***Практическая работа №15 «Получение регрессионных моделей». Регрессионная модель***

**Цель урока:**

1. Образовательные: освоение способов построения по экспериментальным данным регрессионной модели и тренда средствами Ms Excel.

2. Познавательные:

–– формирование умений применять имеющиеся математические знания и знания из курса информатики к решению практических задач;

– развитие внимания, познавательной активности, творческих способностей, логического мышления.

3. Воспитательные:

– воспитание интереса к предмету;

– самостоятельности в принятии решения;

**Используемые программные средства:** табличный процессор Ms Excel.

Тип урока: Комбинированный.

### План урока.

1. Организационная часть.

**2. Постановка цели урока**

Анализ данных - область информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных.

Вопросы ученикам: 1) Как Вы думаете, какое программное обеспечение имеет средства анализа данных? (табличный процессор Excel)

2) Какие именно возможности табличного процессора можно отнести к средствам анализа данных?

К средствам анализа относятся:

• Обработка списка с помощью различных формул и функций;

• Построение диаграмм и использование карт Ms Excel;

• Проверка данных рабочих листов и рабочих книг на наличие ошибок;

• Структуризация рабочих листов;

• Автоматическое подведение итогов (включая мастер частичных сумм);

• Консолидация данных;

• Сводные таблицы;

• Специальные средства анализа выборочных записей и данных - подбор параметра, поиск решения, сценарии и др.

3) В каких областях могут найти практическое применение средства анализа табличного процессора Excel?

Цель нашего урока: научиться строить регрессионные модели средствами Excel..

**3. Актуализация знаний**

В состав Microsoft Excel входит набор средств анализа данных (так называемый пакет анализа), предназначенный для решения сложных статистических и инженерных задач. Для проведения анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры; анализ будет проведен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещен в выходной диапазон. Другие средства позволяют представить результаты анализа в графическом виде.

Статистические данные приводятся в виде длинных и сложных статистических таблиц, поэтому бывает весьма трудно обнаружить в них имеющиеся неточности и ошибки.

Графическое же представление статистических данных помогает легко и быстро выявить ничем не оправданные пики и впадины, явно не соответствующие изображаемым статистическим данным, аномалии и отклонения.

### 4. Теоретическая часть. Объяснение нового материала.

Объяснение нового материала происходит с использованием презентации**.** В управлении и планировании существует целый ряд типовых задач, которые можно переложить на плечи компьютера

Ms Excel – это не просто электронная таблица с данными и формулами для вычислений. Это универсальная система обработки данных, которая может использоваться для анализа и представления данных в наглядной форме.

Мы уже с вами говорили о том, что решение задач планирования и управления постоянно требует учета зависимостей одних факторов от других. Таких примеров мы приводили очень много.

Один из таких примеров: определение зависимости время падения тела на землю от первоначальной высоты. Зависимость эта очевидна. Для её проверки можно провести эксперимент, сбрасывая предметы с разных этажей многоэтажного здания, данные занести в таблицу. Таким образом мы легко создадим табличную модель, на основе её построим график. Кроме этого нам не составит особого труда и составление функциональной зависимости, так как падение тел происходит согласно всем нам известному физическому закону. Тем самым у нас будет и математическая модель по которой мы легко рассчитаем время падения тел даже с очень большой высоты.

|  |  |
| --- | --- |
| Н (м) | t (сек) |
| 6  9  12  15  18  21  24  27  30 | 1,1  1,4  1,6  1,7  1,9  2,1  2,2  2,3  2,5 |

image294,

Но не все зависимости так просты.

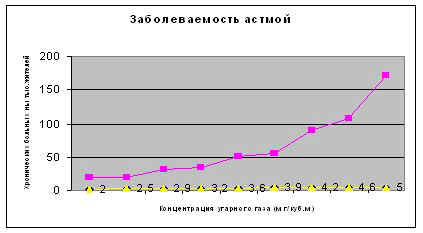
|  |  |
| --- | --- |
| С, мг/куб.м | Р, бол./тыс. |
| 2 | 19 |
| 2,5 | 20 |
| 2,9 | 32 |
| 3,2 | 34 |
| 3,6 | 51 |
| 3,9 | 55 |
| 4,2 | 90 |
| 4,6 | 108 |
| 5 | 171 |

Например, нам необходимо найти зависимость частоты заболеваемости жителей города бронхиальной астмой от качества воздуха.

Любому человеку понятно, что такая зависимость существует. Очевидно, что чем хуже воздух, тем больше больных астмой. Но это качественное заключение. Его недостаточно для того, чтобы управлять этим процессом, нам потребуются более конкретные знания. Нужно установить, какие именно примеси сильнее всего влияют на здоровье людей, как связаны концентрация этих примесей в воздухе с числом заболеваний. Такую зависимость можно установить только путем сбора многочисленных данных, их анализа и обобщения.

*В таких ситуациях на помощь приходит статистика: наука о сборе, изменении и анализе массовых количественных данных***.**

Специалисты по медицинской статистике проводят сбор данных. Они собирают сведения из разных городов о средней концентрации угарного газа в атмосфере и о заболеваемости астмой (число хронических больных на тысячу жителей). Полученные данные можно свести в таблицу, а также представить в виде точечной диаграммы .



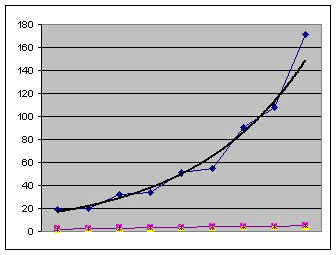
При этом необходимо помнить, что статистические данные всегда являются приближенными, усредненными. Поэтому они носят оценочный характер. Однако, они верно отражают характер зависимости величин. И еще одно важное замечание: для достоверности результатов, полученных путем анализа статистических данных, этих данных должно быть много.

Из полученных данных можно сделать вывод, что при концентрации угарного газа до 3 мг/куб.м его влияние на заболеваемость астмой несильное. С дальнейшим ростом концентрации наступает резкий рост заболеваемости.

Построить табличную модель и графическую по экспериментальным данным

Но нужно ещё и **получить формулу, отражающую эту зависимость. На языке математики это называется функцией зависимости Р от С: Р(С). Вид такой функции неизвестен, её следует искать методом подбора по экспериментальным данным.** Понятно, что график искомой функции должен проходить близко к точкам диаграммы экспериментальных данных. Строить функцию так, чтобы ёе график точно проходил через все данные точки (рисунок 2), не имеет смысла. Во-первых, математический вид такой функции может оказаться слишком сложным. Во-вторых, уже говорилось о том, что экспериментальные значения являются приближенными.

Отсюда следуют основные требования к искомой функции:

* она должна быть достаточно простой для использования её в дальнейших вычислениях;
* график этой функции должен проходить вблизи экспериментальных точек так, чтобы отклонения этих точек от графика были минимальны и равномерны

Полученную функцию, график которой приведен на рисунке, принято называть в статистике регрессионной моделью. ***Регрессионная модель – это функция, описывающая зависимость между количественными характеристиками сложных систем.***

Получение регрессионной модели происходит в два этапа:

1. подбор вида функции;
2. вычисление параметров функции.

Чаще всего выбор производится среди следующих функций:

**y=ax+b – линейная функция;**

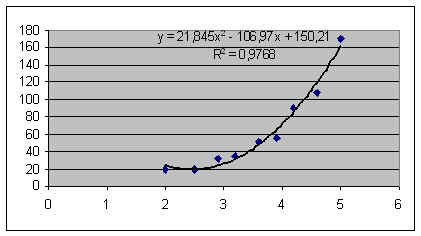
**y=ax2+bx+c – квадратичная функция;**

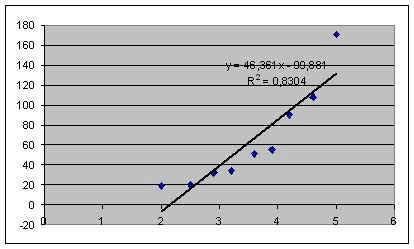
**y=aln(x)+b – логарифмическая функция;**

**y=aebx - экспоненциальная функция;**

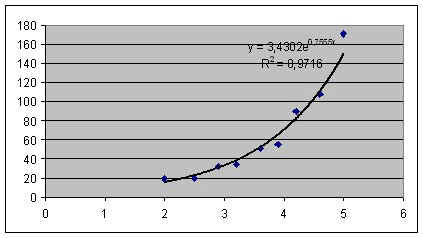
**y=axb - степенная функция.**

Если Вы выбрали (сознательно или наугад) одну из предлагаемых функций, то следующим шагом нужно подобрать параметры (a,b,c и пр.) так, чтобы функция располагалась как можно ближе к экспериментальным точкам. Для этого подходит метод наименьших квадратов (МНК). Суть его заключается в следующем: искомая функция должна быть построена так, чтобы сумма квадратов отклонений у – координат всех экспериментальных точек от у – координат графика функции была бы минимальной.

Важно понимать следующее: методом наименьших квадратов по данному набору экспериментальных точек можно построить любую функцию. А вот будет ли она нас удовлетворять, это уже другой вопрос – вопрос критерия соответствия. На рисунке 4 изображены 3 функции, построенные методом наименьших квадратов.

    Рисунок 4

Данные рисунки получены с помощью Ms Excel. График регрессионной модели называется *трендом* (trend – направление, тенденция).

График линейной функции – это прямая. Полученная по методу МНК прямая отражает факт роста заболеваемости от концентрации угарного газа, но по этому графику трудно что – либо сказать о характере этого роста. А вот квадратичный и экспоненциальный тренды – ведут себя очень правдоподобно.

На графиках присутствует ещё одна величина, полученная в результате построения трендов. Она обозначена как R2. **В статистике эта величина называется *коэффициентом детерминированности.* Именно она определяет, насколько удачной получится регрессионная модель**. Коэффициент детерминированности всегда заключен в диапазоне от 0 до 1. Если он равен 1, то функция точно проходит через табличные значения, если 0, то выбранный вид регрессионной модели неудачен. Чем R2 ближе к 1, тем удачнее регрессионная модель.

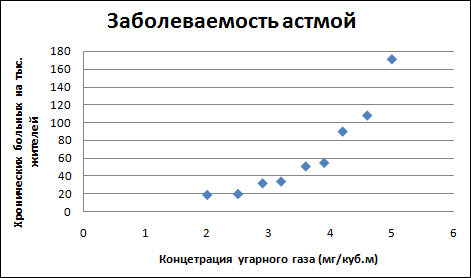
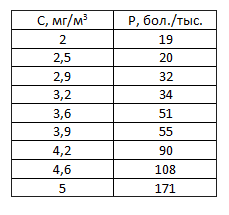
Метод наименьших квадратов используется для вычисления параметров регрессионной модели. Этот метод содержится в математическом арсенале электронных таблиц.

### 5. Практическая часть.

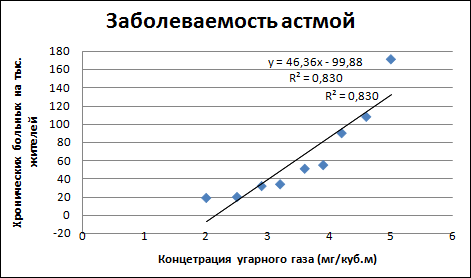
Выполнение лабораторной работы. По предложенной инструкции выполнить практическую работу, оформить отчет

**Задание 1**

   1. Ввести табличные данные зависимости заболеваемости бронхиальной астмой от концентрации угарного газа в атмосфере (см. рисунок).  
   2. Представить зависимость в виде точечной диаграммы (см. рисунок).

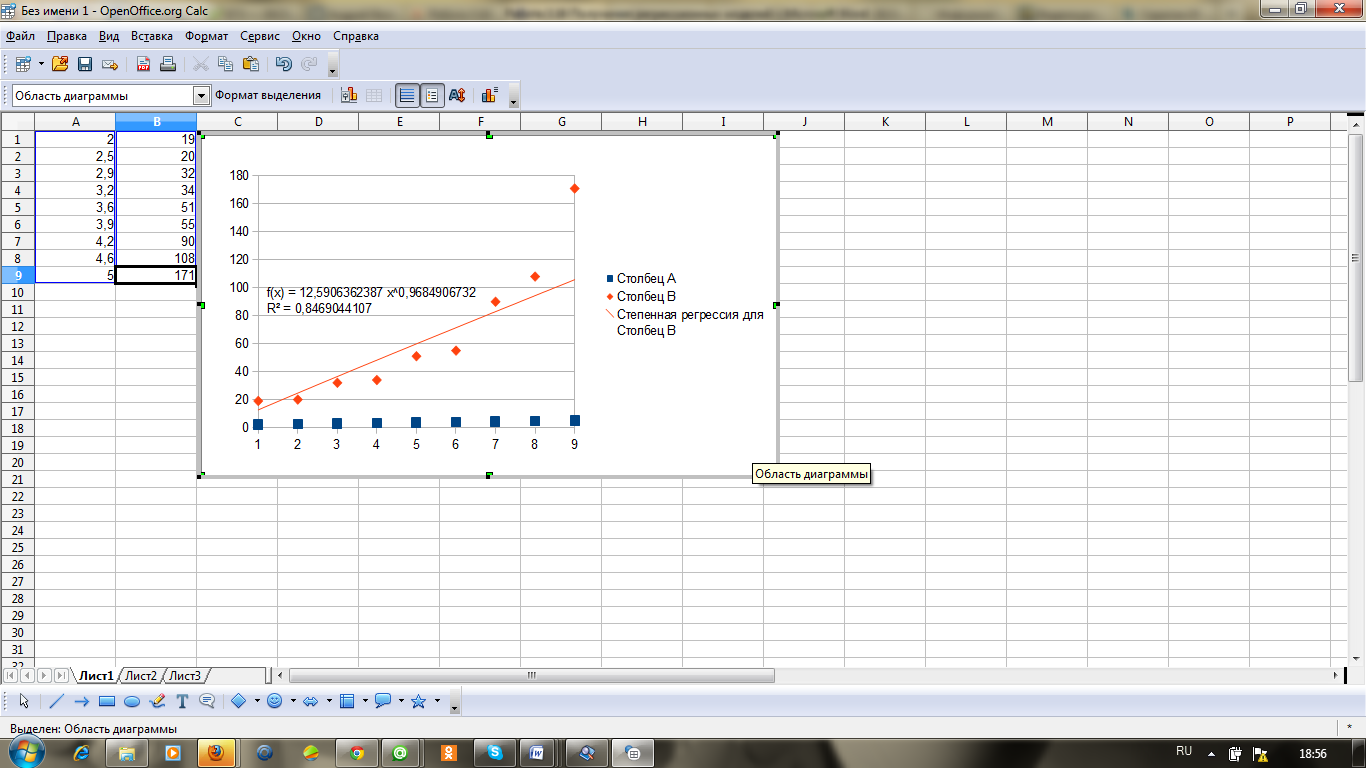
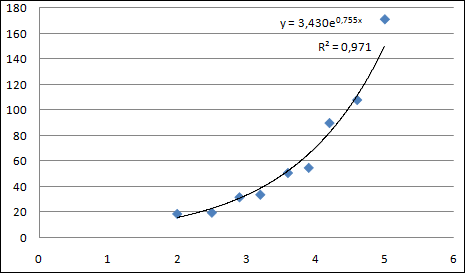


**Задание 2**

   Требуется получить три варианта регрессионных моделей (три графических тренда) зависимости заболеваемости бронхиальной астмой от концентрации угарного газа в атмосфере.

   1. Для получения линейного тренда выполнить следующий алгоритм:  
     => щелкнуть ПКМ на поле диаграммы «Заболеваемость астмой», построенной в предыдущем задании;  
     => выполнить команду **Вставить линию тренда**;  
     => в открывшемся окне на вкладке **Тип** выбрать **Линейный тренд**;  
     => установить галочки на флажках: п**оказывать уравнения на диаграмме и поместить на диаграмму величину достоверности аппроксимации R^2**;  
     => щелкнуть на кнопке **ОК**. Полученная диаграмма представлена на рисунке:

   2. Получить экспоненциальный тренд. Алгоритм аналогичен предыдущему. На закладке **Тип** выбрать **Экспоненциальный тренд**. Результат представлен на рисунке:



   3. Получить степенной тренд. Алгоритм аналогичен предыдущему. На закладке **Тип** выбрать **Степенной тренд**. Результат представлен на рисунке:

### 6. Домашнее задание:

Конспект (тема “Регрессионная модель”); заполнить лист Отчета полностью (ответить на вопросы).

### 7. Подведение итогов. Рефлексия

**Отчет по практической работе**

**«Построение регрессионных моделей с помощью** табличный процессор Ms Excel**»**

Практическую работу выполнял:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |
| --- | --- | --- |
| Тип тренда | Уравнение тренда | R2 |
| Линейный |  |  |
| Квадратичный |  |  |
| Логарифмический |  |  |
| Степенной |  |  |
| Экспоненциальный |  |  |
| Полином третей степени |  |  |

**Вывод**: для графика, полученного по экспериментальным точкам больше всего подходит регрессионная модель, построенная с помощью \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ёе формула имеет вид \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

R2 равен \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

По полученной формуле рассчитайте предполагаемую на 15 число.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Вывод по работе: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_