспомогательные величины:

$$\frac{\bar{u}_t}{y_t}, \log \bar{u}_t, \log \frac{\bar{u}_t}{y_t}, \log \frac{\bar{u}_t}{y_t^2}$$

# Какую кривую выбрать?

Показатель	Характер изменения	Вид кривой
	Постоянный	Прямая
	Линейный	Квадратичная парабола
	Линейный	Кубическая парабола
	Постоянный	Простая экспонента
	Линейный	Модифицированная экспонента
	Линейный	Кривая Гомперца
	Линейный	Логистическая кривая

# Примеры S-образных кривых

- Модель роста населения Швеции с 1850 по 1950 гг.\* (Г. Тинтнер, 1965 г.):

$$y = \frac{10328806}{1 + 2,117 \cdot e^{-0,14t}}$$

(логистическая кривая)

\* - в 2005 г. численность населения Швеции составила 9 млн. чел.

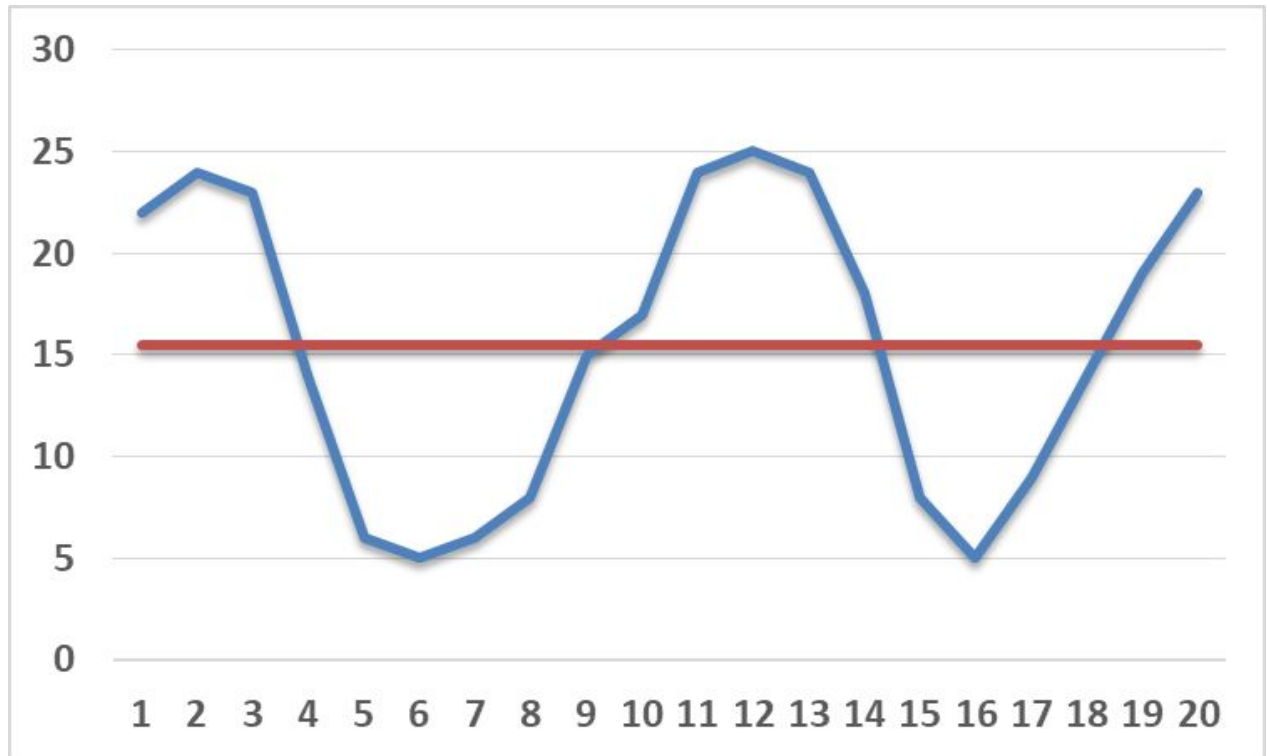
- Затраты на строительство автомобильных дорог (К. Льюис, 1986 г.):

$$y = 4644,5 \cdot 0,0961435^{0,93176t}$$

(кривая Гомперца)

# Периодические колебания

№ месяца	Объем пр-ва
1	22
2	24
3	23
4	14
5	6
6	5
7	6
8	8
9	15
10	17
11	24
12	25
13	24
14	18
15	8
16	5
17	9
18	14
19	19
20	23



• **Стационарный динамический ряд**

**(тенденция отсутствует)**

$$N = 20; \bar{y} = 15,45; \sigma_y^2 = 52,1475$$

$y_t = \bar{y} + A \cos W(t - \Phi)$ , где  $A$  – амплитуда,  
 $W = \frac{2\pi}{P}$  - угловая частота,  $\Phi$  - фаза

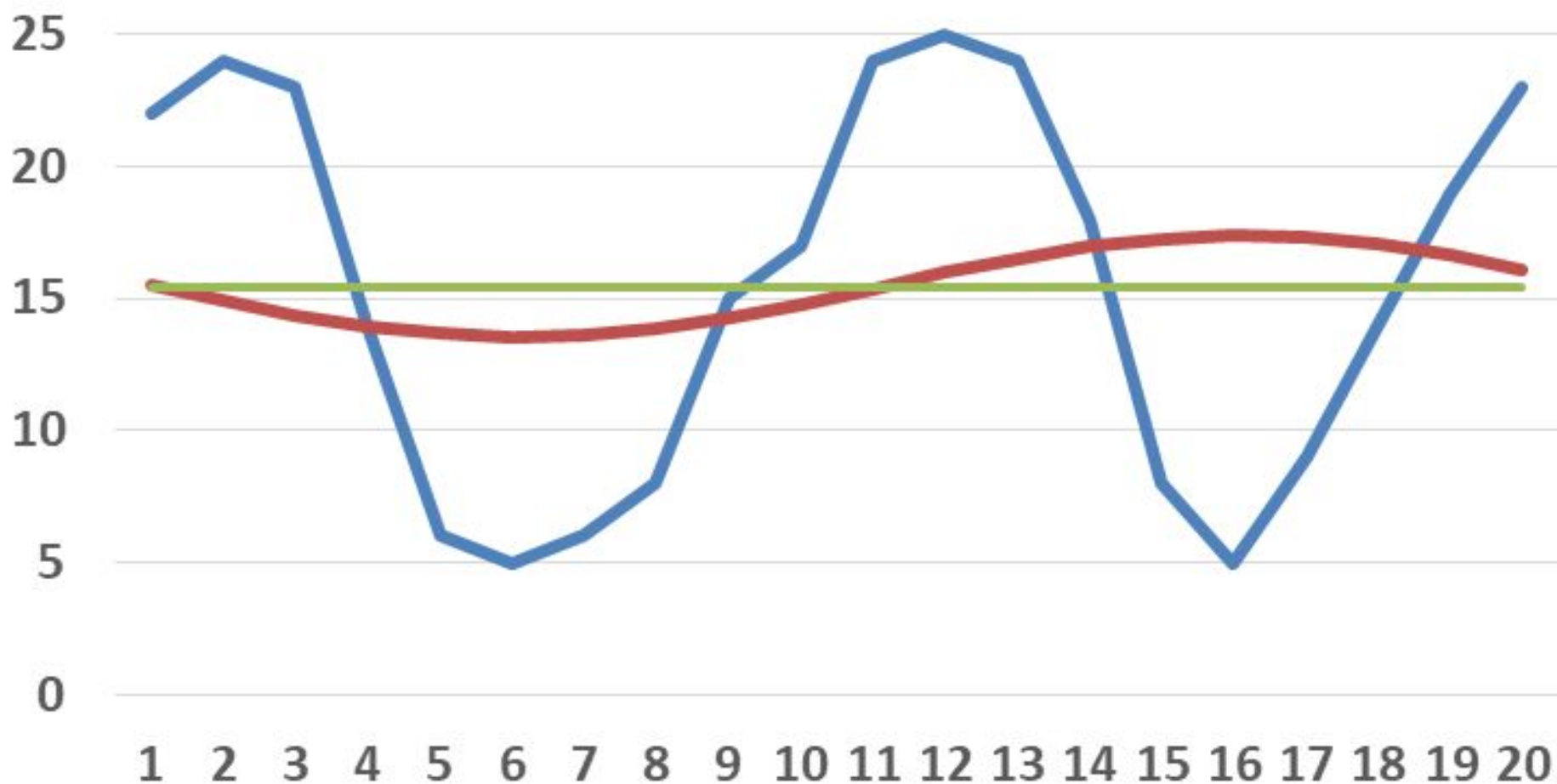
$k$	Уравнение с $k$ гармониками
1	
2	
3	
4	



Уравнение с одной гармоникой:

$$y_t = 15,45 + 0,6667 \cdot \cos\theta - 1,7948 \cdot \sin\theta$$

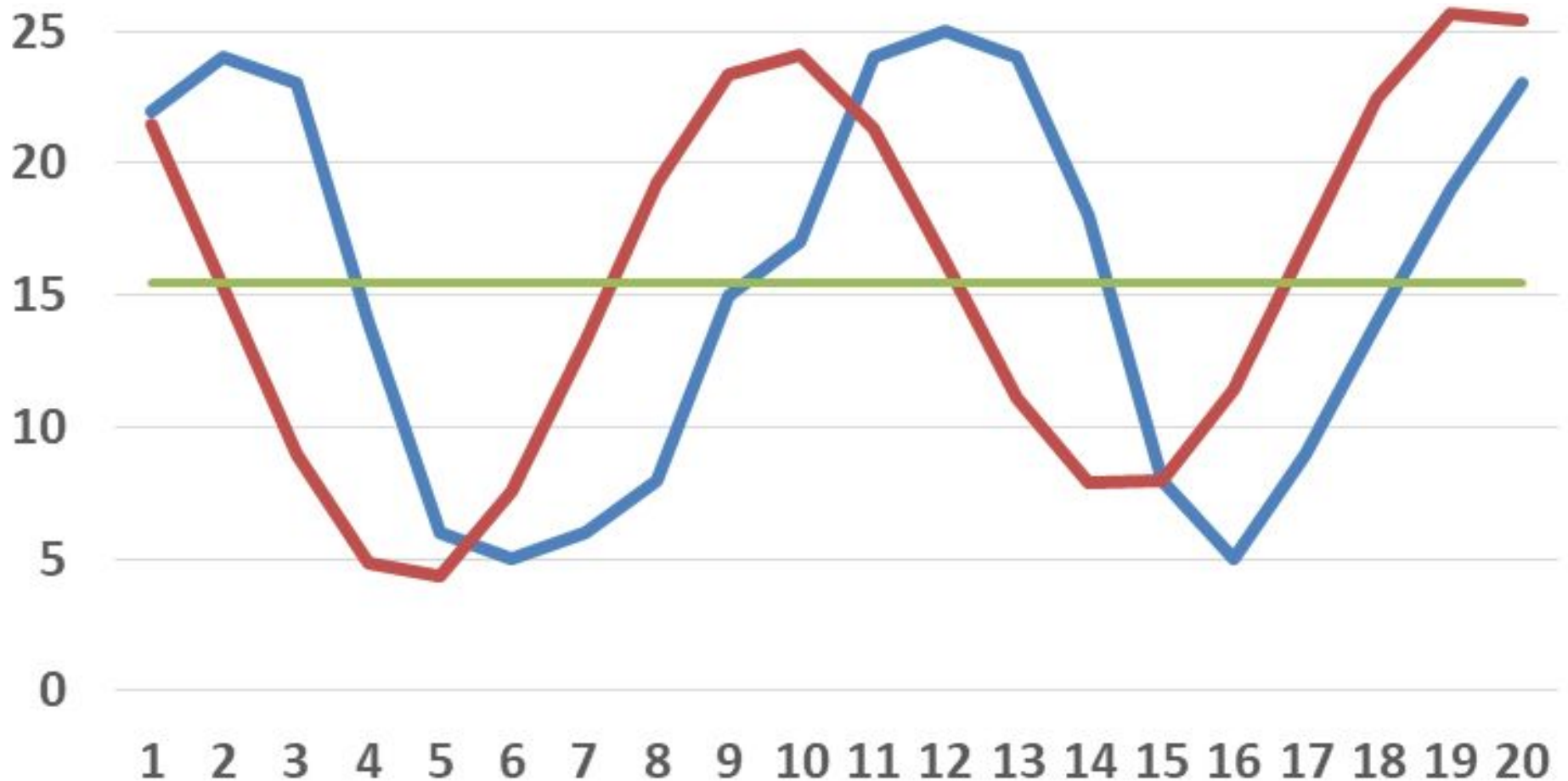
$$\theta = \frac{2\pi}{20}(t - 1), R^2 = 0,0351$$



Уравнение с двумя гармониками:

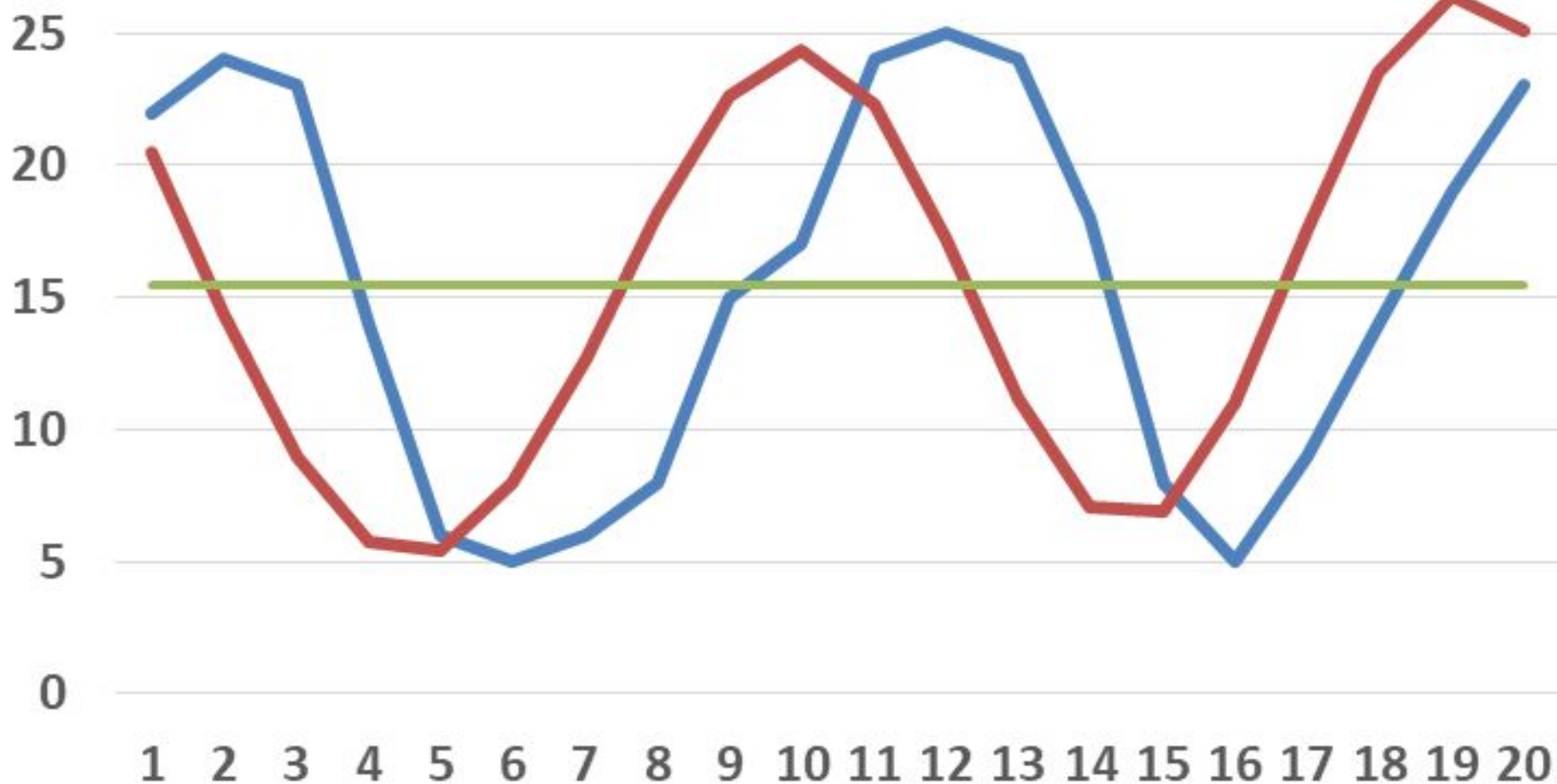
$$y_t = 15,45 + 0,6667 \cdot \cos\theta - 1,7948 \cdot \sin\theta + 9,2883 \cdot \cos 2\theta - 2,6577 \cdot \sin 2\theta$$

$$R^2 = 0,930$$



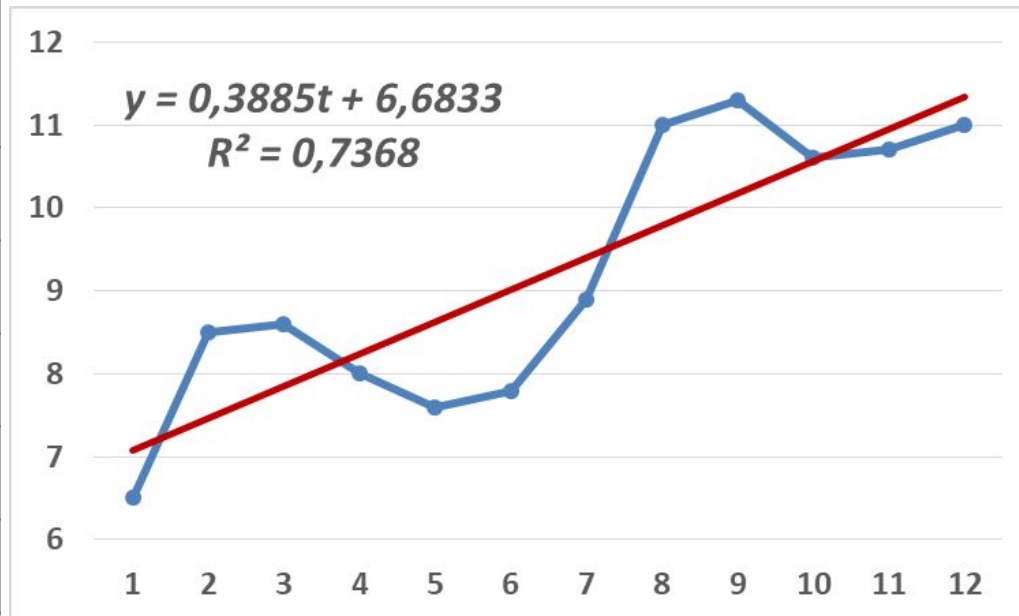
Уравнение с тремя гармониками:

$$y_t = 15,45 + 0,6667 \cdot \cos\theta - 1,7948 \cdot \sin\theta + \\ + 9,2883 \cdot \cos 2\theta - 2,6577 \cdot \sin 2\theta - \\ - 0,2698 \cdot \cos 3\theta - 1,0568 \cdot \sin 3\theta; R^2 = 0,942$$



# Пример: ряд с тенденцией

месяц	цена	урасч.	остатки
1	6,5	7,07	-0,57
2	8,5	7,46	1,04
3	8,6	7,85	0,75
4	8,0	8,24	-0,24
5	7,6	8,63	-1,03
6	7,8	9,01	-1,21
7	8,9	9,40	-0,50
8	11,0	9,79	1,21
9	11,3	10,18	1,12
10	10,6	10,57	0,03
11	10,7	10,96	-0,26
12	11,0	11,35	-0,35



# Модель ряда с двумя гармониками

- $$\begin{cases} y_t = 6,683 + 0,388 \cdot t + \varepsilon_t \\ \varepsilon_t = 0,123 \cdot \cos\theta - 0,296 \cdot \sin\theta - \\ - 0,137 \cdot \cos 2\theta + 1,005 \cdot \sin 2\theta \\ R^2 = 0,9712 \end{cases}$$
- Остатки представляют собой стационарный ряд и хорошо описываются рядом Фурье с двумя гармониками
- Модель линейного тренда имеет  $R^2 = 0,7368$
- Модель с учетом периодических колебаний:  $R^2 = 0,9712$



# **Модели регрессии по временным рядам**

- Появление «ложной корреляции» требует предварительной обработки рядов**
- При построении модели регрессии нужно исключать регулярные компоненты**
- Необходим анализ остатков с помощью автокорреляционной функции, для устранения автокорреляции в остатках применить ОМНК**
- Необходимо выявлять временной лаг**
- Особое внимание уделить проявлениям мультиколлинеарности**

# Учет тенденции при построении модели регрессии

- Методы исключения тенденции:
  - метод последовательных разностей
  - метод отклонений от тренда
- Включение в модель регрессии фактора времени как отдельной независимой переменной:

$$y = a + bx + ct + \varepsilon$$

- с применением МНК для оценок  $a$ ,  $b$ ,  $c$
- последовательным включением в модель линейной тенденции ряда  $y$  и линейной регрессии остатков

$$dy = b \cdot dx + \varepsilon$$



# Пример МПР:

$y_t$  - инвестиции

$x_t$  - прибыль за  
предыдущий год

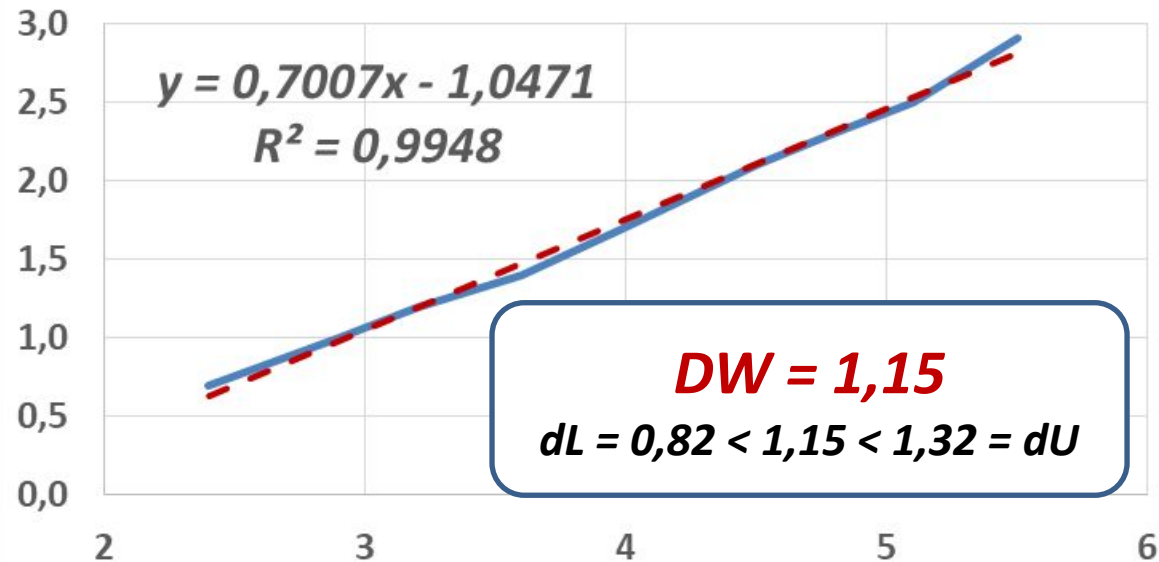
Есть линейный  
тренд для  $x$

$$x_t = 2,075 + 0,385t$$

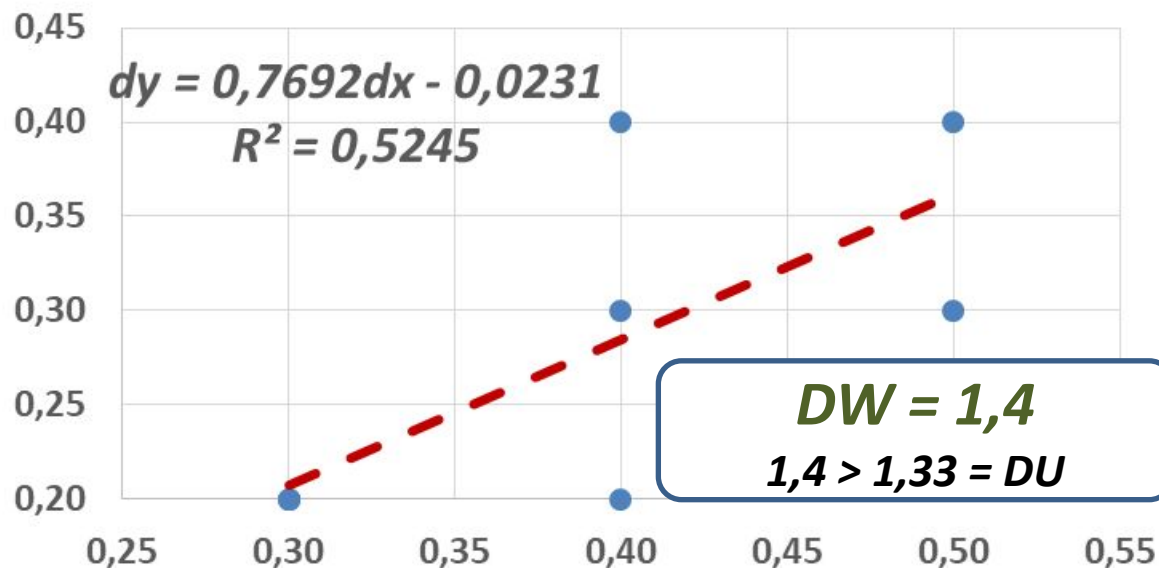
$$R^2 = 0,997$$

ГОДЫ	$x_t$	$y_t$	$dx$	$dy$
1	2,4	0,7	-	-
2	2,9	1,0	0,5	0,3
3	3,2	1,2	0,3	0,2
4	3,6	1,4	0,4	0,2
5	4,0	1,7	0,4	0,3
6	4,5	2,1	0,5	0,4
7	4,8	2,3	0,3	0,2
8	5,1	2,5	0,3	0,2
9	5,5	2,9	0,4	0,4

## Модель регрессии по МНК



## Первые разности





# Метод отклонений от тренда

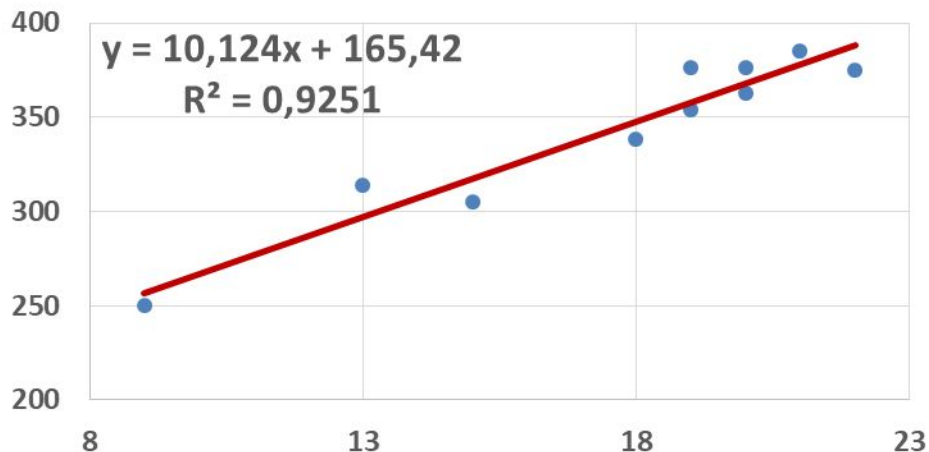
$y$  – прибыль,  $x$  – затраты на охрану труда за 10 мес.

$t$	$yt$	$yt$ <i>расч</i>	$dy$	$xt$	$xt$ <i>расч</i>	$dx$
1	250	259,28	-9,28	9	10,11	-1,11
2	305	294,00	11,00	15	12,86	2,14
3	314	316,43	-2,43	13	14,80	-1,80
4	338	333,37	4,63	18	16,35	1,65
5	354	347,13	6,87	19	17,67	1,33
6	363	358,80	4,20	20	18,82	1,18
7	375	368,97	6,03	22	19,86	2,14
8	376	378,01	-2,01	19	20,80	-1,80
9	376	386,17	-10,17	20	21,67	-1,67
10	385	393,61	-8,61	21	22,47	-1,47

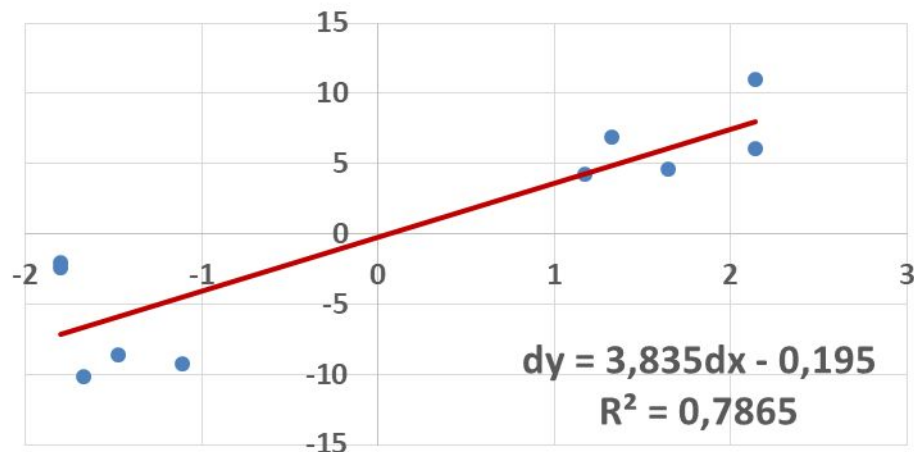
# Метод отклонений от тренда

$y$  – прибыль,  $x$  – затраты на охрану труда за 10 мес.

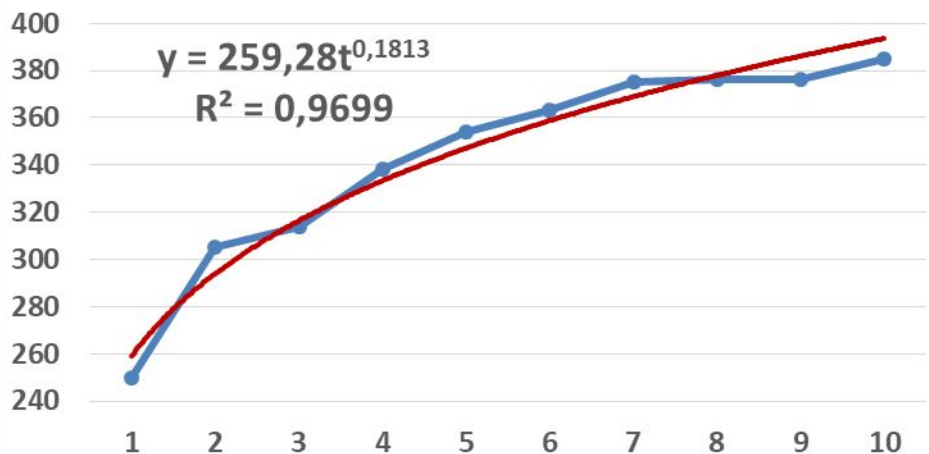
Регрессия  $y(x)$



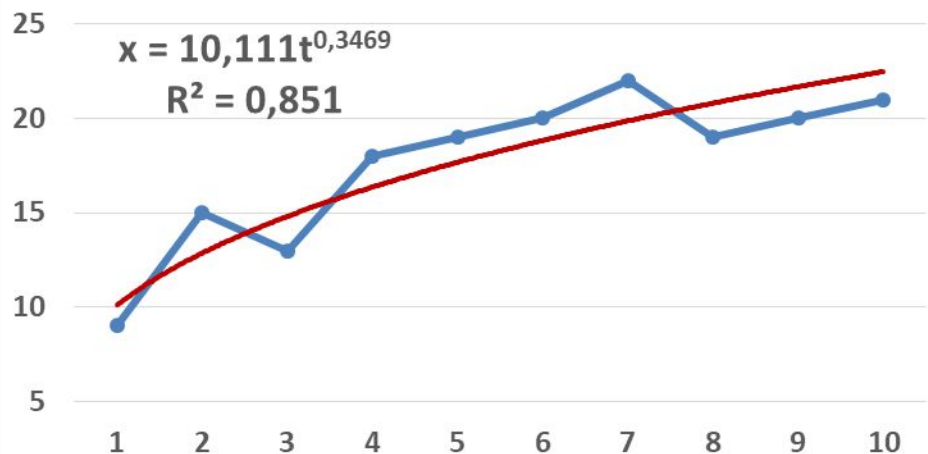
Регрессия  $dy$  по  $dx$



Тренд для  $y(t)$



Тренд для  $x(t)$



# Модели с лаговыми переменными

- Модели с распределенными лагами (с лаговыми объясняющими переменными):

$$y_t = a + b_0x_t + b_1x_{t-1} + \dots + b_kx_{t-k} + \varepsilon_t$$

- Модели авторегрессии (с лаговыми зависимыми переменными):

$$y_t = a + bx_t + c_1y_{t-1} + \dots + c_ky_{t-k} + \varepsilon_t$$

- Авторегрессионные модели с распределенным лагом:

$$y_t = a + b_1y_{t-1} + \dots + b_ky_{t-k} + c_0x_t + c_1x_{t-1} + \dots + c_kx_{t-k} + \varepsilon_t$$

- Основные вопросы:
  - выбор величины лага
  - определение числа лаговых переменных

# Авторегрессионные модели $AR(p)$

- Авторегрессионный процесс порядка  $p$ :

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t$$

- Условие стационарности процесса:

- ряд  $a_1, a_2, \dots, a_p$  - сходится

- все (комплексные) корни характеристического уравнения  $a_1 z + a_2 z^2 + \dots + a_p z^p = 1$  должны удовлетворять условию  $|z| > 1$

- Для авторегрессии  $AR(1)$ :  $\hat{y}_t = a_0 + a_1 y_{t-1}$ ,  
характеристическое уравнение:  $a_1 z = 1$ ,  
поэтому  $-1 < a_1 < 1$

# Пример модели $AR(2)$ :

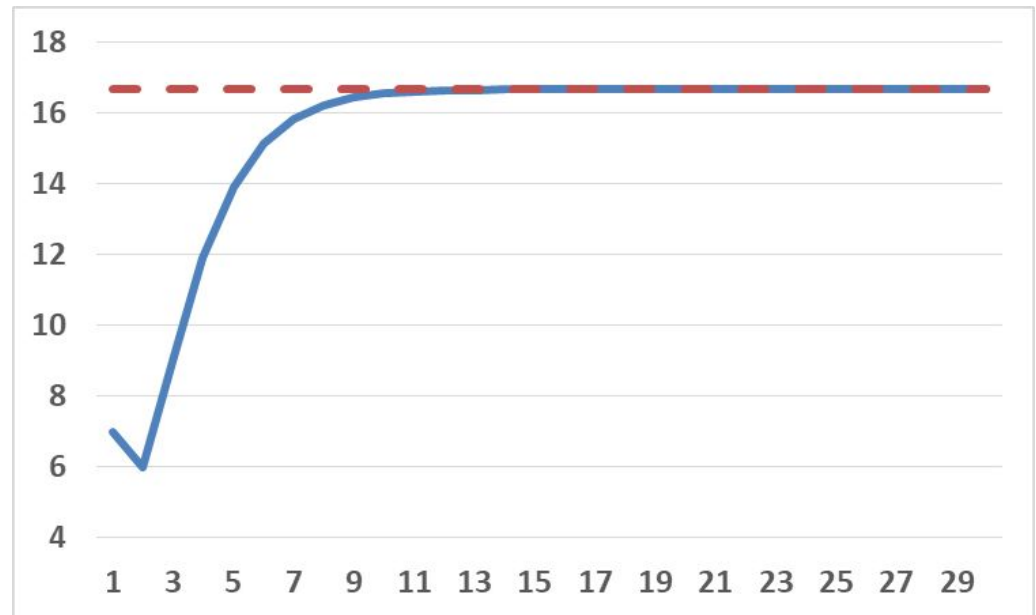
$$y_t = 5 + 0,9y_{t-1} - 0,2y_{t-2} + \varepsilon_t; y_1 = 7, y_2 = 6$$

- Характеристическое уравнение:

$$0,2z^2 - 0,9z + 1 = 0$$

- Корни:  $z_1 = 2,5 > 1$ ;  $z_2 = 2 > 1$ , поэтому процесс стационарный, с асимптотой

$$\begin{aligned} \mu &= \frac{a_0}{1 - a_1 - a_2} = \\ &= \frac{5}{1 - 0,9 + 0,2} \approx \\ &\approx 16,67 \end{aligned}$$



# Модели скользящей средней

## МА

- Для стационарного ряда  $\hat{y}_t$  представляется линейной функцией прошлых ошибок:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

где  $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$  – «белый шум» в текущий и предыдущие периоды

- Процесс скользящего среднего порядка  $q$ :

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

- Модель  $MA(1)$ :  $y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$

- Модель  $MA(2)$ :

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

- При  $q = 0, \mu = 0$  получаем «белый шум»

# Модели $ARMA(p, q)$

*(Auto Regressive – Moving Average)*

- Соединение в одной модели  $AR$  и  $MA$  – авторегрессионный процесс со скользящими средними в остатках
- Пример для  $ARMA(3, 2)$ :

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

- Основная проблема – выбор числа лагов  $p, q$
- Инструмент идентификации  $ARMA(p, q)$  – частная автокорреляционная функция  $PACF$
- Обобщение – модель Бокса-Дженкинса  $ARIMA(p, d, q)$ , где  $d$  – порядок разностей

# Модель $ARIMA(p,d,q)$ Бокса-

## Дженкинса

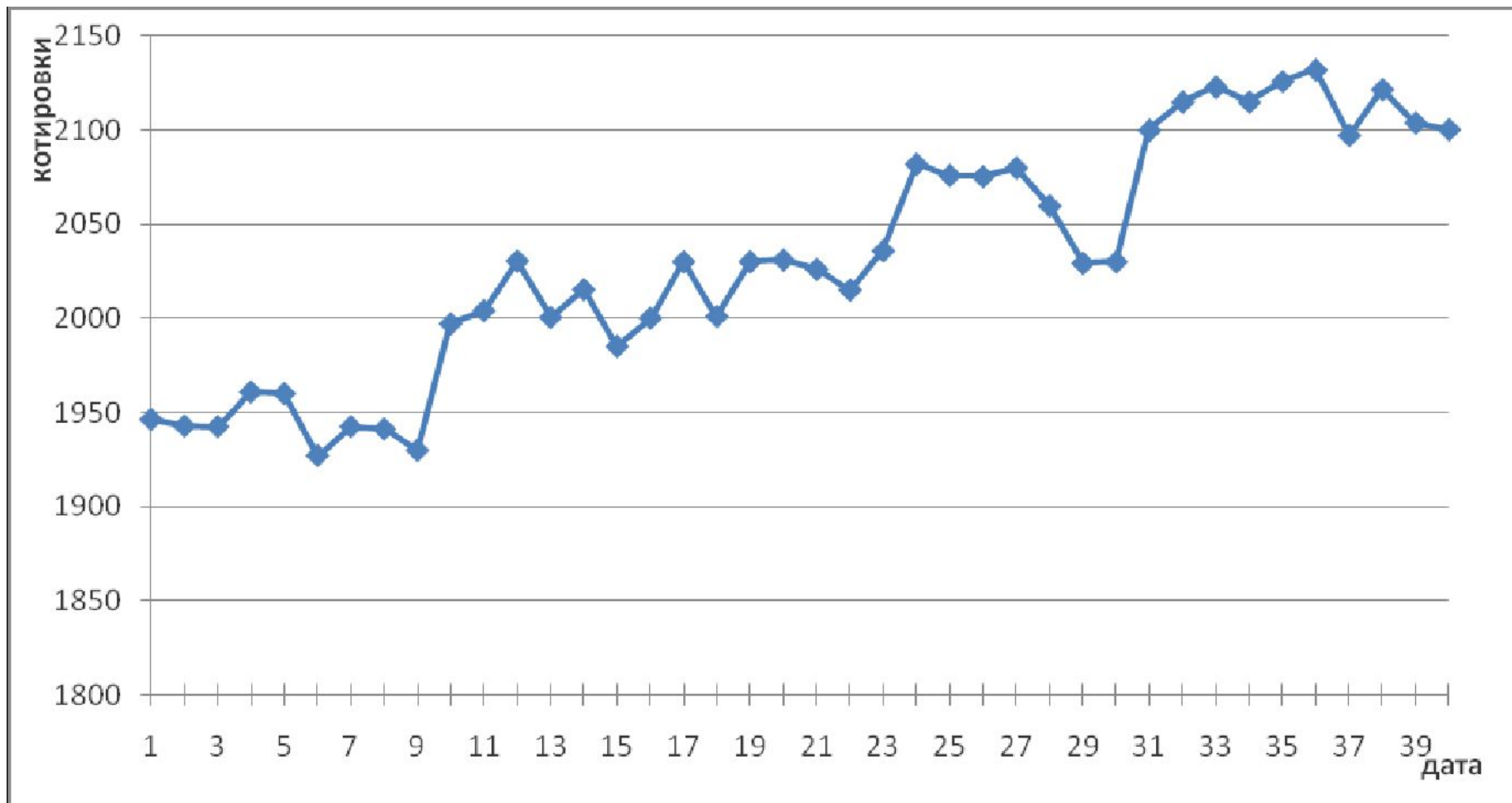
- $$\Delta^d y_t = \mu_1 \Delta^d y_{t-1} + \dots + \mu_p \Delta^d y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$
- $p$  – порядок авторегрессии
- $q$  – порядок скользящего среднего
- $d$  – порядок разностей (интегрирования)
- $\varepsilon_t \sim N(0, \sigma_\varepsilon)$  - процесс типа «белый шум»
- Сначала определяется порядок разностей  $d$
- Интегрирование проводится до получения стационарного ряда
- Для стационарного ряда  $p$  и  $q$  оцениваются как в модели  $ARMA(p,q)$



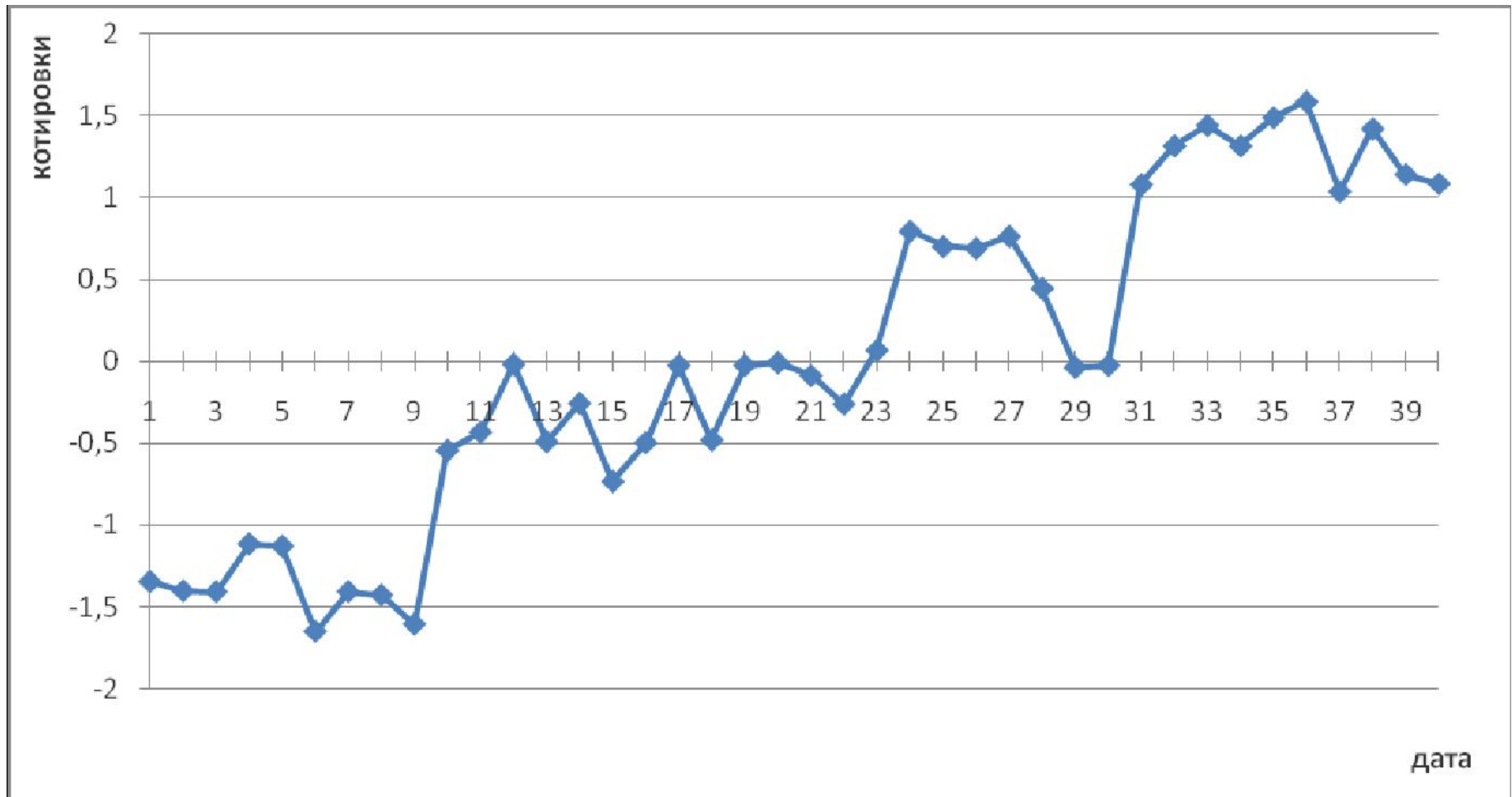
# Проверка остатков

- Соответствие нормальному закону распределения (критерий Колмогорова-Смирнова)
  - вычисляем остатки  $\varepsilon_t = y_t - \hat{y}_t, t = 1, \dots, n$
  - рассчитываем  $\bar{\varepsilon}$  и  $S_\varepsilon^2$
  - находим  $\Phi(z_t)$ , где  $z_t = \frac{\varepsilon_t - \bar{\varepsilon}}{S_\varepsilon}$
  - рассчитываем  $D_n^H = D_n \left( \sqrt{n} - 0,01 + \frac{0,85}{\sqrt{n}} \right)$ , где  $D_n = \max(\max_{1 \leq t \leq n} (\frac{t}{n} - \Phi(z_t)); \max_{1 \leq t \leq n} (\Phi(z_t) - \frac{t-1}{n}))$
  - $D_n^H$  сравниваем с критическим  $D_n^H(\alpha)$
  - если  $D_n^H < D_n^H(\alpha)$  – гипотеза не отклоняется

# Пример: котировки акций Лукойл на рынке *RTS Standard* (23.08.2013-30.10.2013)



# Стандартизованные котировки акций Лукойл на рынке *RTS* *Standard* (23.08.2013-30.10.2013)



# Проверка нормальности распределения остатков

- Критерий Колмогорова-Смирнова:  $D_n^H = 0,524$
- Критическое значение:  $D_n^H(0,05) = 0,895$
- $D_n^H = 0,524 < D_n^H(\alpha) = 0,895$
- Гипотеза о нормальности остатков не отклоняется на уровне значимости 0,05

	0,15	0,10	0,05	0,03	0,01
	0,775	0,819	0,895	0,955	1,035