

ОСНОВЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

Лекция 2



Лекцию читает

к.т.н., профессор

БОБРОВА

ЛЮДМИЛА ВЛАДИМИРОВНА

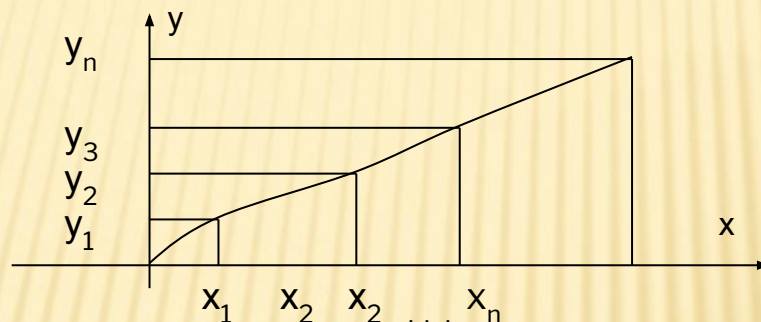


5. РЕГРЕССИОННЫЙ АНАЛИЗ



5.1. Постановка задачи

Пусть в моменты времени x_1, x_2, \dots, x_n измеряются значения некоторой величины y_1, y_2, \dots, y_n .



Предположим, что теперь нужно получить прогноз значения переменной y в момент x_{n+2} . Для этого нужно иметь математическое описание зависимости $y=f(x)$.

При этом возможны два подхода:

Первый – подбирают функцию $f(x)$ так, чтобы она проходила **точно** через узлы $(x_i; y_i)$ – это задача **интерполяции**.

Второй – функция $f(x)$ проходит **как можно ближе** к узлам $(x_i; y_i)$. Это задача **аппроксимации**.

Функция, полученная при этом, называется **функцией регрессии**.



Если связь между переменными x и y **линейная**, регрессия называется **линейной**.

Если переменные связаны нелинейным образом, регрессия будет **нелинейной**.

При линейной связи между переменными, уравнение регрессии имеет вид

$$y = b_0 + b_1 x.$$

Коэффициенты b_0 и b_1 называются **коэффициентами регрессии**.



Если рассматривается зависимость **между двумя** переменными X и y , регрессия называется **парной**.

Если существует связь между **одной** зависимой переменной y и **несколькими неизвестными** переменными X , говорят о **множественной регрессии**, например,

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$



5.2. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (МНК)

Этот метод наиболее часто используется для получения уравнения регрессии. Предположим, между значениями x и y существует линейная зависимость

$$y = b_0 + b_1 x.$$

Нам нужно найти

такую функцию

$$y^* = f(x) = b_0^* + b_1^* x$$

которая проходила бы как можно ближе к функции y .

Будем искать такую функцию $f(x)$, для которой величина

$$S = \sum_{i=1}^n [y_i - y_i^*]^2$$

была бы минимальной.

Это метод **наименьших квадратов**.

Итак, ищем минимум функции

$$S = \sum_{i=1}^n (y_i - b_0^* - b_1^* \cdot x_i)^2$$

Для этого нужно взять частные производные функции S по a^* и по b^* ,

$$\left(\frac{\partial S}{\partial b_0^*}; \frac{\partial S}{\partial b_1^*} \right)$$

приравнять их к нулю и решить полученную систему уравнений.



Получим

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

(1)

$$b_0 = \frac{1}{n} \sum_{i=1}^n y_i - b_1 \frac{1}{n} \sum_{i=1}^n x_i$$

Имея уравнение $y^* = b_0^* + b_1^* x$ и подставляя в него значения x за будущий период, можно осуществить прогнозирование.

Обозначим через e_i разность

$$e_i = y_i - y_i^* = y_i - b_0^* - b_1^* \cdot x_i$$

Случайная величина e_i называется остатком регрессии в i -м наблюдении.

Метод наименьших квадратов (МНК) можно записать в виде:

$$S(b_0, b_1) = \sum_{i=1}^n e_i^2 = \sum (y_i - b_0 - b_1 \cdot x_i)^2 \rightarrow \min$$

Пример 1. (аналог задания 2 Контрольной работы)

Имеются данные о затратах на рекламу (х, усл. ед.) и объеме реализации продукции (у, усл. ед.).

Таблица 1

x_i	2	3	4	5	6
y_i	1,4	1,7	1,8	1,7	1,9

Построить выборочное уравнение линейной парной регрессии для выборки из табл. 1.

Определить прогноз объема продаж при затратах на рекламу $x = 7$ у.е.



Сведем вычисления в табл. 2

Таблица 2

i	x	y	xy	x^2	y^*	e_i	e^2
1	2	1,4					
2	3	1,7					
3	4	1,8					
4	5	1,7					
5	6	1,9					
Сумма	20	8,5					

Таблица 1

x_i	2	3	4	5	6
y_i	1,4	1,7	1,8	1,7	1,9

Находим суммы первых трех столбцов

Таблица 2

i	x	y	xy	x^2	y^*	e_i	e^2
1	2	1,4	2,8				
2	3	1,7	5,1				
3	4	1,8	7,2				
4	5	1,7	8,5				
5	6	1,9	11,4				
Сумма	20	8,5	35				

Вычисляем значения и суммы трех следующих столбцов

Таблица 2

i	x	y	xy	x^2	y^*	e_i	e^2
1	2	1,4	2,8	4			
2	3	1,7	5,1	9			
3	4	1,8	7,2	16			
4	5	1,7	8,5	25			
5	6	1,9	11,4	36			
Сумма	20	8,5	35	90			

Находим коэффициенты регрессии по системе (1)

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{5 \cdot 35 - 20 \cdot 8,5}{5 \cdot 90 - 20^2} = \frac{175 - 170}{450 - 400} = \frac{5}{50} = 0,1.$$

$$b_0 = \frac{\sum_{i=1}^n y_i - b_1 \sum_{i=1}^n x_i}{n} = \frac{8,5 - 0,1 \cdot 20}{5} = \frac{6,5}{5} = 1,3.$$

i	x	y	xy	x^2	y^*	e_i	e^2
1	2	1,4	2,8	4			
2	3	1,7	5,1	9			
3	4	1,8	7,2	16			
4	5	1,7	8,5	25			
5	6	1,9	11,4	36			
Сумма	20	8,5	35	90			

Таким образом, уравнение регрессии имеет вид:

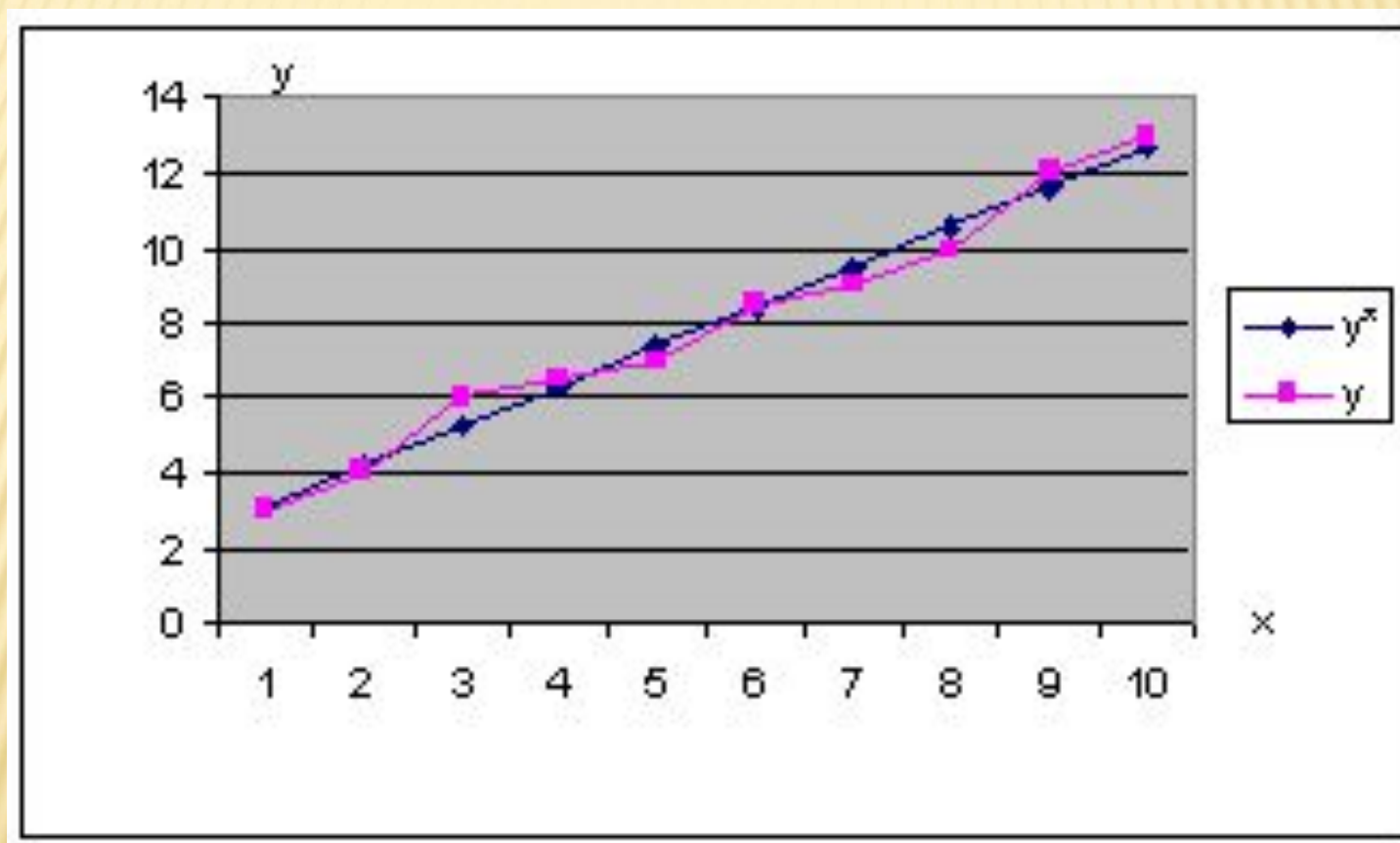
$$y^* = b_0 + b_1 \cdot x = 1,3 + 0,1 \cdot x. \quad (2)$$

Величина коэффициента b_1 показывает, что с увеличением расходов на рекламу на 1 усл. ед. объем продаж увеличится в среднем на 0,1 усл. ед.

Пример 1. Имеются данные о затратах на рекламу (х, усл. ед.) и объеме реализации продукции (у, усл. ед.).



ГРАФИК РЕГРЕССИИ



По уравнению регрессии (2) находим приближенные (теоретические) значения y^* :

$$y^*(x=2) = 1.3 + 0.1 \cdot 2 = 1.5$$

$$y^*(x=3) = 1.3 + 0.1 \cdot 3 = 1.6$$

$$y^*(x=4) = 1.3 + 0.1 \cdot 4 = 1.7$$

$$y^*(x=5) = 1.3 + 0.1 \cdot 5 = 1.8$$

$$y^*(x=6) = 1.3 + 0.1 \cdot 6 = 1.9$$

Таблица 2

i	x	y	xy	x^2	y^2	y^*	e_i	e^2
1	2	1,4	2,8	4	1,96	1,5		
2	3	1,7	5,1	9	2,89	1,6		
3	4	1,8	7,2	16	3,24	1,7		
4	5	1,7	8,5	25	2,89	1,8		
5	6	1,9	11,4	36	3,61	1,9		
Сумма	20	8,5	35	90	14,59			

$$y^* = b_0 + b_1 \cdot x = 1,3 + 0,1 \cdot x. \quad (2)$$

Вычисляем остатки регрессии:

$$e_1 = y_1 - y^*_1 = 1.4 - 1.5 = -0.1$$

$$e_2 = y_2 - y^*_2 = 1.7 - 1.6 = 0.1$$

$$e_3 = y_3 - y^*_3 = 1.8 - 1.7 = 0.1$$

$$e_4 = y_4 - y^*_4 = 1.7 - 1.8 = -0.1$$

$$e_5 = y_5 - y^*_5 = 1.9 - 1.9 = 0$$

Таблица 2

i	x	y	xy	x^2	y^2	y^*	e_i	e^2
1	2	1,4	2,8	4	1,96	1,5	-0,1	0,01
2	3	1,7	5,1	9	2,89	1,6	0,1	0,01
3	4	1,8	7,2	16	3,24	1,7	0,1	0,01
4	5	1,7	8,5	25	2,89	1,8	-0,1	0,01
5	6	1,9	11,4	36	3,61	1,9	0	0
Сумма	20	8,5	35	90	14,59			0,04

Рассчитаем объем продаж

при $x = 7$

$$y^* = b_0 + b_1 \cdot x = 1,3 + 0,1 \cdot x.$$

$$y^* = 1,3 + 0,1x = 1,3 + 0,1 \cdot 7 = \mathbf{2} \text{ (y.e.)}.$$



САМОСТОЯТЕЛЬНАЯ РАБОТА 1

Задание.

При установлении взаимосвязи между уровнем инфляции (x) и ценой на некоторый товар (y) получено уравнение регрессии

$$y^* = 3x - 4$$

Осуществите прогнозирование цены на товар при значении инфляции 10.

A. 26.

B. 30

C. 40

D. 37.



САМОСТОЯТЕЛЬНАЯ РАБОТА 2

Задание.

При обработке статистических данных получены коэффициенты линейной регрессии:

$$b_0^* = 3,5; \quad b_1^* = -0,8.$$

Осуществите прогноз исследуемого процесса для $X = 20$.



Пример 2

Для уравнения парной регрессии :

$$y^* = 200 - 78x$$

вычислить отклонение выборочного (фактического) значения от расчетного (теоретического) для точки с координатами (2;50).

Решение

$$x = 2;$$

$$y^* = b_0 + b_1 X = 200 - 78 \cdot 2 = 200 - 156 = 44$$

Итак, расчетное значение; $x = 2$; $y = 44$.

Фактическое: $x = 2$; $y = 50$.

Отклонение: $y^* - y = 50 - 44 = 6$.

САМОСТОЯТЕЛЬНАЯ РАБОТА 3

Задание.

Для уравнения парной регрессии :

$$y = -8 + 13x$$

вычислить отклонение выборочного
(фактического) значения от расчетного
для точки с координатами (3;36).

A. -4.

B. 4

C. 5

D. 36.



Регрессионный анализ в Excel осуществляется:

- в **Пакете анализа**,
- **графически**,
- а также функциями:
**ЛИНЕЙН, ТЕНДЕНЦИЯ,
ЛГРФПРИБЛ, РОСТ,
ПРЕДСКАЗ, ОТРЕЗОК,
НАКЛОН, СТОШУХ** и другими.



5.3. НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Нелинейные регрессии делятся на два класса:

- регрессии, **нелинейные** относительно включенных в анализ **объясняющих переменных**, но линейные по оцениваемым параметрам,
- регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по объясняющим переменным*:

- полиномы разных степеней $y = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3$;
- равносторонняя гиперболола $y = a + \frac{b}{x}$;

НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Регрессии, *нелинейные по оцениваемым параметрам*:

- степенная $y = a \cdot x^b$;
- показательная $y = a \cdot b^x$;
- экспоненциальная $y = e^{a+bx}$.

Построение уравнения регрессии сводится к оценке ее параметров.

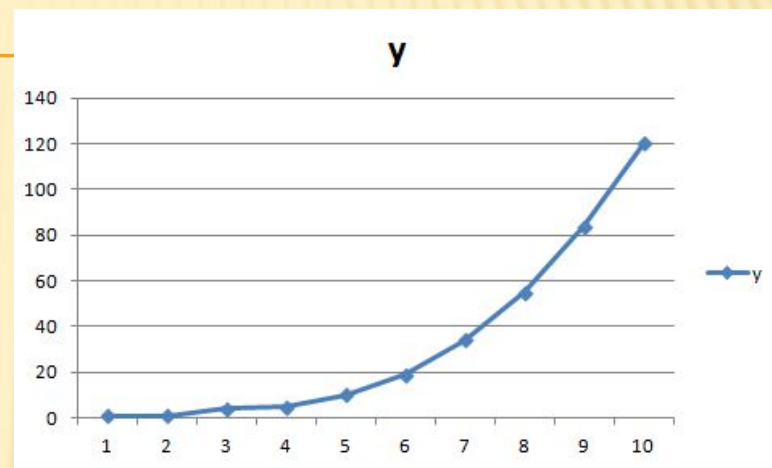
Для оценки параметров регрессий, линейных по параметрам, используют *метод наименьших квадратов (МНК)*

Пример 1. Осуществить аппроксимацию статистических данных, используя **Линию тренда** графического пакета Excel

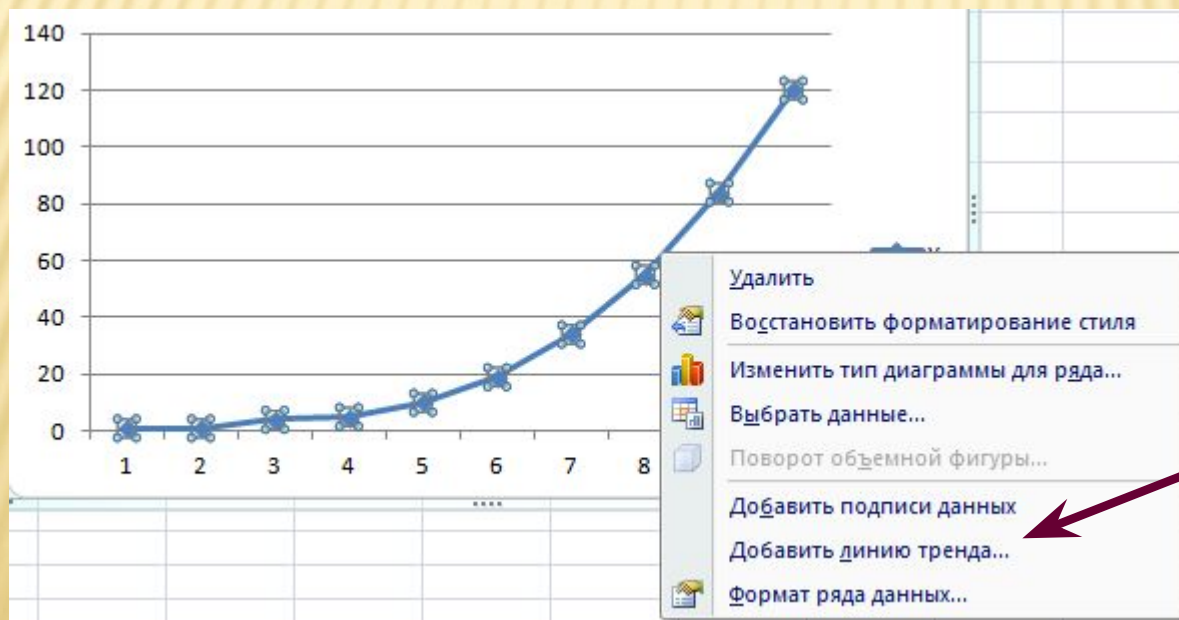
	А	В
1	Спрос на книгу	Неделя
2	1	1
3	1	2
4	4	3
5	5	4
6	10	5
7	19	6
8	34	7
9	55	8
10	84	9
11	120	10

Решение

1. Строим график процесса: выделить значения у - **Вставка – График – График с маркерами**



2. Щелкнуть ПРАВОЙ клавишей мыши точно по линии графика



3. Выбрать в контекстном меню

Добавить линию тренда

Окно линии тренда

Формат линии тренда


Параметры линии тренда


Цвет линии


Тип линии


Тень


Построение линии тренда (аппроксимация и сглаживание)


 ☐ Экспоненциальная

 ☒ Линейная

 ☐ Логарифмическая

 ☐ Полиномиальная Степень: 2

 ☐ Степенная

 ☐ Линейная фильтрация Точки: 2

Название аппроксимирующей (сглаженной) кривой

☒ автоматическое: Линейная (y)

☐ другое:

Прогноз

вперед на: 0,0 периодов

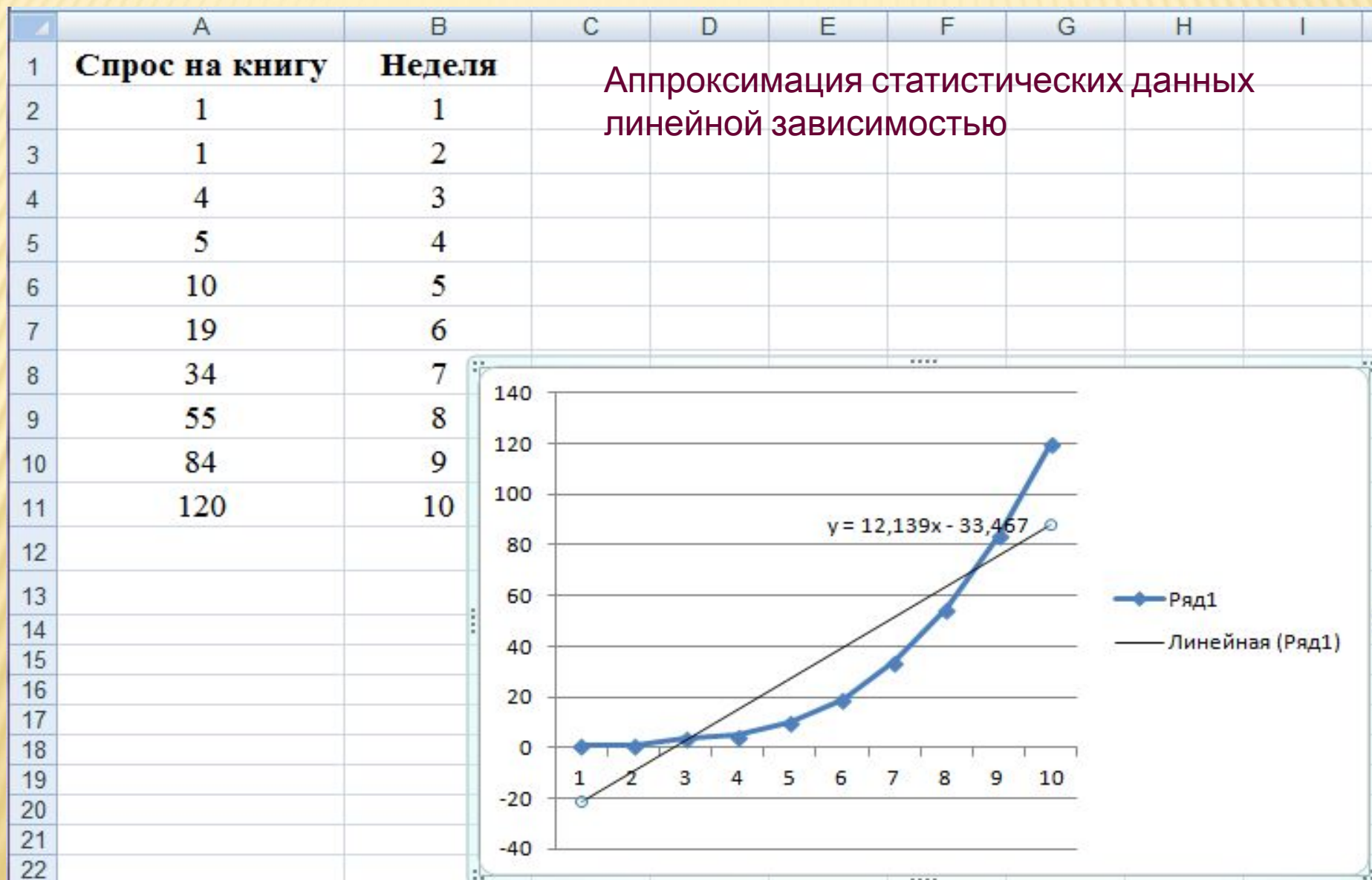
назад на: 0,0 периодов

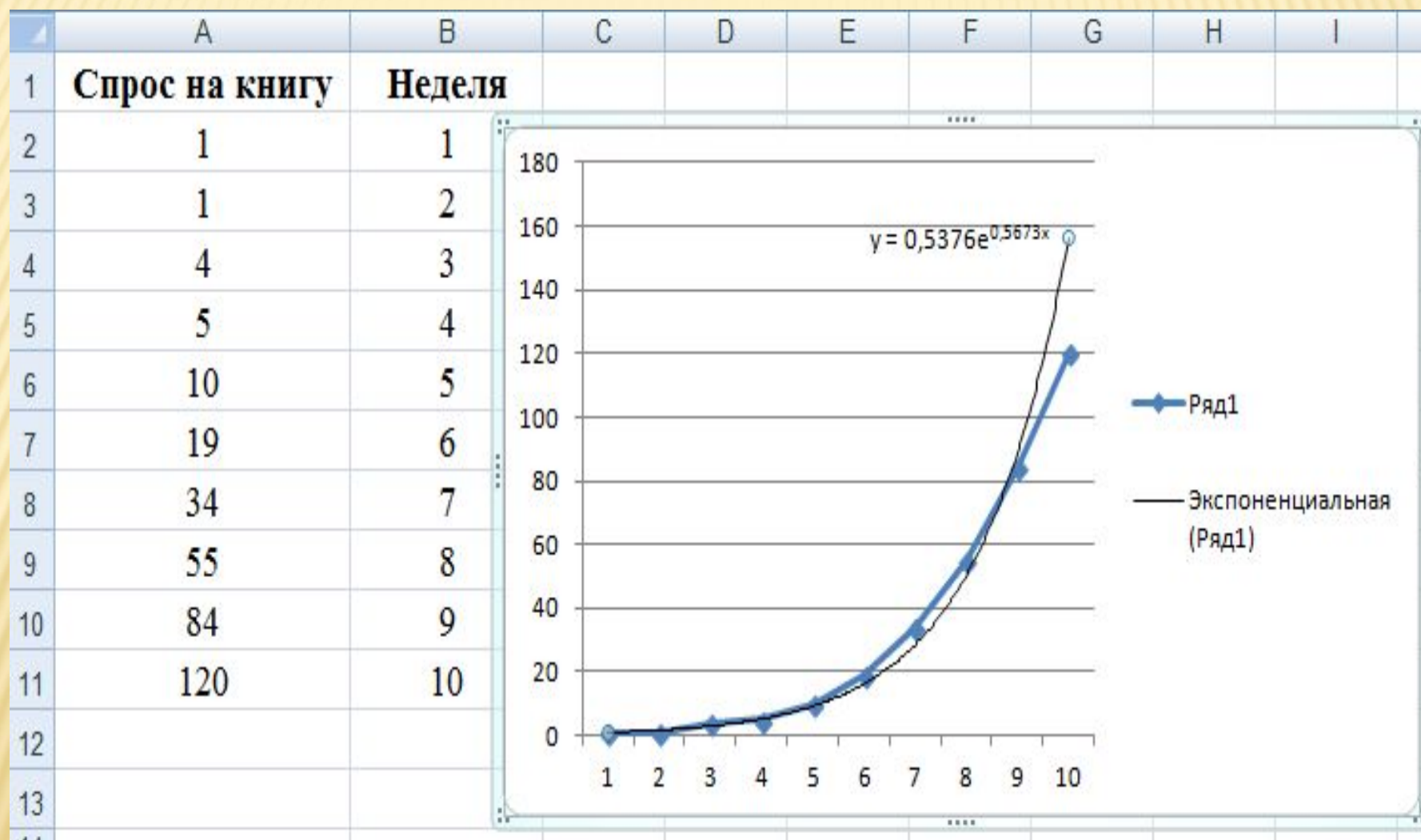
☐ пересечение кривой с осью Y в точке: 0,0

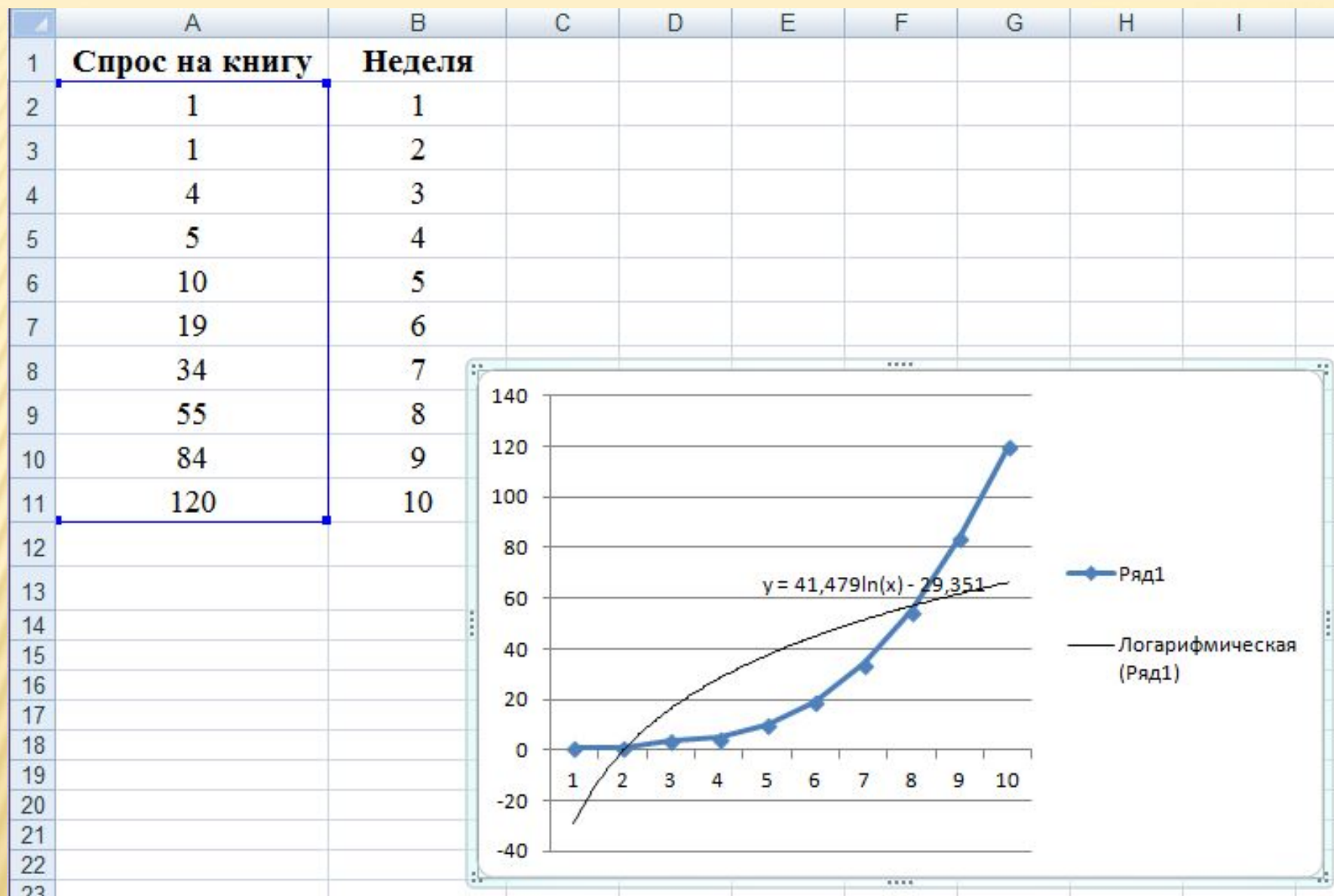
☒ показывать уравнение на диаграмме

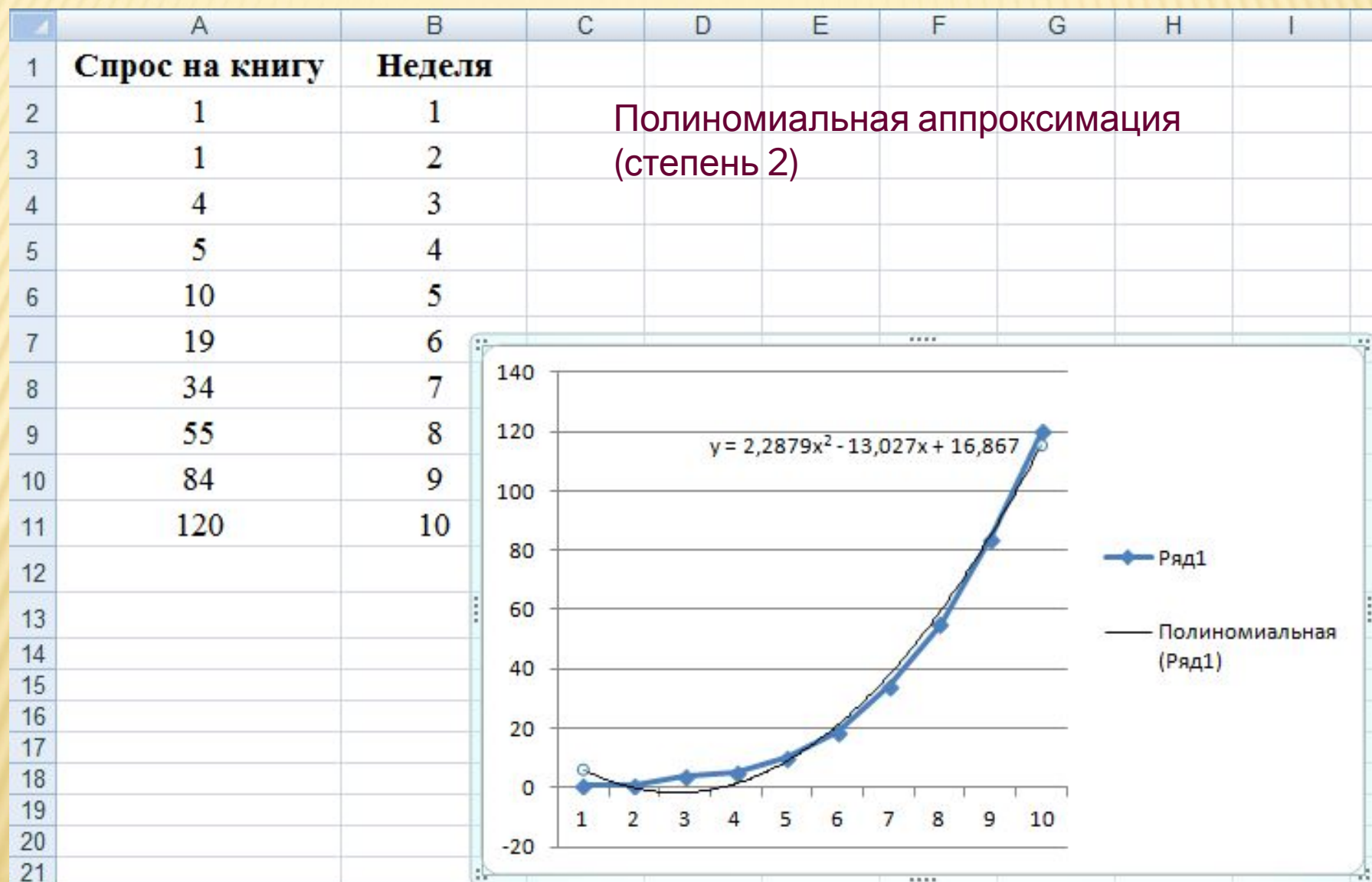
☐ поместить на диаграмму величину достоверности аппроксимации (R^2)

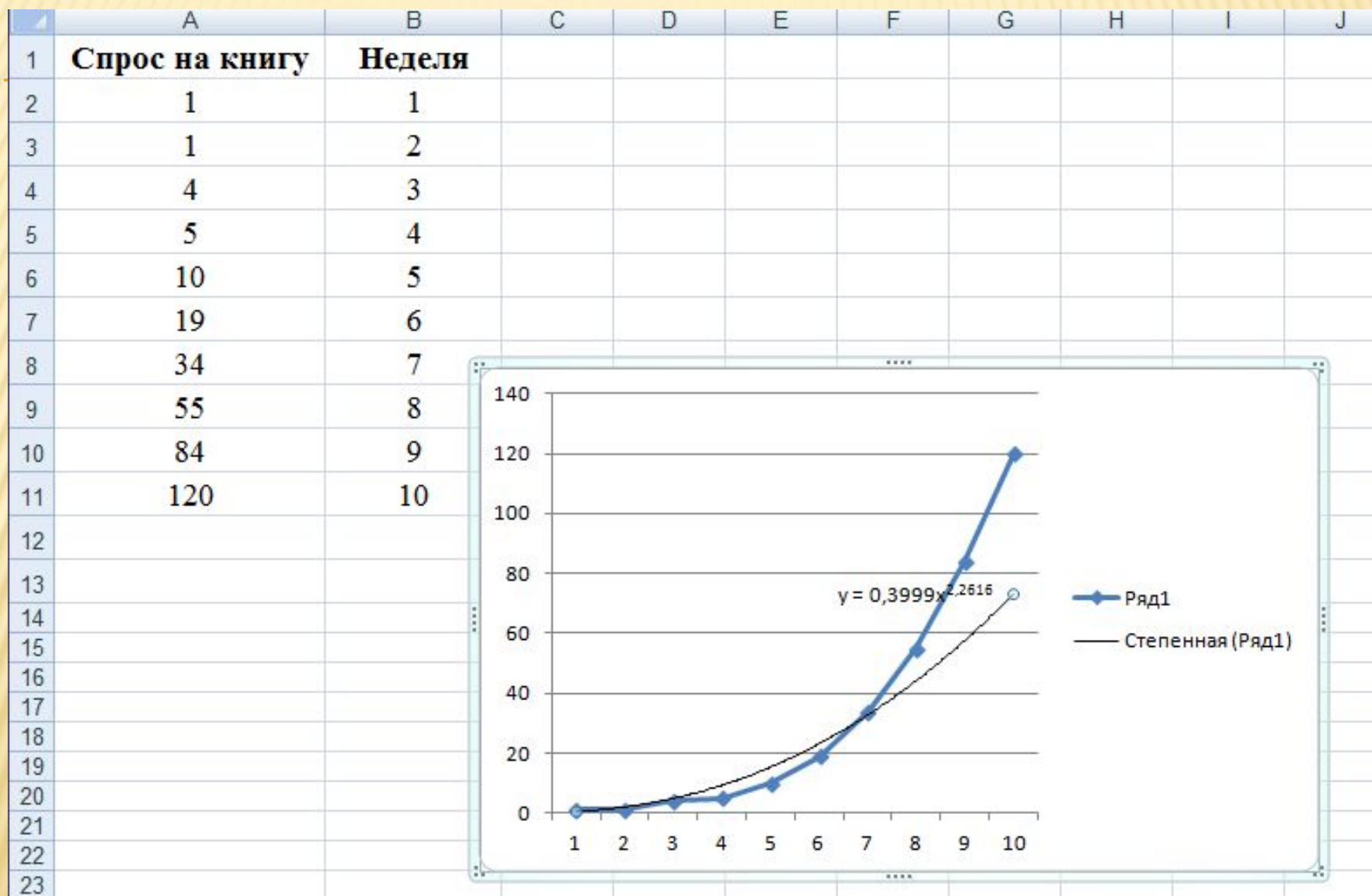
Заккрыть



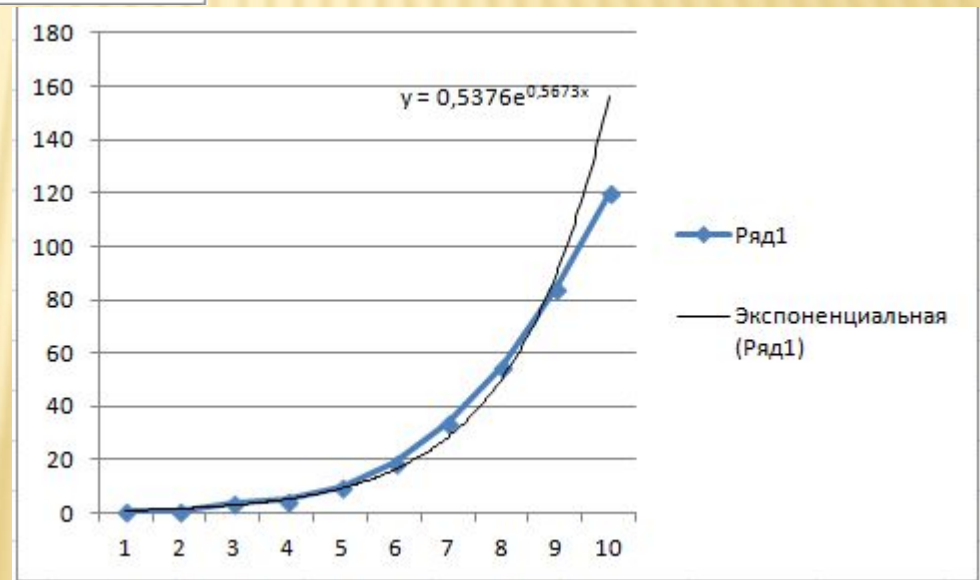
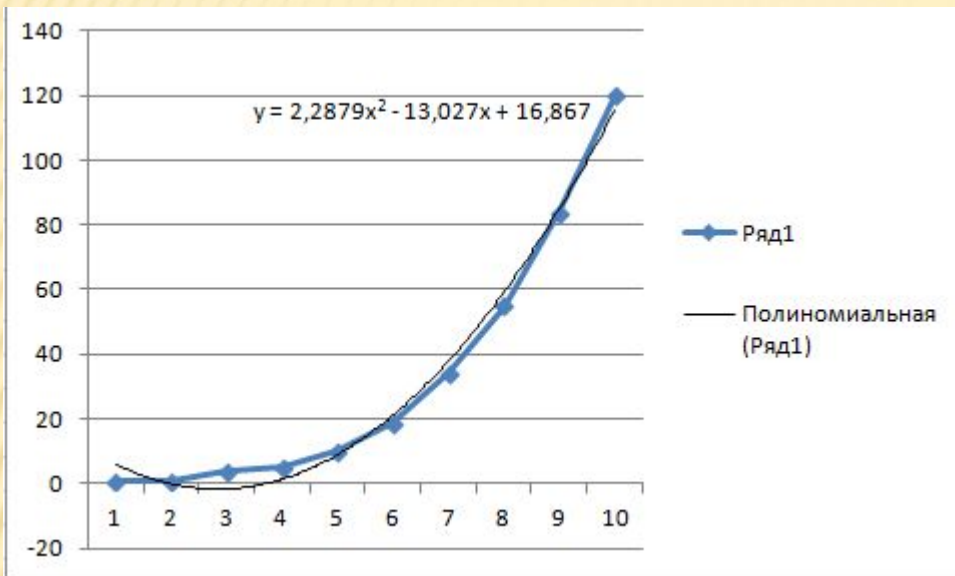








Наиболее удачная аппроксимация



5.4. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Множественная регрессия - уравнение связи переменной y с несколькими независимыми переменными:

$$y=f(x_1,x_2,...,x_p),$$

где y - зависимая переменная (результативный признак);
 $x_1, x_2, ..., x_p$ - независимые переменные (факторы).

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов* (МНК).

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Для построения уравнения множественной регрессии чаще используются следующие функции:

- линейная - $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p;$

- степенная - $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p};$

- экспонента -

$$y = e^{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p};$$

- гипербола -

$$y = \frac{1}{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p}$$

Пример 2

Задание 1. Построить регрессионную модель для предсказания изменения уровня заболевания органов дыхания (Y) в зависимости от содержания в воздухе двуокиси углерода CO_2 (x_1) и степени запыленности (x_2) по данным таблицы.

x_1 Степень запыленности	x_2 Содержание в воздухе CO_2	y Уровень заболевания
1160	1,2	1
1155	1,3	1
1158	1,4	1,1
1157	1,4	1,1
1160	1,5	1,1
1161	1,5	1,1
1157	1,4	1
1159	1,5	1
1256	1,6	1,2
1260	1,7	1,2
1040	1	0,6
1039	1	0,6
1039	1,1	0,7
1040	1,15	0,7
1140	1,4	0,8
1138	1,4	0,8

Задание 2. Определить уровень заболеваемости для величины степени запыленности **1280** и содержания в воздухе CO_2 , равного **1,7**.

1. Используя процедуру **Регрессия** Пакета анализа, находим коэффициенты регрессии

	А	В	С	Д	Е
1	Степень запылен- ности	Содержание в воздухе CO ₂	Уровень заболева- емости	ВЫВОДИТОГОВ	
2	1160	1,2	1		
3	1155	1,3	1	Регрессионная статистика	
4	1158	1,4	1,1	Множественный R	0,917837
5	1157	1,4	1,1	R-квадрат	0,842426
6	1160	1,5	1,1	Нормирован-ный R- квадрат	0,818183
7	1161	1,5	1,1	Стандартная ошибка	0,087905
8	1157	1,4	1	Наблюдения	16
9	1159	1,5	1		
10	1256	1,6	1,2	Дисперсионный анализ	
11	1260	1,7	1,2		df
12	1040	1	0,6	Регрессия	2
13	1039	1	0,6	Остаток	13
14	1039	1,1	0,7	Итого	15
15	1040	1,15	0,7		
16	1140	1,4	0,8		Коэффициенты
17	1138	1,4	0,8	Y-пересечение	-1,86357
18				Переменная X 1	0,002245
19				Переменная X 2	0,181461

a

b₁

b₂

$$y = a + b_1 x_1 + b_2 x_2$$

2. Составляем уравнение

	A	B	C	D	E
	Степень запылен- ности	Содержание в воздухе CO ₂	Уровень заболева- емости	ВЫВОД ИТОГОВ	
1					
2	1160	1,2	1		
3	1155	1,3	1	Регрессионная статистика	
4	1158	1,4	1,1	Множественный R	0,917837
5	1157	1,4	1,1	R-квадрат	0,842426
6	1160	1,5	1,1	Нормирован-ный R- квадрат	0,818183
7	1161	1,5	1,1	Стандартная ошибка	0,087905
8	1157	1,4	1	Наблюдения	16
9	1159	1,5	1		
10	1256	1,6	1,2	Дисперсионный анализ	
11	1260	1,7	1,2		df
12	1040	1	0,6	Регрессия	2
13	1039	1	0,6	Остаток	13
14	1039	1,1	0,7	Итого	15
15	1040	1,15	0,7		
16	1140	1,4	0,8		Коэффициенты
17	1138	1,4	0,8	Y-пересечение	-1,86357
18				Переменная X 1	0,002245
19				Переменная X 2	0,181461

a

b₂

b₁

$$y = a + b_1 x_1 + b_2 x_2 = -1,863 + 0,1812 * x_1 + 0,002 * x_2.$$

3. Подставим в уравнение п.2, данные из задания 2.

Получим

$$y = a + b_1 \cdot 1,7 + b_2 \cdot 1280 = -1,863 + 0,1812 \cdot 1,7 + 0,0022 \cdot 1280 = 1,26.$$

x_1 Степень запыленности	x_2 Содержание в воздухе CO ₂	y Уровень заболевания
1160	1,2	1
1155	1,3	1
1158	1,4	1,1
1157	1,4	1,1
1160	1,5	1,1
1161	1,5	1,1
1157	1,4	1
1159	1,5	1
1256	1,6	1,2
1260	1,7	1,2
1040	1	0,6
1039	1	0,6
1039	1,1	0,7
1040	1,15	0,7
1140	1,4	0,8
1138	1,4	0,8

$$y = a + b_1 x_1 + b_2 x_2 = -1,863 + 0,1812 \cdot x_1 + 0,002 \cdot x_2.$$

Задание 2. Определить уровень заболеваемости для величины степени запыленности **1280** и содержания в воздухе CO₂, равного **1,7**.

ПРИМЕР 3

При обработке статистических данных получены коэффициенты регрессии:

$$a^* = 2,9; \quad b_1^* = 4,6. ; \quad b_2^* = -1,6.$$

Написать уравнение линейной регрессии.

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p;$$

В нашем случае

$$y = 2,9 + 4,6x_1 - 1,6x_2 .$$



САМОСТОЯТЕЛЬНАЯ РАБОТА 4

Задание.

При обработке статистических данных получены коэффициенты линейной регрессии:

$$a^* = 2,9; \quad b_1^* = 4,6; \quad b_2^* = -1,6.$$

Осуществите прогноз исследуемого процесса для $X_1 = 1; X_2 = 2$.

1. 7,5
2. 4,3
3. 5,9
4. 10,7.



6. МОДЕЛИ ВРЕМЕННЫХ РЯДОВ



Одним из основных типов экономических моделей являются **модели временных рядов**.

Необходимость их **исследования** возникает в случае, когда:

- объектом исследования является набор данных;
- данные регистрируются на одном и том же объекте в течение ряда моментов времени.

Такие данные называются временно'й выборкой или **временными рядами** данных



Модели временных рядов

активно применяются в исследованиях **динамики** множества **реальных процессов**:

- пассажиропотоков;
- складских запасов;
- спроса на различные виды продукции;
- миграционных процессов в человеческом и биологических сообществах;
- финансовых рынков;
- в анализе финансовых показателей;
- прогнозировании курсов акций;
- прогнозировании курсов валют и т. п.



Временные ряды служат основой при моделировании природных явлений:

- ☐ динамики числа солнечных пятен;
- ☐ природных катастроф и многих других процессов.



Характерная особенность социально-экономических процессов, представленных **временными рядами** - ярко выраженная **периодичность**:

```
graph TD; A[Характерная особенность социально-экономических процессов, представленных временными рядами - ярко выраженная периодичность:] --> B[Интенсивность транспортных поездок значительно выше в летние месяцы]; A --> C[Интенсивность посещений зрелищных мероприятий резко возрастает в зимний период,]; A --> D[В РФ спрос на валюту (и, соответственно, ее цена) обычно растет к концу года, когда подходят сроки платежей];
```

Интенсивность транспортных поездок значительно выше в летние месяцы

Интенсивность посещений зрелищных мероприятий резко возрастает в зимний период,

В РФ спрос на валюту (и, соответственно, ее цена) обычно растет к концу года, когда подходят сроки платежей

Экономические показатели деятельности предприятия за определенный период также являются **временными рядами**.

В этом случае исключительно важной является задача прогнозирования - возможность предсказать, как будет развиваться экономический процесс в дальнейшем.

Основным средством для прогнозирования поведения временных рядов является **метод моделирования**.



6.1. Основные определения

Последовательность наблюдений одного показателя (признака), упорядоченных в зависимости от последовательно возрастающих или убывающих значений другого показателя (признака), называют *динамическим рядом*, или рядом динамики.

Если в качестве признака, в зависимости от которого происходит упорядочение, **выступает время**, то такой динамический ряд называется *временным рядом*.



Время, прошедшее от начального момента наблюдения до конечного, называют *длиной* временного ряда.

Значение показателя в каждый конкретный момент времени - *уровень* временного ряда.

Обозначим уровни ряда y_t ($t=1,2,\dots,n$),
где n — длина ряда.

Временной ряд обычно обозначают $Y(t)$, или

$$y_1, y_2, \dots, y_n$$



В модели временного ряда принято выделять две основные составляющие:

- детерминированную (систематическую);
- случайную.



Детерминированная составляющая может содержать следующие структурные компоненты:

Тренд, или **тенденция** — плавно меняющаяся с течением времени составляющая, описывающая влияние долговременных факторов;

Сезонная компонента связана с наличием факторов, действующих с заранее известной периодичностью (под “сезоном” можно понимать и день, и неделю, и месяц, и квартал);

Циклическая компонента отражает повторяемость экономических процессов в течение длительных периодов.

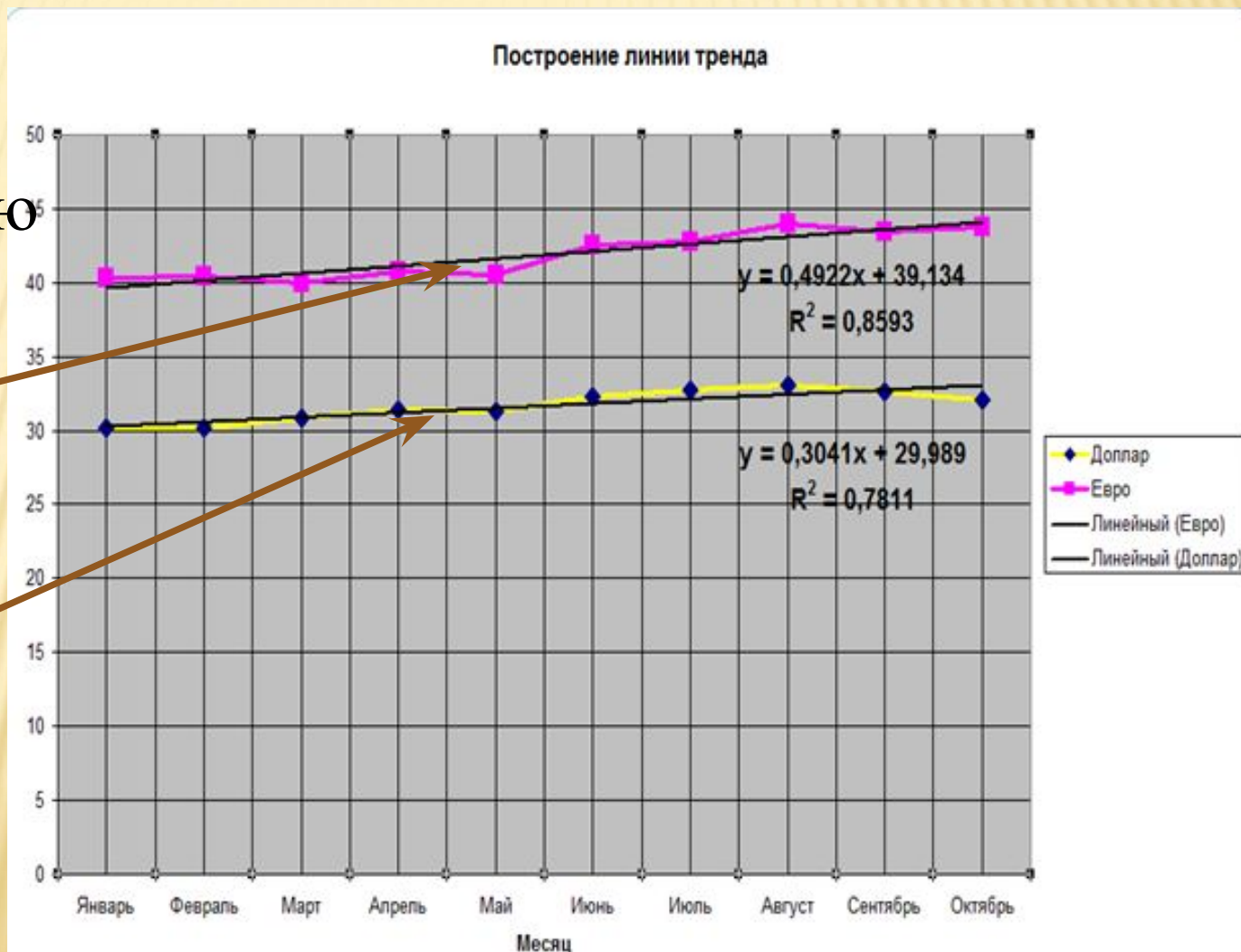
Если во временном ряду проявляется длительная **тенденция** изменения экономического показателя, то говорят, что имеет место **тренд** .



Диаграмма
временного ряда
"Численность
населения
Австралии, млн.
чел." с
преобладанием
возрастающей
тенденции

Обычно эта тенденция описывается с помощью той или иной неслучайной функции $T(t)$ (аргументом которой является t - время), как правило, монотонной.

Эту функцию называют функцией тренда или просто — трендом



6.2. Моделирование тенденции временного ряда

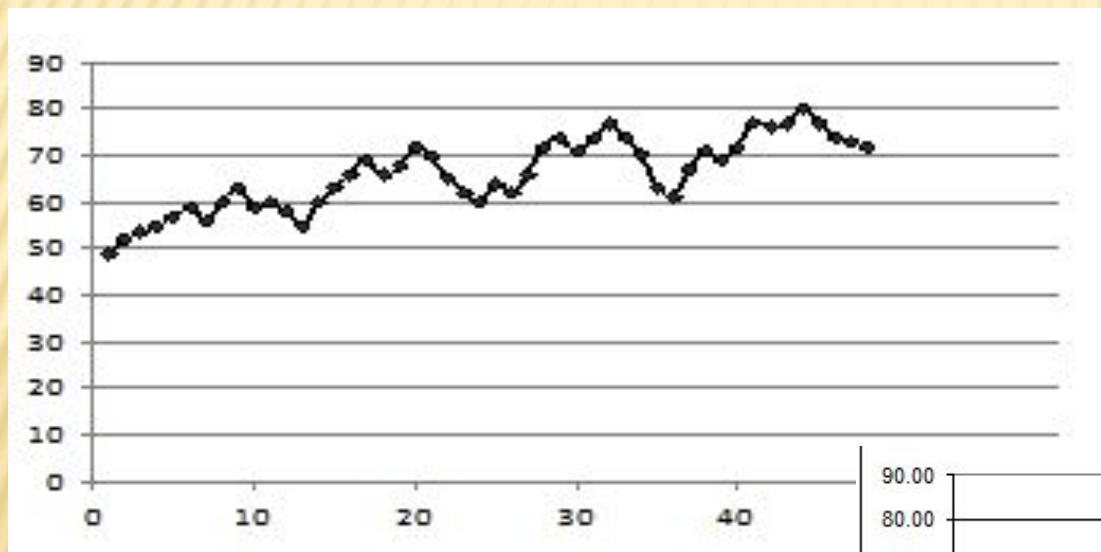
Для выявления трендовой составляющей временного ряда и построения ее модели необходимо выполнить предварительную обработку ряда путем его **сглаживания** и **выравнивания**.

Наиболее распространенные методы сглаживания:

- ☐ метод укрупнения интервалов;
- ☐ метод скользящей средней.

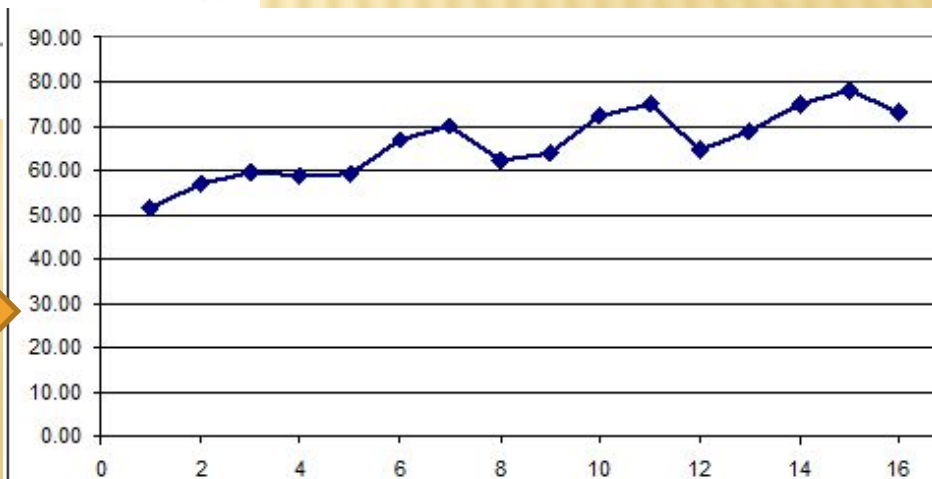


6.2.1. Метод укрупнения интервалов заключается в преобразовании временного ряда в ряд с укрупненными временными интервалами (например, ряд с месячными данными преобразуется в ряд с квартальными данными, квартальные – в годовые и т.



← Помесечный
объем продаж

Поквартальный
объем продаж →



Пример 1. Имеется статистика ежемесячной продажи туристической фирмой путевок

53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47, 49, 47,
51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48, 46, 50, 47, 47,
44, 48, 47, 49, 50, 45, 46, 46, 47, 45, 48, 49, 50, 50, 44,
47, 45, 47, 51, 47, 52.

Провести сглаживание
временного ряда
**методом укрупнения
интервалов
с интервалом
сглаживания 3**



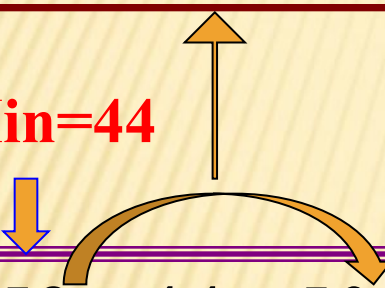
Решение

1. Преобразуем помесечные данные в поквартальные

$$(53+44+50)/3=49$$

Min=44

Max=52



53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

Решение

Преобразуем помесечные данные в поквартальные

$$(53+44+50)/3=49$$

$$(49+45+47)/3=47$$

Min=44

Max=52

53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

Решение

Преобразуем помесечные данные в поквартальные

$$(53+44+50)/3=49$$

$$(50+47+47)/3=48$$

$$(49+45+47)/3=47$$

Min=44

Max=52

53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

Самостоятельная работа 1

Преобразуйте помесячные данные в поквартальные ля
следующих трех месяцев

$$(53+44+50)/3=49$$

$$(50+47+47)/3=48$$

$$(49+45+47)/3=47$$

53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

A. 47

B. 48

C. 49

D. 50

Решение

Преобразуем помесечные данные в поквартальные

$$(53+44+50)/3=49$$

$$(50+47+47)/3=48$$

$$(49+45+47)/3=47$$

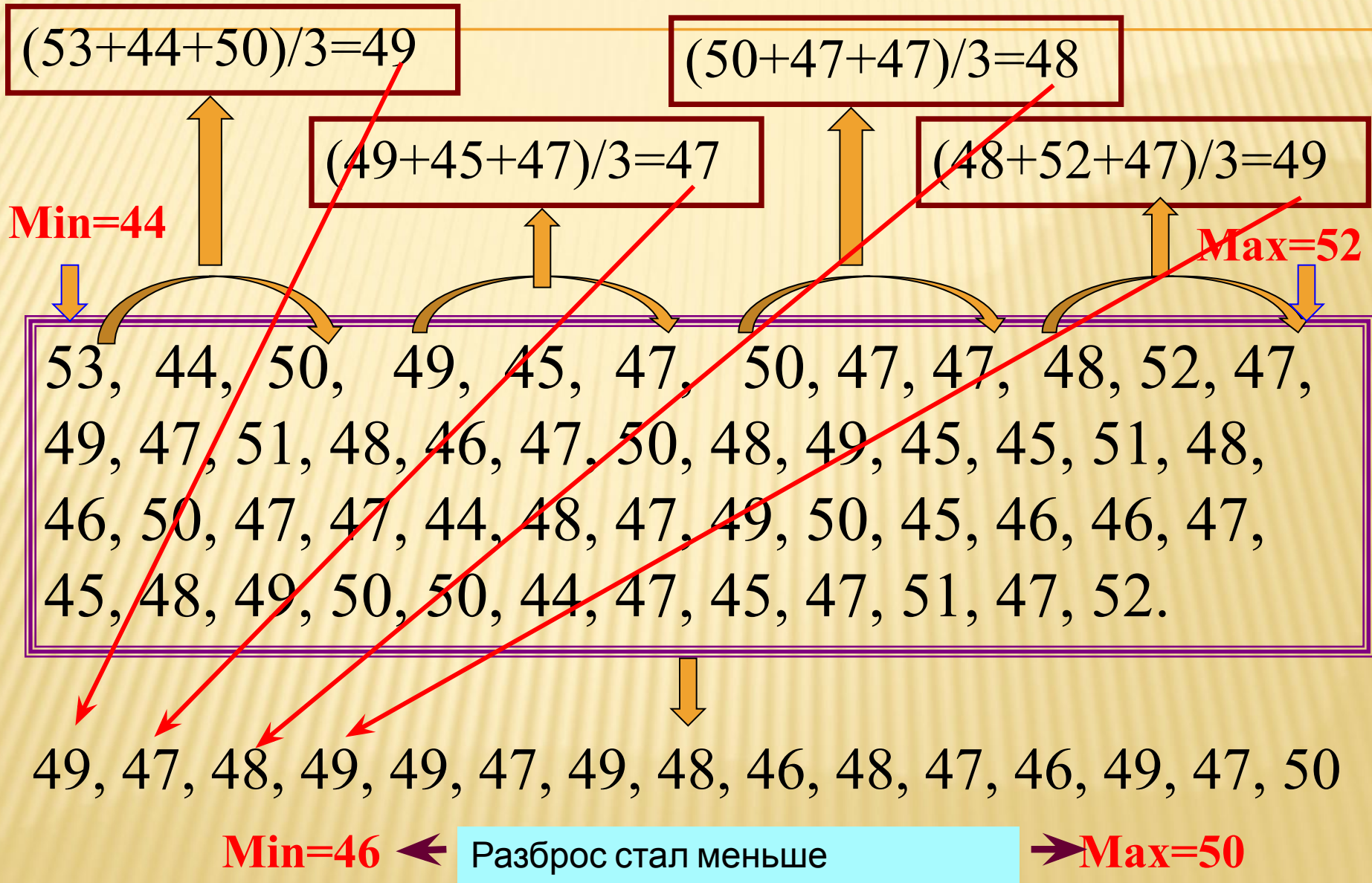
$$(48+52+47)/3=49$$

Min=44

Max=52

53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

2. Составим новую выборку из укрупненных данных



6.2.2. Сглаживание методом скользящей средней

Для исследования временных рядов нужно «погасить» случайные колебания, которые искажают общую картину. Для этих целей используется метод **скользящей средней**.

Его суть состоит в том, что фактические уровни временного ряда заменяются их средними значениями, а основная тенденция развития выражается в виде некоторой плавной кривой.



Сглаживание можно проводить по любому числу уровней. Например, при сглаживании **по трем соседним уровням** теоретический уровень рассчитывается по формуле

$$\bar{y}_i = \frac{y_{i-1} + y_i + y_{i+1}}{3}, \quad i = 2, 3, \dots, n - 1.$$

При этом средние значения уровня вычисляются за определенные периоды времени путем последовательного передвижения начала периода на единицу времени, принятую при построении данного временного ряда.

Пример 2. Имеется статистика ежемесячной продажи туристической фирмой путевок

53, 44, 50, 49, 45, 47, 52, 47, 47, 48, 52, 47, 49, 47,
51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48, 46, 50, 47, 47,
44, 48, 47, 49, 50, 45, 46, 46, 47, 45, 48, 49, 50, 50, 44,
47, 45, 47, 51, 47, 52.

Провести сглаживание
временного ряда

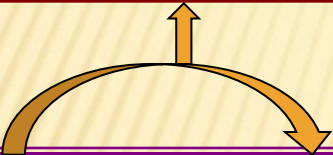
**методом скользящей
средней с интервалом
сглаживания 3**



Решение

1. Находим среднее арифметическое трех первых слагаемых

$$(53+44+50)/3=49$$



53, 44, 50, 49, 45, 47, 52, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

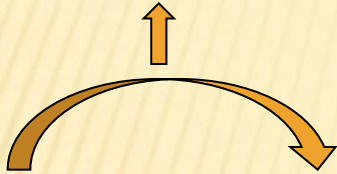
и записываем первый результат:



49,

2. Сдвигаемся **вправо на один период** и находим среднее арифметическое трех следующих слагаемых

$$(44+50+49)/3=47,7$$



53, 44, 50, 49, 45, 47, 52, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

и записываем новый результат:

49, 47.7,

3. Сдвигаемся **вправо** еще **на один период** и находим среднее арифметическое трех следующих слагаемых

$$(50+49+45)/3=48$$



53, 44, 50, 49, 45, 47, 52, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

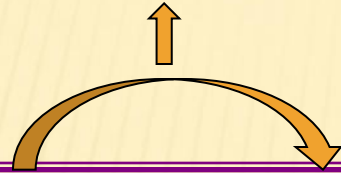
и записываем третий результат:



49, 47.7, 48,

4. Сдвигаемся **вправо** еще **на один период** и находим среднее арифметическое трех следующих слагаемых

$$(49+45+47)/3=47$$



53, 44, 50, 49, 45, 47, 52, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

и записываем четвертый результат:



49, 47.7, 48, 47,

Самостоятельная работа 2

Определите пятый результат сглаживания

53, 44, 50, 49, 45, 47, 52, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

A. 48,5

B. 47

C. 47,5

D. 48

5. Сдвигаемся **вправо** еще **на один период** и находим среднее арифметическое трех следующих слагаемых

$$(45+47+52)/3=48$$



53, 44, 50, 49, 45, 47, 50, 47, 47, 48, 52, 47,
49, 47, 51, 48, 46, 47, 50, 48, 49, 45, 45, 51, 48,
46, 50, 47, 47, 44, 48, 47, 49, 50, 45, 46, 46, 47,
45, 48, 49, 50, 50, 44, 47, 45, 47, 51, 47, 52.

и записываем пятый результат:



49, 47.7, 48, 47, 48,

Самостоятельная работа 3

Задание. Для значений временного ряда
 $y_1=6$, $y_2=14$, $y_3=7$, $y_4=12$, $y_5=9$
определите значение уровня ряда, сглаженного при
помощи простой скользящей средней с **интервалом**
сглаживания равным 3, на момент $t=3$.

Варианты

A. 10.

ответов:

B. 9.

C. 11.

D. 12.

Самостоятельная работа 4

Задание. Для значений временного ряда
 $y_1=6$, $y_2=14$, $y_3=7$, $y_4=12$, $y_5=9$
определите значение уровня ряда, сглаженного при
помощи простой скользящей средней с **интервалом**
сглаживания равным 3, на момент $t=4$.

Варианты

A. 10.

ответов:

B. 9.

C. 11.

D. 12.

Пример 3. Исключить случайную составляющую при изучении динамика производства подсолнечника на Украине.

Таблица 1

		A	B	C	D	E	F	G	H	I	J	K	L
1	1	Динамика урожайности подсолнечника в Украине											
2	2	Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
3	3	Урожайность, ц/га	15,8	14,6	13	12,7	9,1	12,2	10,5	11,5	9,3	10	12,2


В Excel метод скользящей средней реализован в **Пакете анализа**. Чтобы вычислить скользящие средние для временного ряда из табл.1 выполним следующие действия:

1. Команды **Сервис – Анализ данных – Скользящее среднее –Ок.**

2. В окне Скользящее среднее выполним установки

Скользящее среднее


Входные данные

Входной интервал: 

☒ Метки в первой строке

Интервал:

Параметры вывода

Выходной интервал: 

Новый рабочий лист:

Новая рабочая книга

☒ Вывод графика ☒ Стандартные погрешности

OK
Отмена
Справка

	A	B	C	D	E	F	G	H	I	J	K	L
1	Динамика урожайности подсолнечника в Украине											
2	Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
3	Урожайность, ц/га	15,8	14,6	13	12,7	9,1	12,2	10,5	11,5	9,3	10	12,2

	A	B	C	D	E	F	G	H	I	J	K	L
1	Динамика урожайности подсолнечника в Украине											
2	Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
3	Урожайность, ц/га	15,8	14,6	13	12,7	9,1	12,2	10,5	11,5	9,3	10	12,2
4		#Н/Д	#Н/Д	13,4	11,6	11,33	10,6	11,4	10,43	10,27	10,5	10,4



Применение в анализе временных рядов методов укрупнения интервалов и скользящей средней позволяет лишь выявить тренд.

Способ, позволяющий построить аналитическую функцию (тренда), характеризующую зависимость уровней ряда от времени, называется *аналитическим выравниванием временного ряда*.



Для получения первого представления о возможном характере тренда используется **графическое представление временного ряда**

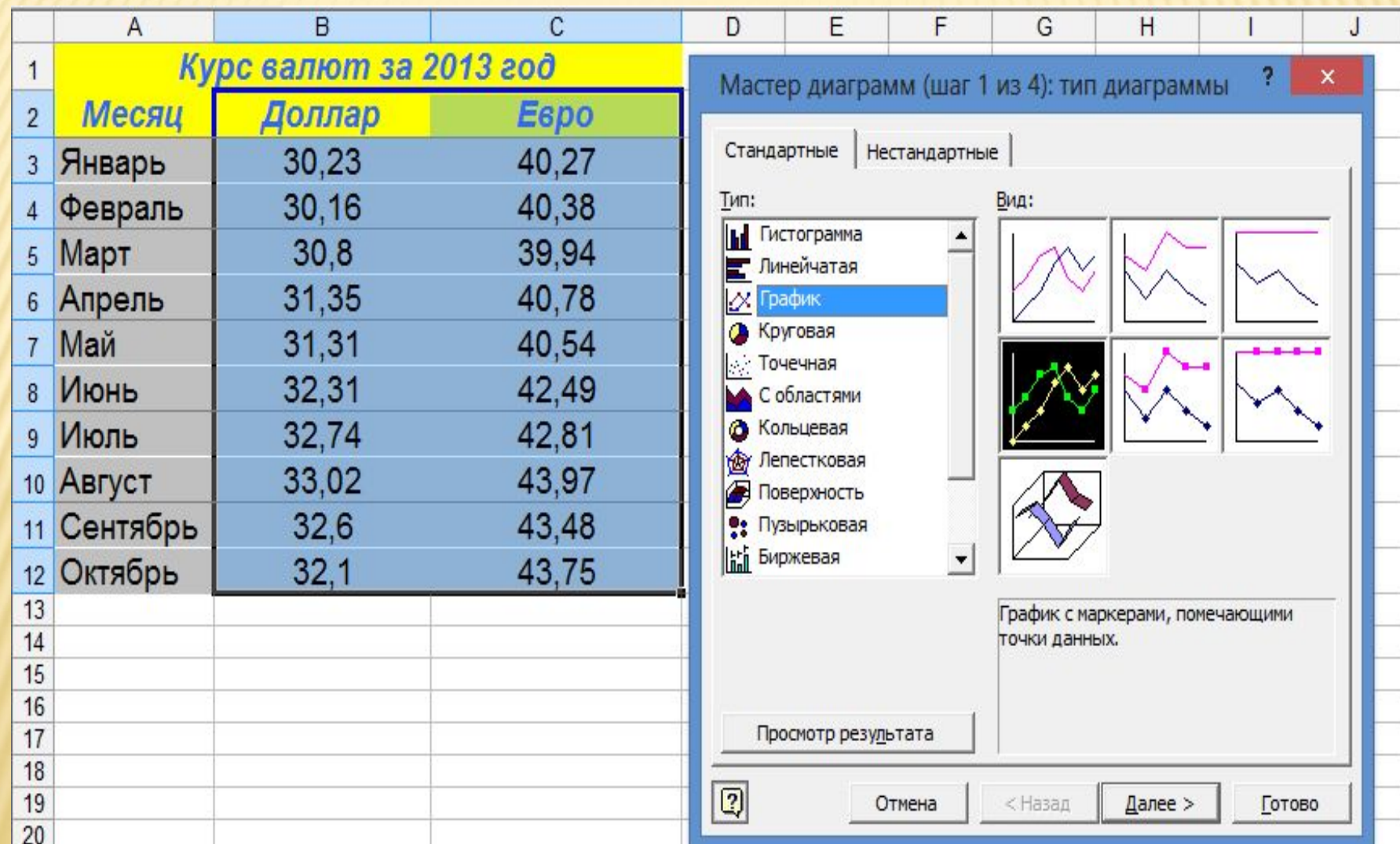
Пример. Построить линию тренда для временных рядов валют

Дата	Курс доллара	Курс евро
Октябрь, 2013	32.10	43.75
Сентябрь, 2013	32.60	43.48
Август, 2013	33.02	43.97
Июль, 2013	32.74	42.81
Июнь, 2013	32.31	42.59
Май, 2013	31.31	40.54
Апрель, 2013	31.35	40.78
Март, 2013	30.80	39.94
Февраль, 2013	30.16	40.38
Январь, 2013	30.23	40.27

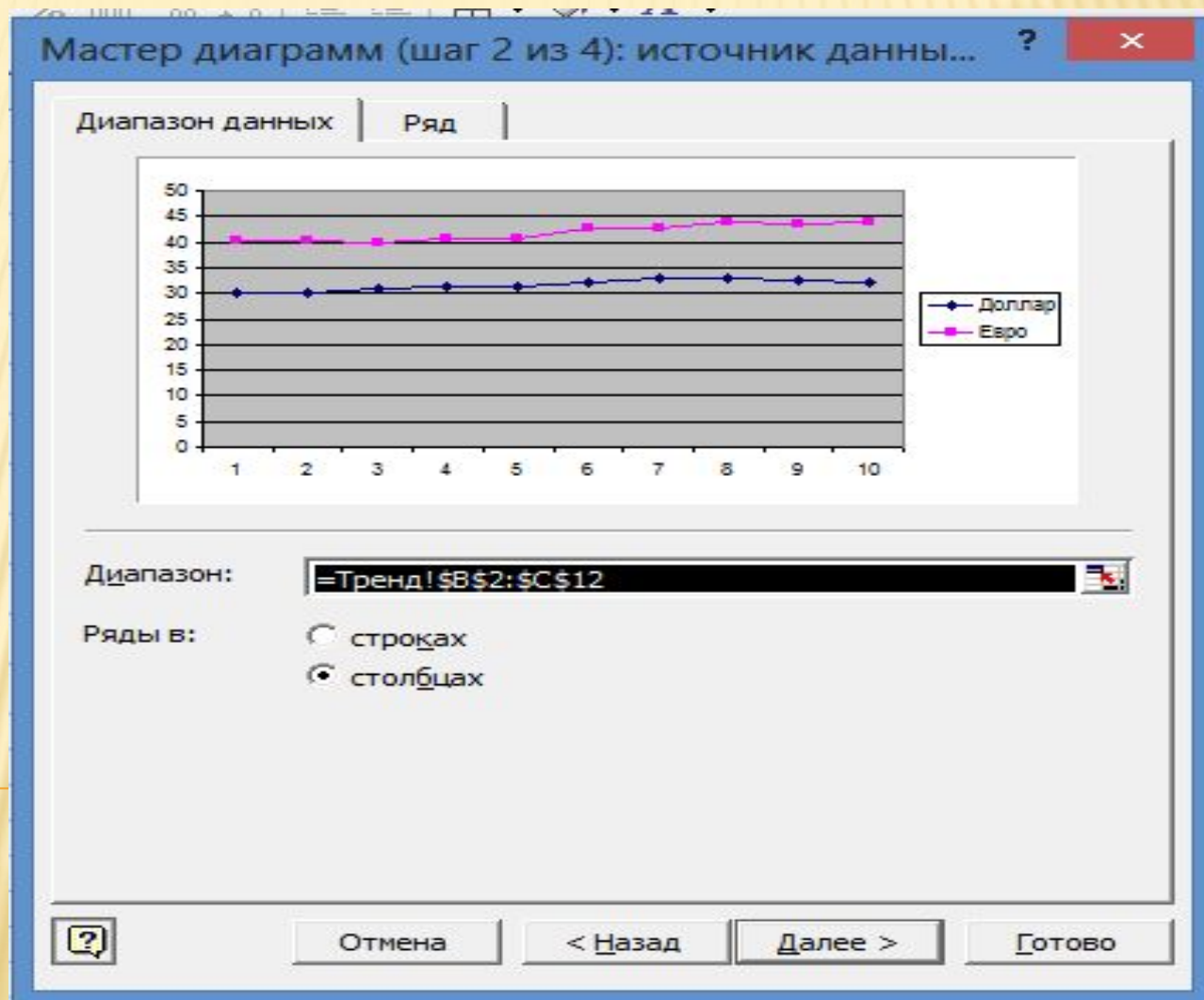
1. ПОСТРОЕНИЕ ЭЛЕКТРОННОЙ ТАБЛИЦЫ ПО ДАННЫМ ЦБ КУРСА ВАЛЮТ ЗА 2013 ГОД

	A	B	C
1	Курс валют за 2013 год		
2	Месяц	Доллар	Евро
3	Январь	30,23	40,27
4	Февраль	30,16	40,38
5	Март	30,8	39,94
6	Апрель	31,35	40,78
7	Май	31,31	40,54
8	Июнь	32,31	42,49
9	Июль	32,74	42,81
10	Август	33,02	43,97
11	Сентябрь	32,6	43,48
12	Октябрь	32,1	43,75

2. ПОСТРОЕНИЕ ДИАГРАММЫ. ШАГ 1

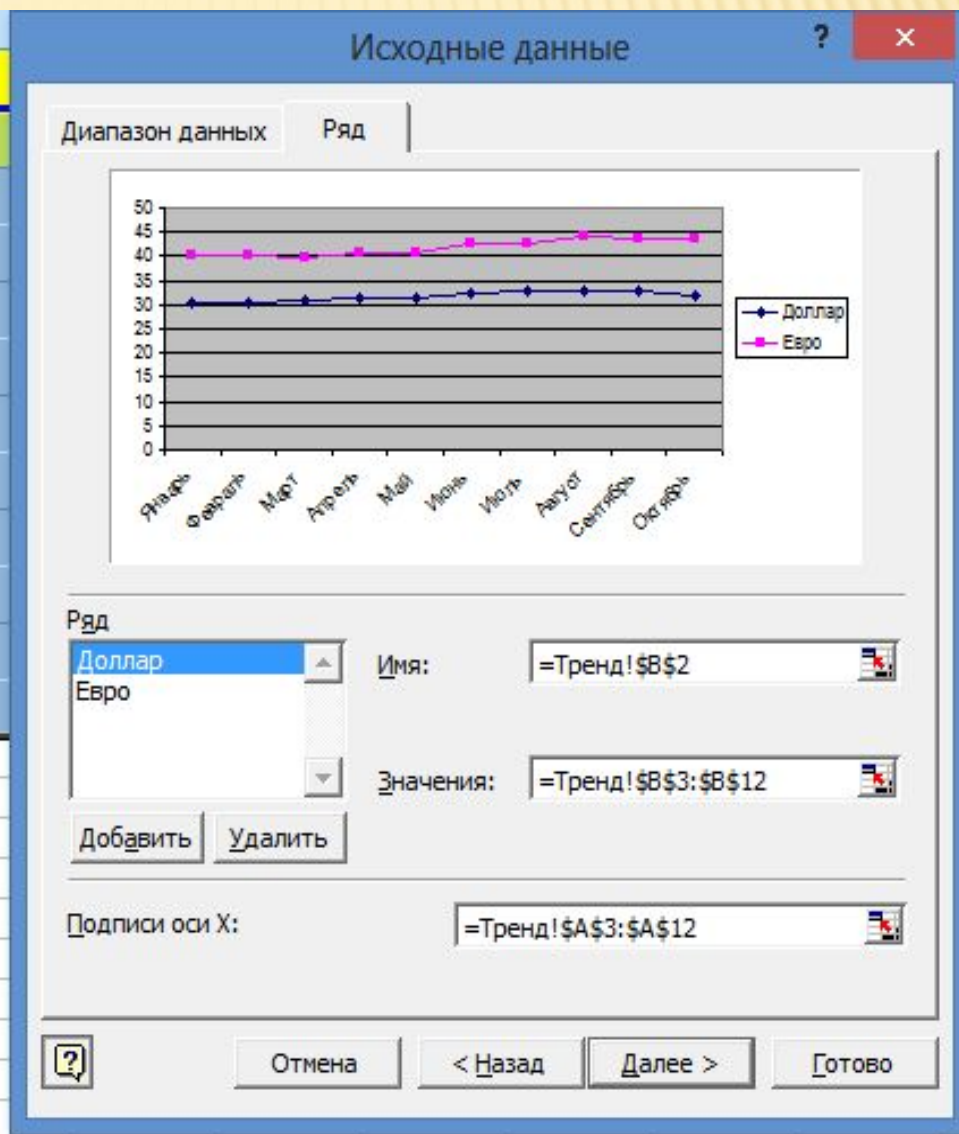


ВТОРОЙ ШАГ – ОПРЕДЕЛЕНИЕ ИСТОЧНИКА ДАННЫХ



ВВОД ПОДПИСИ ПО ОСИ X

	A	B	C
1	Курс валют за 2013 год		
2	Месяц	Доллар	Евро
3	Январь	30,23	40,27
4	Февраль	30,16	40,38
5	Март	30,8	39,94
6	Апрель	31,35	40,78
7	Май	31,31	40,54
8	Июнь	32,31	42,49
9	Июль	32,74	42,81
10	Август	33,02	43,97
11	Сентябрь	32,6	43,48
12	Октябрь	32,1	43,75
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			



ВВОД НАЗВАНИЯ ДИАГРАММЫ – ШАГ 3

Мастер диаграмм (шаг 3 из 4): параметры диаграммы

Подписи данных | Таблица данных

Заголовки | Оси | Линии сетки | Легенда

Название диаграммы:

Ось X (категорий):

Ось Y (значений):

Вторая ось X (категорий):

Вторая ось Y (значений):

Построение линии тренда

Месяц	Доллар	Евро
Январь	30	42
Февраль	30	42
Март	31	41
Апрель	32	42
Май	32	41
Июнь	33	43
Июль	33	43
Август	33	44
Сентябрь	33	44
Октябрь	32	45

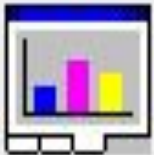
Отображение: Доллар, Евро

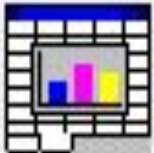
Отмена < Назад Далее > Готово


РАЗМЕЩЕНИЕ ДИАГРАММЫ – ШАГ 4

Мастер диаграмм (шаг 4 из 4): размещение диагра... ? X

Поместить диаграмму на листе: _____

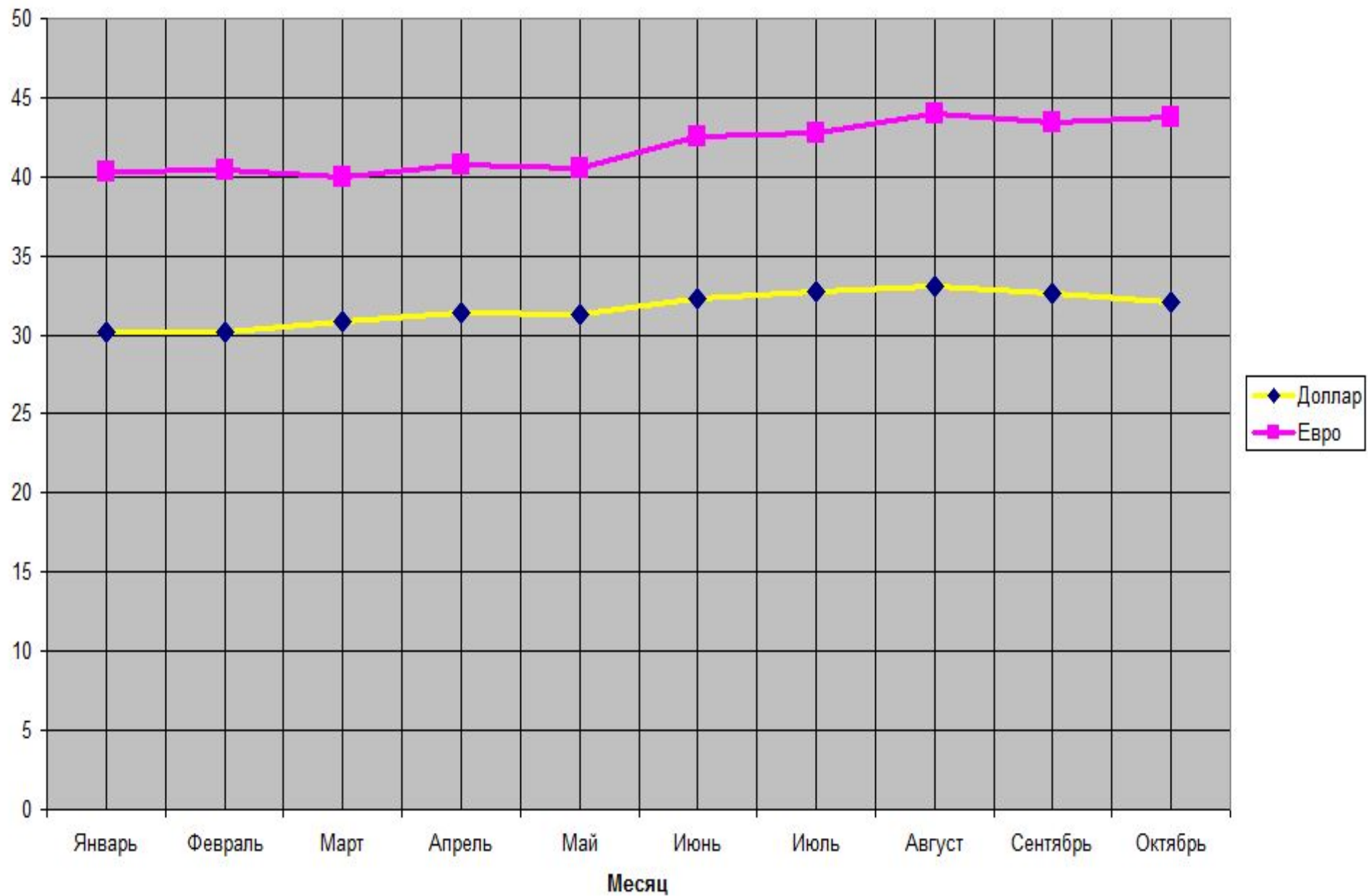
 ☒ отдельном:

 ☐ имеющемся:

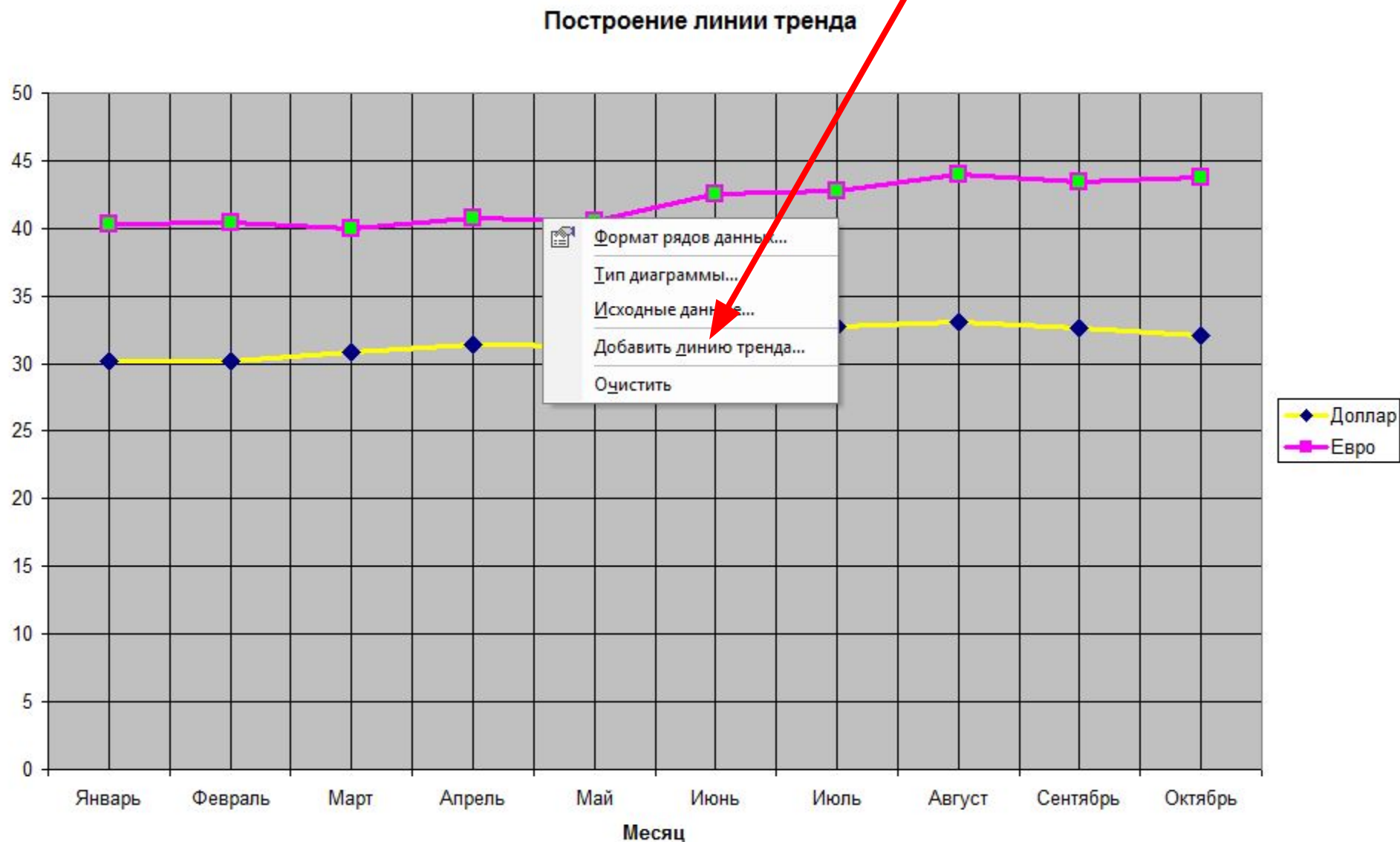


Отмена < Назад Далее > Готово

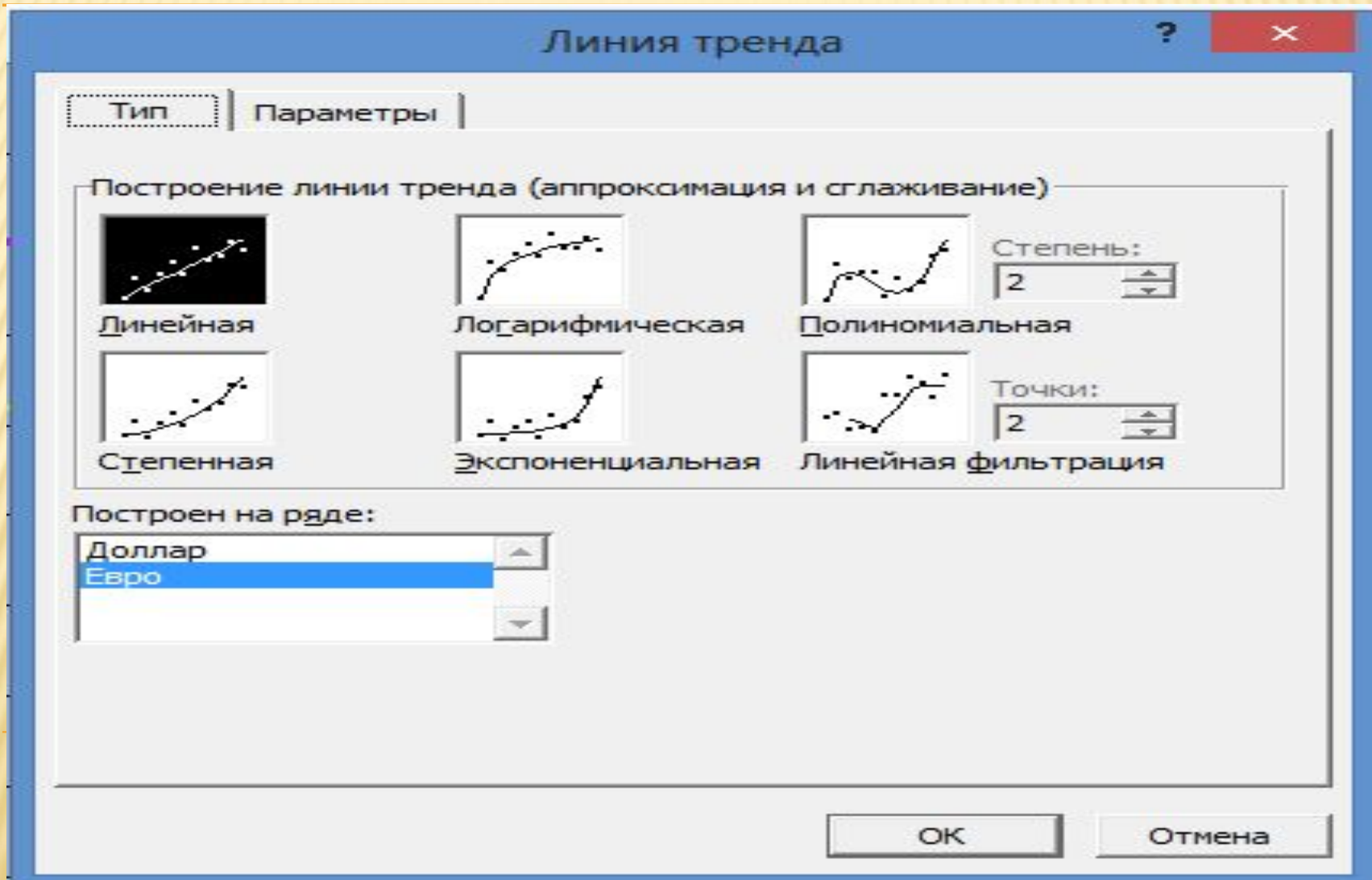
Динамика курса валют



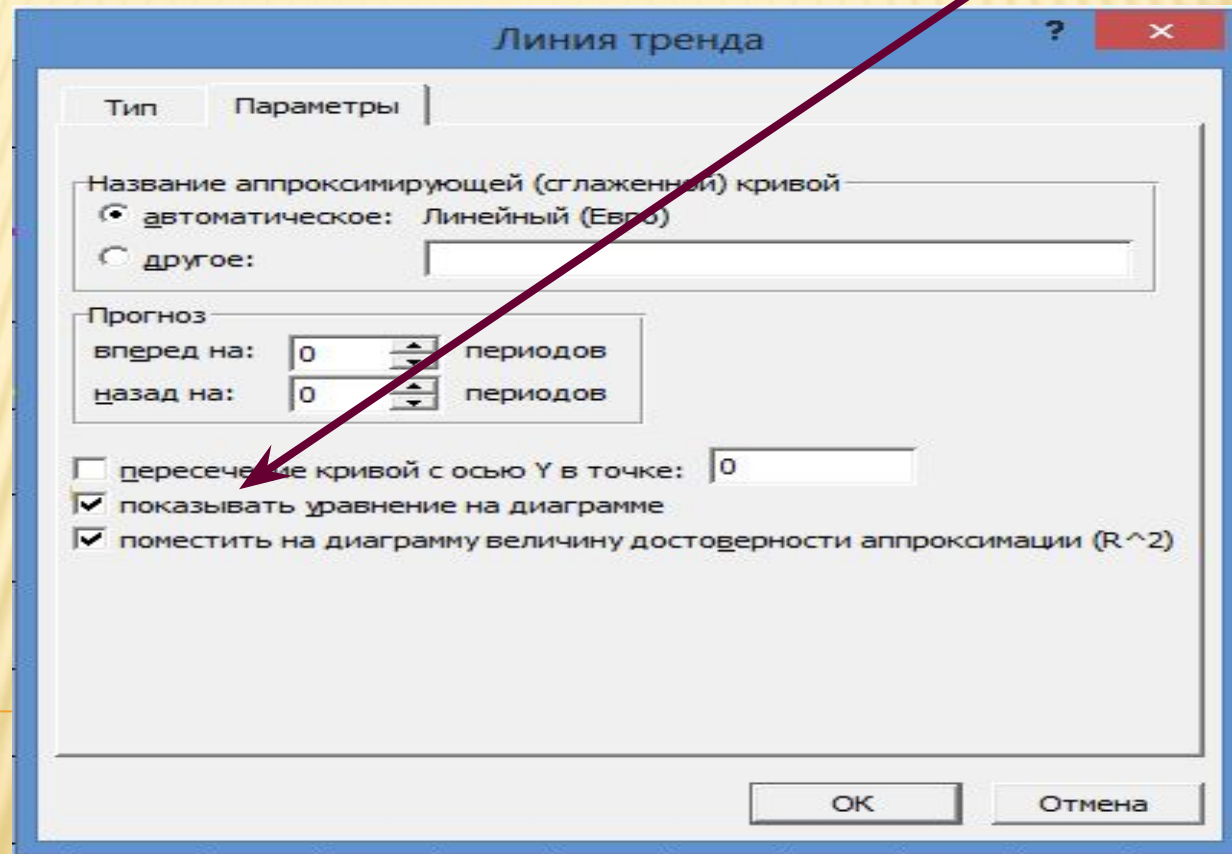
3. ДОБАВЛЕНИЕ ЛИНИИ ТРЕНДА НА ГРАФИК



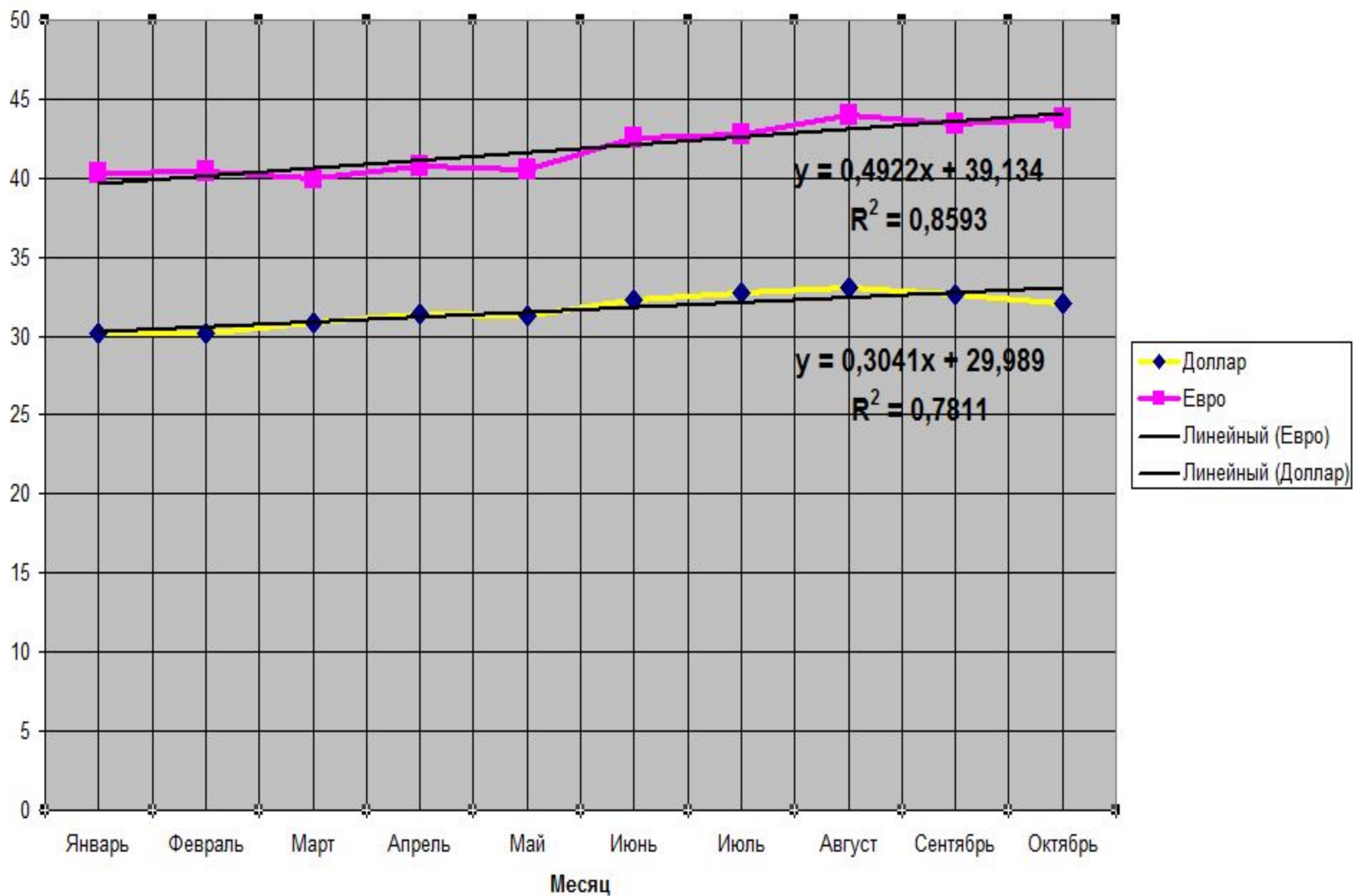
4. ВЫБОР ТИПА ЛИНИИ ТРЕНДА



В окне *Линия тренда* открыть вкладку *Параметры* и установить флажок *Показывать уравнение на диаграмме*



Построение линии тренда

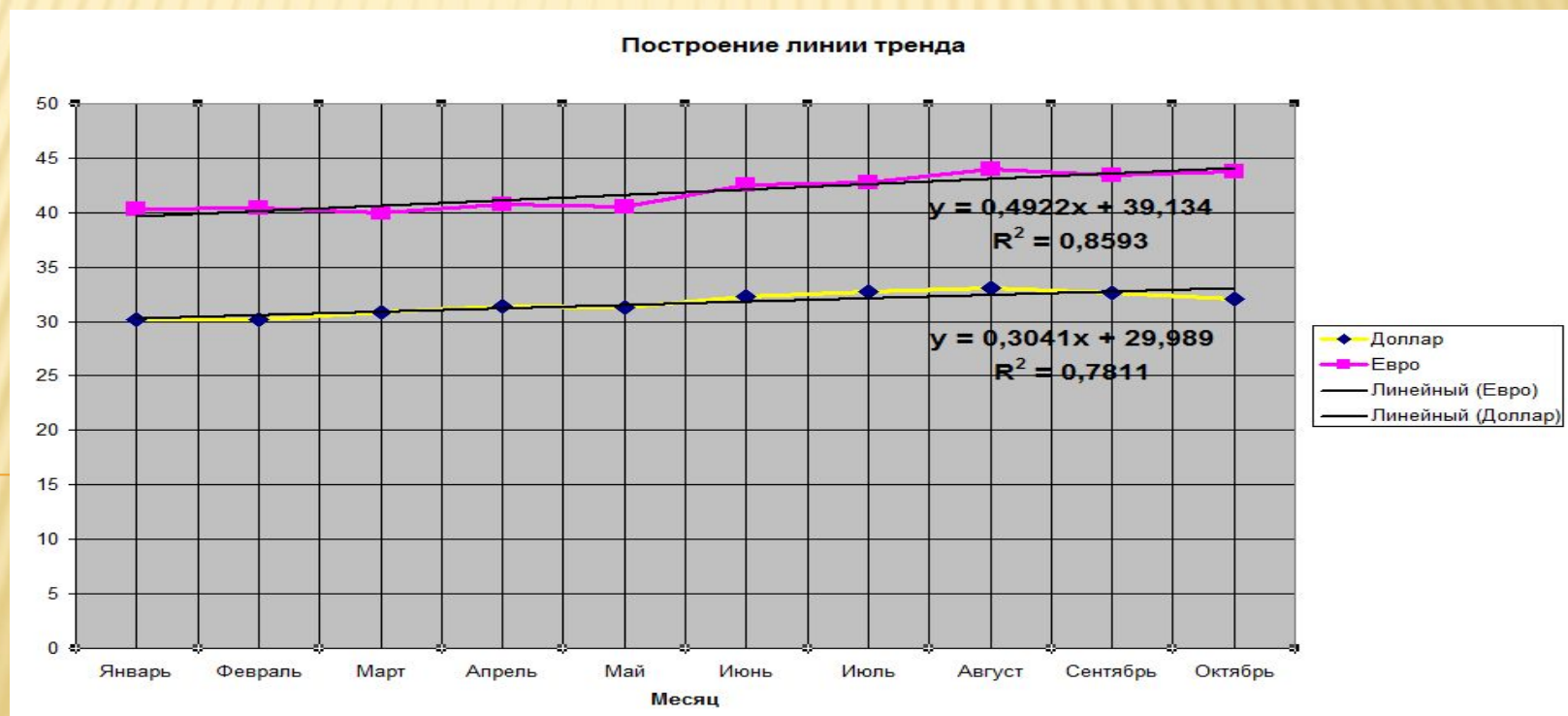


ПРОГНОЗ КУРСА ВАЛЮТ НА НОЯБРЬ 2013Г.

Уравнение курса евро и коэффициент достоверности (его максимальное значение равно 1):

$$y = 0,4922x + 39,134;$$

$$R^2 = 0,8593.$$

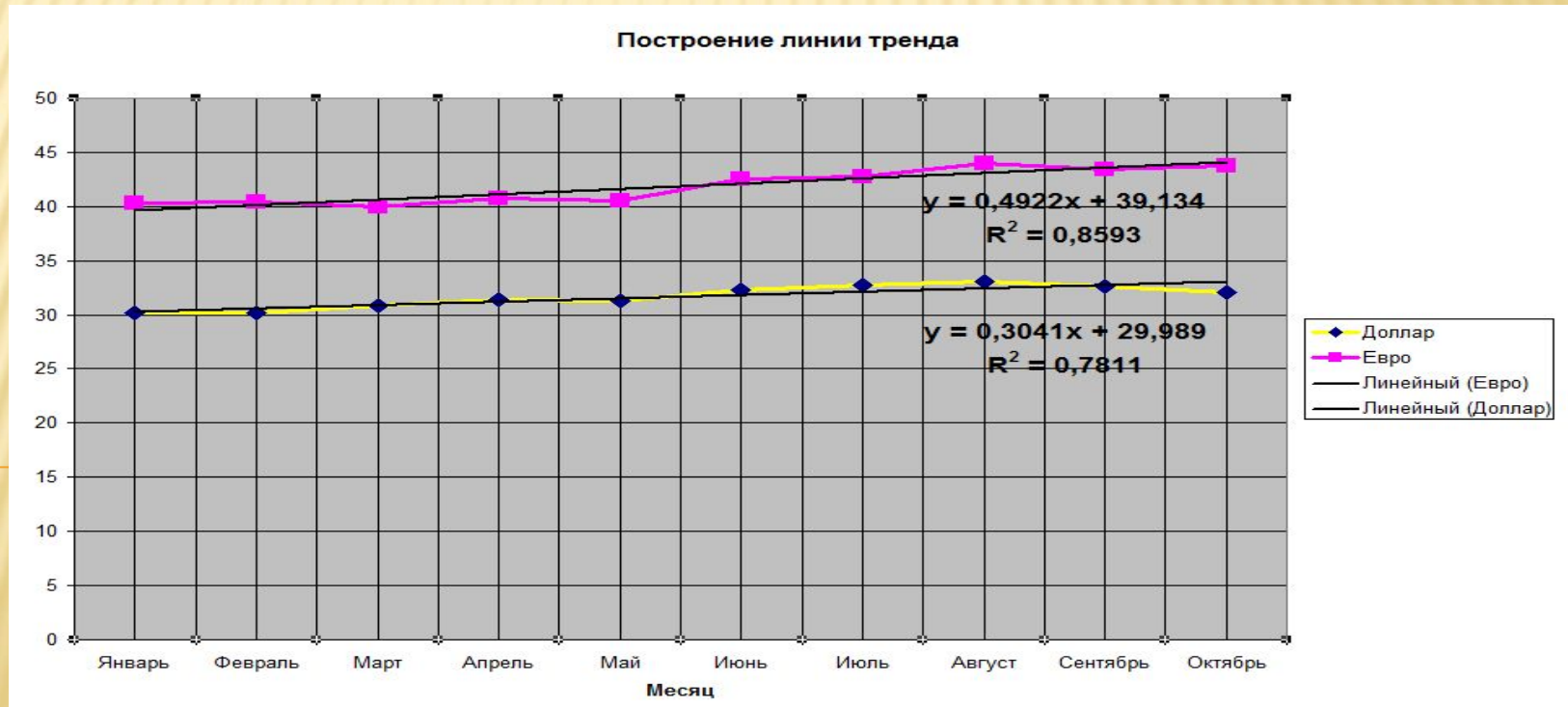


5. ПРОГНОЗ КУРСА ВАЛЮТ НА НОЯБРЬ 2013Г.

Уравнение курса доллара и коэффициент достоверности:

$$y = 0,3041x + 29,989;$$

$$R^2 = 0,7811.$$



ВЫЧИСЛЕНИЯ В ЭТ

В полученные с помощью линии тренда уравнения вместо X подставляем номер месяца для прогноза (для ноября $X=11$)

Таблица в режиме показа формул

E	F
№ месяца	Уравнение для евро
11	$y = 0,4922x + 39,134$
Прогноз на ноябрь	$=0,4922*E2 + 39,135$
№ месяца	Уравнение для доллара
11	$y = 0,3041x + 29,989$
Прогноз на ноябрь	$=0,3041*E6 + 29,99$

Таблица в режиме показа вычислений

E	F
№ месяца	Уравнение для евро
11	$y = 0,4922x + 39,134$
Прогноз на ноябрь	44,55
№ месяца	Уравнение для доллара
11	$y = 0,3041x + 29,989$
Прогноз на ноябрь	33,34

Самостоятельная работа 3

Задание. Определите курс евро на декабрь

- A. 48
- B. 76
- C. 65
- D. 67



Уравнение тренда для курса евро 2014

$$y = 0,5x + 61$$

Самостоятельная работа 4

Задание. Определите курс доллара на декабрь (с точностью до целых)

- A. 44
- B. 37
- C. 32
- D. 34



Уравнение тренда для курса доллара

$$y = 0,3041x + 29,989$$

Спасибо за
внимание

Спасибо за
ВНИМАНИЕ

Спасибо за
внимание

Спасибо за
внимание

СПАСИБО ЗА
ВНИМАНИЕ

Спасибо за
внимание

Спасибо за
внимание

Спасибо за
ВНИМАНИЕ

Спасибо за
внимание