



**ПОЛИТЕХ**  
Санкт-Петербургский  
политехнический университет  
Петра Великого

# Статистика, часть 6

*Николай Вячеславович Павлов*

*[pavlov@kafedrapik.ru](mailto:pavlov@kafedrapik.ru)*

# Условные обозначения

Это самое важное, надо знать на  
100%!!!

$$2*2=4$$

**Быть или не быть?**

Это надо решить и записать!!!



# Временные ряды

- Это история в цифрах

- Это ряд значений определенного показателя, расположенных в хронологическом порядке.

Ряд динамики всегда состоит из 2-х элементов:

- **y** – уровень динамического ряда = измерение = отсчёт - значение показателя в конкретный **момент** или **период** времени. Уровень ряда принято обозначать **y**;

- **t** - момент или период времени, в котором зафиксировано значение показателя.

Правильнее задавать номер периода



- Достоверны
- Исчислены за одинаковые периоды времени или на одинаковые даты;
- Сопоставимы по содержанию = исчислены **по единой методологии**;
- Приведены в одних и тех же единицах измерения.

## • Первоначальные

- Интервальные (показатели за период)
- Моментные (показатели на момент времени)
  - С равными промежутками времени
  - С неравными промежутками времени

## • Производные

- Средние значения
- Относительные базисные величины
- Относительные цепные величины

*Относительная величина = сравниваемая величина / базис*

*[коэффициент, процент, промилле]*

1. Численность населения (на конец года), млн. человек
2. Среднегодовая численность занятых в экономике, тыс. человек
3. Среднедушевые денежные доходы населения в месяц, руб.
4. Валовой внутренний продукт, млрд. руб.
5. Основные фонды в экономике (по полной учетной стоимости; на конец года), млрд. руб.
6. Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по видам экономической деятельности, млрд. руб.:
  - 6.1. добыча полезных ископаемых
  - 6.2. обрабатывающие производства
  - 6.3. производство и распределение электроэнергии, газа и воды
  - 6.4. продукция сельского хозяйства млрд. руб.
7. Индекс потребительских цен (декабрь к декабрю предыдущего года), процентов

- по **методике** расчета показателей (если с течением времени методика расчета показателей менялась, то анализировать их как единый временной ряд нельзя);
- по **территории**, по которой рассчитывается показатель (к несопоставимости могут приводить изменения границ регионов, стран, организаций: объединение, дробление и т.п.);
- по **охвату единиц** (при оценке масштабов деятельности малого бизнеса показатель может быть рассчитан только по малым предприятиям, а может включать и микропредприятия);

# Сопоставимость уровней ряда

- по **единицам измерения** (все уровни ряда должны быть измерены в одних единицах измерения);
- по **периоду времени** (ряды должны быть **полными**)

**Полный ряд = ряд с одинаковыми интервалами между моментами измерения**

- по **моменту времени** (например, все значения на конец года)



- Показатели **изменения** уровней временных рядов.
- **Средние** характеристики рядов динамики.
- Основная **тенденция** изменения уровней ряда, трендовая модель.
- **Автокорреляция** уровней + **авторегрессионная** модель.
- **Цикличность**, например, сезонность
- **Взаимосвязь** между временными рядами, корреляция рядов динамики.
- **Прогнозирование на основе моделей временных рядов.**





## Показатели изменения уровней ряда

# Цепные и базисные показатели

Период	Произведено, тыс. т.	Абсолютные приросты, тыс. т.		Коэффициенты роста		Темпы роста, %		Темпы прироста, %		Значение 1% прироста, тыс. т.
		цеп- ные	базис- ные	цеп- ные	базис- ные	цеп- ные	базис- ные	цеп- ные	базис- ные	
1	200	-	-	-	-	-	-	-	-	-
2	210	10	10	1,05	1,05	105,0	105	5,0	5,0	2,00
3	218	8	18	1,04	1,09	103,8	109	3,8	9,0	2,10
4	230	12	30	1,06	1,15	105,5	115	5,5	15,0	2,18
5	234	4	34	1,02	1,17	101,7	117	1,7	17,0	2,30

- цепной:

$$\Delta_{\text{ц}} = y_t - y_{t-1} ;$$

- базисный:

$$\Delta_{\text{б}} = y_t - y_0 .$$

где

$y_t$  - значение уровня ряда в момент (за период)  $t$ ;

$y_{t-1}$  - значение уровня ряда в предшествующий момент (период) времени;

$y_0$  - значение уровня, выбранного за базу сравнения.

Показывает, отличие данного уровня ряда от предшествующего (базисного) уровня.

Измеряется в тех же единицах, что и исходные данные.

- цепной

$$K_{\text{P}}^{\text{Ц}} = \frac{y_t}{y_{t-1}}$$

- базисный

$$K_{\text{P}}^{\text{Б}} = \frac{y_t}{y_0}$$

**Смысл?**

$$T_P^Ц = \frac{y_t}{y_{t-1}} \times 100\% \quad \text{или} \quad T_P^Ц = K_P^Ц \times 100\%$$

$$T_P^Б = \frac{y_t}{y_0} \times 100\% \quad \text{или} \quad T_P^Б = K_P^Б \times 100\%$$

$$T_{\text{пр}} = T_{\text{р}} - 100\%$$

ИЛИ

$$T_{\text{пр}} = (y_t - y_{t-1}) / y_{t-1} * 100\%$$



$$y_0 = 1000$$

$$y_1 = 1000$$

- Темп роста = ?
- Темп прироста = ?



## Средние характеристики рядов

# Средние значения рядов

- Интервальные ряды
  - Среднее арифметическое

Годы	Продано сахара, тыс. тонн
2015	2905
2016	2585
2017	2647

$$\bar{y} = \frac{\sum y}{n} = \frac{2905 + 2585 + 2647}{3} = \frac{8137}{3} = 2712 \text{ тыс. тонн}$$

- Моментные ряды
  - Среднее **хронологическое**

Число работников	
на 1 января	150
на 1 февраля	145
на 1 марта	162
на 1 апреля	166

$$\bar{y} = \frac{150/2 + 145 + 162 + 166/2}{4 - 1} = \frac{465}{3} = 155 \text{ человек.}$$

## • Производный ряд

Число работников	
на 1 января	150
на 1 февраля	145
на 1 марта	162
на 1 апреля	166



Месяцы	Расчет	Среднесписочная численность работников
Январь	$(150+145)/2$	147,5
Февраль	$(145+162)/2$	153,5
Март	$(162+166)/2$	164,0

$$\bar{\Delta} = \frac{(y_2 - y_1) + (y_3 - y_2) + \dots}{n - 1} = \frac{y_n - y_1}{n - 1}$$

$$\bar{K}_p = \sqrt[n-1]{\frac{y_2}{y_1} \times \frac{y_3}{y_2} \times \dots \times \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

$$\bar{T}_p = \bar{K}_p \times 100$$

$$\bar{T}_{np.} = \bar{T}_p - 100$$



- Обязательно должно быть указание на период времени!
- Примеры:
  - за последние **семь лет** **ежегодный** прирост объема продаж в среднем составлял 10%;
  - в **текущем году** **ежемесячно** уровень безработицы снижался в среднем на 0,5% .

# Средний темп роста, пример

Ставка а	Сумма
12%	$1 * (1,12)$
10%	$1 * (1,12) * (1,1)$
5%	$1 * (1,12) * (1,1) * (1,05)$
0,1%	$1 * (1,12) * (1,1) * (1,05) * (1,001)$
7%	$1 * (1,12) * (1,1) * (1,05) * (1,001) * (1,07) =$
<del>34%</del>	<b>= 1,385536152</b>
<del>=6.8%</del>	<b>=1,067391088</b>

оплата 0.0.24 = ?

# Компоненты временного ряда



Уровень ряда может быть представлен как функция четырех компонент:

$$y=f (T,S,C,E)$$

где T – трендовая компонента;

S – сезонная компонента;

C – циклическая компонента;

E – случайная компонента.

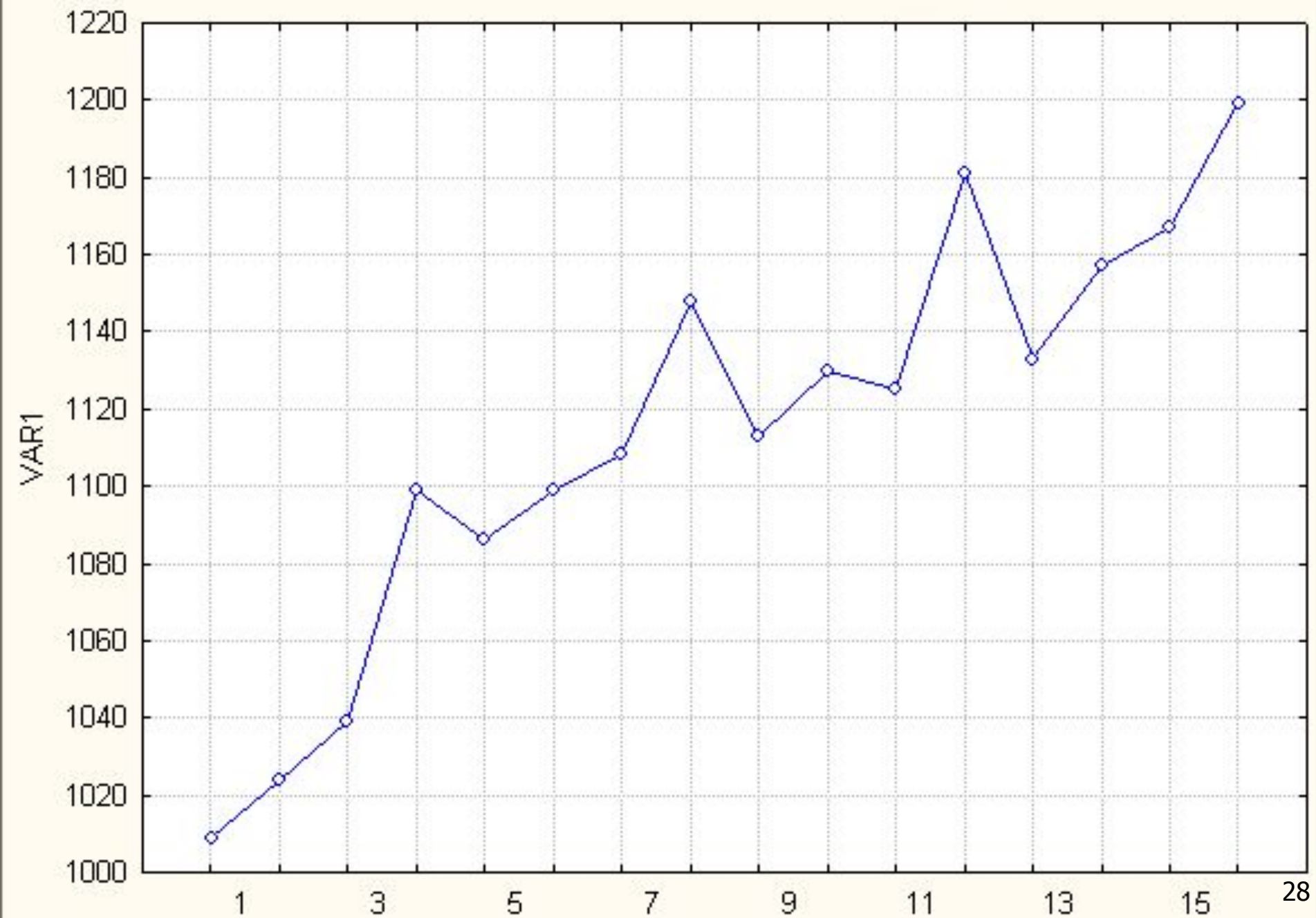


**ПОЛИТЕХ**

Санкт-Петербургский  
политехнический университет  
Петра Великого

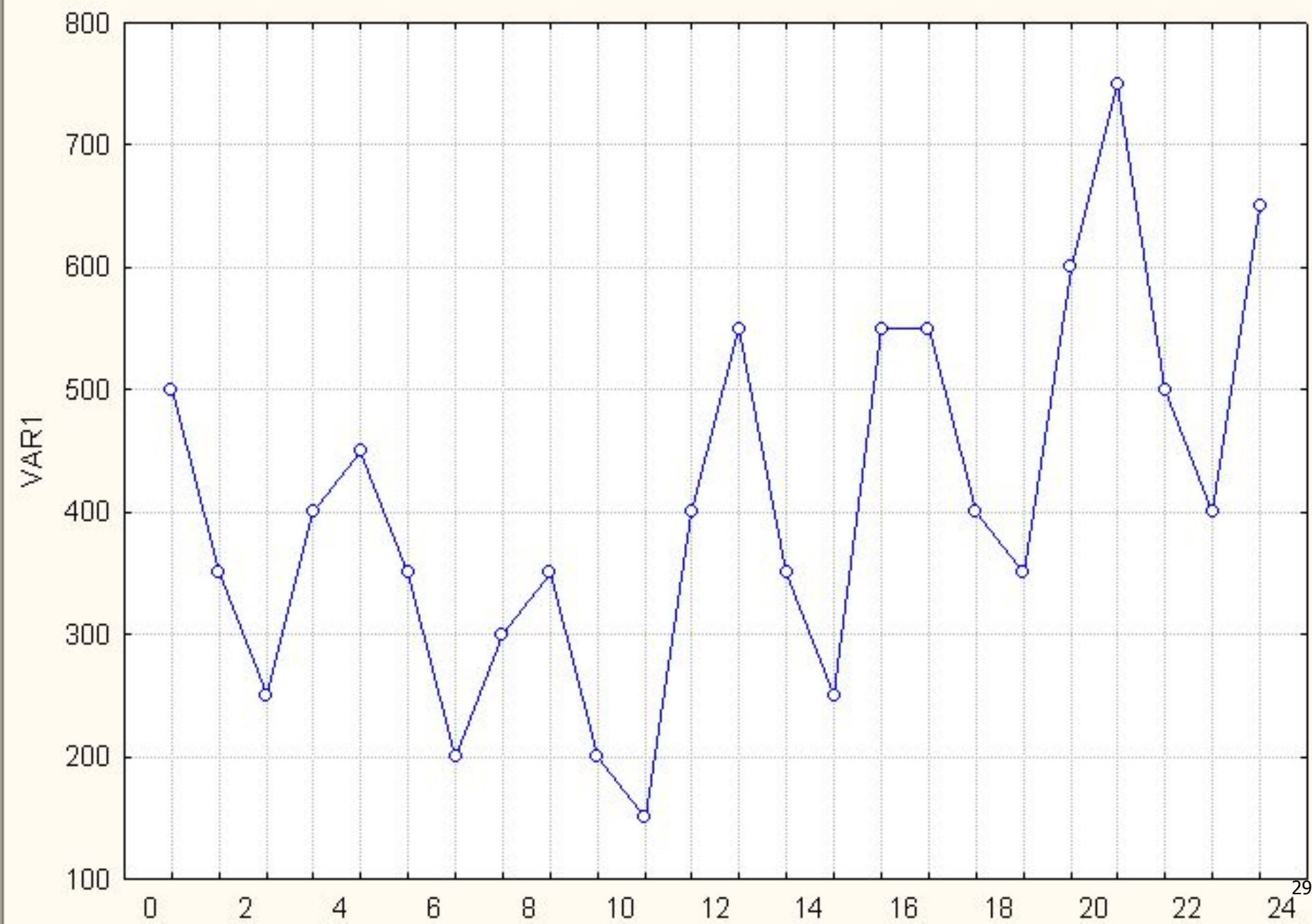
# Примеры графического отображения рядов с разной структурой

Line Plot (Spreadsheet в 171с. Мар. эксп. сглаж 2v\*16с)

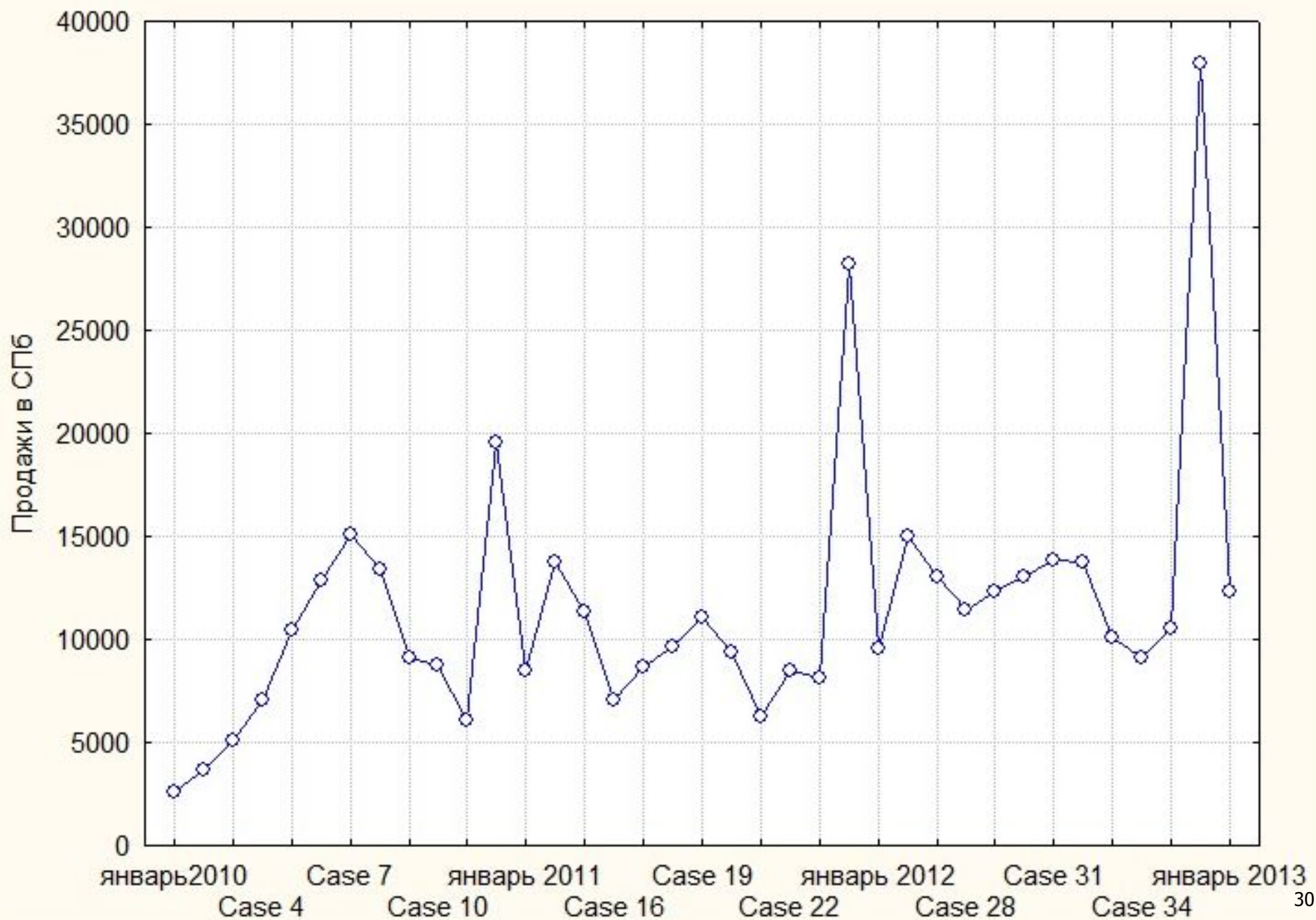




Line Plot (Spreadsheet в динамика 4v\*25с)



Line Plot (Spreadsheet2 9v\*37c)



# Выравнивание рядов

- Цель — **устранить** влияние **случайных** факторов и выявить тенденцию = тренд

## Методы

Механическое  
выравнивание

Аналитическое  
выравнивание

**Метод  
укрупнения  
интервалов.**  
Аналогично  
группировке

**Метод скользящих средних.**  
При определении средних значений случайные отклонения погашаются. Фактические значения ряда заменяются средними значениями, которые характеризуют срединную точку периода скользяния

Выравнивание объемов продаж при  $L = 3$ ,  $L = 5$



Исходный ряд;  
 Среднее по 3 точкам;  
 Среднее по 7 точкам  
**Сколько точек у  
 сглаженного ряда?**

# Аналитическое выравнивание

Это описание формы тренда = **уравнение (модель) тренда.**

Уравнение тренда – это уравнение парной регрессии, фактор - время  $t$ .

Переменная  $t$  задается простой последовательностью чисел от 1 до  $n$ .

Линейный тренд:

$$y = a_0 + a_1 \times t$$

где  $y$  – среднее значение уровней временного ряда;  $a_0$  и  $a_1$  – параметры уравнения тренда;  $t$  – время (независимая переменная, фактор)



# Применимость аналитического выравнивания



Модель можно строить только для стационарного ряда!

- Стохастический процесс называется **стационарным**, если его основные свойства остаются неизменными во времени:
  - **среднее = ?**
  - **дисперсия = ?**
  - **автокорреляция = ?**

**Периодизация** – это деление ряда на периоды, различающиеся направленностью тенденций или интенсивностью изменения уровней ряда.

- Обосновать необходимость периодизации временного ряда можно на основе оценки существенности различия дисперсий и средних значений уровней по выделяемым отрезкам ряда.

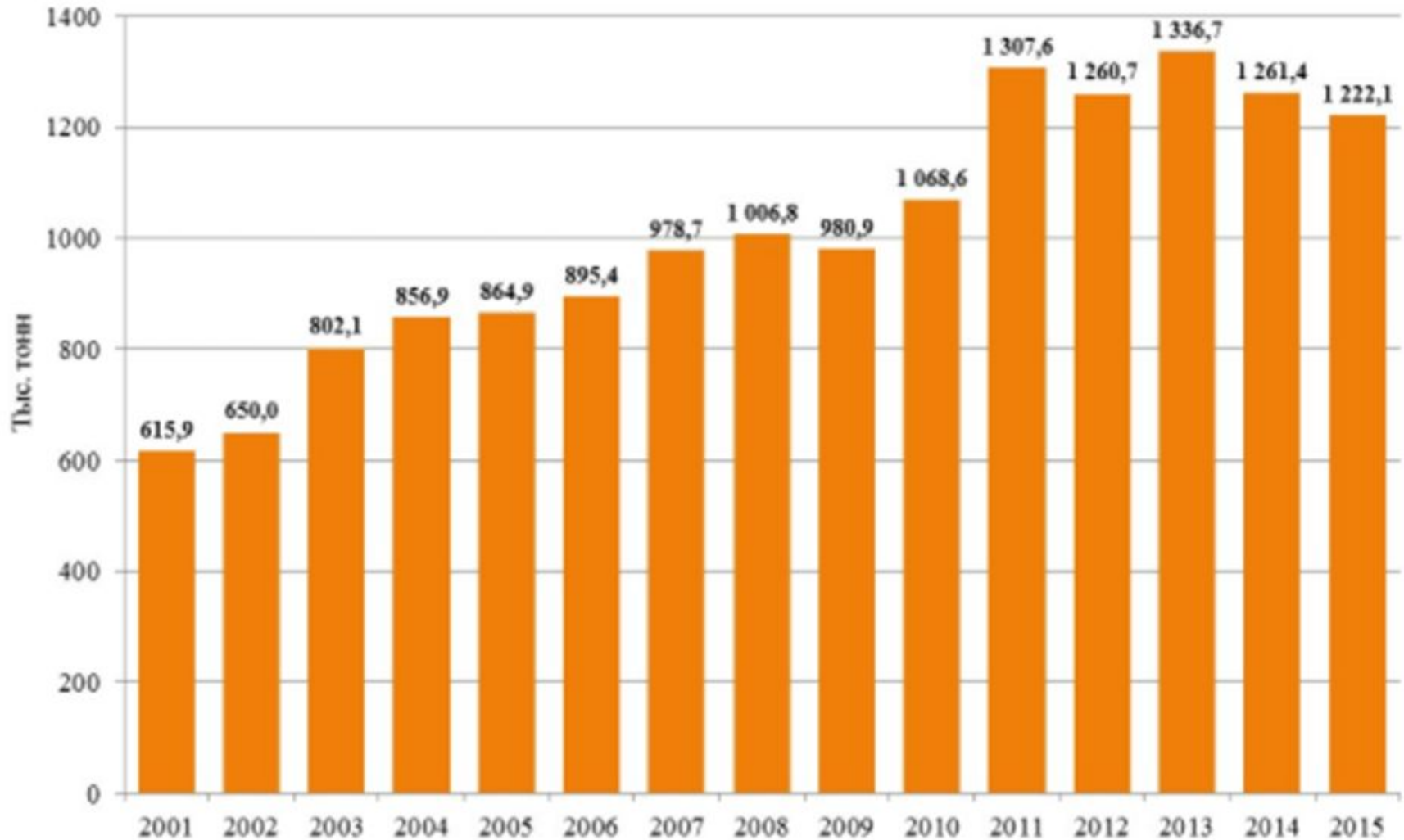
- (См. Трамп)



Динамика импорта бананов в Россию  
в 2001-2015 гг., тыс. тонн



ЭКСПЕРТНО-АНАЛИТИЧЕСКИЙ  
ЦЕНТР  
АГРОБИЗНЕСА  
[www.ab-centre.ru](http://www.ab-centre.ru)

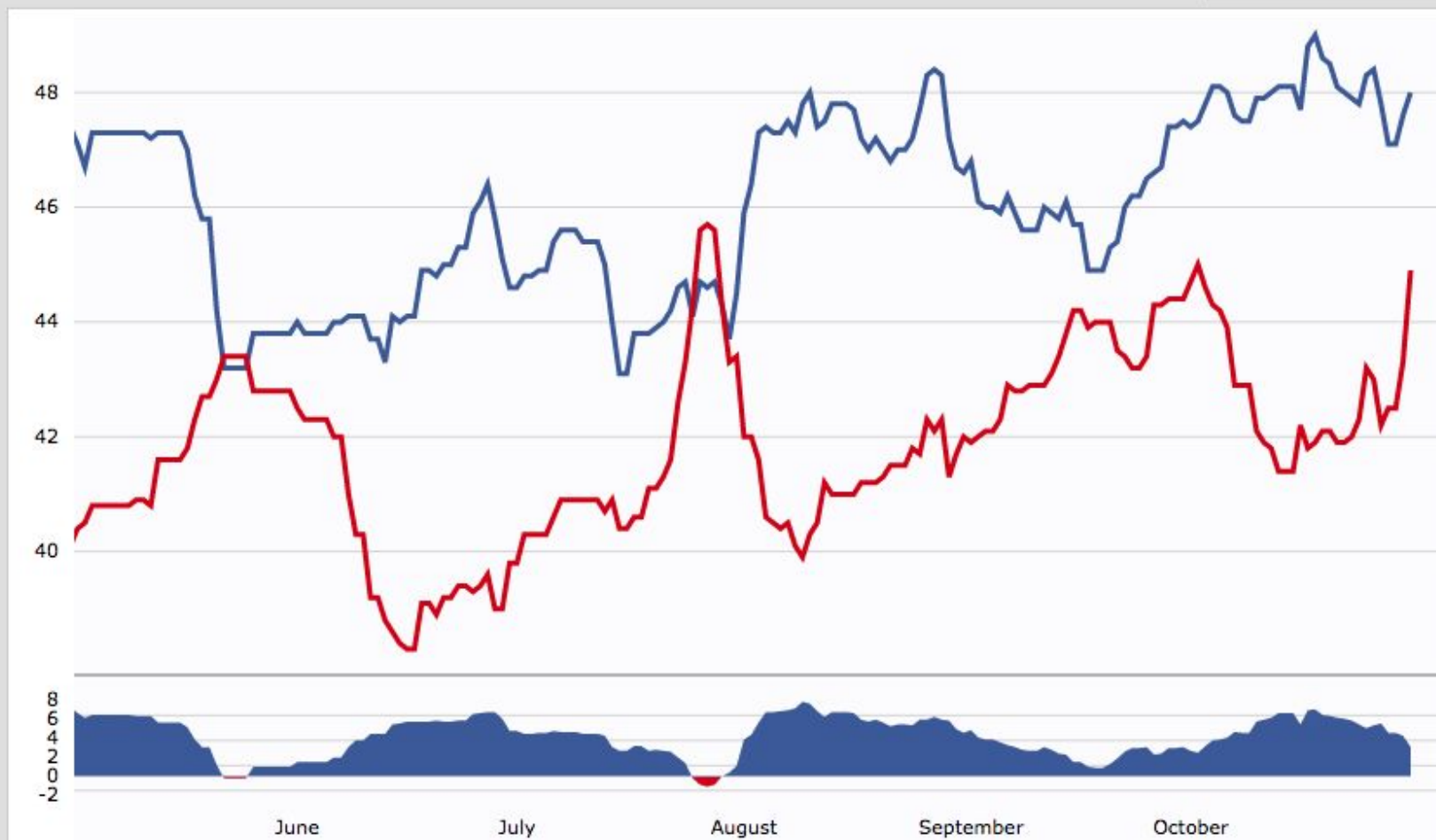


# Оценка существенности различия средних по двум отрезкам



RCP POLL AVERAGE  
General Election: Trump vs. Clinton

48.0 Clinton (D) +3.1  
44.9 Trump (R)



# Оценка существенности различия средних по двум отрезкам

Если есть тренд, то должно иметься существенное различие между средними величинами и дисперсиями двух отрезков временного ряда.

- Оценка существенности различий средних значений осуществляется на основе t-статистики: **Как это сделать?**

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

**Автокорреляция** — статистическая взаимосвязь между последовательностями величин одного динамического ряда, взятыми со сдвигом по времени.

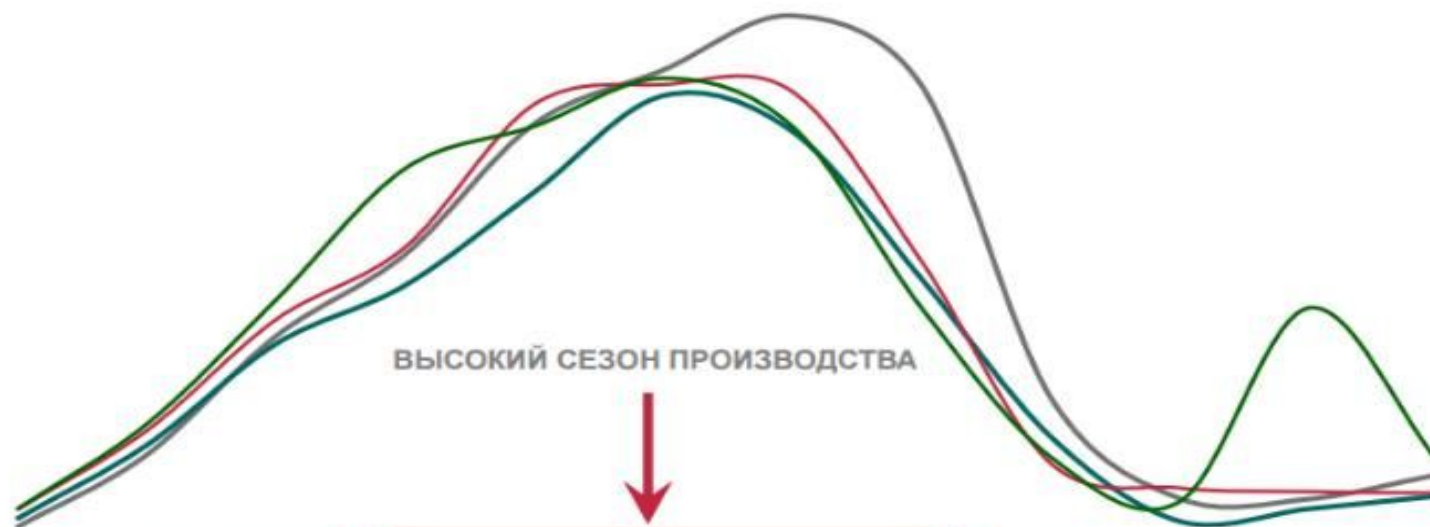
# Автокорреляция

график со сдвигом на 12 периодов



## СЕЗОННОСТЬ ПРОИЗВОДСТВА ПИЛОМАТЕРИАЛОВ

— 2010  
— 2011  
— 2012  
— 2013

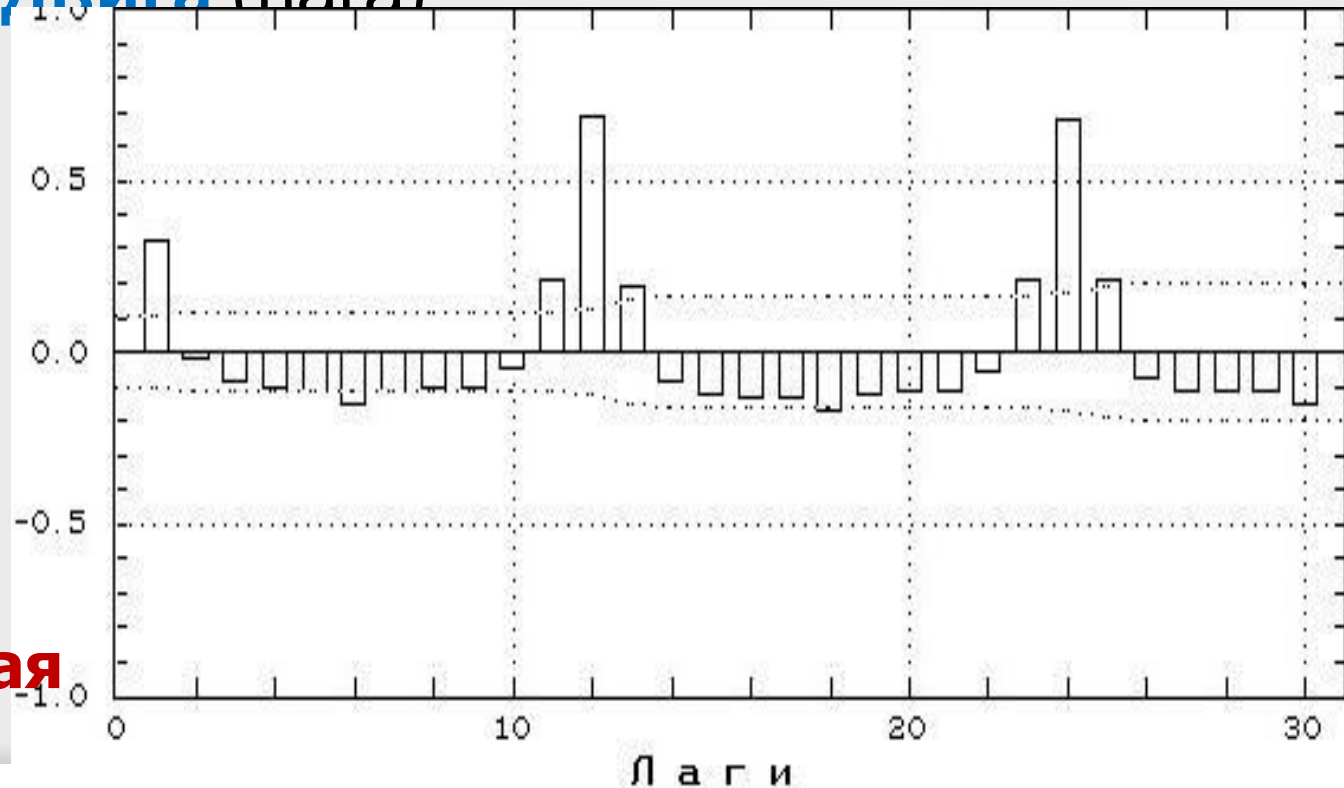


ВЫСОКИЙ СЕЗОН ПРОИЗВОДСТВА

тыс. м3.	Янв	Фев	Мар	Апр	Май	Июн	Июл	Авг	Сен	Окт	Ноя	Дек
<b>2010</b>	31,23	53,28	89,46	114,39	154,23	170,70	187,26	165,84	70,65	39,84	39,69	47,13
<b>2011</b>	6,77	11,25	17,42	20,99	26,69	32,56	30,37	21,37	11,74	6,58	7,34	8,12
<b>2012</b>	16,25	26,96	42,01	51,83	71,21	73,87	72,91	49,85	22,48	19,13	18,64	18,45
<b>2013</b>	18,41	31,34	50,30	70,31	76,76	84,08	76,68	48,98	26,24	19,47	49,02	25,59
<b>2014</b>	23,61	35,48	51,51	65,37								

Автокорреляционная функция —

**зависимость взаимосвязи** между функцией (сигналом) и ее сдвинутой копией **от величины временного сдвига (лага)**



«Феноменальная сезонность»(?)

# Оценка наличия тренда

Для оценки наличия тренда может быть использован **анализ автокорреляционной функции**.

Если коэффициент автокорреляции первого порядка (лаг = 1) **статистически значим** и автокорреляционная функция медленно убывает, то это говорит **о присутствии тренда**.

$r$  – коэффициент автокорреляции, **лаг = 1**.  
 $\sigma_r$  – стандартная ошибка коэффициента автокорреляции  
 $n$  – число измерений ряда

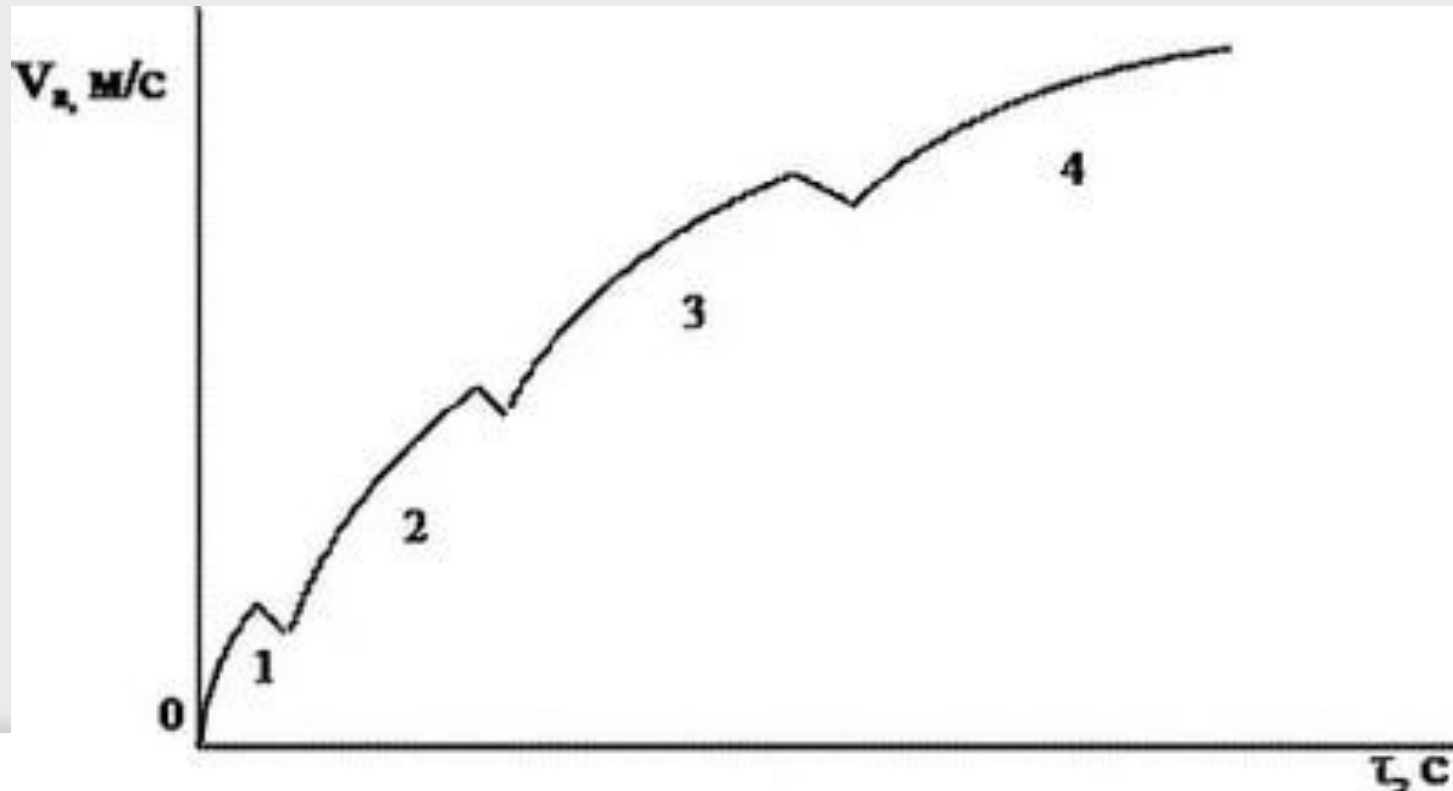
$$z = \frac{r}{\sigma_r}$$

$$\sigma_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

# Оценка наличия тренда

Разгон автомобиля, измерения каждые 1/10 секунды

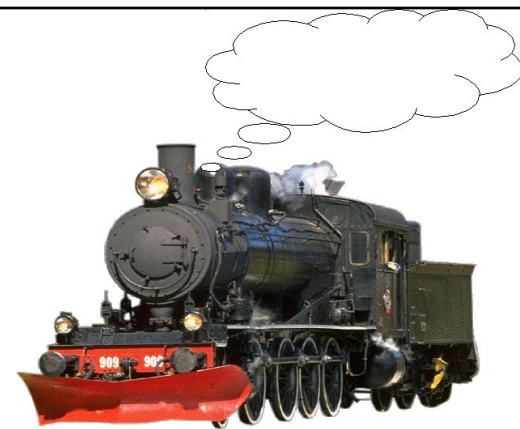
**Автокорреляция с лагом 1 = ?**





# Автокорреляционная функция

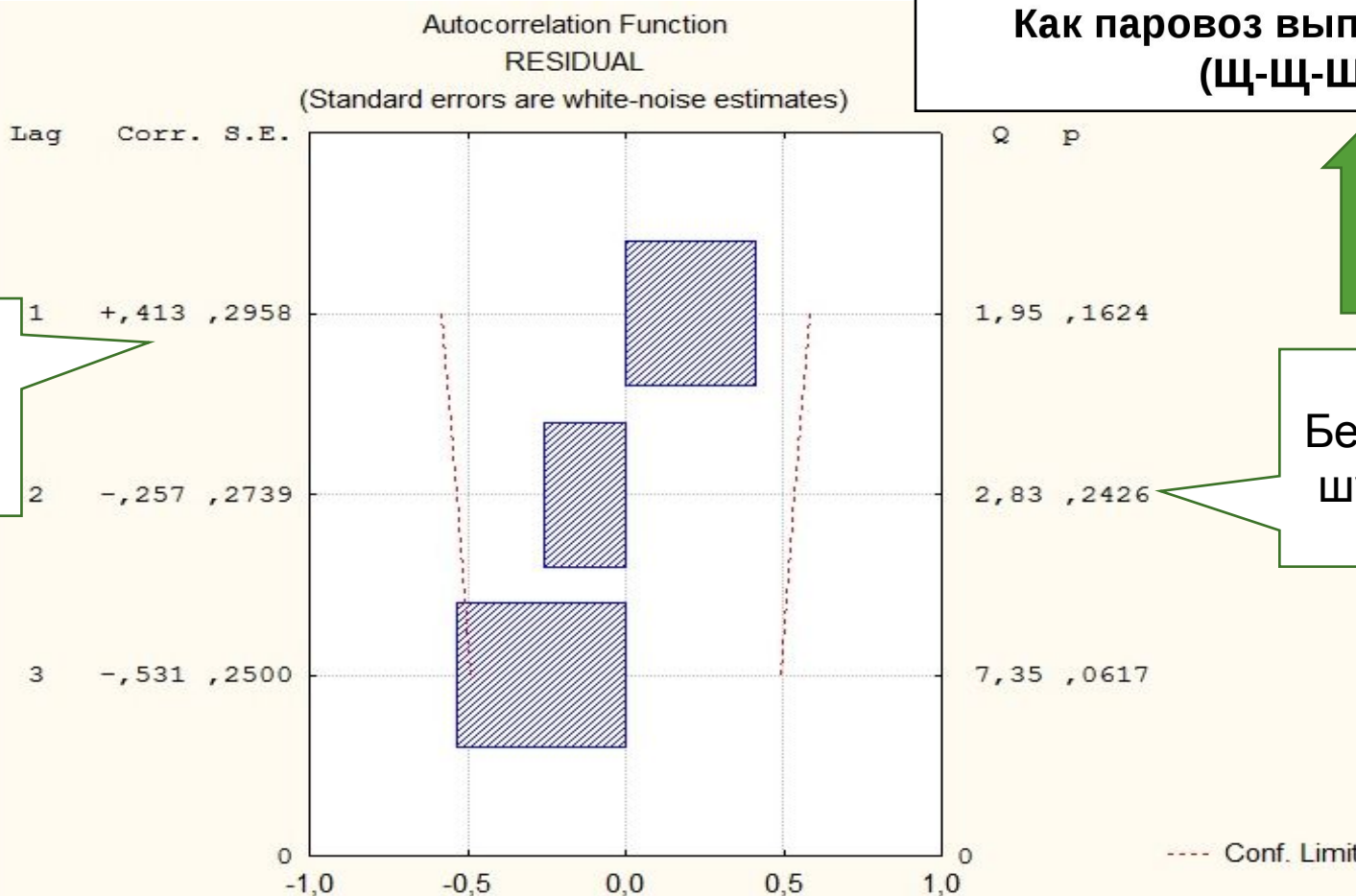
Corr.  $\rho$



Как паровоз выпускает пар?  
(Щ-Щ-Щ!..)



Белый шум?



Автокорреляция



## Выбор вида уравнения тренда

# Примеры уравнений тренда

• Полином к-й степени: 1)  $\hat{y} = a_0 + a_1t + a_2t^2 + \dots + a_k t^k$

• Показательная функция:

$$2) \hat{y} = a_0^{a_1 t}$$

• Гипербола:

$$3) \hat{y} = a_0 + \frac{a_1}{t}$$

• Логистические кривые (Перла-Рида, Гомперца, другие S-образные кривые)

$$4) \hat{y} = \frac{1}{c + a_0 e^{a_1 t}}$$

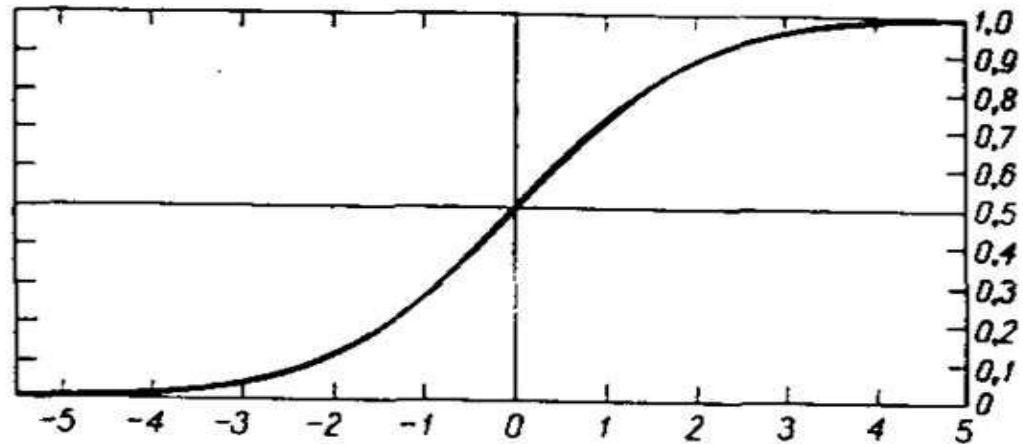
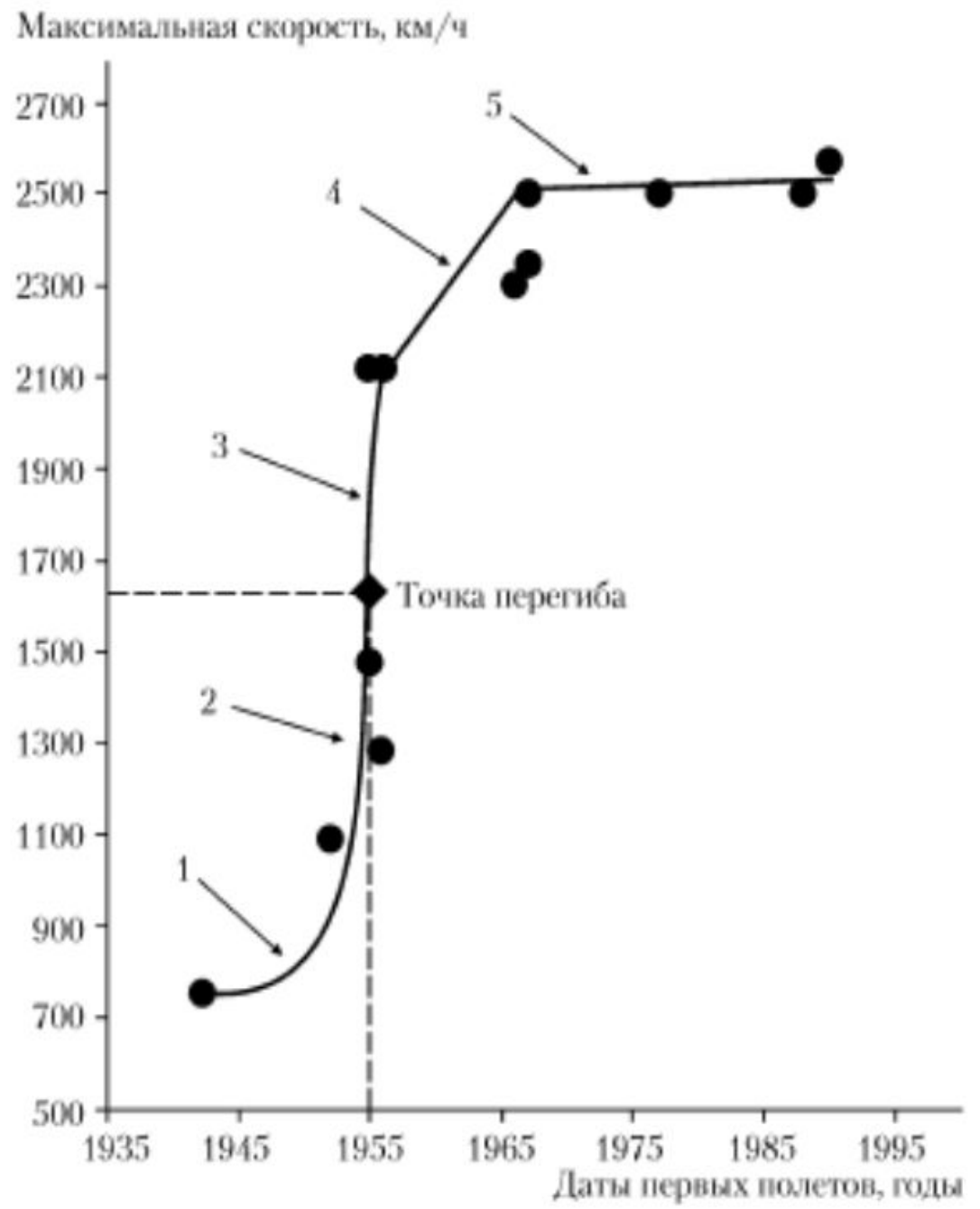
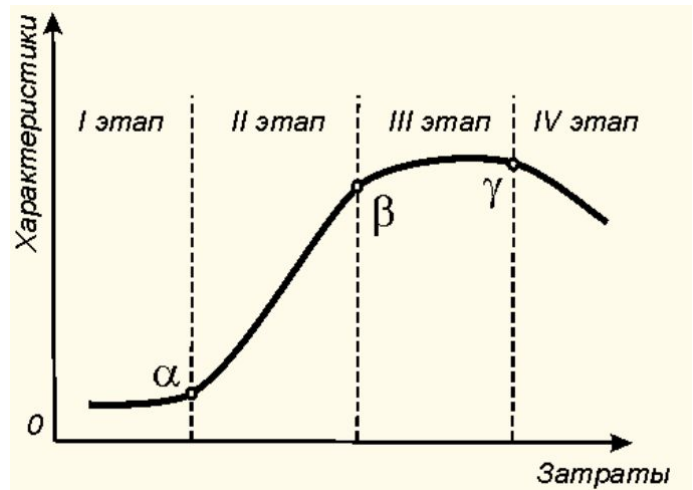
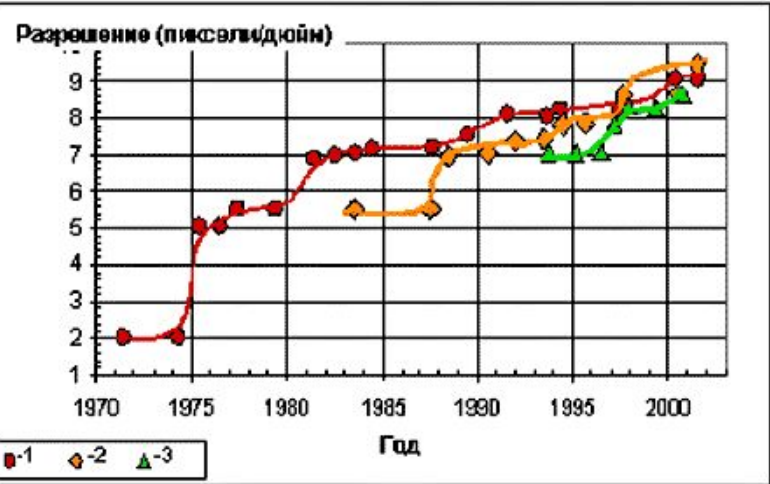


Рис. 27. Кривая Перла

# Примеры уравнений тренда



# Способы выбора

- Графическое представление ряда (видно на глаз)
- Метод последовательных разниц (Абсолютные приросты **примерно** одинаковы => линейный тренд).
- Формализованный подход (рассчитываем несколько вариантов, смотрим, который **«лучше»**)

# Графическое представление

Динамика глобального потепления, зафиксированная метеостанциями



Минимизация суммы квадратов отклонений фактических значений уровней от значений, полученных по уравнению тренда:

$$\sum (y_t - \hat{y}_t)^2 \rightarrow \min$$

где  $y_t$  – фактический уровень ряда периода  $t$ ;

$\hat{y}_t$  – теоретический уровень ряда периода  $t$ , полученный на основе уравнения тренда.

**См. выше – как построили кривые?** 51

Максимальное  
значение  
критерия  
Фишера:

$$F \rightarrow \max$$

Максимальное  
значение  
коэффициента  
детерминации:

$$R^2 \rightarrow \max$$

Минимальное  
значение  
остаточной  
дисперсии.

$$\sigma^2_{ост.} \rightarrow \min$$



- Минимальное значение среднеквадратической ошибки уравнения тренда:

$$\sigma = \sqrt{\sigma_{\text{остат.}}^2}$$

- Минимальное значение средней ошибки аппроксимации:

$$E = \frac{1}{n} \sum \frac{|y_t - \hat{y}_t|}{y_t} \rightarrow \min$$



## Оценка результатов

- Компьютеру задаются:
  - исходные данные – значения ряда
  - вид уравнения
- Компьютер рассчитывает:
  - параметры уравнения тренда
  - статистическую значимость параметров уравнения тренда (t-статистика)
  - статистическую значимость уравнения в целом (F-критерий)

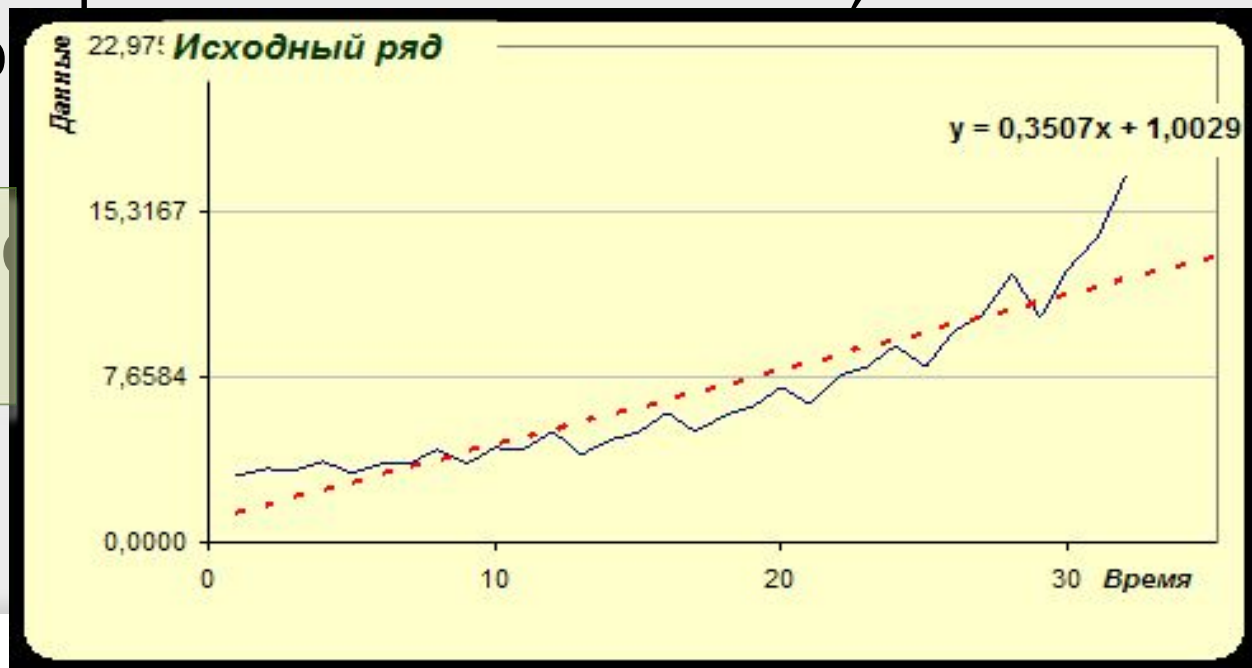
Если значимость отсутствует,  
уравнение непригодно

**И что тогда делать?**

- **Остатки** – разность между фактическими значениями уровней ряда и значениями, полученными по уравнению регрессии.
- **Автокорреляция остатков** – это зависимость остатков периода  $t$  от остатков предшествующего периода  $(t-i)$ .
- Если автокорреляция в остатках значима, то не все закономерно учтены моделью

Должен остаться белый шум

**Прогноз  
похож на  
правду?**

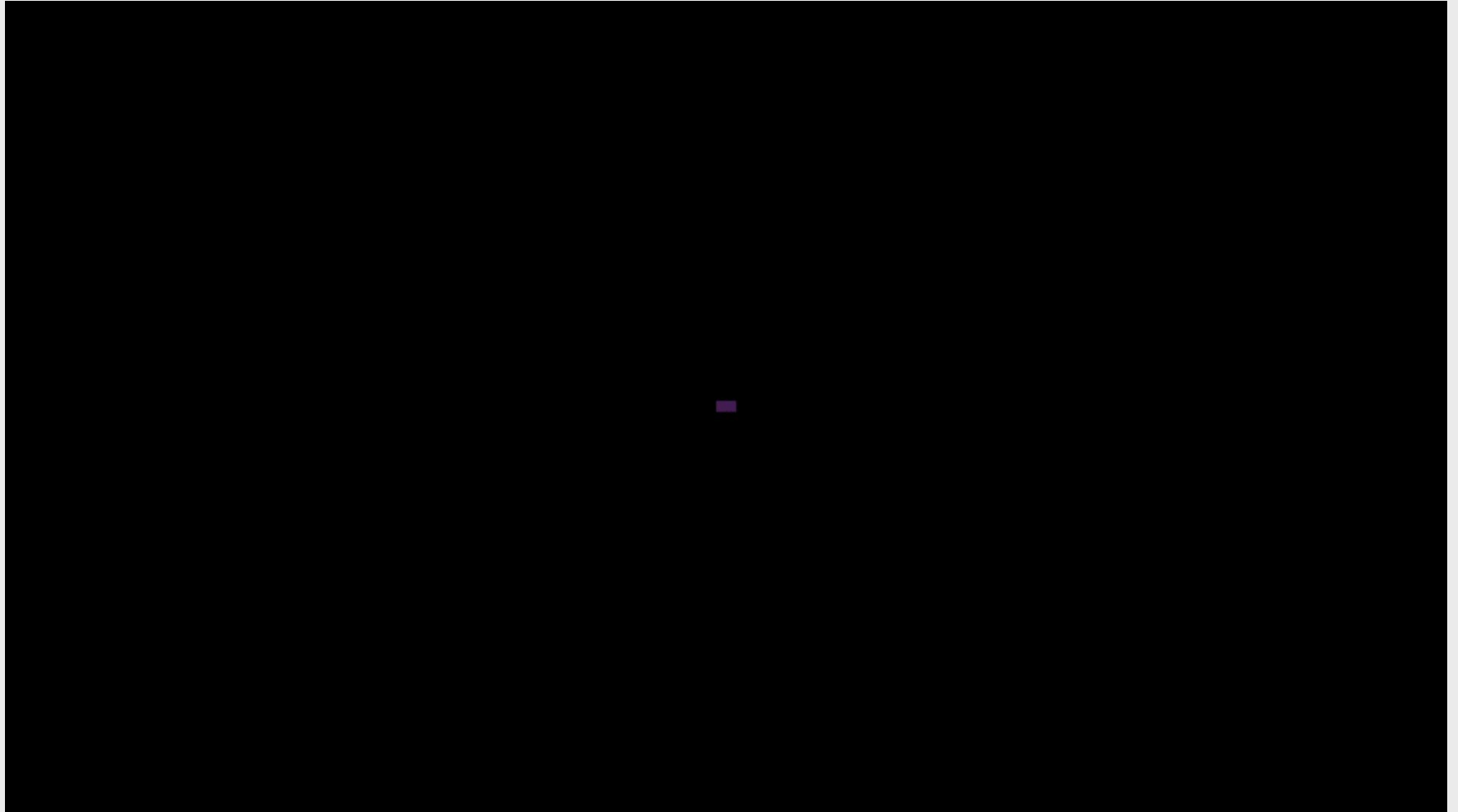




**ПОЛИТЕХ**

Санкт-Петербургский  
политехнический университет  
Петра Великого

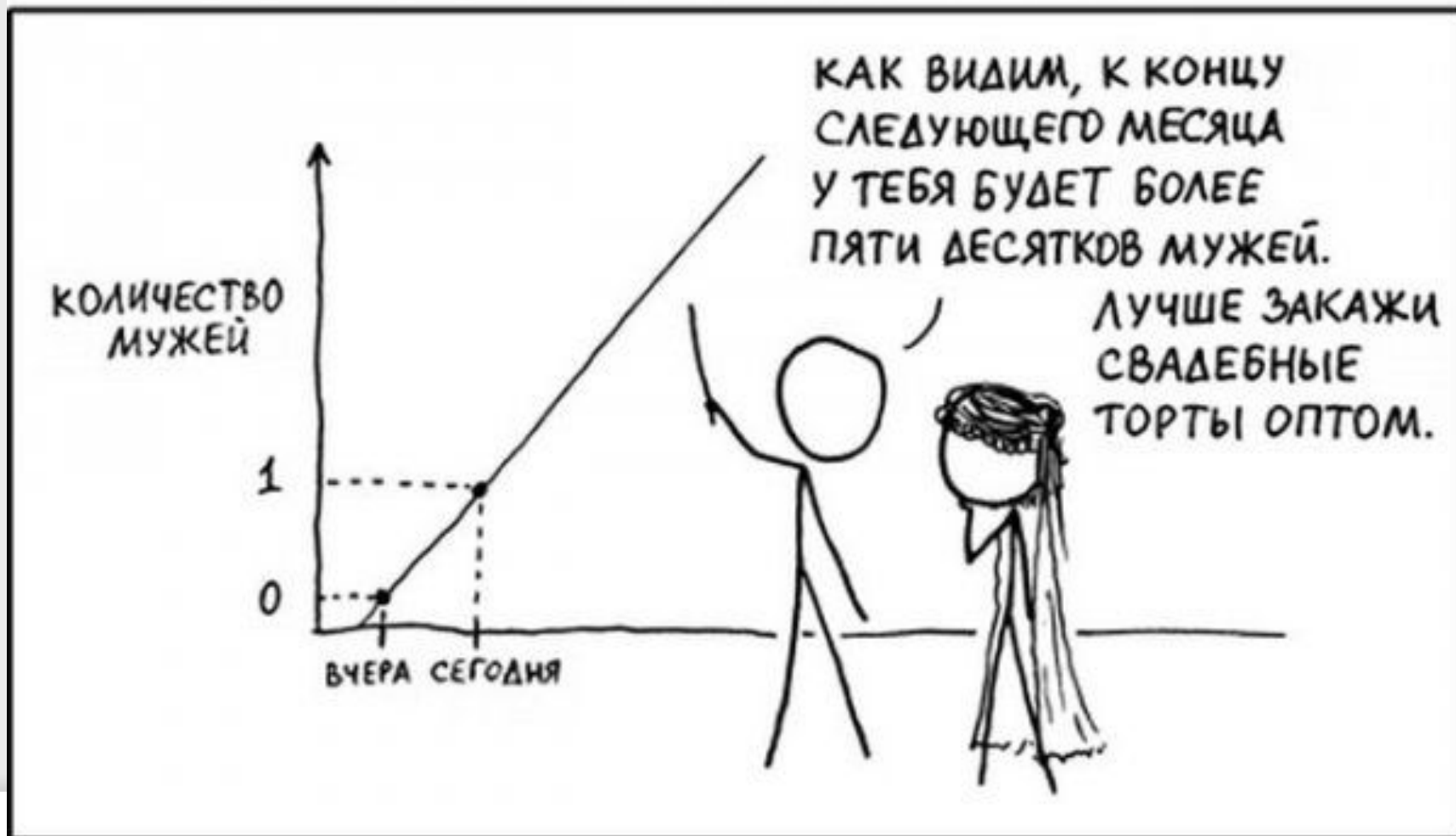
# Прогнозирование методом экстраполяции

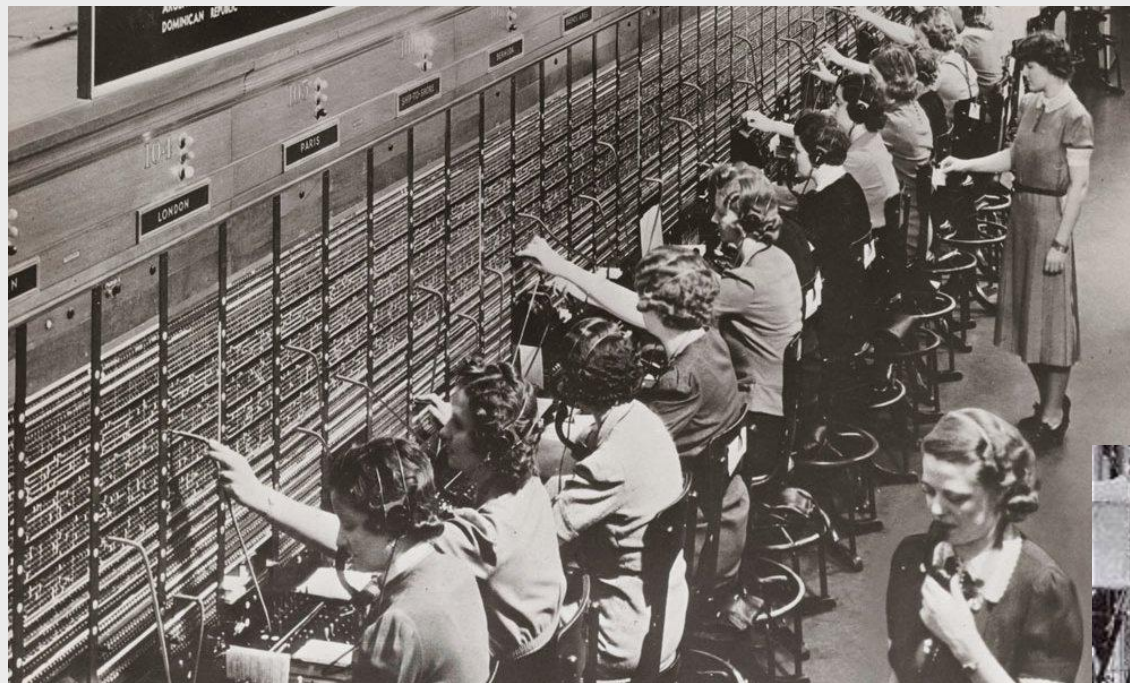


# Экстраполяция = прогнозирование

Экстраполяция - это продление в будущее тенденции, сложившейся в прошлом.

МОЁ ХОББИ: ЭКСТРАПОЛИРОВАТЬ

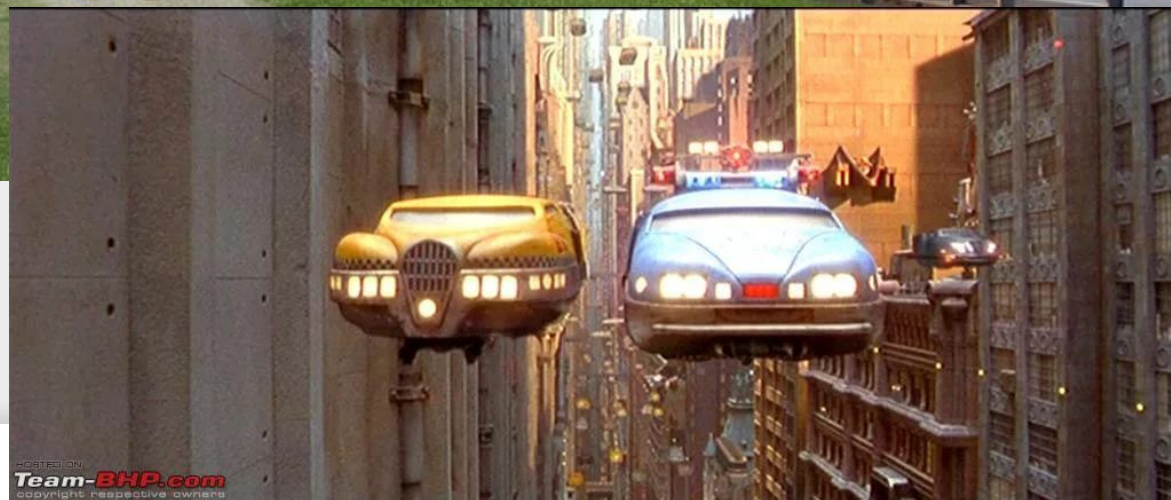
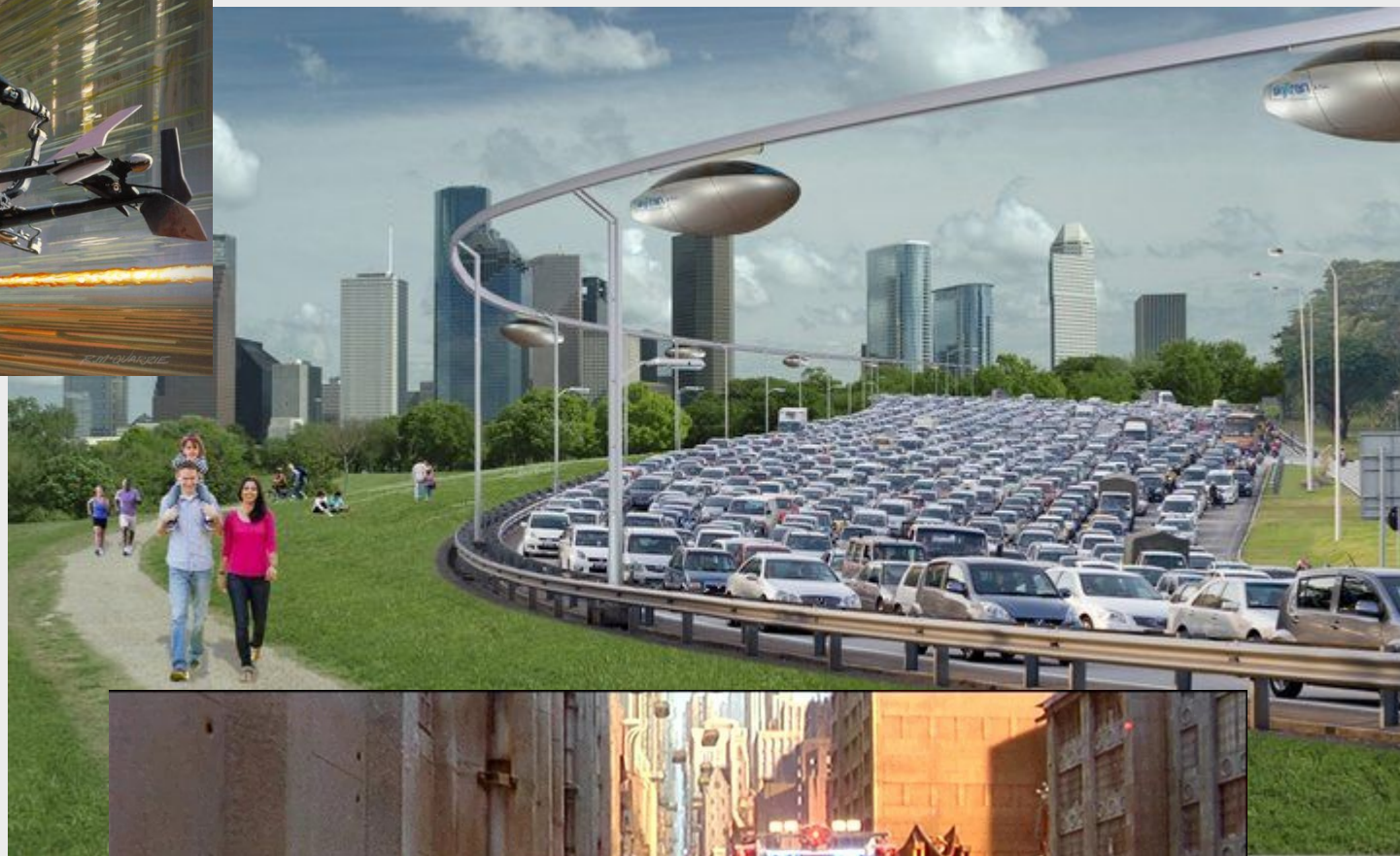




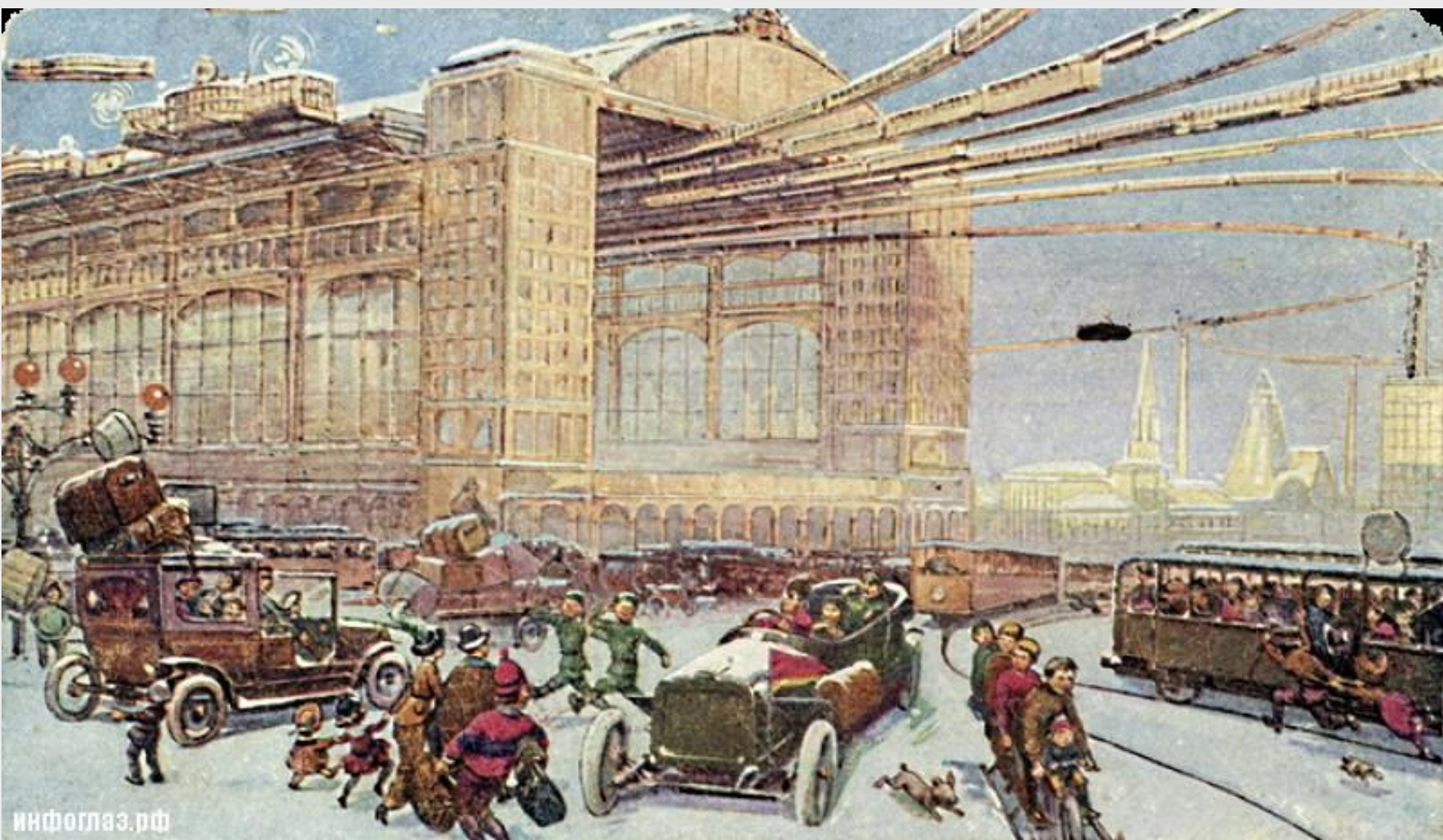
**Нью-Йорк 1887**



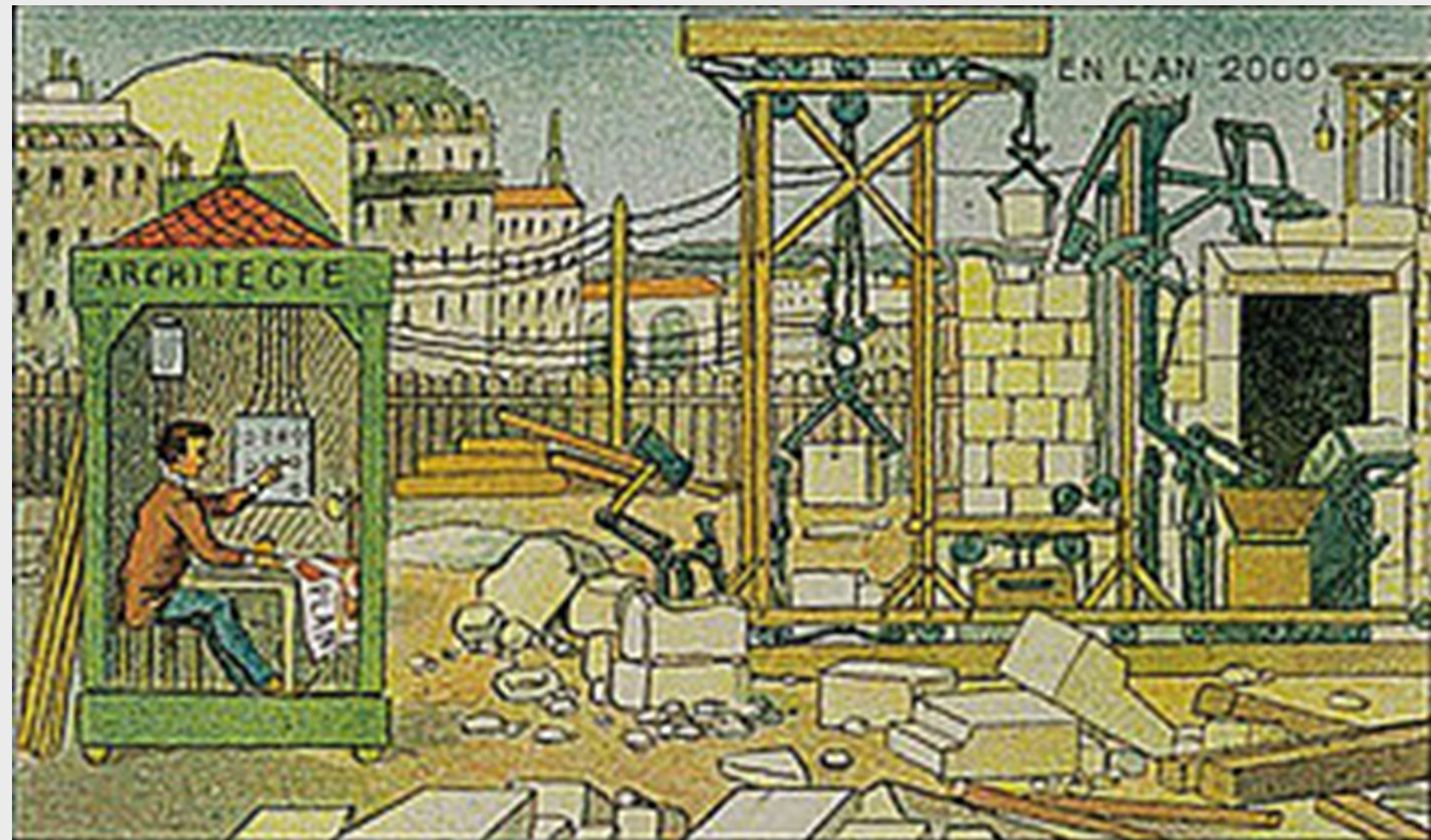
# Прогноз будущего в начале XXI века





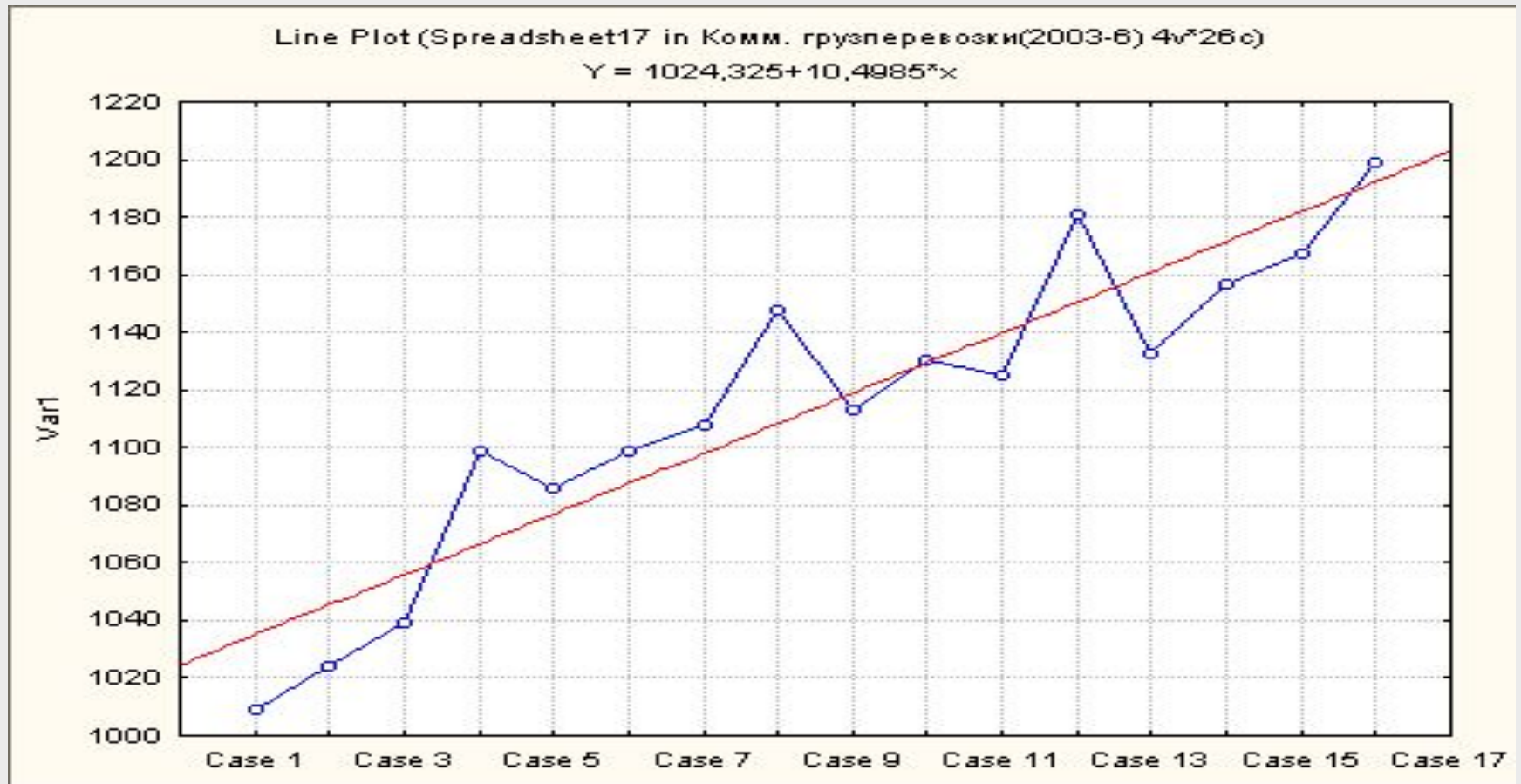






- Отсутствует значимая автокорреляция в остатках;
- Параметры уравнения тренда статистически значимы;
- Уравнение тренда статистически значимо;
- Действие факторов, сформировавших тенденцию (закономерность) развития, в будущем сохранится неизменным.

# Динамика коммерческого грузооборота в России (поквартально, млрд. т-км.)



Line Plot (Spreadsheet17 in Комм. грузперевозки(2003-6) 4v\*26с)



# Уравнение тренда динамики коммерческого грузооборота

Regression Summary for Dependent Variable: Var1 (Spreadsheet17 in							
R= ,91990888 R <sup>2</sup> = ,84623236 Adjusted R <sup>2</sup> = ,83524895							
F(1,14)=77,046 p<,00000 Std.Error of estimate: 22,054							
N=16	Beta	Std.Err. of Beta	B	Std.Err. of B	t(14)	p-level	
Intercept			1024,325	11,56533	88,56861	0,000000	
NewVar	0,919909	0,104802	10,499	1,19606	8,77761	0,000000	

# Прогнозирование по выбранному уравнению регрессии

1. Делается **точечный прогноз** путем подстановки значения порядкового номера периода, на который делается прогноз, в уравнение тренда.
2. Рассчитывается **доверительный интервал** прогноза с учетом ошибки оценивания.

# Точечный прогноз коммерческого грузооборота

$$Y = 1024,33 + 10,50 * 17 =$$
$$= 1202,83 \text{ (млрд. тонно-км.)}$$

**Что еще требуется?**



## Доверительный интервал прогноза

$$1202,83 - 2*22,05 < Y < 1202,83 + 2*22,05$$

$$1158,73 < Y < 1246,93$$

достоверность 95%

(факт оказался = 1167 млрд.тонно-км.)

Прогноз на небольшой срок (**1/10** длины ряда)



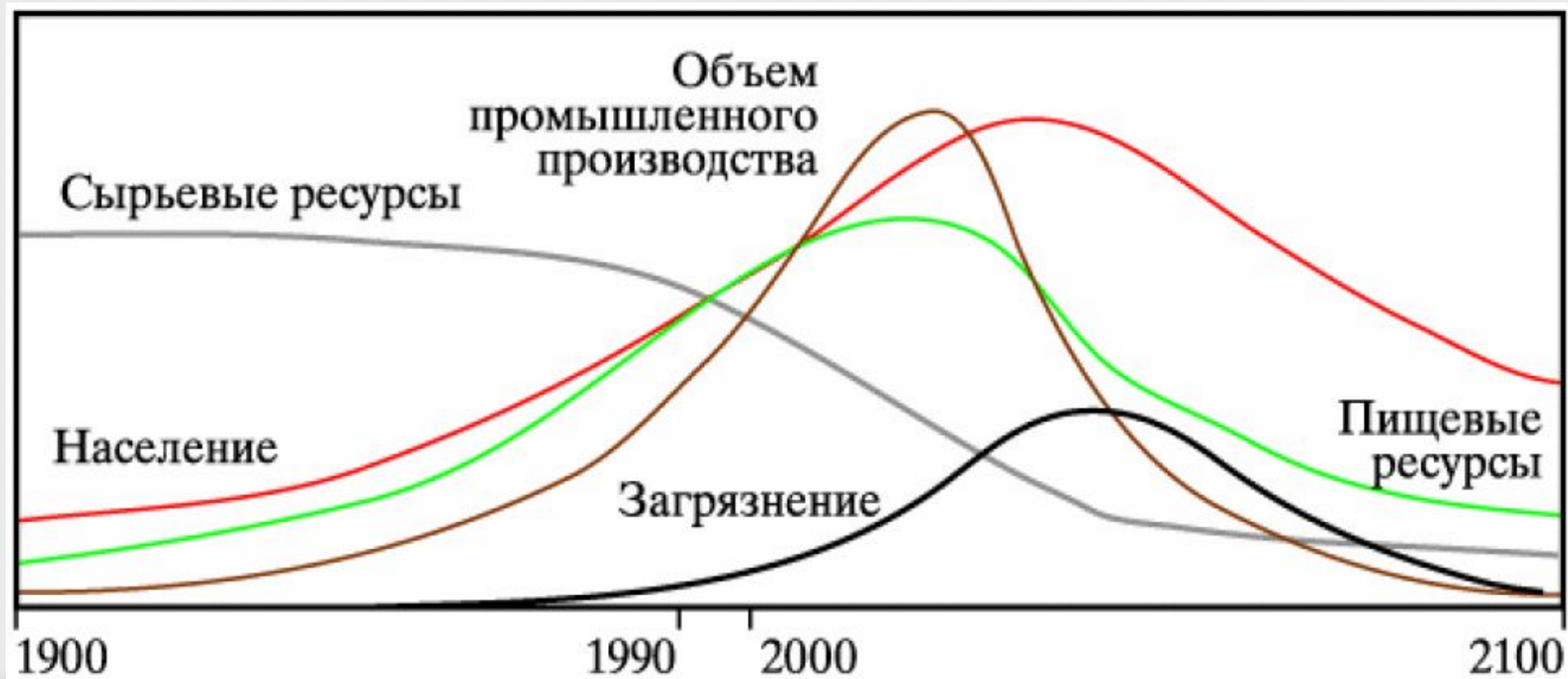
## Другие методы прогноза

- **Экстраполяция**
- Моделирование
- Сценарии
- Историческая аналогия
- Экспертные оценки
- «Адвокат дьявола»

# Динамическая модель

Дифференциальные уравнения: величина зависит от скорости изменения другой величины или наоборот (скорость убывания пищевых ресурсов зависит от численности населения)

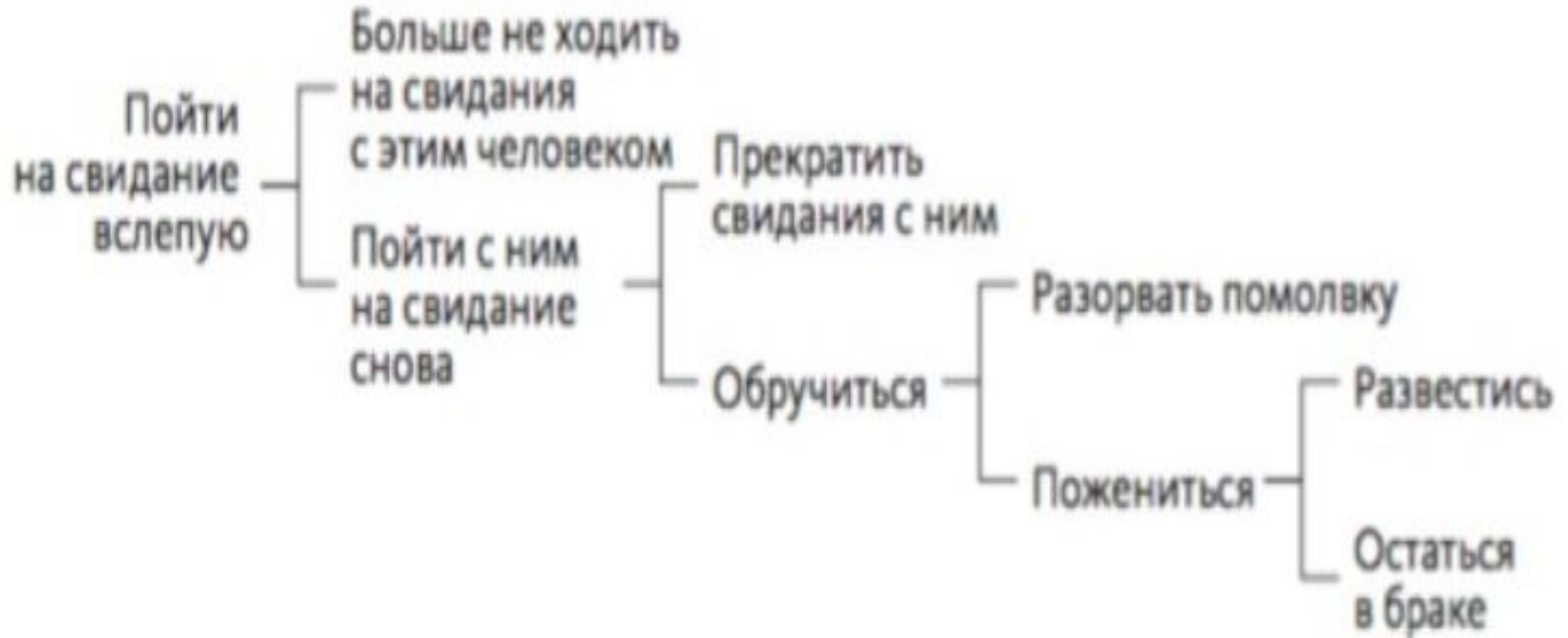
**В чем главный минус?**





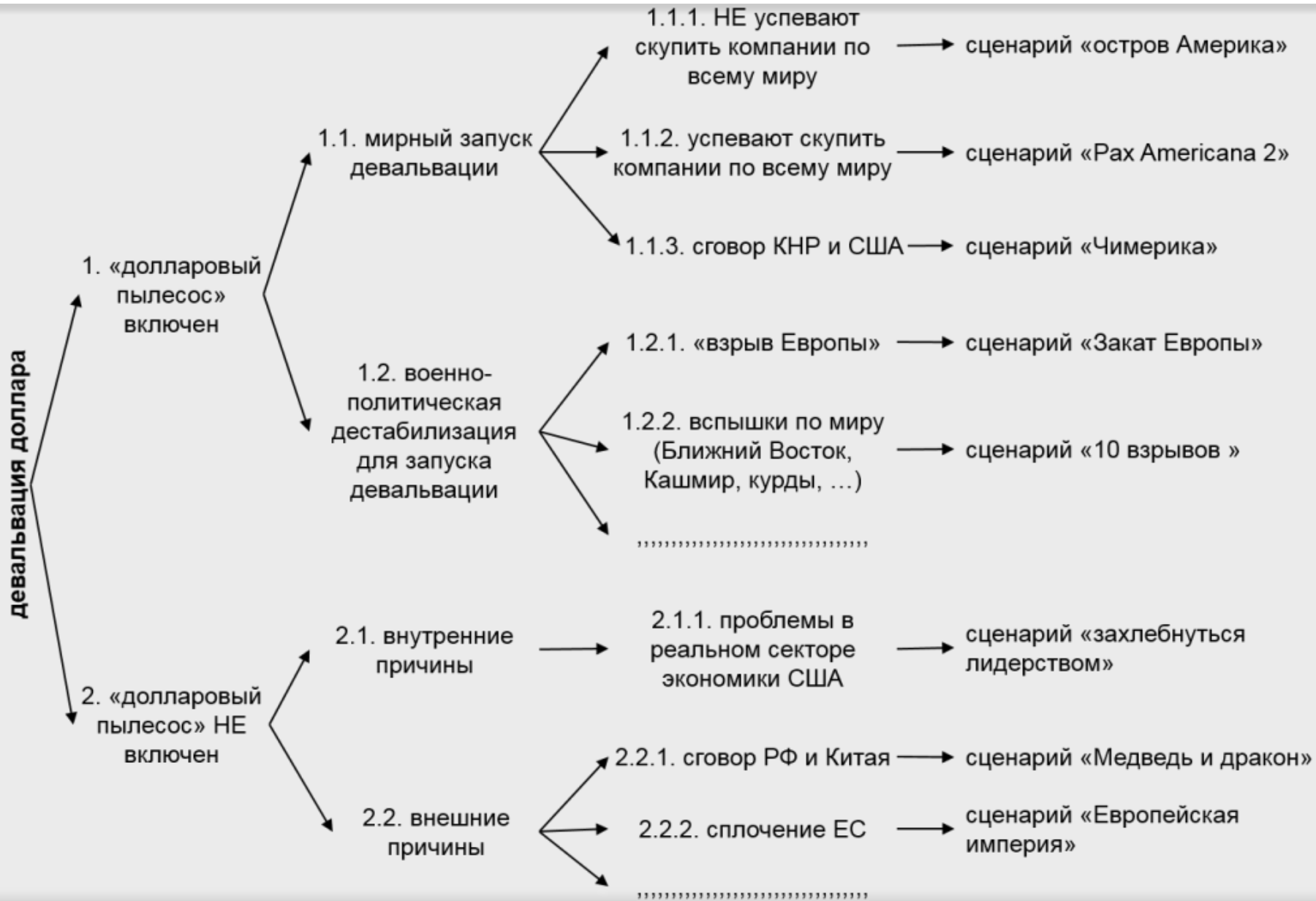
## Структура содержательной части сценария







# Сценарии будущего





- Выявляется сходство с другим объектом.
- Допущение: основные события прошлых лет повторятся в будущем, если факторы, их порождающие, сохранят свое значение.
- Трудности
  - отличить истинную причину от мнимой
  - обеспечить репрезентативность наблюдений для отражения возможных ситуаций в будущем
- Самолет «Илья Муромец» – 1913. Ремни безопасности в самолетах – 1930. Разница 17 лет.
- Серийный ВАЗ – 1970. Ремни – 1975. Разница 5 лет.
- Сапсан – 2009. Ремни = ?
- Разницы: 17 лет, 5 лет, среднее: 11 лет. Это и есть прогноз разницы.
- **Ремень в Сапсане будет в 2020 году.**

- Форсайт – фактически, это мозговой штурм



- Есть много приемов, например, Уолта Диснея
- **Метод Дельфи** – достижение консенсуса (CM Internet)

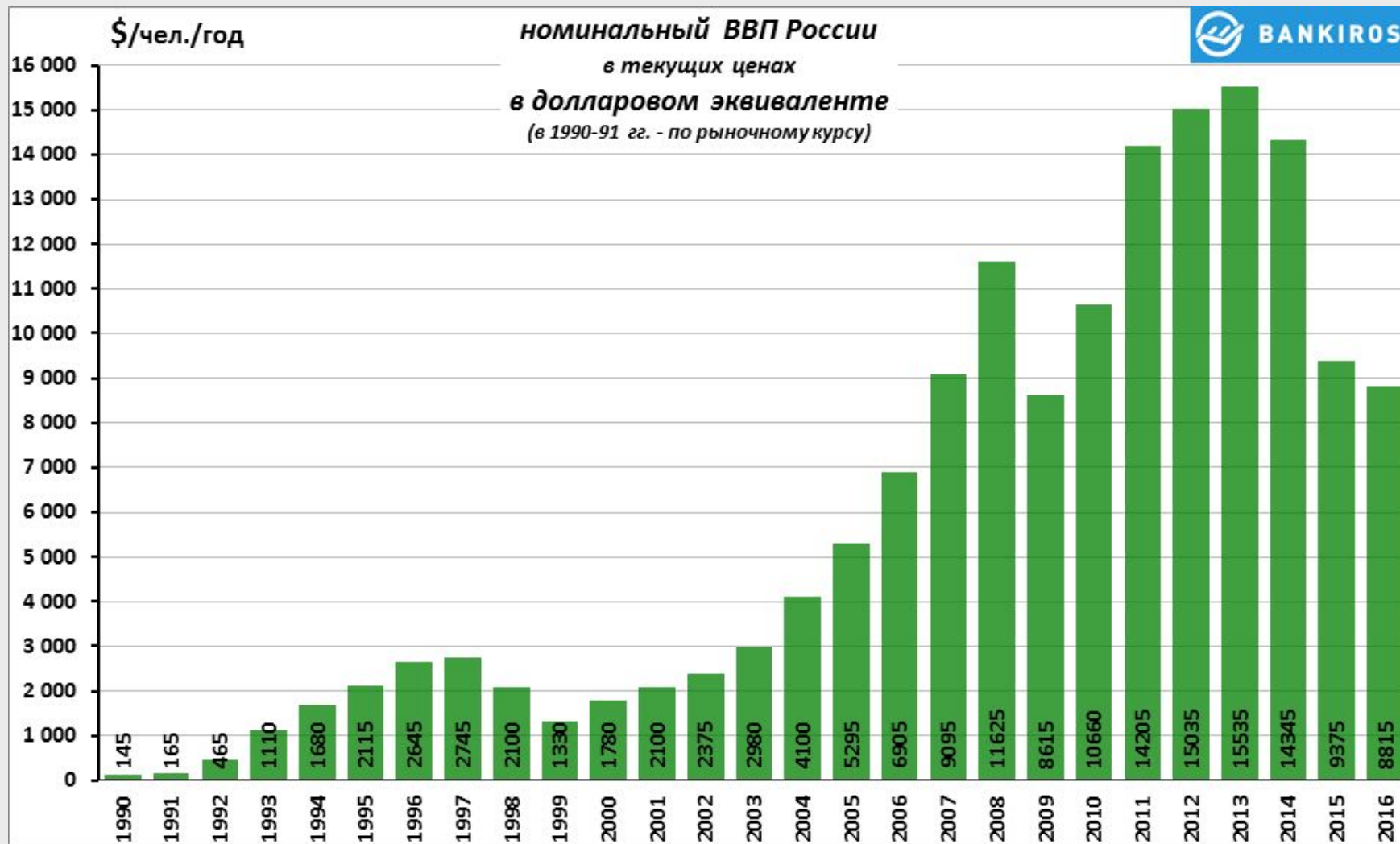
Метод суда.  
Делается прогноз.  
Назначается адвокат, прокурор,  
судья, присяжные.  
Идет процесс.  
Рассматриваются все «за» и  
«против».



## В текущих ценах







# Реальный ВВП РФ

Реальный ВВП = Номинальный ВВП / Общий уровень цен





**Конец части 6**