

Тема 4. Обчислення коефіцієнта детермінації

Одним з показників, що описують якість побудованої моделі в статистиці, є коефіцієнт детермінації (R^2), який ще називають величиною достовірності апроксимації. З його допомогою можна визначити рівень точності прогнозу. Давайте дізнаємося, як можна зробити розрахунок даного показника за допомогою різних інструментів програми Excel.

Обчислення коефіцієнта детермінації

Залежно від рівня коефіцієнта детермінації, прийнято розділяти моделі на три групи:

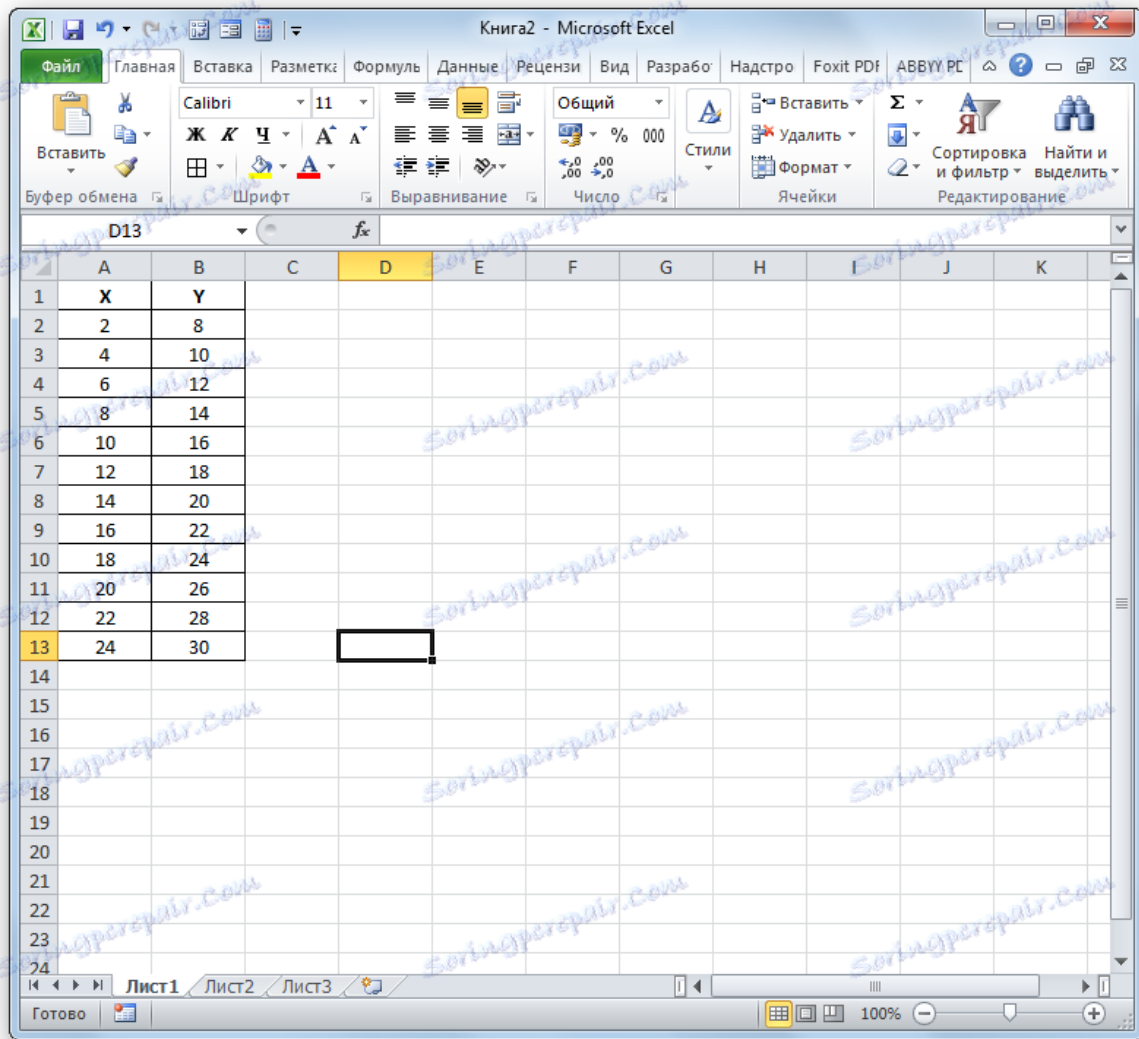
- 0,8 - 1 - модель гарної якості;
- 0,5 - 0,8 - модель прийнятної якості;
- 0 - 0,5 - модель поганої якості.

В останньому випадку якість моделі говорить про неможливість її використання для прогнозу.

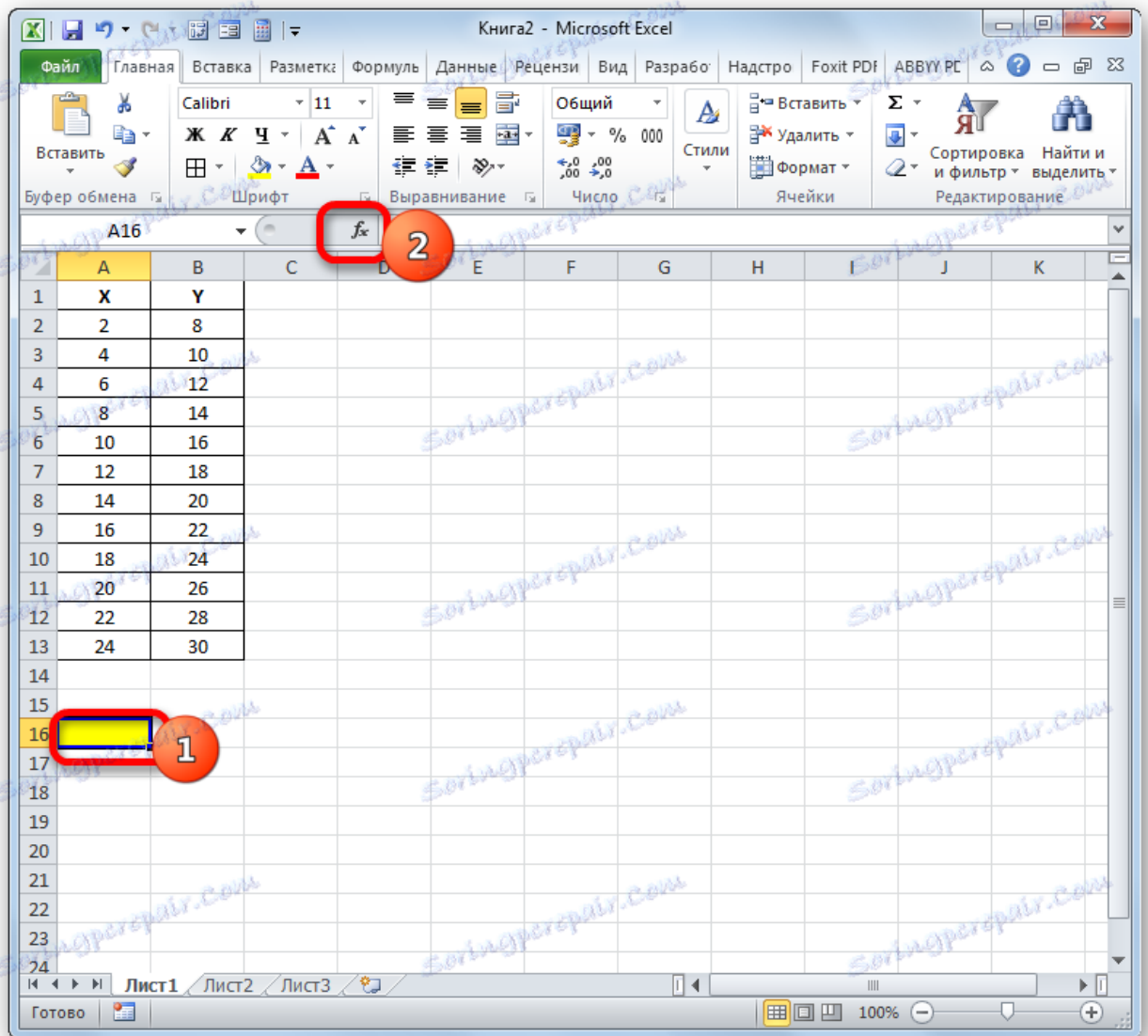
Вибір способу обчислення зазначеного значення в Excel залежить від того, чи є регресія лінійною чи ні. У першому випадку можна використовувати функцію *КВПІРСОН*, а в другому доведеться скористатися спеціальним інструментом з пакета аналізу.

Спосіб 1: обчислення коефіцієнта детермінації за лінійною функцією

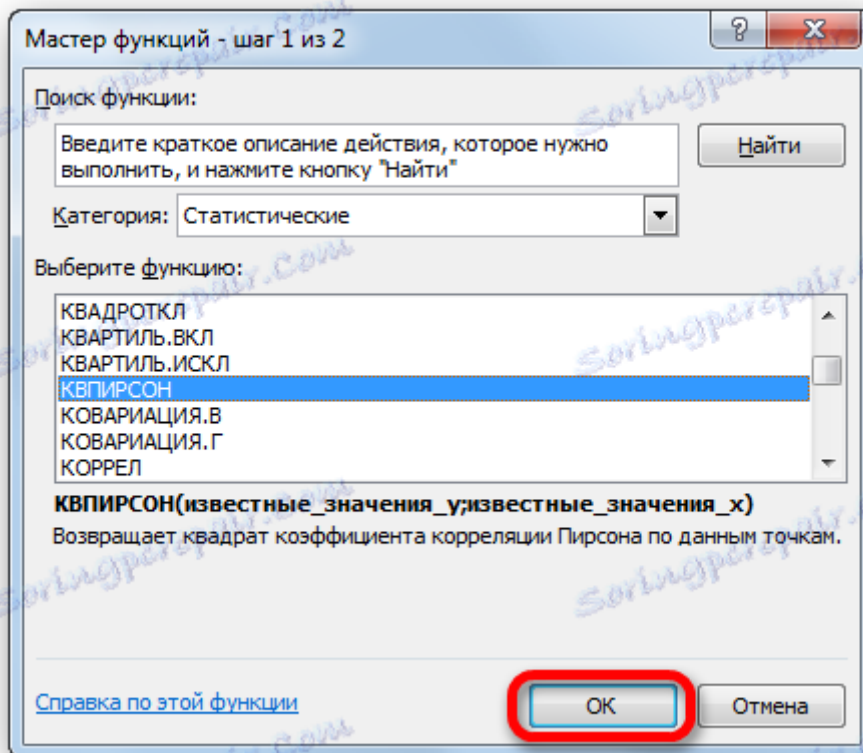
Перш за все, з'ясуємо, як знайти коефіцієнт детермінації за лінійною функцією. У цьому випадку даний показник буде дорівнювати квадрату коефіцієнта кореляції. Зробимо його розрахунок за допомогою вбудованої функції Excel на прикладі конкретної таблиці, яка наведена нижче.



1. Виділяємо осередок, де буде проведений висновок коефіцієнта детермінації після його розрахунку, і клацаємо по піктограмі «Вставити функцію».



2. Запускається *Майстер функцій*. Переміщаємося в його категорію «*Статистичні*» і відзначаємо найменування «*КВПІРСОН*». Далі клацають по кнопці «*ОК*».



3. Відбувається запуск вікна аргументів функції *КВПИРСОН*. Даний оператор з статистичної групи призначений для обчислення квадрата коефіцієнта кореляції функції Пірсона, тобто, лінійної функції. А як ми пам'ятаємо, за лінійною функцією коефіцієнт детермінації якраз дорівнює квадрату коефіцієнта кореляції.

Синтаксис цього оператора такий:

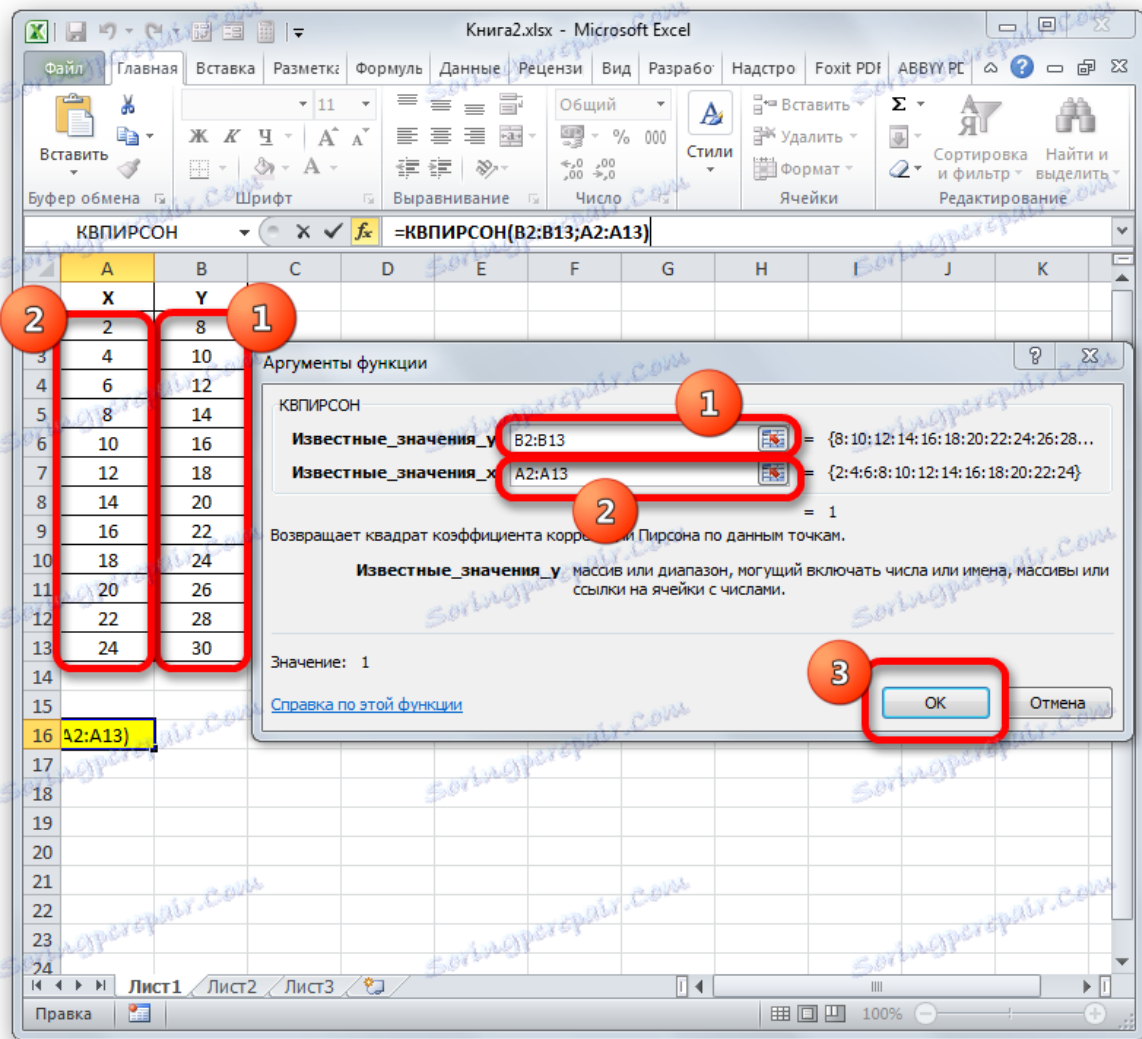
=КВПИРСОН(известные_значения_у;известные_значения_х)

Таким чином, функція має два оператора, один з яких представляє собою перелік значень функції, а другий - аргументів. Оператори можуть бути представлені, як безпосередньо у вигляді значень, перерахованих через крапку з комою (;), так і у вигляді посилань на діапазони, де вони розташовані. Саме останній варіант і буде використаний нами в даному прикладі.

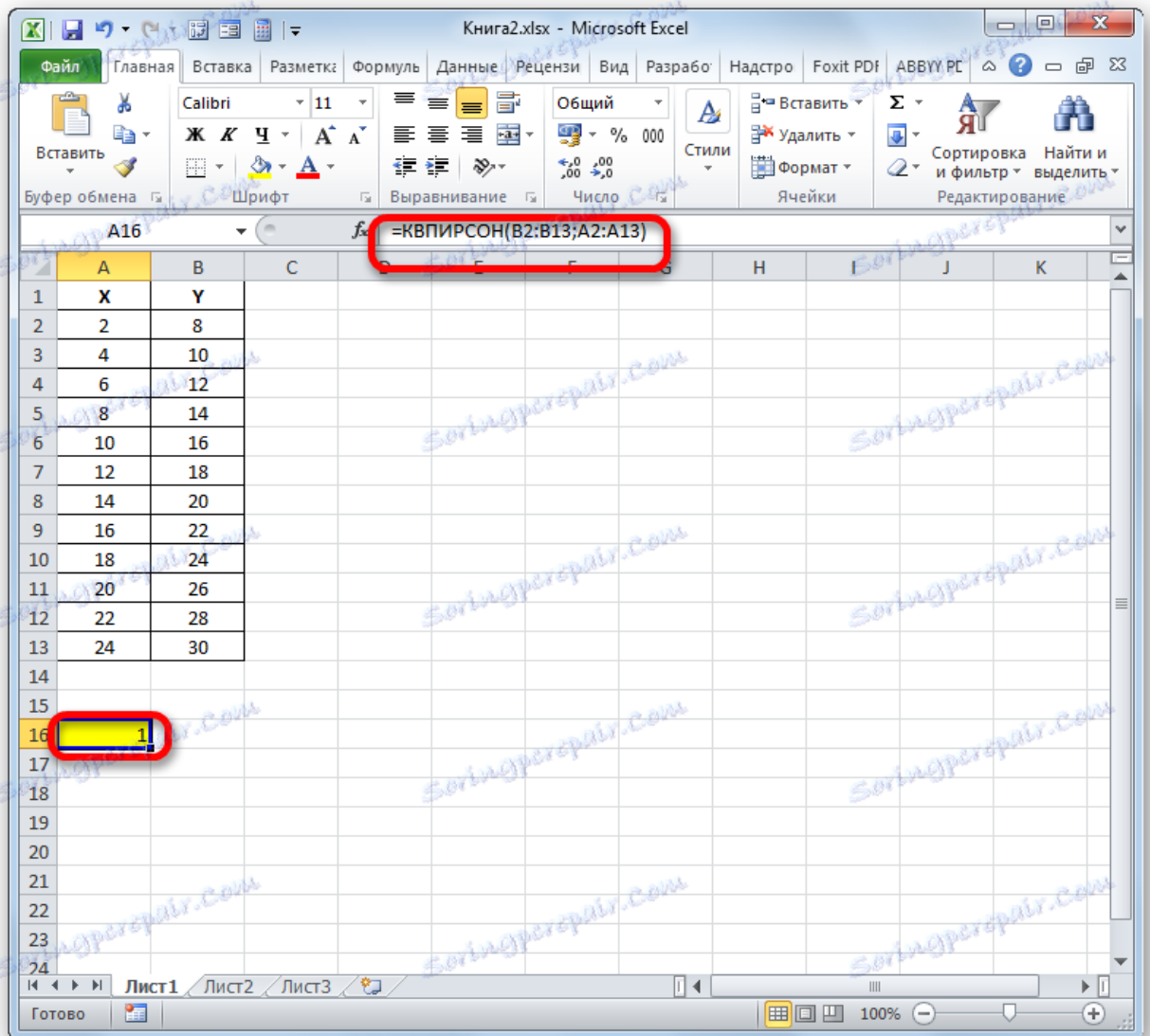
Встановлюємо курсор в поле «Відомі значення у». Виконуємо затиск лівої кнопки мишки і виробляємо виділення вмісту стовпця «У» таблиці. Як бачимо, адреса зазначеного масиву даних тут же відображається у вікні.

Аналогічним чином заповнюємо поле «Відомі значення х». Ставимо курсор в цьому полі, але на цей раз виділяємо значення стовпця «Х».

Після того, як всі дані були відображені у вікні аргументів *КВПИРСОН*, клацають по кнопці «OK», розташованій в самому його низу.



4. Як бачимо, слідом за цим програма проводить розрахунок коефіцієнта детермінації і видає результат у ту клітинку, яка була виділена ще перед викликом *Майстра функцій*. У нашому прикладі значення обчислюваного показника вийшло рівним 1. Це означає, що представлена модель абсолютно достовірна, тобто, виключає похибка.

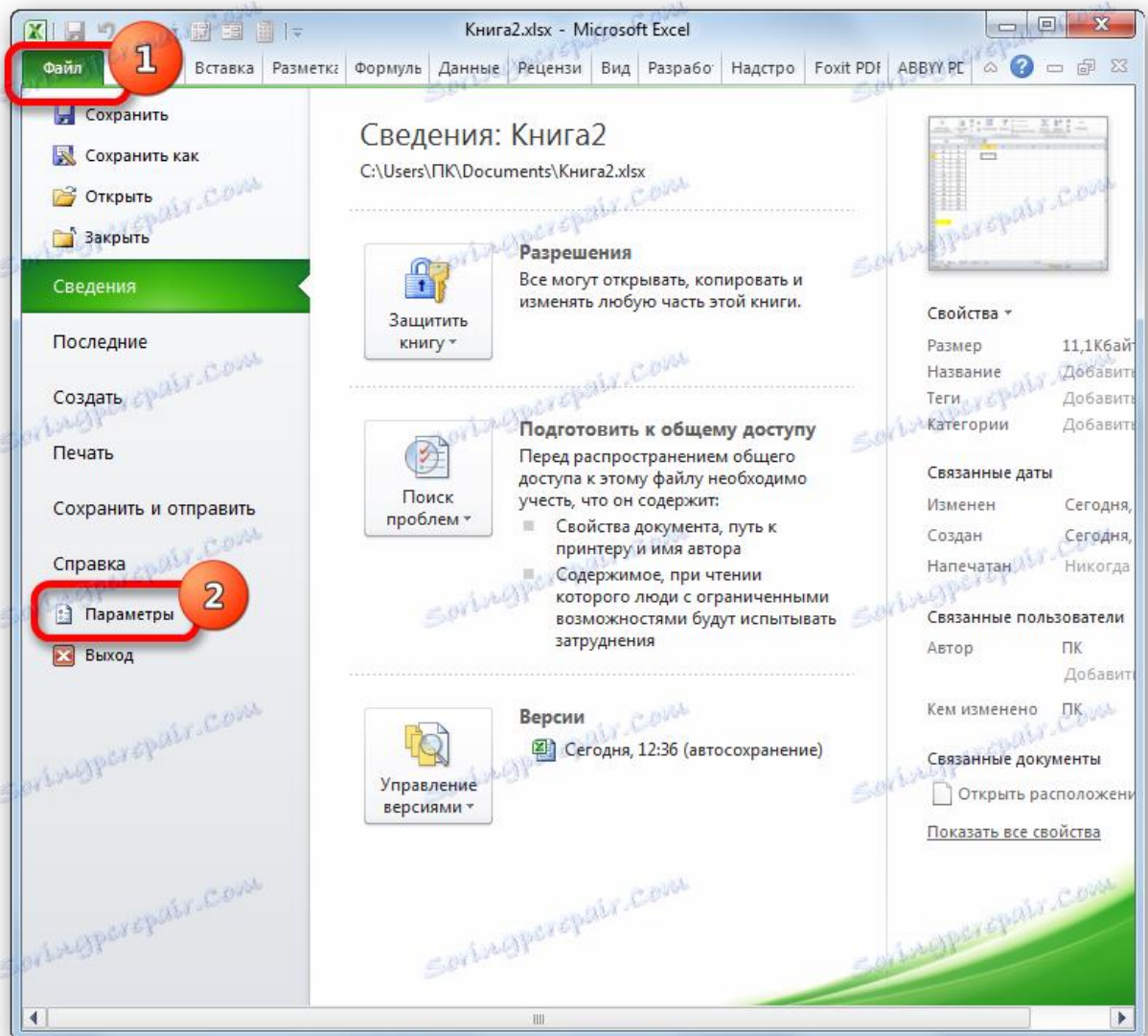


урок: [Майстер функцій в Microsoft Excel](#)

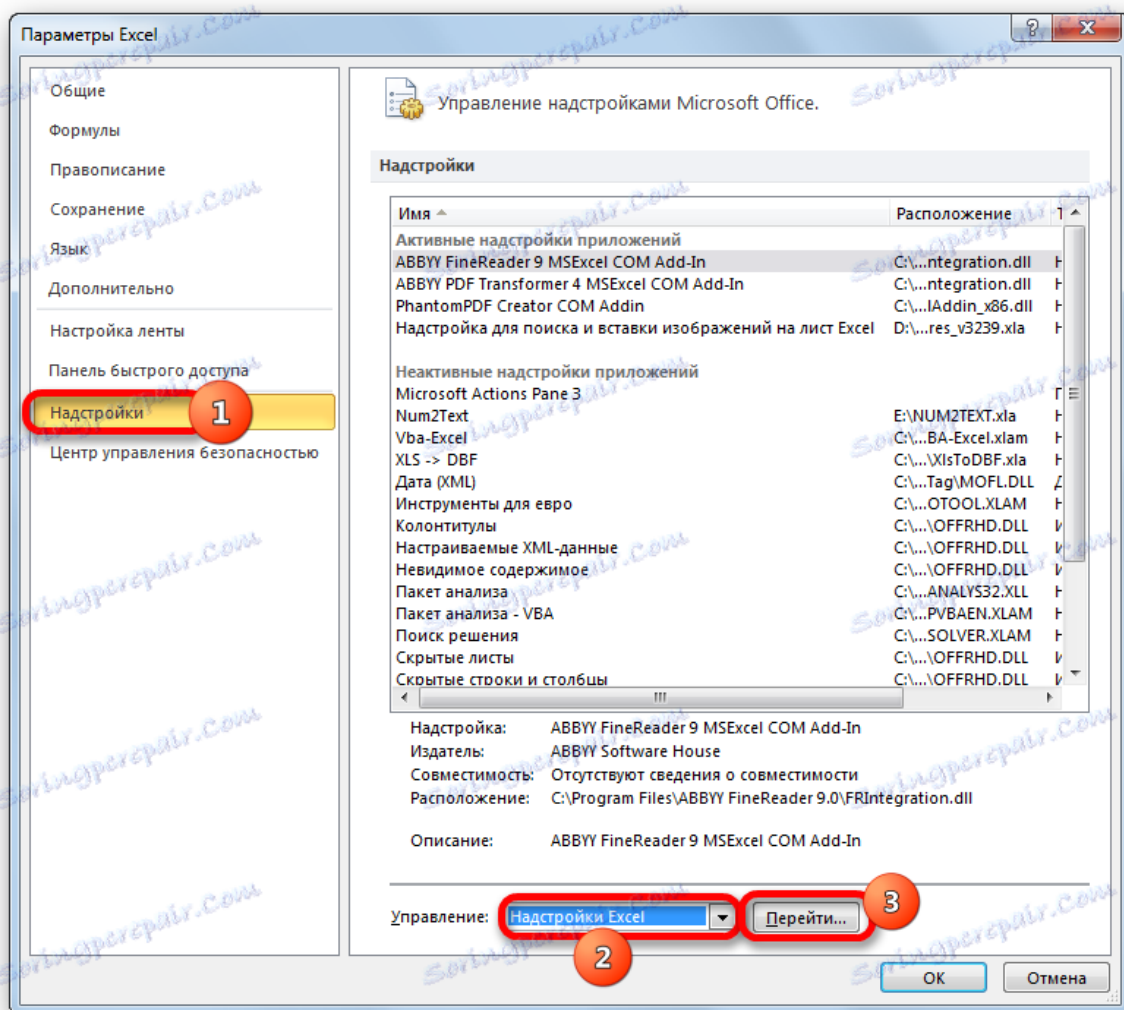
Спосіб 2: обчислення коефіцієнта детермінації в нелінійних функціях

Але зазначений вище варіант розрахунку шуканого значення можна застосовувати тільки до лінійних функцій. Що ж робити, щоб зробити його розрахунок в нелінійній функції? У Ексель є і така можливість. Її можна здійснити за допомогою інструменту «Регресія», який є складовою частиною пакету «Аналіз даних».

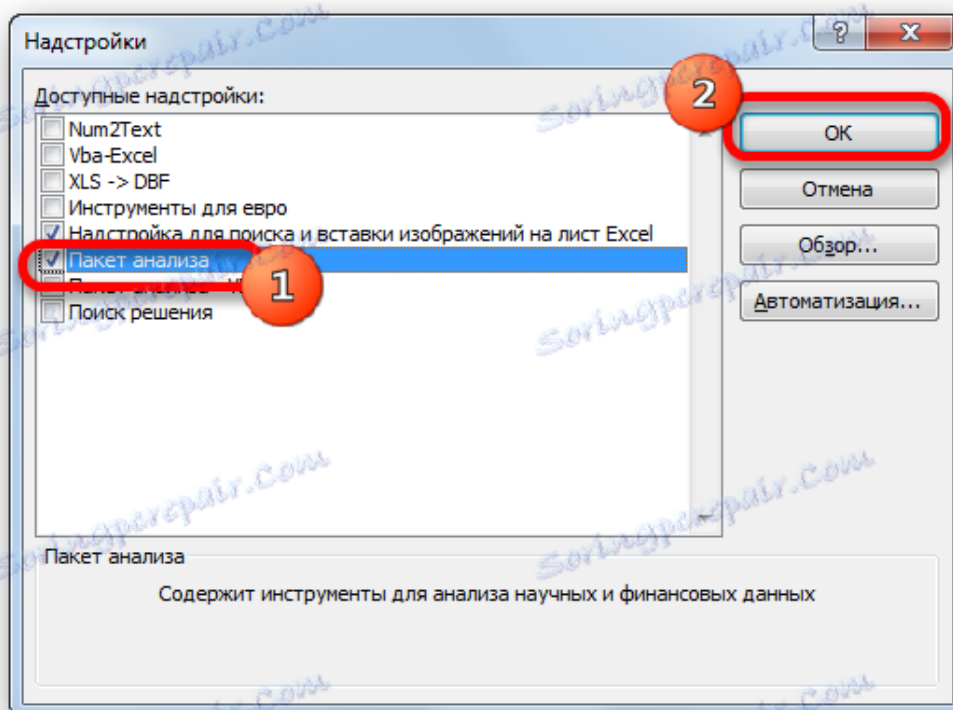
1. Але перш, ніж скористатися зазначеним інструментом, слід активувати сам «Пакет аналізу», який за замовчуванням в Ексель відключений. Переміщаємося у вкладку «Файл», а потім переходимо по пункту «Параметри».



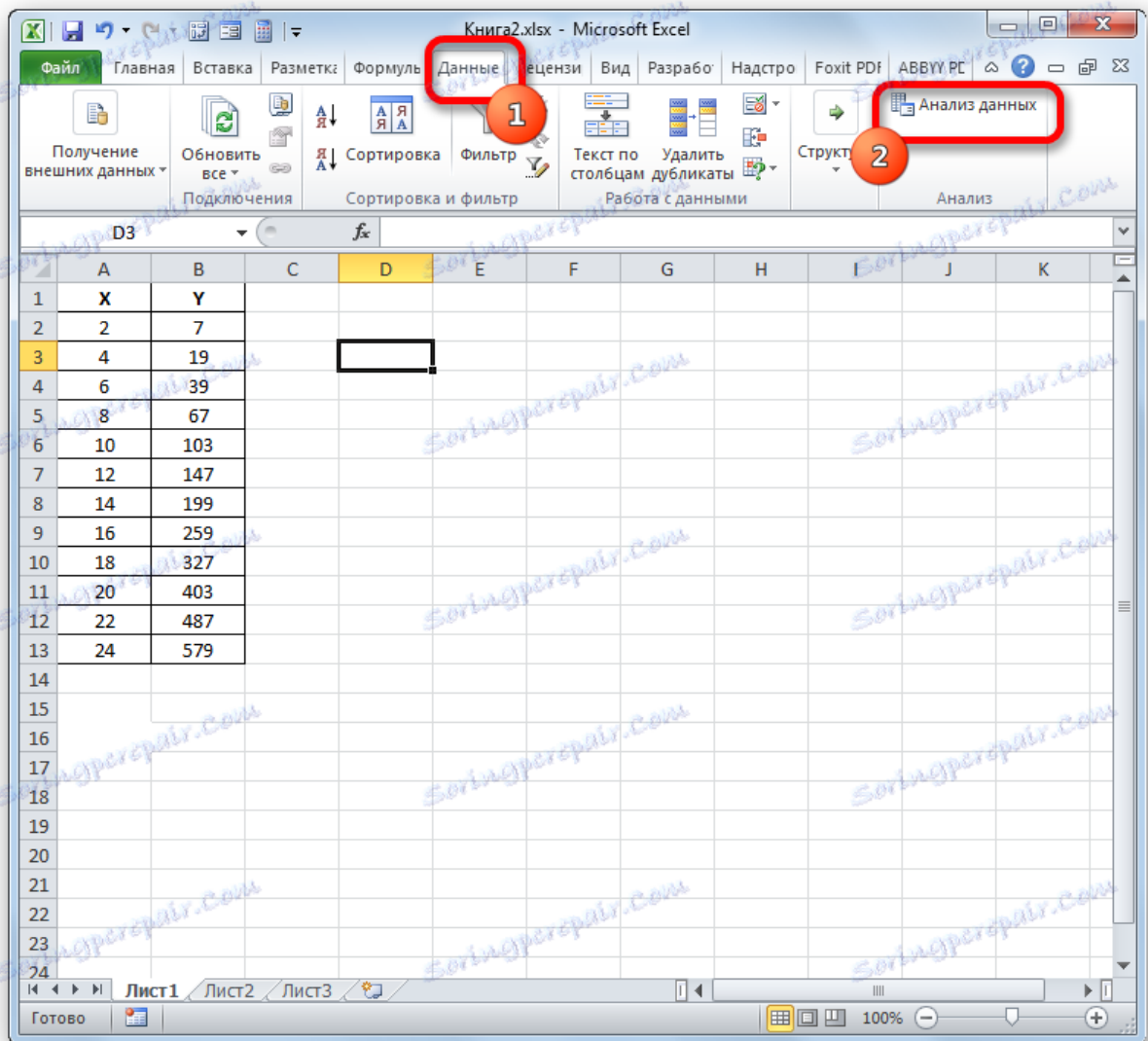
2. У вікні, виробляємо переміщення в розділ «компонентами» за допомогою навігації по лівому вертикальному меню. У нижній частині правої області вікна розташовується поле «Управління». Зі списку доступних там підрозділів вибираємо найменування «Надбудови Excel ...», а потім клацаємо по кнопці «Перейти ...», розташованій праворуч від поля.



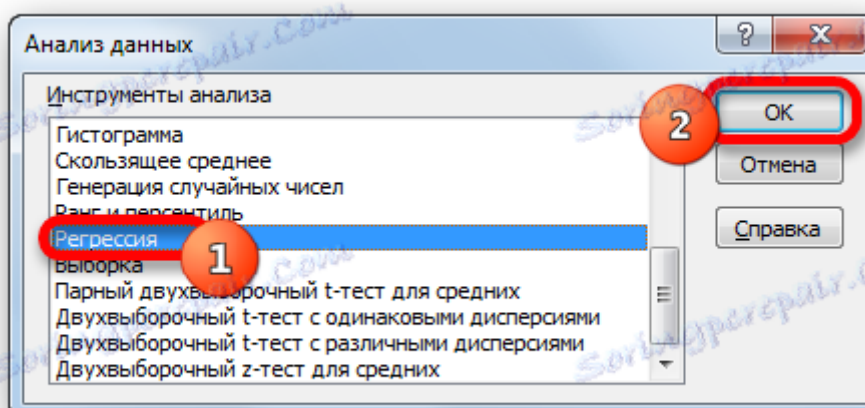
3. Виробляється запуск вікна надбудов. У центральній його частині розташований список доступних надбудов. Встановлюємо прапорець біля позиції «Пакет аналізу». Слідом за цим потрібно клацнути по кнопці «ОК» у правій частині інтерфейсу вікна.



4. Пакет інструментів «Аналіз даних» в поточному екземплярі Excel буде активований. Доступ до нього розташовується на стрічці у вкладці «Дані». Переміщаємося в зазначену вкладку і клацають по кнопці «Аналіз даних» в групі налаштувань «Аналіз».



5. Активується вікнече «Аналіз даних» зі списком профільних інструментів обробки інформації. Виділяємо з цього переліку пункт «Регресія» і клацають по кнопці «ОК».



6. Потім відкривається вікно інструменту «Регресія». Перший блок налаштувань - «Вхідні дані». Тут в двох полях потрібно вказати адреси діапазонів, де знаходяться значення аргументу і функції. Ставимо курсор в поле «Вхідний інтервал Y» і виділяємо на аркуші

вміст колонки «Y». Після того, як адресу масиву відобразився у вікні «Регресія», ставимо курсор в полі «Вихідний інтервал Y» і точно таким же чином виділяємо комірки стовпчика «X».

Близько параметрів «Метка» і «Константа-нуль» прапорці не ставимо. Прапорець можна встановити близько параметра «Рівень надійності» і в полі навпроти вказати бажану величину відповідного показника (за замовчуванням 95%).

У групі «Параметри виводу» потрібно вказати, в якій області буде відображатися результат обчислення. Існує три варіанти:

- Область на поточному аркуші;
- Інший лист;
- Інша книга (новий файл).

Зупинимо свій вибір на першому варіанті, щоб вихідні дані і результат розміщувалися на одному робочому аркуші. Ставимо перемикач близько параметра «Вихідний інтервал». В поле навпроти даного пункту ставимо курсор. Клацаємо лівою кнопкою миші по порожньому елементу на аркуші, який покликаний стати лівої верхньої осередком таблиці виведення підсумків розрахунку. Адреса цього елемента повинен висвітитися в поле вікна «Регресія».

Групи параметрів «Залишки» і «Нормальна ймовірність» ігноруємо, так як для вирішення поставленого завдання вони не важливі. Після цього клацають по кнопці «ОК», яка розміщена в правому верхньому куті вікна «Регресія».

Книга2.xlsx - Microsoft Excel

Регрессия

Входные данные

Входной интервал Y: **1** \$B\$2:\$B\$13 **5** OK

Входной интервал X: **2** \$A\$2:\$A\$13

Метки Константа - ноль

Уровень надежности: 97% **3**

Параметры вывода

4 Выходной интервал: \$D\$3

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

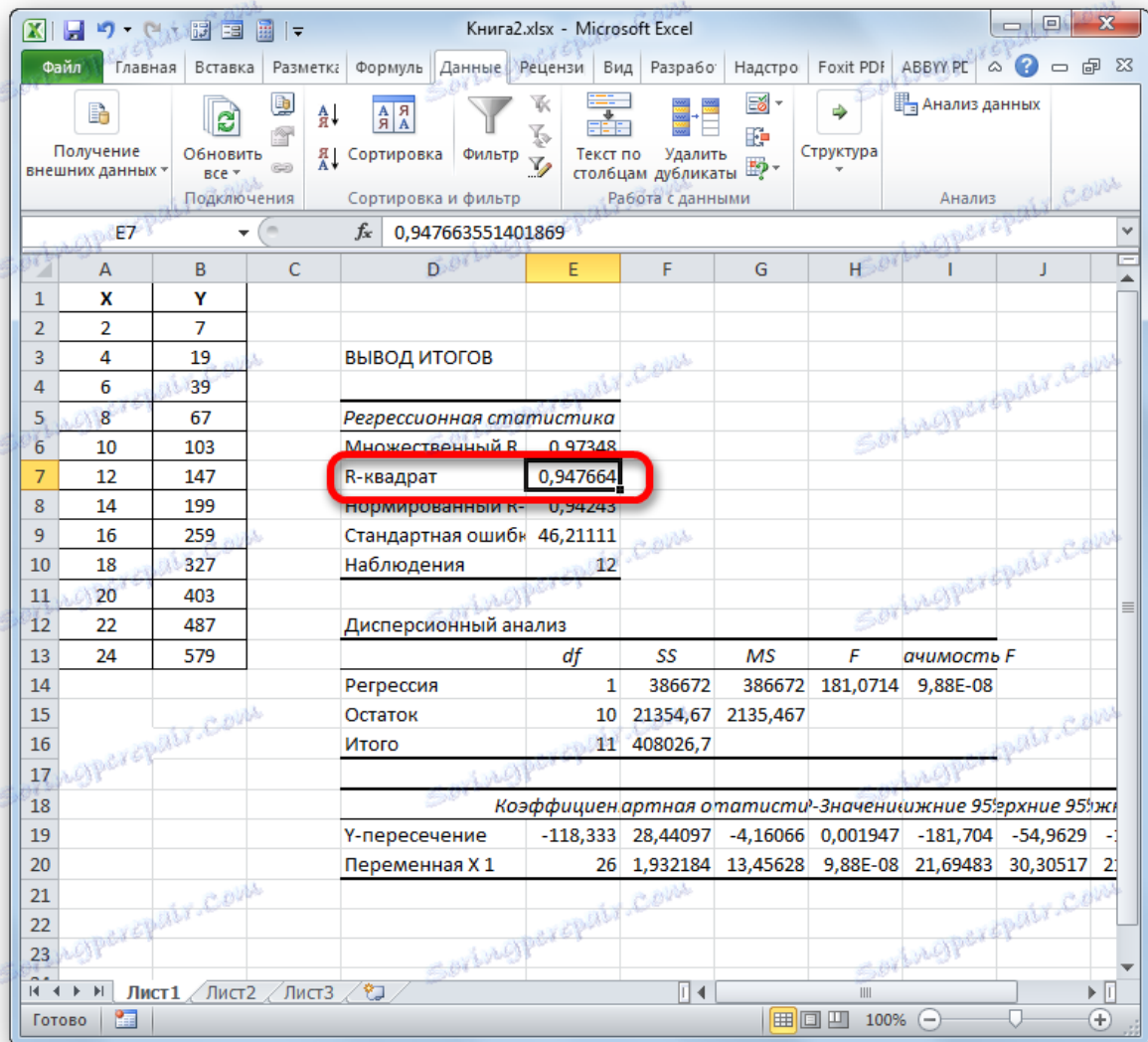
Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

	X	Y
2	2	7
3	4	19
4	6	39
5	8	67
6	10	103
7	12	147
8	14	199
9	16	259
10	18	327
11	20	403
12	22	487
13	24	579

7. Програма проводить розрахунок на основі раніше введених даних і виводить результат у вказаний діапазон. Як бачимо, даний інструмент виводить на лист досить велика кількість результатів за різними параметрами. Але в контексті поточного уроку нас цікавить показник «*R-квадрат*». В даному випадку він дорівнює 0,947664, що характеризує обрану модель, як модель хорошої якості.



Крім того, може бути надано: графік підбору – порівняльна діаграма, що містить емпіричну і теоретичну лінії регресії; таблиця залишків – різниця емпіричних і теоретичних значень Y

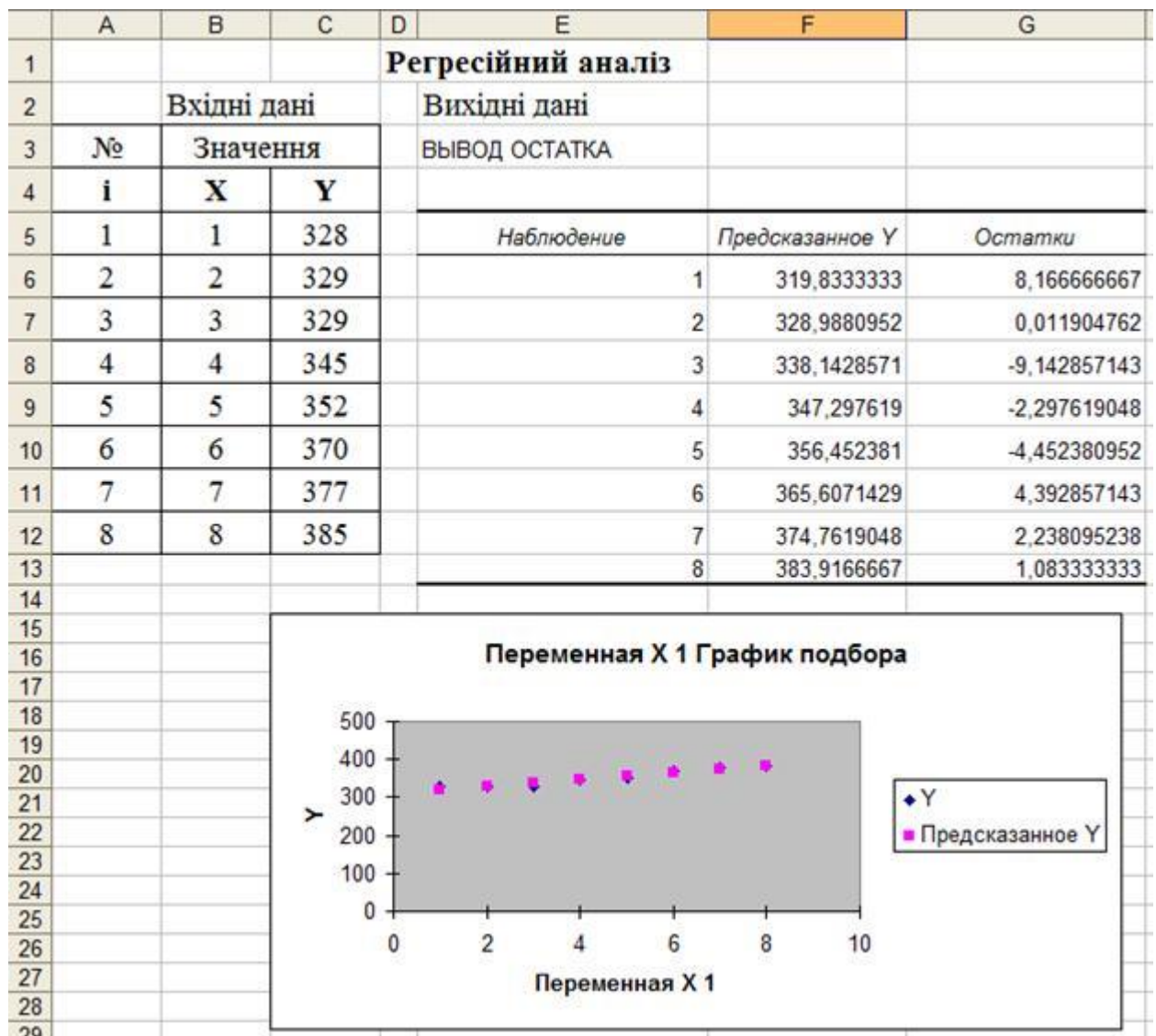
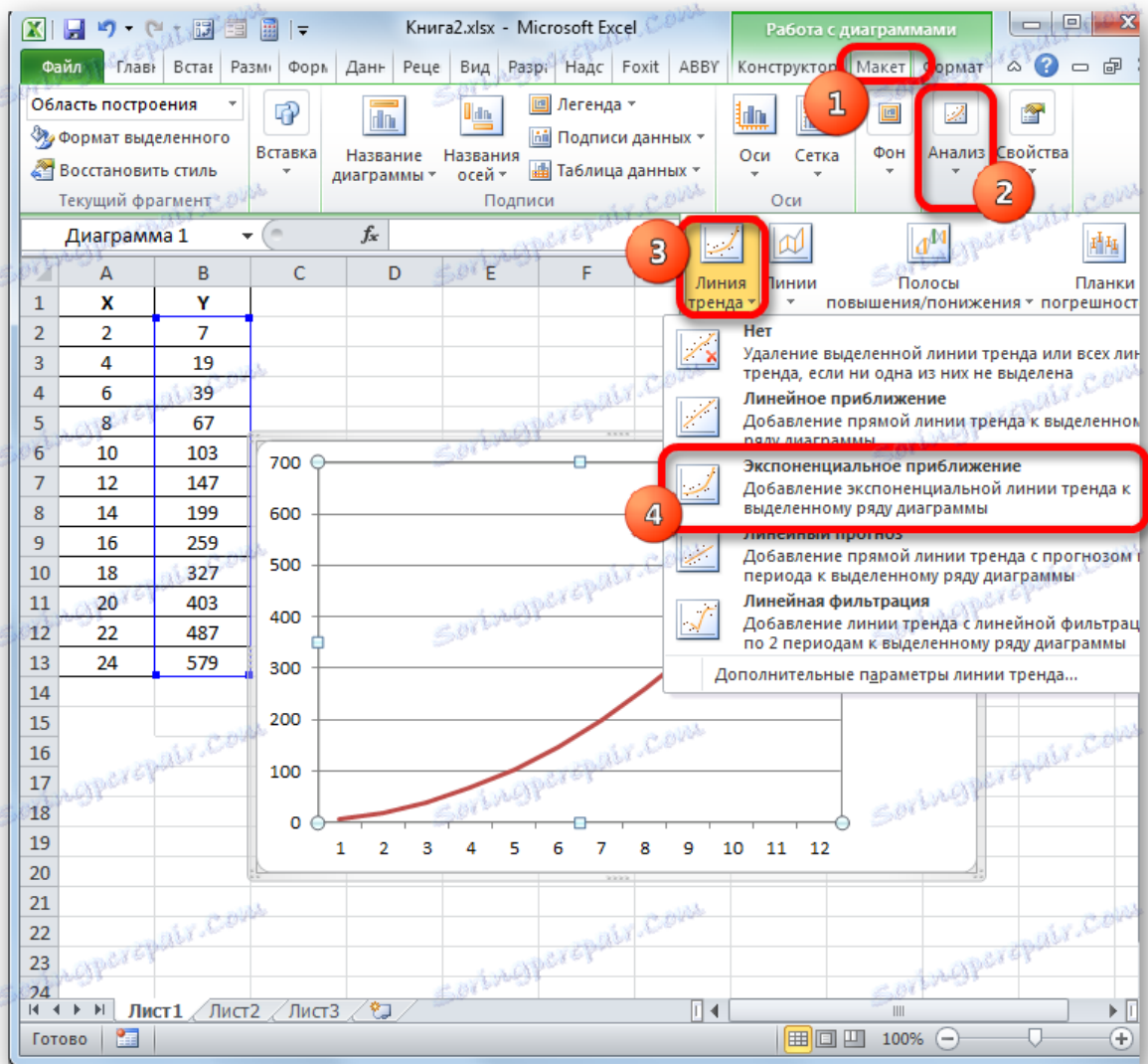


Рис. Додаткові результати регресійного аналізу

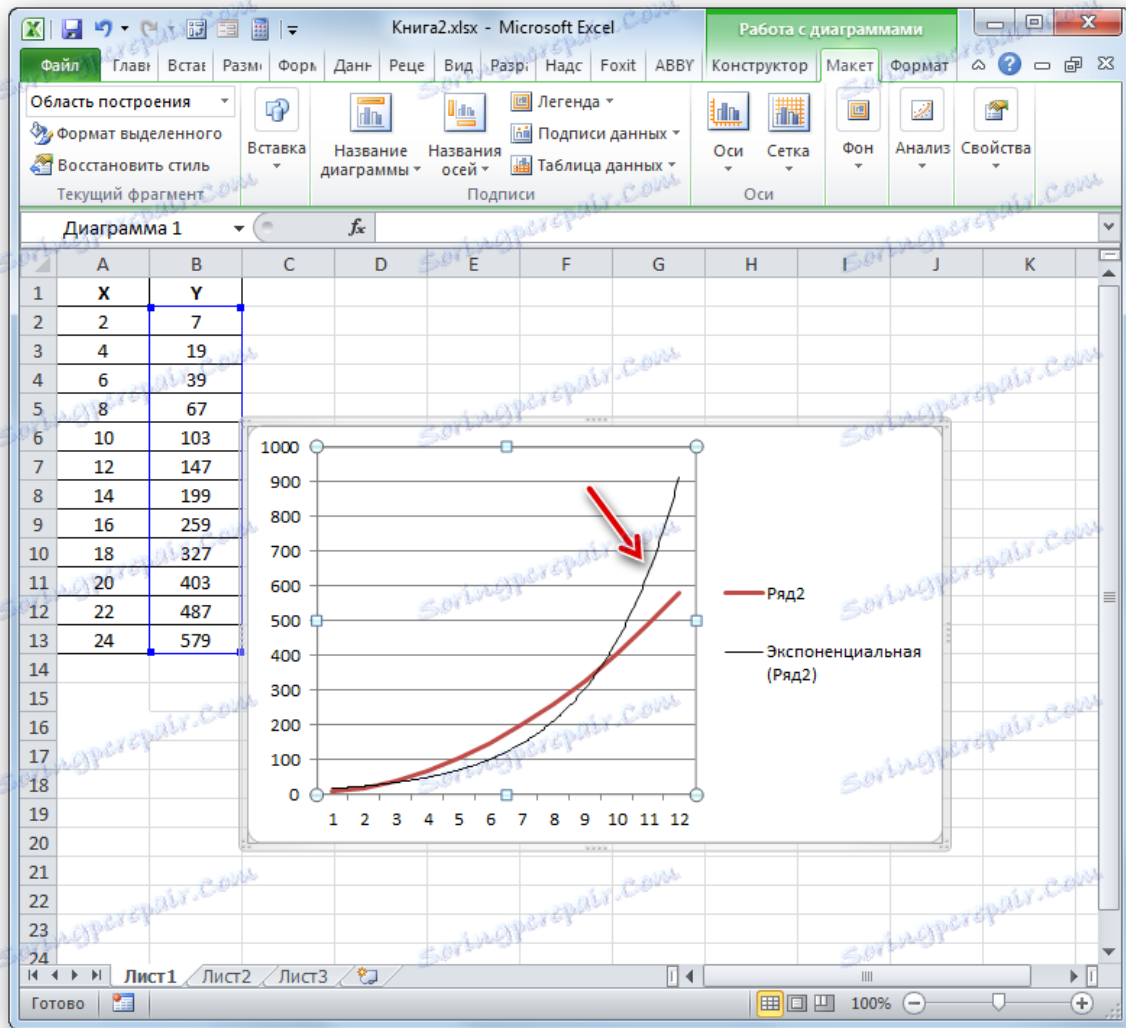
Спосіб 3: коефіцієнт детермінації для лінії тренду

Крім зазначених вище варіантів, коефіцієнт детермінації можна відобразити безпосередньо для лінії тренду в графіку, побудованому на аркуші Excel. З'ясуємо, як це можна зробити на конкретному прикладі.

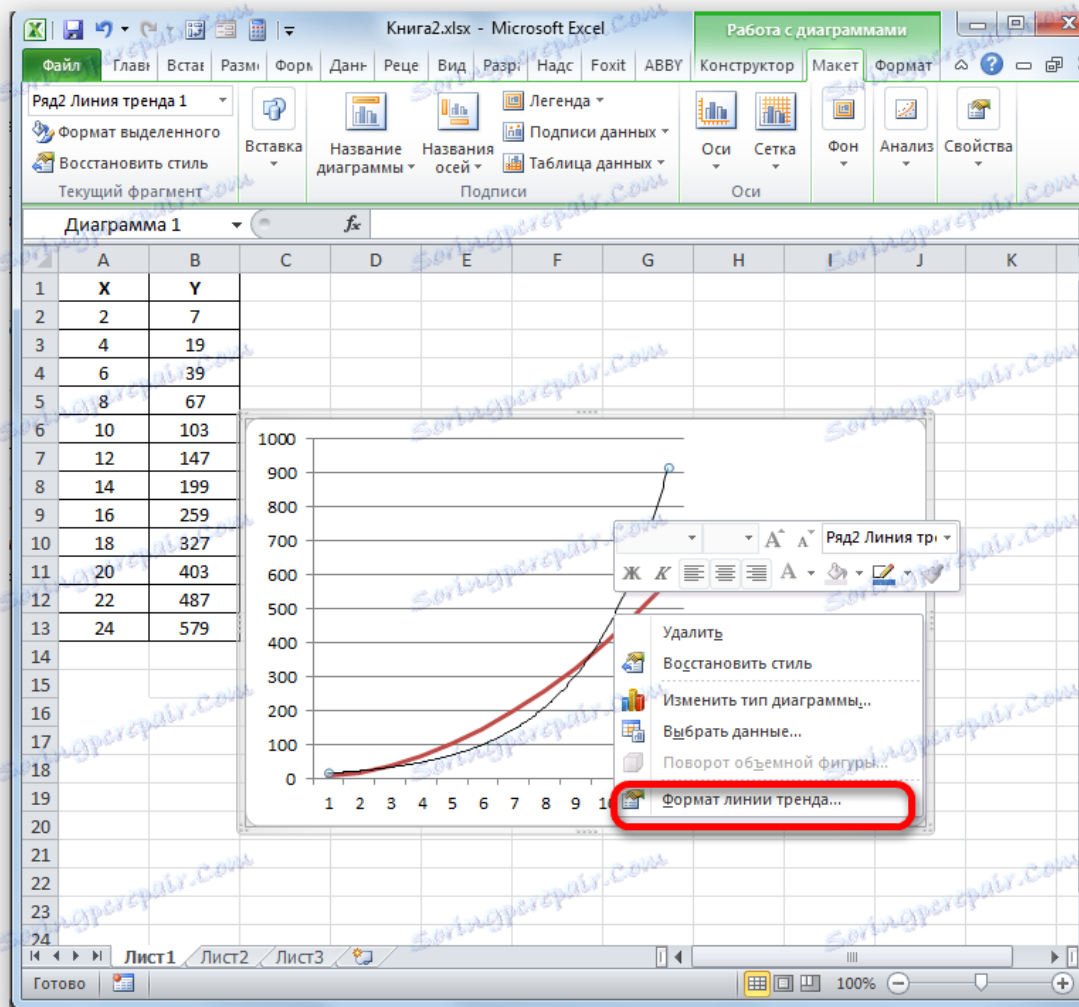
1. Ми маємо графік, побудований на основі таблиці аргументів і значень функції, яка була використана для попереднього прикладу. Зробимо побудова до нього лінії тренда. Кількома по будь-якого місця області побудови, на якій розміщений графік, лівою кнопкою миші. При цьому на стрічці з'являється додатковий набір вкладок - «Робота з діаграмами». Переходимо у вкладку «Макет». Клацають по кнопці «Лінія тренда», яка розміщена в блоці інструментів «Аналіз». З'являється меню з вибором типу лінії тренда. Зупиняємо вибір на тому типі, який відповідає конкретній задачі. Давайте для нашого прикладу виберемо варіант «Експоненціальне наближення».



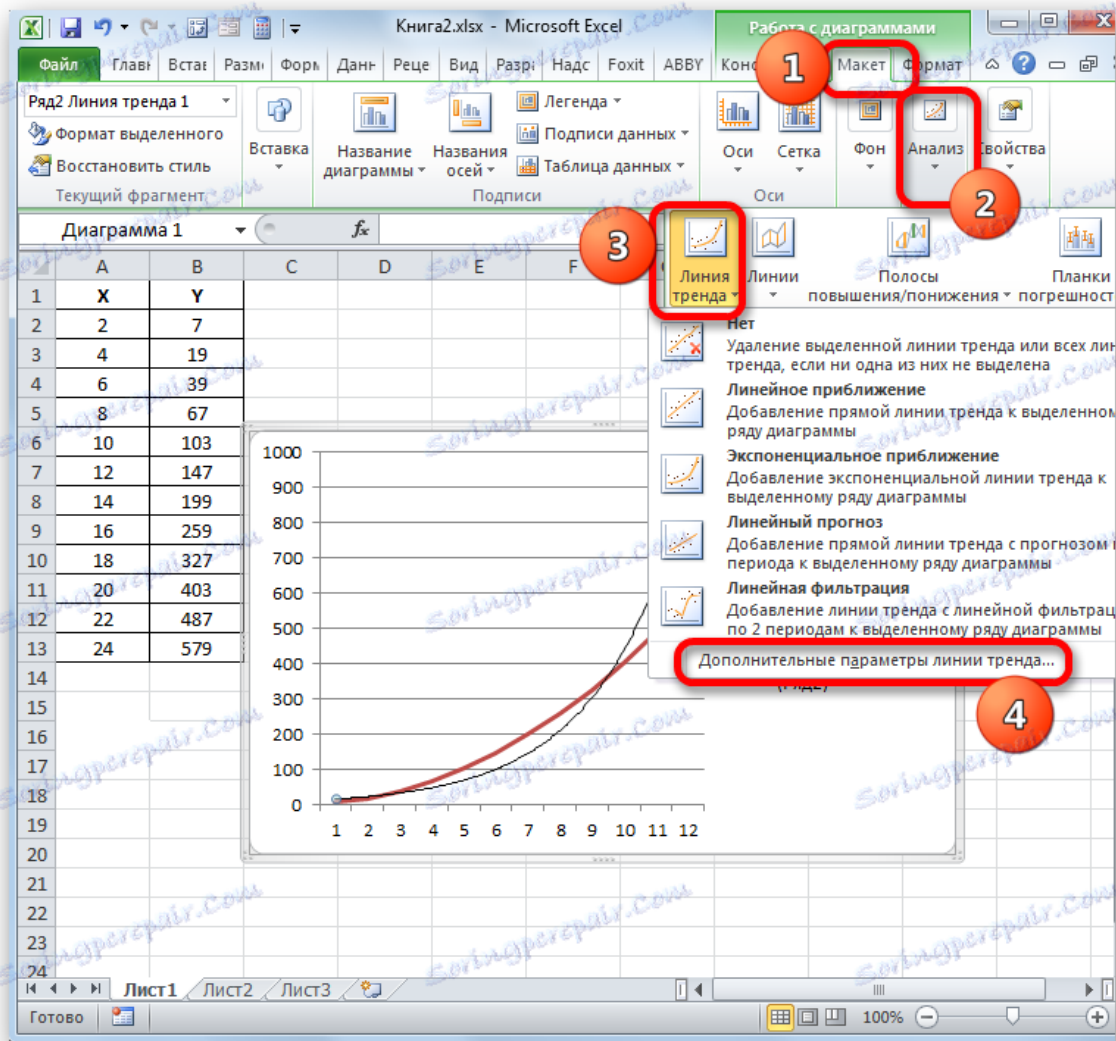
2. Ексель буде прямо на площині побудови графіка лінію тренда у вигляді додаткової чорної кривої.



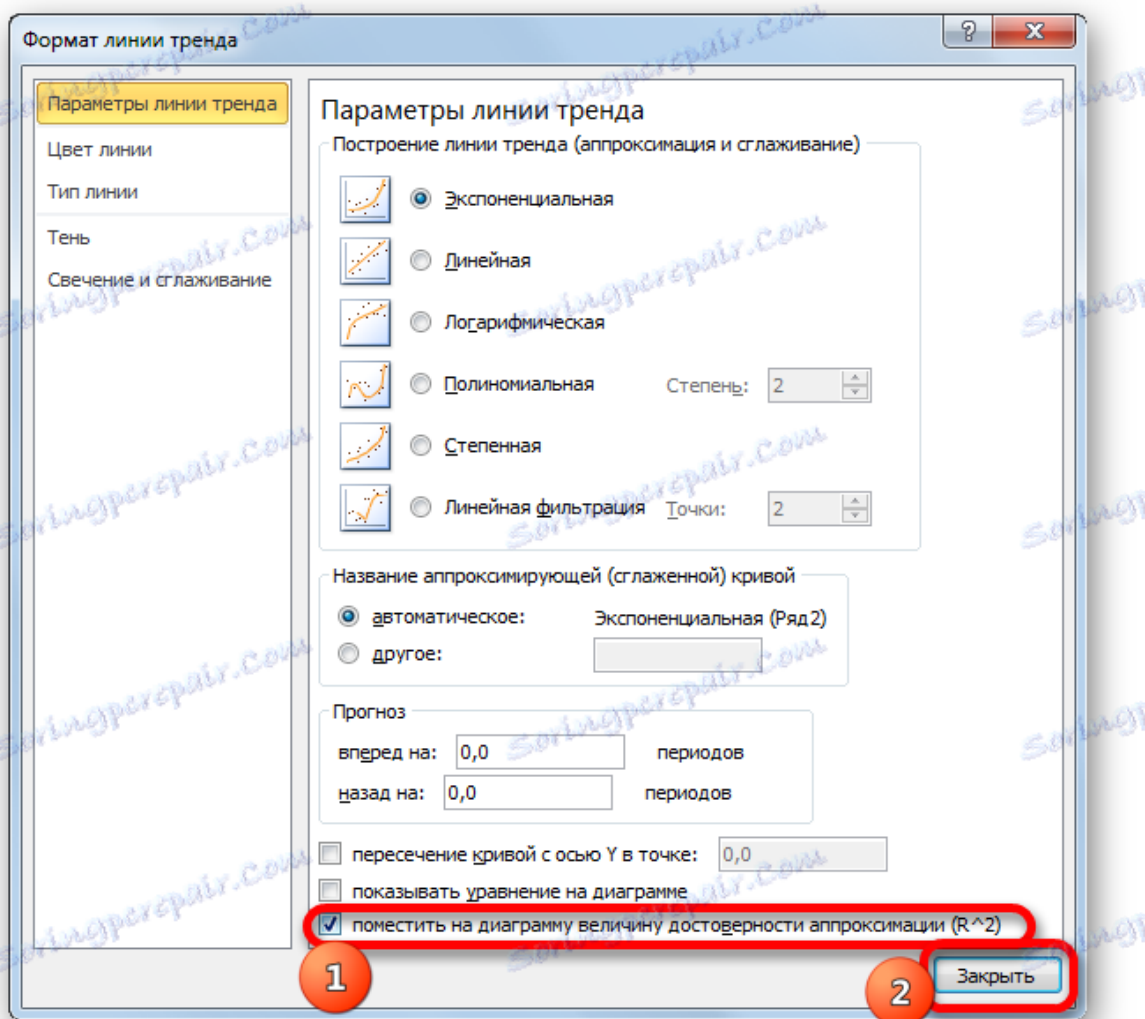
- Тепер нашим завданням є відобразити власне коефіцієнт детермінації. Натискаємо правою кнопкою миші по лінії тренда. Активується контекстне меню. Зупиняємо вибір на ньому на пункті «Формат лінії тренда ...».



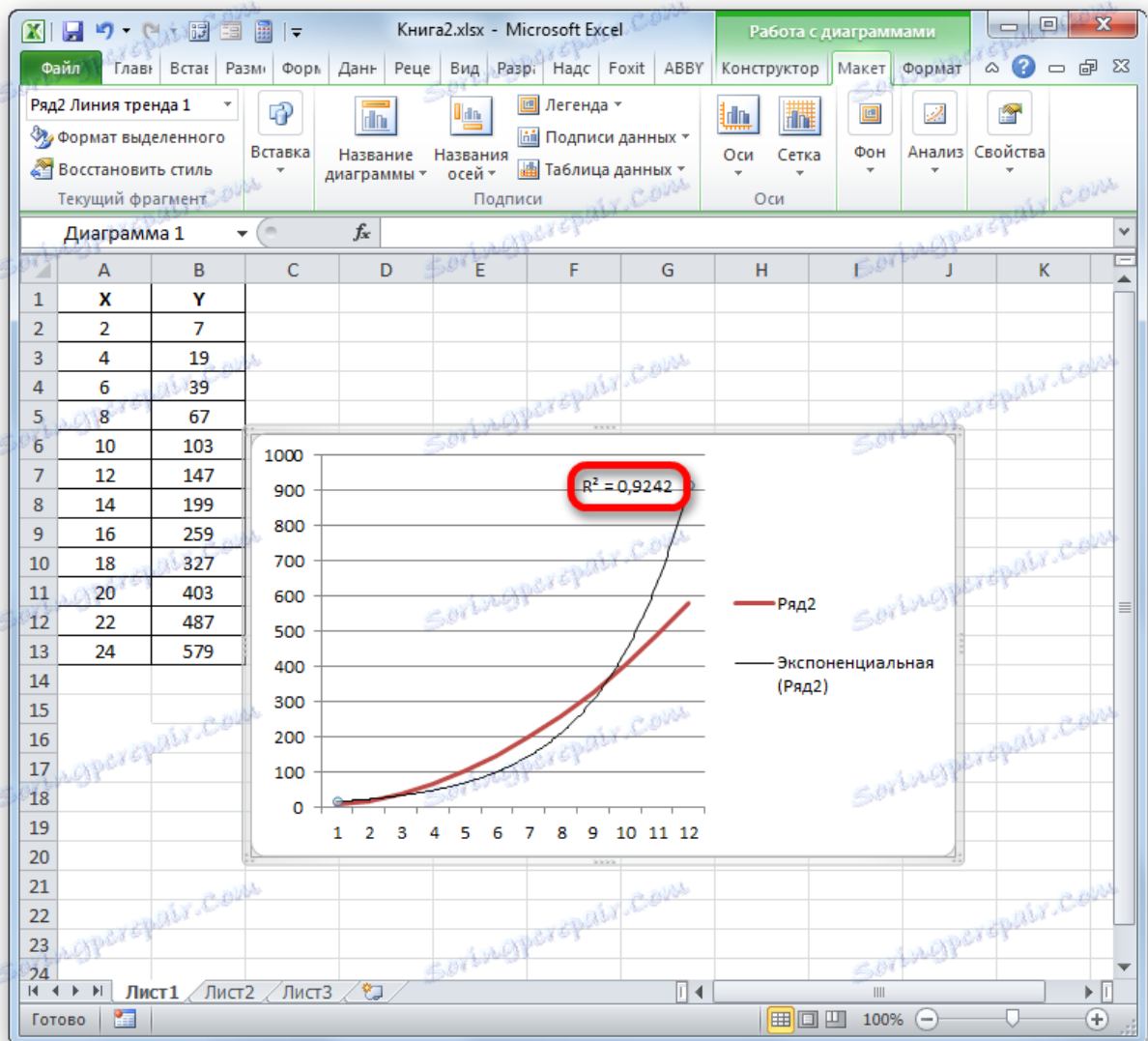
Для виконання переходу в вікно формату лінії тренда можна виконати альтернативне дію. Виділяємо лінію тренда кліком по ній лівою кнопки миші. Переміщаємося у вкладку «Макет». Клацають по кнопці «Лінія тренда» в блоці «Аналіз». У списку, клацають по самому останньому пункту переліку дій - «Додаткові параметри лінії тренда ...».



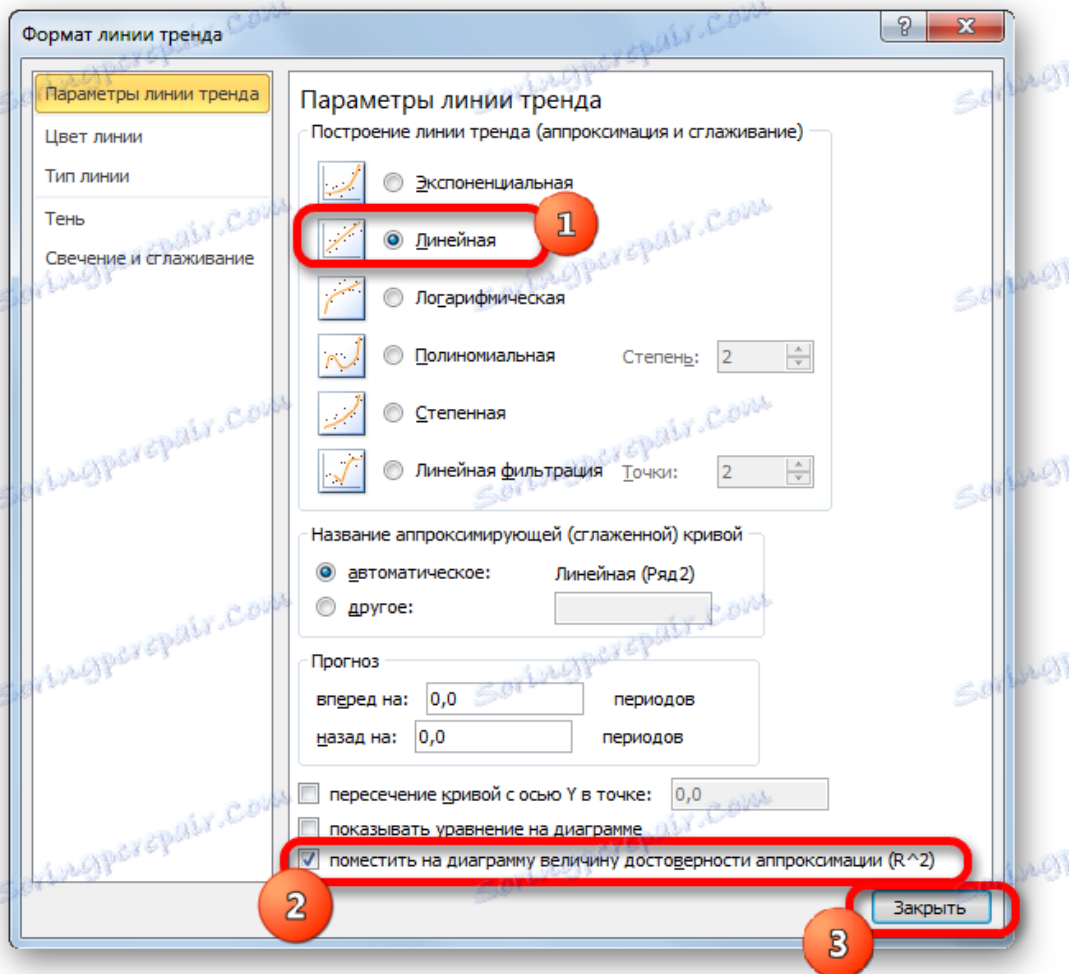
4. Після будь-якого з двох вищевказаних дій запускається вікно формату, в якому можна провести додаткові настройки. Зокрема, для виконання нашого завдання необхідно встановити прапорець навпроти пункту «Помістити на діаграму величину достовірності апроксимації (R^2)». Він розміщений в самому низу вікна. Тобто, таким чином ми включаємо відображення коефіцієнта детермінації на області побудови. Потім не забуваємо натиснути на кнопку «Закрити» внизу поточного вікна.



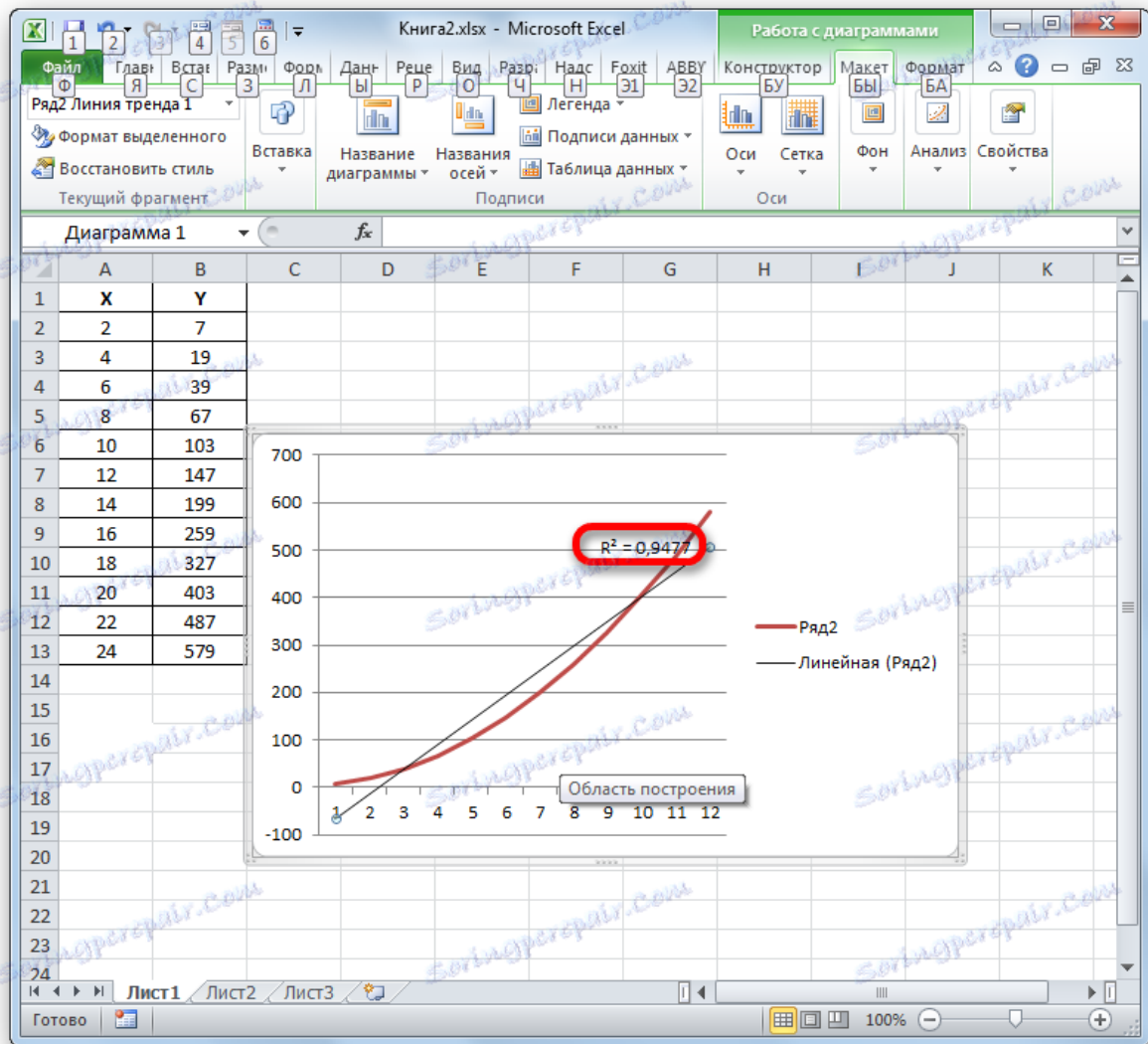
5. Значення достовірності апроксимації, тобто, величина коефіцієнта детермінації, буде відображено на аркуші в області побудови. В даному випадку ця величина, як бачимо, дорівнює 0,9242, що характеризує апроксимацію, як модель хорошої якості.



6. Абсолютно точно таким чином можна встановлювати показ коефіцієнта детермінації для будь-якого іншого типу лінії тренда. Можна міняти тип лінії тренда, зробивши перехід через кнопку на стрічці або контекстне меню у вікно її властивостей, як було показано вище. Потім вже в самому вікні в групі «Побудова лінії тренду» можна переключитися на інший тип. Не забуваємо при цьому контролювати, щоб біля пункту «Помістити на діаграму величину достовірності апроксимації» був встановлений прапорець. Завершивши вищевказані дії, клацаємо по кнопці «Закрити» в нижньому правому куті вікна.

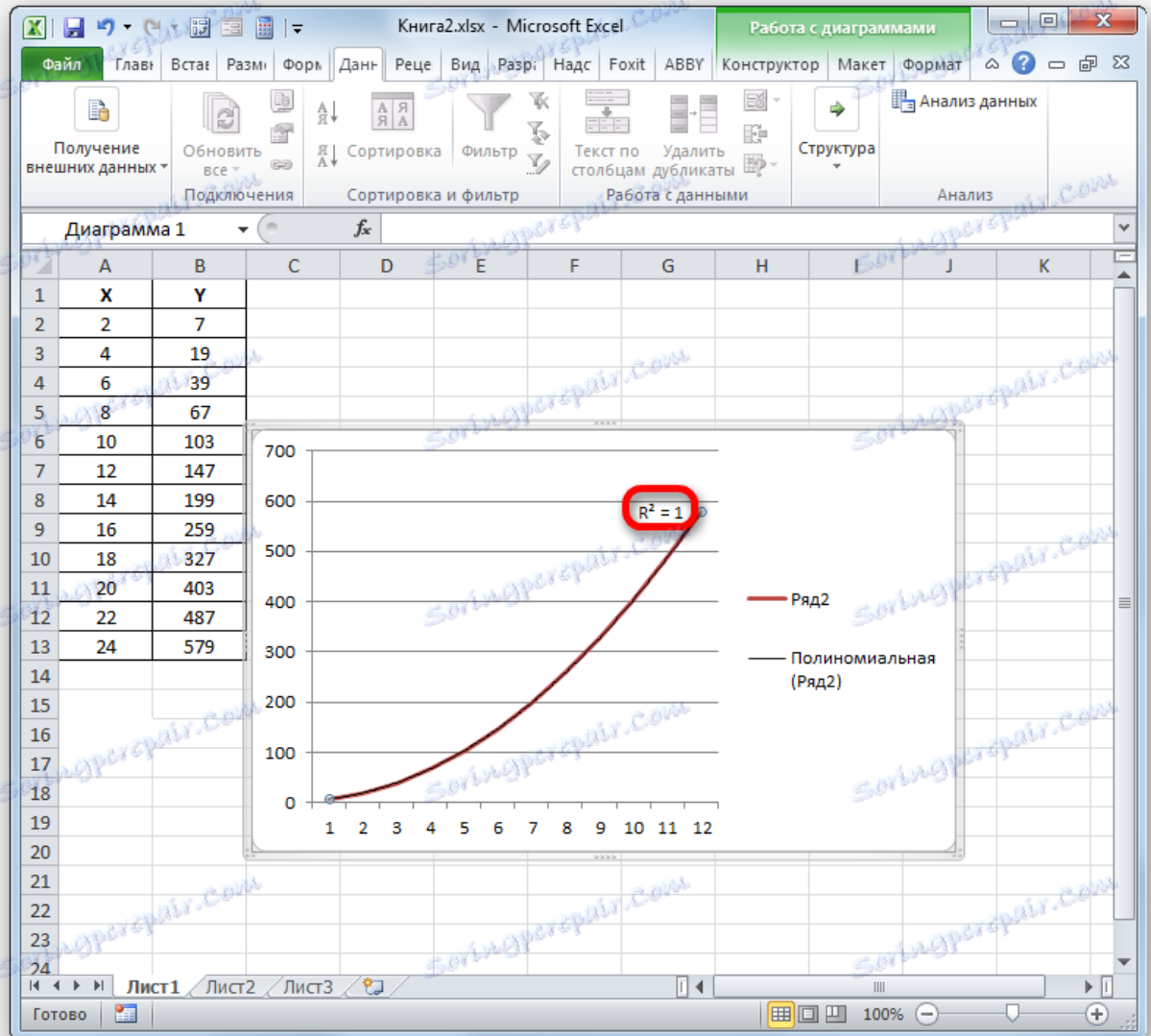


7. При лінійному типі лінія тренда вже має значення достовірності апроксимації дорівнює 0,9477, що характеризує цю модель, як ще більш достовірну, ніж розглянуту нами раніше лінію тренда експоненціального типу.



8. Таким чином, перемикаючись між різними типами лінії тренда і порівнюючи їх значення достовірності апроксимації (коефіцієнт детермінації), можна знайти той варіант, модель якого найбільш точно описує представлений графік. Варіант з найвищим показником коефіцієнта детермінації буде найбільш достовірним. На його основі можна будувати найточніший прогноз.

Наприклад, для нашого випадку дослідним шляхом вдалося встановити, що найвищий рівень достовірності має поліноміальний тип лінії тренду другого ступеня. Коефіцієнт детермінації в даному випадку дорівнює 1. Це говорить про те, що зазначена модель абсолютно достовірна, що означає повне виключення похибок.



Але, в той же час, це зовсім не означає, що для іншого графіка теж найбільш достовірним виявиться саме цей тип лінії тренда. Оптимальний вибір типу лінії тренду залежить від типу функції, на підставі якої був побудований графік. Якщо користувач не володіє достатнім обсягом знань, щоб «на око» прикинути найбільш якісний варіант, то єдиним виходом визначення кращого прогнозу є якраз порівняння коефіцієнтів детермінації.

У Ексель існують два основні варіанти обчислення коефіцієнта детермінації: використання оператора *КВПІРСОН* і застосування інструменту «Регресія» з пакету інструментів «Аналіз даних». При цьому перший з цих варіантів призначений для використання тільки в процесі обробки лінійної функції, а інший варіант можна використовувати практично у всіх ситуаціях. Крім того, існує можливість відображення коефіцієнта детермінації для лінії трендів графіків як величини достовірності апроксимації. За допомогою даного показника є можливість визначити тип лінії тренду, який має в своєму розпорядженні найвищим рівнем достовірності для конкретної функції.