**Раздел III. Статистические методы исследований**

**Тема 1. Понятие выборочного исследования**

**Практическая работа по теме «Создание линий тренда»**

* 1. **Основные понятия**

Обычно данные, отображаемые на диаграммах, являются наблюдаемыми или получены в результате эксперимента. Их можно аппроксимировать, т.е. получить уравнение, описывающее представленную исходную зависимость, и оценить степень приближения. Простейшим способом аппроксимации в Excel являются линии тренда. Линии тренда позволяют также прогнозировать данные.

Для создания линии тренда необходимо:

1. построить точечную диаграмму по табличным данным,
2. добавить на диаграмму линию тренда.

Линиями тренда можно дополнить практически все диаграммы, кроме объемных, круговых, кольцевых.

Линия тренда характеризуется следующими параметрами:

1. **уравнением** (функциональной зависимостью),
2. **величиной достоверности аппроксимации R2**. R2 [0, 1] – число, которое отражает близость значения линии тренда к фактическим данным. Чем ближе к 1 величина этого показателя, тем достовернее линия тренда.

Существует пять различных типов линий тренда (функциональных зависимостей):

1. Линейная *y=ax+b*
2. Полиномиальная *y=a0+a1x+a2x2+ +anxn*, для n 6
3. Логарифмическая *y*
4. Экспоненциальная *y*

*aln x b*

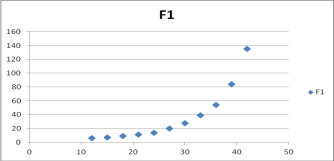
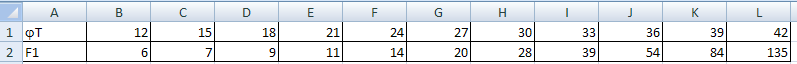
*aebx*

1. Степенная *y axb*

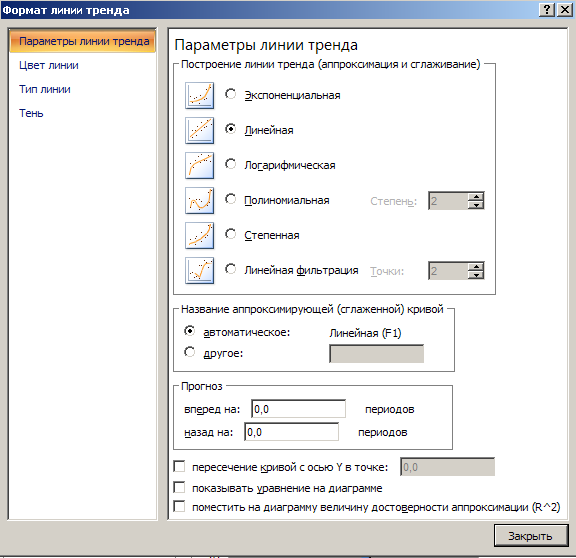
Для одних и тех же данных можно построить различные типы линий тренда, а затем выбрать наиболее подходящую функциональную зависимость, анализируя величину достоверности аппроксимации R2.

# Построение линии тренда

1. Построить точечную диаграмму для заданной таблицы данных.



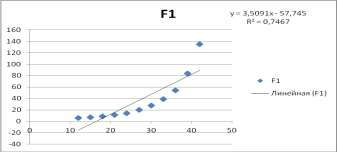
1. Выделить ряд данных и в контекстном меню выбрать команду **Добавить линию тренда**.



1. В диалоговом окне **Линия тренда** задать следующие данные:

 выбрать тип линии тренда (для полиномиального тренда задать дополнительно степень -от 2до 6);

 включить флажки для размещения на диаграмме уравнения и R2.

Ниже представлена диаграмма с добавленной линией тренда

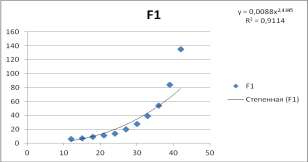
# Использование линии тренда для аппроксимации функции, заданной таблично

1. Построить по табличным данным диаграмму **точечного** типа.
2. Добавить линию тренда заданного типа или, если тип неизвестен, создать несколько трендов различных типов и выбрать тренд с максимальным R2 (рекомендуется каждый тренд создавать на отдельной диаграмме).
3. Используя уравнение выбранной линии тренда, можно получать значение функции для аргумента, не заданного в таблице.

**Пример.** Для табличных данных из предыдущего примера подобрать наиболее подходящую линию тренда. Вычислить значение функции для φ=22, φ =35, φ =40. Отобразить полученное значение графически.

Порядок выполнения:

1. По табличным данным построить точечную диаграмму, скопировать ее несколько раз и добавить е линии тренда:



140

120

F1

100

f1

80

f2

60 f3

40 Полиномиальная (F1)

20

0

0 10 20 30 40 50

y = 0,0114x3 - 0,6968x2 + 14,491x - 89,767

R² = 0,9934

160

50

40

30

20

10

F1

Логарифмическая

(F1)

160

140

120

100

80

60

40

20

0

-20 0

-40

y = 77,51ln(x) - 213,04

R² = 0,6091

**F1**

0 10 20 30 40 50

40

20

0

F1

Полиномиальная (F1)

160

140

120

100

80

60

y = 0,2227x2 - 8,5189x + 84,586

R² = 0,9579

**F1**

0 10 20 30 40 50

40

20

0

Экспоненциальная

(F1)

F1

100

80

60

y = 1,3862e0,1034x

R² = 0,9825

160

140

120

**F1**

50

40

30

20

10

F1

Линейная (F1)

160

140

120

100

80

60

40

20

0

-20 0

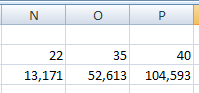
-40

y = 3,5091x - 57,745

R² = 0,7467

**F1**

1. Анализируя величину R2 для каждой линии тренда, можно сделать вывод, что наиболее подходящей аппроксимирующей функцией является полином степени 3 (R2=0,9934).
2. Для вычисления значения функции по заданному значению аргумента φ
3. в отдельную ячейку занести значения аргументов,



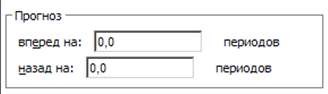
1. скопировать уравнение линии тренда (начиная со знака "=") в ячейки ниже, внося необходимые изменения (в частности, придется вставить знаки \* и ^, удалить лишний пробел в начале формулы и заменить символ х на адрес ячейки).



Для отображения точки на графике выполнить добавление ряда данных, задавая в качестве нового ряда ячейки, заполненные на предыдущем шаге. Отформатировать полученный ряд маркерами "х".

# Использование линии тренда для прогнозирования

Линии тренда позволяют графически и аналитически отображать тенденции данных и прогнозировать данные. Продлив линию тренда в диаграмме за пределы реальных данных, можно предсказать будущие значения. Для этого в диалоговом окне при задании линии тренда использовать область **Прогноз** указывая, на сколько единиц вперед или назад он выполняется.



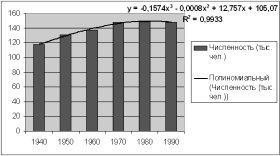
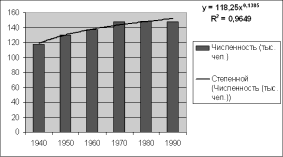
Прогнозирование данных обычно используется для решения социально-экономических задач, поэтому для отображения данных используются диаграммы – гистограммы, графики, линейчатые и др., т.е. диаграммы, по оси значений которых откладываются категории. Соответственно, в уравнении линии тренда значением аргумента является **номер** точки диаграммы (т.е. 1, 2, 3 и т.д.).

**Пример.** Имеется таблица зависимости численности населения страны от года.

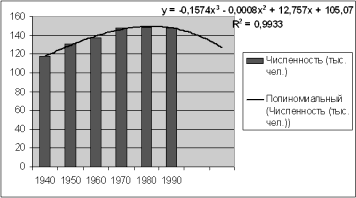
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Год** | **1940** | **1950** | **1960** | **1970** | **1980** | **1990** |
| **Численность**  **(тыс. чел.)** | 117,5 | 130,1 | 137,6 | 147,4 | 148,5 | 147,7 |

Определить функциональную зависимость численности населения от года. Определить, какая численность населения будет в 2010 году, если сохранится такая же тенденция изменения численности. Определить, какая численность населения была в 1965 году.

Порядок выполнения.

1. Как правило, численность населения характеризуется степенной или полиномиальной зависимостью. Проверим, какая из зависимостей является наиболее подходящей.

Из анализа величины R2 ясно, что наиболее подходящей является полиномиальная зависимость.

1. Для прогнозирования выполнить действия:
2. щелкнуть по линии тренда правой кнопкой мыши и выбрать вариант **Формат линии тренда,**
3. в области прогноза установить значение . В результате получим диаграмму:
4. Для вычисления численности населения в 2010 году, занесем в ячейки следующие данные:

|  |  |  |
| --- | --- | --- |
|  | **A** | **B** |
| **4** | 8 | = -0,1574\*A4^3 - 0,0008\*A4^2 + 12,757\*A4 + 105,07 |



Номер категории для 2010 года (номер 1 имеет 1940 год, 2 – 1950 год и т.д.)

В результате в ячейке **В4** получим значение **126,486**.

1. Для вычисления численности населения в 1965 году, занесем в ячейки следующие данные:

|  |  |  |
| --- | --- | --- |
|  | **A** | **B** |
| **5** | 3,5 | = -0,1574\*A5^3 - 0,0008\*A5^2 + 12,757\*A5 + 105,07 |

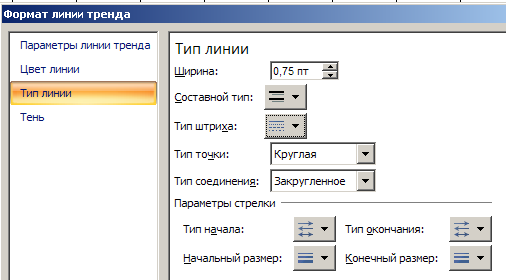
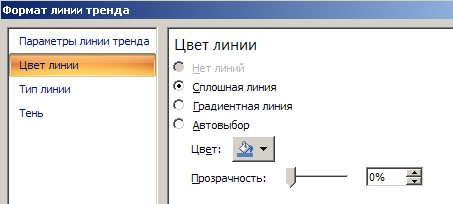
Номер категории для 1965 года (номер 1 имеет 1940 год, 2 – 1950

год, 3 – 1960, следовательно, 3,5 –1965 год).

В результате в ячейке **В4** получим значение **142,9612**.

# Форматирование и редактирование линии тренда

Форматирование линий тренда выполняется стандартными средствами форматирования диаграмм. Для редактирования линии ее необходимо выделить и использовать контекстное меню **Формат линии тренда**. Для удаления линии тренда ее необходимо выделить и нажать клавишу



**Delete**.

**ЗАДАНИЕ.** На сайте Росстата необходимо найти данные по производству продукции или услуг отрасли региона, области, края, республики (по выбору студента, можно по данным для магистерской диссертации). Построить ряд динамики. Использовать линию тренда для аппроксимации функции, выбрав наиболее подходящую модель. Построить прогноз на два шага вперед по выбранной модели.