

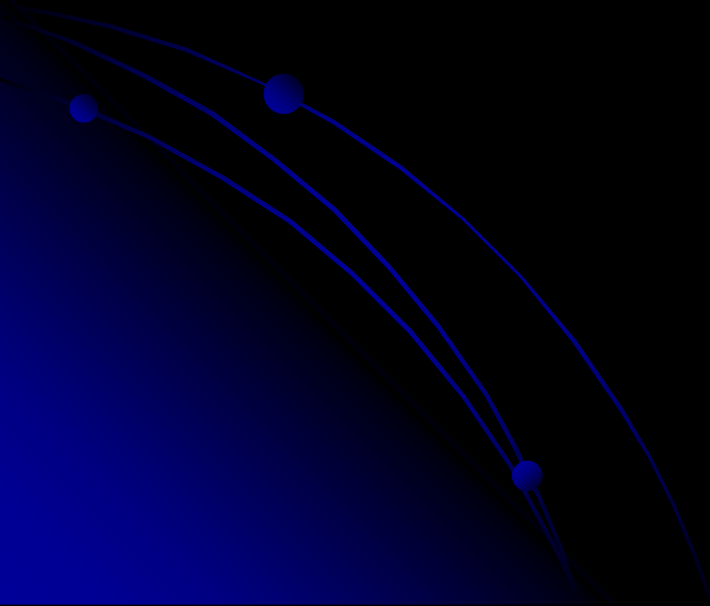
Лекція 11. Аналіз даних в Excel.

Обробка експериментальних даних

- **1.** Метод найменших квадратів та його реалізація в Excel
- **2.** Методи лінійного та експоненціального згладжування
- **3.** Інструментарій статистичного аналізу даних в Excel

Навчальна література:

1. Ларин Рональд У. Инженерные расчёты в Excel: Пер. с англ. – М.: Издательский дом «Вильямс», 2002.-544с.:ил.
2. Саймон Джин жер. Анализ данных в Excel. Пер. с англ. – М.: Издательский дом «Вильямс», 2004. – 528 с.:ил.

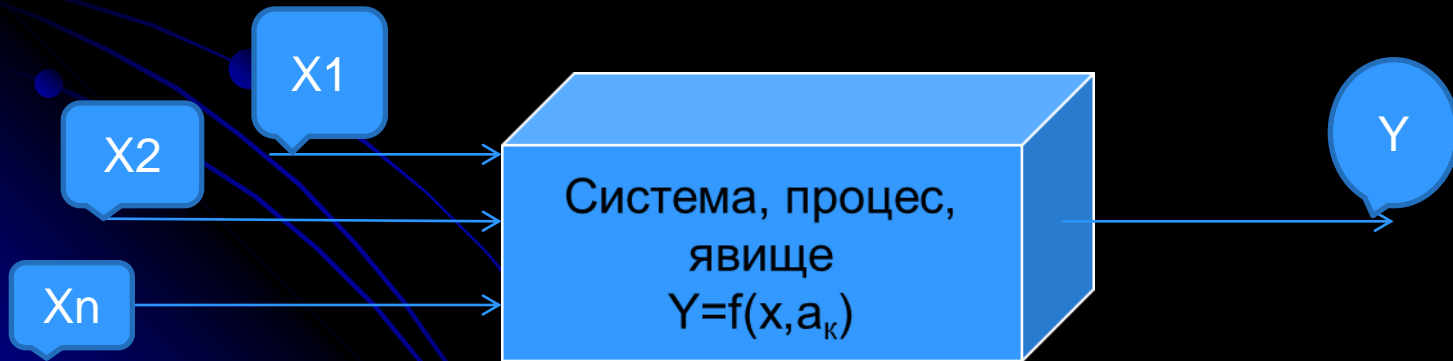
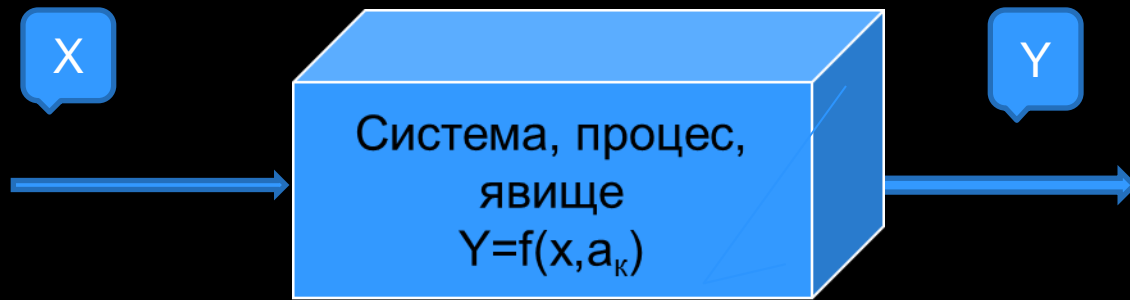


Обробка експериментальних даних (апроксимація даних) в залежності від мети передбачає вирішування наступних задач:

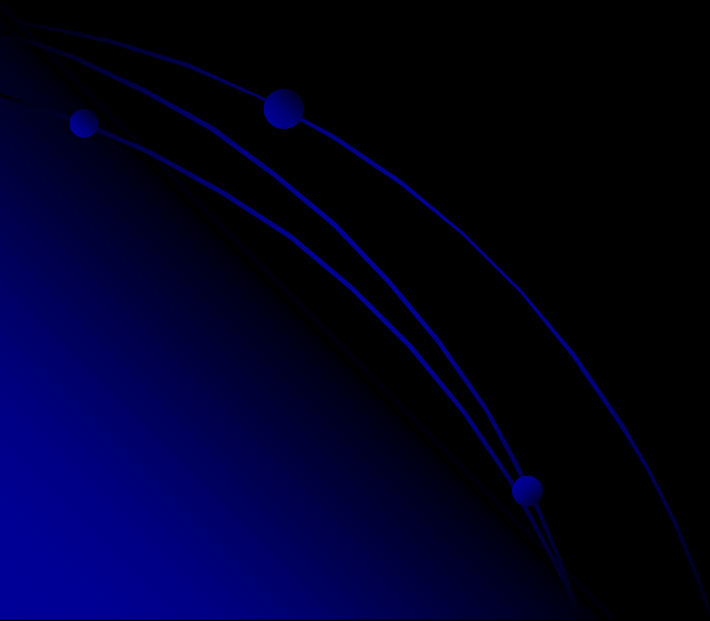
- задача інтерполяції – побудова безперервної функції $f(x_i)$, що з'єднує всі експериментальні точки $y_i(x_i)$;
- задача екстраполяції – побудова функції $f(x_{i+k})$ за границями відомого інтервалу значень – прогноз;
- задача регресії – побудова наближеної (усередненої) функції $\hat{f}(x_i)$;
- задача фільтрації – побудова апроксимуючої функції $\tilde{f}(x_i)$ для зниження систематичної похибки експериментальних даних.

Задачі 3 і 4 іноді називають згладжуванням даних.

1. Метод найменших квадратів та його релізація в Excel



2. Методи лінійного та експоненціального згладжування



Анализ данных

Инструменты анализа

Гистограмма
Скользящее среднее
Генерация случайных чисел
Ранг и перцентиль
Регрессия
Выборка
Парный двухвыборочный t-тест для средних
Двухвыборочный t-тест с одинаковыми дисперсиями
Двухвыборочный t-тест с различными дисперсиями
Двухвыборочный z-тест для средних

ОК

Отмена

Справка

Экспоненциальное сглаживание

Входные данные

Входной интервал:

\$B\$1:\$B\$17

Фактор затухания:

0,7

Метки

ОК

Отмена

Справка

Параметры вывода

Выходной интервал:

\$C\$2

Новый рабочий лист:

Новая рабочая книга

Вывод графика

Стандартные погрешности

Скользящее среднее

Входные данные

Входной интервал:

\$B\$1:\$B\$17

Метки в первой строке

Интервал:

3

ОК

Отмена

Справка

Параметры вывода

Выходной интервал:

\$C\$2

Новый рабочий лист:

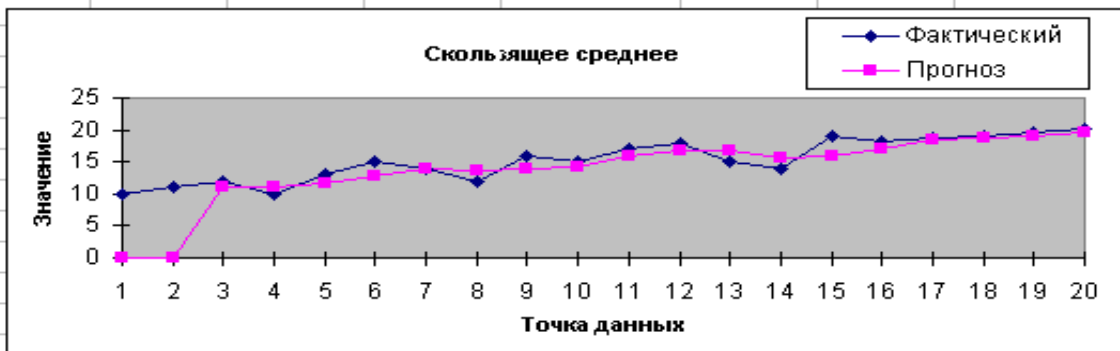
Новая рабочая книга

Вывод графика

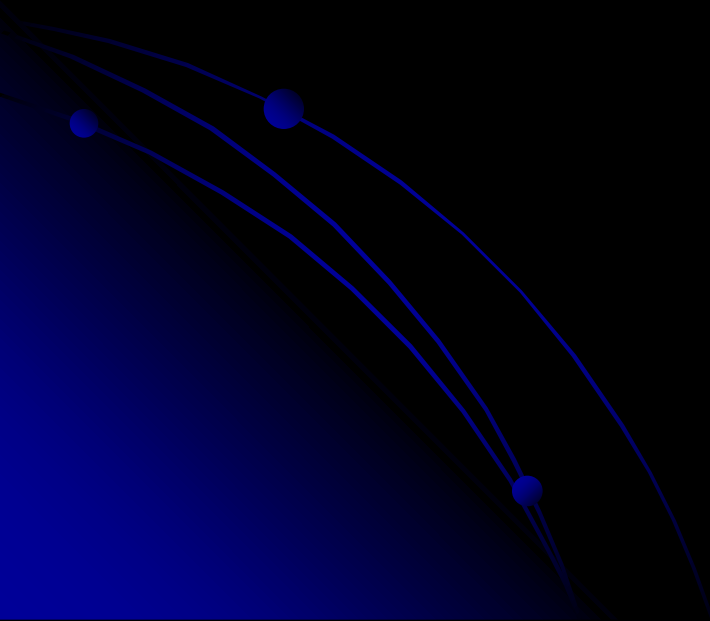
Стандартные погрешности

$$G4 = 0,3 * B3 + 0,7 * G3$$

	A	B	C	D	E	F	G	H
1	x	y		Уковз.сер.	Стандартна похибка		Уэксп.згл.	Стандартна похибка
2	1	10		#И/Д	#И/Д		#И/Д	#И/Д
3	2	11		#И/Д	#И/Д		10	#И/Д
4	3	12		11	#И/Д		10,3	#И/Д
5	4	10		11	#И/Д		10,81	#И/Д
6	5	13		11,66667	1,122167215		10,567	1,231002302
7	6	15		12,66667	1,65551827		11,2969	1,776287233
8	7	14		14	1,551582227		12,40783	2,600547289
9	8	12		13,66667	1,65551827		12,88548	2,718298728
10	9	16		14	1,503083251		12,61984	2,382717742
11	10	15		14,33333	1,551582227		13,63389	2,216948792
12	11	17		16	1,347150628		14,04372	2,16609086
13	12	18		16,66667	1,03637545		14,9306	2,709942421
14	13	15		16,66667	1,360827635		15,85142	2,583734489
15	14	14		15,66667	1,56347192		15,596	2,509030635
16	15	19		16	2,202691956		15,1172	2,056965057
17	16	18,12381		17,04127	2,077632462		16,28204	2,473073775
18	17	18,63095		18,58492	1,841558315		16,83457	2,646726481
19	18	19,1381		18,63095	0,690701326		17,37348	2,689192672
20	19	19,64524		19,1381	0,414932395		17,90287	1,801200105
21	20	20,15238		19,64524	0,507142857		18,42558	1,767927042



3. Інструментарій статистичного аналізу даних в Excel



Інструментарій програми Excel

Задачі інтерполяції, регресії, фільтрації (згладжування):

Алгоритм МНК

Функції категорії
“статистичні”

Побудова “лінії тренду”
(графічний спосіб)

“Пакет аналізу”

Інші алгоритми

ковзне середнє

експоненціальне
згладжування

Аргументы функции



ОТРЕЗОК

Известные_значения_y



= массив

Известные_значения_x



= массив

=

Возвращает отрезок, отсекаемый на оси линией линейной регрессии.

Известные_значения_y зависимое множество наблюдений или данных - числа, массивы или ссылки на ячейки, содержащие числа.

[Справка по этой функции](#)

Значение:

OK

Отмена

Функции Структура

Результат функции

Категория

Статистический

Функция

- GAMMAINV
- GAMMALN
- GAUSS
- GEOMEAN
- HARMEAN
- HYPGEOMDIST
- INTERCEPT**
- KURT
- LARGE
- LOGINV
- LOGNORMDIST
- MAX
- MAXA
- MEDIAN
- MIN
- MINA

INTERCEPT

INTERCEPT(данные_Y; данные_X)

Вычисляет отрезок, отсекаемый линией линейной регрессии.

Формула

Результат

#NULL!

Массив

Справка

Отменить

<< Назад

Далее >>

OK

Аргументы функции



НАКЛОН

Известные_значения_y



= МАССИВ

Известные_значения_x



= МАССИВ

=

Возвращает наклон линии линейной регрессии.

Известные_значения_y массив или диапазон, содержащий числовые зависимые элементы данных.

[Справка по этой функции](#)

Значение:

OK

Отмена

Функции Структура

Результат функции

Категория

Статистический

Функция

PERCENTRANK
PERMUT
PERMUTATIONA
PHI
POISSON
PROB
QUARTILE
RANK
RSQ
SKEW
SKEWP
SLOPE
SMALL
STANDARDIZE
STDEV
STDEVA

SLOPE

SLOPE(данные_Y; данные_X)

Возвращает наклон линии линейной регрессии.

Формула

Результат #NULL!

=

 Массив

Справка

Отменить

<< Назад

Далее >>

OK

Аргументы функции



КВПИРСОН

Известные_значения_y



= массив

Известные_значения_x



= массив

=

Возвращает квадрат коэффициента корреляции Пирсона по данным точкам.

Известные_значения_y массив или диапазон, могущий включать числа или имена, массивы или ссылки на ячейки с числами.

[Справка по этой функции](#)

Значение:

OK

Отмена

Мастер функций



Функции Структура

Результат функции

Категория

Статистический

Функция

- PERCENTILE
- PERCENTRANK
- PERMUT
- PERMUTATIONA
- PHI
- POISSON
- PROB
- QUARTILE
- RANK
- RSQ**
- SKEW
- SKEWP
- SLOPE
- SMALL
- STANDARDIZE
- STDEV

RSQ

RSQ(данные_Y; данные_X)

Возвращает квадрат коэффициента корреляции смещённого момента Пирсона.

Формула

Результат #NULL!

=

Массив

Справка Отменить << Назад Далее >> ОК

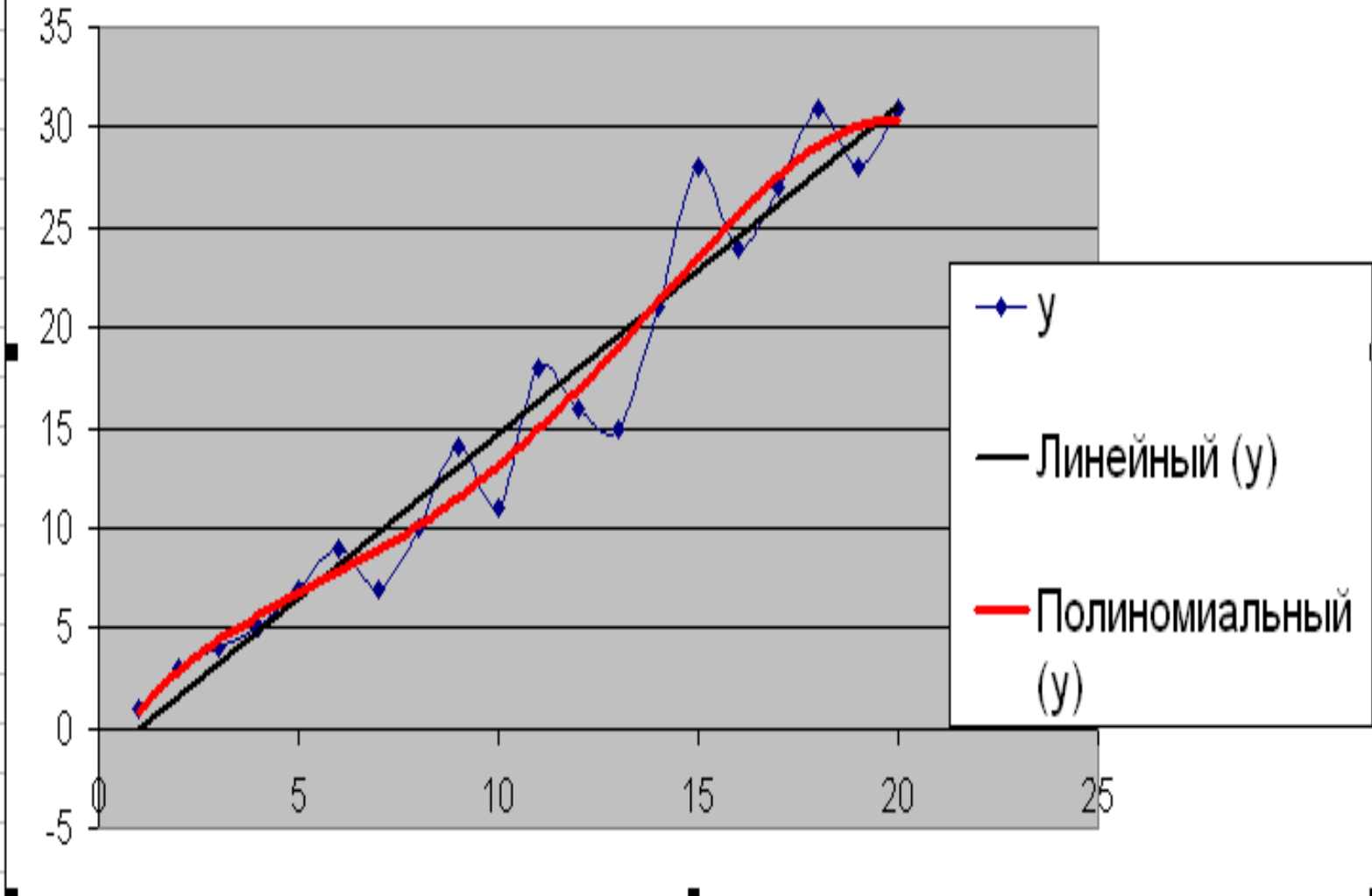
x	y
1	1
2	3
3	4
4	5
5	7
6	9
7	7
8	10
9	14
10	11
11	18
12	16
13	15
14	21
15	28
16	24
17	27
18	31
19	28
20	31

$$y = -0,0009x^4 + 0,0351x^3 - 0,4081x^2 + 2,9571x - 1,7168$$

$$R^2 = 0,9603$$

$$y = 1,6391x - 1,7105$$

$$R^2 = 0,9488$$



Аргументы функции



ЛИНЕЙН

Известные_значения_y



= ссылка

Известные_значения_x



= ссылка

Конст



= логическое

Статистика



= логическое

=

Возвращает параметры линейного приближения по методу наименьших квадратов.

Статистика логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии (ИСТИНА) или только коэффициенты m и константу b (ЛОЖЬ или отсутствие значения).

Значение:

[Справка по этой функции](#)

OK

Отмена

Функции Структура

Категория

Массив

Функция

- FREQUENCY
- GROWTH
- LINEST**
- LOGEST
- MDETERM
- MINVERSE
- MMULT
- MUNIT
- SUMPRODUCT
- SUMX2MY2
- SUMX2PY2
- SUMXMY2
- TRANSPOSE
- TREND

LINEST

Результат функции Ошибка:511

Вычисляет параметры линейной регрессии в виде массива.

Тип_линии (дополнительно)

Если тип = 0, то линии проходят через начало координат, иначе - смещаются.

данные_Y

fx



данные_X

fx



Тип_линии

fx



статистика

fx



Формула

Результат Ошибка:511

=LINEST(;;)

Массив

Справка

Отменить

<< Назад

Далее >>

OK

Аргументы функции



ЛГРФПРИБЛ

Известные_значения_y



= ссылка

Известные_значения_x



= ссылка

Конст



= логическое

Статистика



= логическое

=

Возвращает параметры экспоненциального приближения.

Известные_значения_y множество значений y , которые уже известны для соотношения $y=b^*m^x$.

Значение:

[Справка по этой функции](#)

OK

Отмена

Функции Структура

Категория

Массив

Функция

- FREQUENCY
- GROWTH
- LINEST
- LOGEST**
- MDETERM
- MINVERSE
- MMULT
- MUNIT
- SUMPRODUCT
- SUMX2MY2
- SUMX2PY2
- SUMXMY2
- TRANSPOSE
- TREND

LOGEST

Результат функции Ошибка:511

Вычисляет параметры кривой экспоненциальной регрессии ($y=b*m^x$) в виде массива.

Тип_функции (дополнительно)

Если тип = 0, то функции вычисляются по формуле $y=m^x$, при наличии значения - по формуле $y=b*m^x$.

данные_Y

данные_X

Тип_функции

статистика

Формула

Результат Ошибка:511

=LOGEST(;;)

Массив

Справка

Отменить

<< Назад

Далее >>

OK

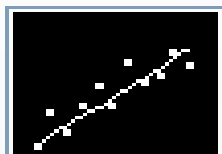
Линия тренда



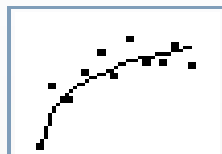
Тип

Параметры

Построение линии тренда (аппроксимация и сглаживание)



Линейная



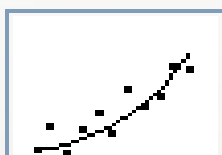
Логарифмическая



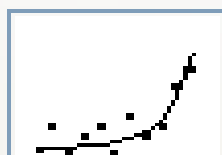
Полиномиальная

Степень:

2



Степенная



Экспоненциальная



Линейная фильтрация

Точки:

2



Построен на ряде:

y

OK

Отмена

Линия тренда



Тип

Параметры

Название аппроксимирующей (сглаженной) кривой

автоматическое: Линейный (y)

другое:

Прогноз

вперед на: единиц

назад на: единиц

пересечение кривой с осью Y в точке:

показывать уравнение на диаграмме

поместить на диаграмму величину достоверности аппроксимации (R^2)

OK

Отмена

Линия тренда для рядов данных «у»



Тип **Линии**

Тип регрессии



Линейная



Логарифмическая



Экспоненциальная



Степенная

Уравнение

Показать уравнение

Показать коэффициент детерминации (R^2)

ОК

Отменить

Справка

Восстановить

Регрессия



Входные данные

Входной интервал Y:



Входной интервал X:



Метки

Константа - ноль

Уровень надежности:

95

%

ОК

Отмена

Справка

Параметры вывода

Выходной интервал:



Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки

График остатков

Стандартизованные остатки

График подбора

Нормальная вероятность

График нормальной вероятности

Задача екстраполяції (прогнозування)

Алгоритм МНК

Функції категорії “статистичні”
ТЕНДЕНЦІЯ, ПРЕДСКАЗ, РОСТ

Побудова “лінії тренду” (графічний спосіб)

“Пакет аналізу”

“Маркер заповнення” (ЛКМ) – лінійне наближення

“Маркер заповнення” (ПКМ) – лінійне наближення або експоненціальне згладжування

- Функції «Тенденція» та «Предказ» апроксимують прямою лінією, а функція «Рост» експонентою (за методом найменших квадратів)

$$\begin{aligned} \text{"Тенденція"} &\Rightarrow \bar{y} = a_0 + a_1^* \bar{x}; \\ \text{"Рост"} &\Rightarrow \bar{y} = a_0^* \exp(a_1 \bar{x}). \end{aligned}$$

- Параметри функцій однакові (для обох функцій):
ТЕНДЕНЦІЯ (значення Y ; старі значення X ; нові значення X для прогнозу; логічна константа: якщо «1» - a_0 обчислюється, якщо «0» - $a_0 = 0$).

Аргументы функции



ТЕНДЕНЦИЯ

Известные_значения_y

= ссылка

Известные_значения_x

= ссылка

Новые_значения_x

= ссылка

Конст

= логическое

=

Возвращает значения в соответствии с линейной аппроксимацией по методу наименьших квадратов.

Известные_значения_y множество значений y , для которых уже известно соотношение $y = mx + b$.

Значение:

[Справка по этой функции](#)

OK

Отмена



Функции Структура

Категория

Массив

Функция

- FREQUENCY
- GROWTH
- LINEST
- LOGEST
- MDETERM
- MINVERSE
- MMULT
- MUNIT
- SUMPRODUCT
- SUMX2MY2
- SUMX2PY2
- SUMXMY2
- TRANSPOSE
- TREND**

TREND

Результат функции Ошибка:511

Вычисляет точки вдоль линии регрессии.

Тип_линии (дополнительно)

Если тип = 0, то линии проходят через начало координат, иначе - смещаются.

данные_Y *fx*

данные_X *fx*

новые_данные_X *fx*

Тип_линии *fx*

Формула

Результат Ошибка:511

=TREND(;;:;:)

Массив

Справка

Отменить

<< Назад

Далее >>

OK

Аргументы функции



РОСТ

Известные_значения_y



= ссылка

Известные_значения_x



= ссылка

Новые_значения_x



= ссылка

Конст



= логическое

=

Возвращает значения в соответствии с экспоненциальным трендом.

Известные_значения_y множество значений y, которые уже известны для соотношения $y=b*m^x$, массив или диапазон положительных чисел.

Значение:

[Справка по этой функции](#)

OK

Отмена



Функции Структура

Категория

Массив

Функция

- FREQUENCY
- GROWTH**
- LINEST
- LOGEST
- MDETERM
- MINVERSE
- MMULT
- MUNIT
- SUMPRODUCT
- SUMX2MY2
- SUMX2PY2
- SUMXMY2
- TRANSPOSE
- TREND

GROWTH

Результат функции Ошибка:511

Вычисляет точки экспоненциальной регрессии.

данные_Y (обязательно)

Массив данных Y.

данные_Y f_x

данные_X f_x

новые_данные_X f_x

Тип_функции f_x

Формула

Результат Ошибка:511

=GROWTH()

Массив

Справка

Отменить

<< Назад

Далее >>

OK

Аргументы функции



ПРЕДСКАЗ

X



= число

Известные_значения_y



= массив

Известные_значения_x



= массив

=

Возвращает значение линейного тренда, значение проекции по линейному приближению.

X элемент данных, для которого предсказывается значение.

Значение:

[Справка по этой функции](#)

OK

Отмена

Функции Структура

Категория

Статистический

Функция

- CONFIDENCE
- CORREL
- COUNT
- COUNTA
- COVAR
- CRITBINOM
- DEVSQ
- EXPONDIST
- FDIST
- FINV
- FISHER
- FISHERINV
- FORECAST**
- FTEST
- GAMMA
- GAMMADIST

FORECAST

Результат функции Ошибка:511

Возвращается значение по линейной регрессии

Значение (обязательно)

Значение X, для которого предсказывается значение Y по линейной регрессии.

Значение	<i>f_x</i>	<input type="text"/>	
данные_Y	<i>f_x</i>	<input type="text"/>	
данные_X	<i>f_x</i>	<input type="text"/>	

Формула

Результат Ошибка:511

=FORECAST()

Массив

Справка

Отменить

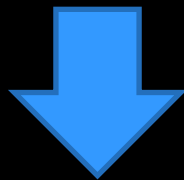
<< Назад

Далее >>

OK

Висновки:

1. *Обробка експериментальних даних в залежності від мети передбачає вирішування задач інтерполяції, екстраполяції та згладжування.*
2. *Інструментарій апроксимації даних програми Excel розроблений на основі метода найменших квадратів.*
3. *Для розв'язування однофакторної задачі інтерполяції (згладжування) $y=f(x)$ застосовують: функції категорії статистичні (отрезок, наклон), графічний алгоритм побудови лінії тренду, пакет аналізу.*
4. *Для багатofакторної задачі $y=f(x_1, x_2, \dots, x_n)$ використовують функції категорії статистичні (линейн, лгрфприбл) та пакет аналізу.*
5. *Екстраполяцію даних можна виконати за допомогою функцій (тенденция, предсказ, рост), засобів автоматизації розрахунків (маркер автозаповнення, контекстне меню), комбінацією інструментів регресії, інтерполяції та згладжування.*



Висновок:

6. В Excel прогноз (екстраполяція) ґрунтується тільки на статистичних методах. Спочатку вивчається стратегія процесу за минулий період, а далі на її основі будується прогноз. Цю стратегію ще називають базовою лінією даних.

- починати побудову базової лінії потрібно з результатів самих ранніх спостережень і закінчувати останніми;
- всі часові періоди спостережень повинні бути однакові. Не слід змішувати дані за різні періоди. Наприклад, дані за день з середніми показниками за декілька днів.
- спостереження повинні фіксуватися в один і той же момент кожного часового періоду. Наприклад, при складанні базової лінії на основі добових (щоденних) спостережень фіксувати дані в один і той же час;
- пропущення даних не допускається. Тому, якщо відсутні дані спостережень за незначний період, доцільно буде їх доповнити наближеними даними.