

## КУРСОВАЯ РАБОТА

По дисциплине

ИНФОРМАТИКА

(наименование учебной дисциплины согласно учебному плану)

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Тема: Аппроксимация функций методом наименьших квадратов

Автор: студент гр. ИЗ-99-1

/ \_\_\_\_\_ /  
(подпись)

Брук Б.М.  
(Ф.И.О.)

**ОЦЕНКА:** \_\_\_\_\_

Дата: \_\_\_\_\_

### ПРОВЕРИЛ

Руководитель проекта ст. преподаватель  
(должность)

/ \_\_\_\_\_ /  
(подпись) (Ф.И.О.)

Быкова Е.В.

УТВЕРЖДАЮ  
Заведующий кафедрой  
/\_\_\_\_\_/

доц. /Прудинский Г.А./

"\_\_" \_\_\_\_\_ 2000 г.

Кафедра Информатики и компьютерных технологий

## КУРСОВАЯ РАБОТА

По дисциплине

ИНФОРМАТИКА

(наименование учебной дисциплины согласно учебному плану)

## ЗАДАНИЕ

Студенту группы

ИЗ-99-1  
(шифр группы)

Брук Б.М.  
(Ф.И.О.)

1. Тема проекта: Использование информационных технологий для решения прикладных задач на примере построения аппроксимации функции методом наименьших квадратов.
2. Исходные данные к проекту: Вариант №22, Задана таблица значений двух наблюдаемых переменных «X» и «Y».
3. Содержание пояснительной записки: Пояснительная записка включает в себя задание на выполнение работы, титульный лист, аннотацию, оглавление, введение, собственно текст пояснительной записки, выводы, библиографический список.
4. Перечень графического материала: Представление результатов в виде графиков, блок-схема.
5. Срок сдачи законченного проекта: 1.12.00

Руководитель проекта ст. преподаватель  
(должность)

/\_\_\_\_\_/

(подпись)

Быкова Е.В.  
(Ф.И.О.)

Дата выдачи задания: 7.09.00

Санкт-Петербург  
2000 год

### **Аннотация.**

Пояснительная записка представляет собой отчет о выполнении курсовой работы. В ней рассматриваются вопросы по получению эмпирических формул методом наименьших квадратов (МНК). Расчеты проведены средствами пакета Microsoft Excel, в Turbo Pascal 7.0.

Страниц - 32, таблиц - 8, рис.5.

### **The Summary**

The explanatory note presents a report: in which we discuss questions of the construction of the empirical formulas using method of the least squares in Microsoft Excel. Also this task is presented in Turbo Pascal 7.0.

Pages - 32, tables - 8, pic.5.

## Оглавление.

Введение.....	4
1. Постановка задачи.....	6
2. Расчетные формулы.....	7
2.1 Построение эмпирических формул методом наименьших квадратов.....	7
2.2 Линеаризация экспоненциальной зависимости.....	9
2.3 Элементы теории корреляции.....	10
3. Расчет коэффициентов аппроксимации в Microsoft Excel.....	13
4. Построение графиков в Excel и использование функции ЛИНЕЙН.....	21
5. Программа на языке Pascal.....	24
5.1. Блок-схема.....	24
5.2. Результаты расчета Pascal.....	29
Заключение.....	30
Список литературы.....	31

## Введение.

Аппроксимация (от латинского "approximate" - "приближаться")- приближенное выражение каких-либо математических объектов (например, чисел или функций) через другие более простые, более удобные в пользовании или просто более известные. В научных исследованиях аппроксимация применяется для описания, анализа, обобщения и дальнейшего использования эмпирических результатов.

Как известно, между величинами может существовать точная (функциональная) связь, когда одному значению аргумента соответствует одно определенное значение, и менее точная (корреляционная) связь, когда одному конкретному значению аргумента соответствует приближенное значение или некоторое множество значений функции, в той или иной степени близких друг к другу. При ведении научных исследований, обработке результатов наблюдения или эксперимента обычно приходится сталкиваться со вторым вариантом. При изучении количественных зависимостей различных показателей, значения которых определяются эмпирически, как правило, имеется некоторая их вариабельность. Частично она задается неоднородностью самих изучаемых объектов неживой и, особенно, живой природы, частично обуславливается погрешностью наблюдения и количественной обработке материалов. Последнюю составляющую не всегда удастся исключить полностью, можно лишь минимизировать ее тщательным выбором адекватного метода исследования и аккуратностью работы. Поэтому при выполнении любой научно-исследовательской работы возникает проблема выявления подлинного характера зависимости изучаемых показателей, этой или иной степени замаскированных неучтенностью вариабельности значений. Для этого и применяется аппроксимация - приближенное описание корреляционной зависимости переменных подходящим уравнением функциональной зависимости, передающим основную тенденцию зависимости (или ее "тренд").

При выборе аппроксимации следует исходить из конкретной задачи исследования. Обычно, чем более простое уравнение используется для аппроксимации, тем более приблизительно получаемое описание зависимости. Поэтому важно считывать, насколько существенны и чем обусловлены отклонения конкретных значений от получаемого тренда. При описании зависимости эмпирически определенных значений можно добиться и гораздо большей точности, используя какое-либо более сложное, много параметрическое уравнение. Однако нет никакого смысла стремиться с максимальной точностью передать случайные

отклонения величин в конкретных рядах эмпирических данных. Гораздо важнее уловить общую закономерность, которая в данном случае наиболее логично и с приемлемой точностью выражается именно двухпараметрическим уравнением степенной функции. Таким образом, выбирая метод аппроксимации, исследователь всегда идет на компромисс: решает, в какой степени в данном случае целесообразно и уместно «пожертвовать» деталями и, соответственно, насколько обобщенно следует выразить зависимость сопоставляемых переменных. Наряду с выявлением закономерностей, замаскированных случайными отклонениями эмпирических данных от общей закономерности, аппроксимация позволяет также решать много других важных задач: формализовать найденную зависимость; найти неизвестные значения зависимой переменной путем интерполяции или, если это допустимо, экстраполяции.

## 1. Постановка задачи.

Во всех вариантах требуется:

1. Используя метод наименьших квадратов функцию  $y = f(x)$ , заданную таблично, аппроксимировать
  - а) многочленом первой степени  $y = P_1(x) \equiv a_1 + a_2x$  ;
  - б) многочленом второй степени  $y = P_2(x) \equiv a_1 + a_2x + a_3x^2$  ;
  - в) экспоненциальной зависимостью  $y = a_1e^{a_2x}$  .
2. Для каждой зависимости вычислить коэффициент детерминированности.
3. Вычислить коэффициент корреляции (только в случае а).
4. Для каждой зависимости построить линию тренда.
5. Используя функцию ЛИНЕЙН вычислить числовые характеристики зависимости  $y$  от  $x$ .
6. Сравнить свои вычисления с результатами, полученными при помощи функции ЛИНЕЙН.
7. Сделать вывод, какая из полученных формул наилучшим образом аппроксимирует функцию  $y = f(x)$  .
8. Написать программу на одном из языков программирования и сравнить результаты счета с полученными выше.

## 2. Расчетные формулы.

### 2.1 Построение эмпирических формул методом наименьших квадратов

Очень часто, особенно при анализе эмпирических данных возникает необходимость найти в явном виде функциональную зависимость между величинами  $x$  и  $y$ , которые получены в результате измерений.

При аналитическом исследовании взаимосвязи между двумя величинами  $x$  и  $y$  производят ряд наблюдений и в результате получается таблица значений:

$x$	$x_1$	$x_2$	...	$x_i$	...	$x_n$
$y$	$y_1$	$y_2$	...	$y_i$	...	$y_n$

Эта таблица обычно получается как итог каких-либо экспериментов, в которых  $x_i$  (независимая величина) задается экспериментатором, а  $y_i$  получается в результате опыта. Поэтому эти значения  $y_i$  будем называть эмпирическими или опытными значениями.

Между величинами  $x$  и  $y$  существует функциональная зависимость, но ее аналитический вид обычно неизвестен, поэтому возникает практически важная задача - найти эмпирическую формулу

$$y = f(x; a_1, a_2, \dots, a_m), \quad (2.1.1)$$

(где  $a_1, a_2, \dots, a_m$  - параметры), значения которой при  $x = x_i$  возможно мало отличались бы от опытных значений  $y_i$  ( $i = 1, 2, \dots, n$ ).

Обычно указывают класс функций (например, множество линейных, степенных, показательных и т.п.) из которого выбирается функция  $f(x)$ , и далее определяются наилучшие значения параметров.

Если в эмпирическую формулу (2.1.1) подставить исходные  $x_i$ , то получим теоретические значения  $y_i^T = f(x_i; a_1, a_2, \dots, a_m)$ , где  $i = 1, 2, \dots, n$ .

Разности  $y_i^T - y_i$  называются отклонениями и представляют собой расстояния по вертикали от точек  $M_i$  до графика эмпирической функции.

Согласно методу наименьших квадратов наилучшими коэффициентами  $a_1, a_2, \dots, a_m$  считаются те, для которых сумма квадратов отклонений найденной эмпирической функции от заданных значений функции



$$S(a_1, a_2, \dots, a_m) = \sum_{i=1}^n [f(x_i; a_1, a_2, \dots, a_m) - y_i]^2 \quad (2.1.2)$$

будет минимальной.

Поясним геометрический смысл метода наименьших квадратов.

Каждая пара чисел  $(x_i, y_i)$  из исходной таблицы определяет точку  $M_i$  на плоскости  $XOY$ . Используя формулу (2.1.1) при различных значениях коэффициентов  $a_1, a_2, \dots, a_m$  можно построить ряд кривых, которые являются графиками функции (2.1.1). Задача состоит в определении коэффициентов  $a_1, a_2, \dots, a_m$  таким образом, чтобы сумма квадратов расстояний по вертикали от точек  $M_i(x_i, y_i)$  до графика функции (2.1.1) была наименьшей.

Построение эмпирической формулы состоит из двух этапов: выяснение общего вида этой формулы и определение ее наилучших параметров.

Если неизвестен характер зависимости между данными величинами  $x$  и  $y$ , то вид эмпирической зависимости является произвольным. Предпочтение отдается простым формулам, обладающим хорошей точностью. Удачный выбор эмпирической формулы в значительной мере зависит от знаний исследователя в предметной области, используя которые он может указать класс функций из теоретических соображений. Большое значение имеет изображение полученных данных в декартовых или в специальных системах координат (полулогарифмической, логарифмической и т.д.). По положению точек можно примерно угадать общий вид зависимости путем установления сходства между построенным графиком и образцами известных кривых.

Определение наилучших коэффициентов  $a_1, a_2, \dots, a_m$  входящих в эмпирическую формулу производят хорошо известными аналитическими методами.

Для того, чтобы найти набор коэффициентов  $a_1, a_2, \dots, a_m$ , которые доставляют минимум функции  $S$ , определяемой формулой (2.1.2), используем необходимое условие экстремума функции нескольких переменных - равенство нулю частных производных. В результате получим нормальную систему для определения коэффициентов  $a_i$  ( $i = 1, 2, \dots, m$ ):

$$\frac{\partial S}{\partial a_1} = 0; \quad \frac{\partial S}{\partial a_2} = 0; \quad \dots; \quad \frac{\partial S}{\partial a_m} = 0. \quad (2.1.3)$$

Таким образом, нахождение коэффициентов  $a_i$  сводится к решению системы (2.1.3).

Эта система упрощается, если эмпирическая формула (2.1.1) линейна относительно параметров  $a_i$ , тогда система (2.1.3) - будет линейной.

Конкретный вид системы (2.1.3) зависит от того, из какого класса эмпирических формул мы ищем зависимость (2.1.1). В случае линейной зависимости  $y = a_1 + a_2x$  система (2.1.3) примет вид:

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad , \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad . \end{cases} \quad (2.1.4)$$

Эта линейная система может быть решена любым известным методом (методом Гаусса, простых итераций, формулами Крамера).

В случае квадратичной зависимости  $y = a_1 + a_2x + a_3x^2$  система (2.1.3) примет вид:

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i + a_3 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i \quad , \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 + a_3 \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i \quad , \\ a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 + a_3 \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i \quad . \end{cases} \quad (2.1.5)$$

## 2.2 Линеаризация экспоненциальной зависимости.

В ряде случаев в качестве эмпирической формулы берут функцию в которую неопределенные коэффициенты входят нелинейно. При этом иногда задачу удается линеаризовать, т.е. свести к линейной. К числу таких зависимостей относится экспоненциальная зависимость

$$y = a_1 \cdot e^{a_2 x} \quad , \quad (2.2.1)$$

где  $a_1$  и  $a_2$  неопределенные коэффициенты.

Линеаризация достигается путем логарифмирования равенства (2.2.1), после чего получаем соотношение

$$\ln y = \ln a_1 + a_2 x \quad (2.2.2)$$

Обозначим  $\ln y$  и  $\ln a_1$  соответственно через  $t$  и  $\zeta$ , тогда зависимость (2.2.1) может быть записана в виде  $t = a_1 + a_2 x$ , что позволяет применить формулы (2.1.4) с заменой  $a_1$  на  $\zeta$  и  $y_i$  на  $t_i$ .

### 2.3 Элементы теории корреляции.

График восстановленной функциональной зависимости  $y(x)$  по результатам измерений  $(x_i, y_i)$ ,  $i=1,2,\dots,n$  называется кривой регрессии. Для проверки согласия построенной кривой регрессии с результатами эксперимента обычно вводят следующие числовые характеристики: коэффициент корреляции (линейная зависимость), корреляционное отношение и коэффициент детерминированности. При этом результаты обычно группируют и представляют в форме корреляционной таблицы. В каждой клетке этой таблицы приводятся численности  $n_{ij}$  тех пар  $(x, y)$ , компоненты которых попадают в соответствующие интервалы группировки по каждой переменной. Предполагая длины интервалов группировки (по каждой переменной) равными между собой, выбирают центры  $x_i$  (соответственно  $y_i$ ) этих интервалов и числа  $n_{ij}$  в качестве основы для расчетов.

Коэффициент корреляции является мерой линейной связи между зависимыми случайными величинами: он показывает, насколько хорошо в среднем может быть представлена одна из величин в виде линейной функции от другой.

Коэффициент корреляции вычисляется по формуле:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2.3.1)$$

где  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ ,  $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ ,  $\bar{x}$  и  $\bar{y}$  – среднее арифметическое значение соответственно

по  $x$  и  $y$ .

Коэффициент корреляции между случайными величинами по абсолютной величине не превосходит 1. Чем ближе  $|\rho|$  к 1, тем теснее линейная связь между  $x$  и  $y$ .

В случае нелинейной корреляционной связи условные средние значения располагаются около кривой линии. В этом случае в качестве характеристики силы

связи рекомендуется использовать корреляционное отношение, интерпретация которого не зависит от вида исследуемой зависимости.

Корреляционное отношение вычисляется по формуле:

$$\eta_{y|x}^2 = \frac{\frac{1}{n} \sum_i n_i \cdot (\bar{y}_i - \bar{y})^2}{\frac{1}{n} \sum_j n_j \cdot (y_j - \bar{y})^2}, \quad (2.3.2)$$

где  $n_i = \sum_j n_{ij}$ ,  $n_j = \sum_i n_{ij}$ , а числитель характеризует рассеяние условных средних  $\bar{y}_i$  около безусловного среднего  $\bar{y}$ .

Всегда  $0 \leq \eta_{y|x}^2 \leq 1$ . Равенство  $\eta_{y|x}^2 = 0$  соответствует некоррелированным случайным величинам;  $\eta_{y|x}^2 = 1$  тогда и только тогда, когда имеется точная функциональная связь между  $y$  и  $x$ . В случае линейной зависимости  $y$  от  $x$  корреляционное отношение совпадает с квадратом коэффициента корреляции. Величина  $\eta_{y|x}^2 - \rho^2$  используется в качестве индикатора отклонения регрессии от линейной.

Корреляционное отношение является мерой корреляционной связи  $y$  с  $x$  в какой угодно форме, но не может дать представления о степени приближенности эмпирических данных к специальной форме. Чтобы выяснить насколько точно построенная кривая отражает эмпирические данные вводится еще одна характеристика – коэффициент детерминированности.

Для его описания рассмотрим следующие величины.  $S_{\text{полн}} = \sum_{i=1}^n (y_i - \bar{y})^2$  -

полная сумма квадратов, где  $\bar{y}$  среднее значение  $y_i$ .

Можно доказать следующее равенство

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - y_i^{\text{т}})^2 + \sum_{i=1}^n (y_i^{\text{т}} - \bar{y})^2.$$

Первое слагаемое равно  $S_{\text{ост}} = \sum_{i=1}^n (y_i - y_i^{\text{т}})^2$  и называется остаточной суммой квадратов. Оно характеризует отклонение экспериментальных данных от теоретических.

Второе слагаемое равно  $S_{\text{регр}} = \sum_{i=1}^n (y_i^{\text{т}} - \bar{y})^2$  и называется регрессионной суммой квадратов и оно характеризует разброс данных.

Очевидно, что справедливо следующее равенство  $S_{\text{полн}} = S_{\text{ост}} + S_{\text{регр}}$ .

Коэффициент детерминированности определяется по формуле:

$$r^2 = 1 - \frac{S_{\text{ост}}}{S_{\text{полн}}}. \quad (2.3.3)$$

Чем меньше остаточная сумма квадратов по сравнению с общей суммой квадратов, тем больше значение коэффициента детерминированности  $r^2$ , который показывает, насколько хорошо уравнение, полученное с помощью регрессионного анализа, объясняет взаимосвязи между переменными. Если он равен 1, то имеет место полная корреляция с моделью, т.е. нет различия между фактическим и оценочным значениями  $y$ . В противоположном случае, если коэффициент детерминированности равен 0, то уравнение регрессии неудачно для предсказания значений  $y$ .

Коэффициент детерминированности всегда не превосходит корреляционное отношение. В случае когда выполняется равенство  $r^2 = \eta_{y|x}^2$  то можно считать, что построенная эмпирическая формула наиболее точно отражает эмпирические данные.

### 3. Расчет коэффициентов аппроксимации в Microsoft Excel.

Вариант №22

Функция  $y=f(x)$  задана таблицей 1

Таблица 1

Исходные данные.

$x_i$	$y_i$	$x_i$	$y_i$	$x_i$	$y_i$	$x_i$	$y_i$	$x_i$	$y_i$
12.85	154.77	9.65	81.43	7.74	55.86	5.02	24.98	1.86	3.91
12.32	145.59	9.63	80.97	7.32	47.63	4.65	22.87	1.76	3.22
11.43	108.37	9.22	79.04	7.08	48.03	4.53	20.32	1.11	1.22
10.59	100.76	8.44	61.76	6.87	36.85	3.24	9.06	0.99	1.10
10.21	98.32	8.07	60.54	5.23	25.65	2.55	6.23	0.72	0.53

Требуется выяснить - какая из функций - линейная, квадратичная или экспоненциальная наилучшим образом аппроксимирует функцию заданную таблицей 1.

Решение.

Поскольку в данном примере каждая пара значений  $(x_i, y_i)$  встречается один раз, то между  $y$  и  $x$  существует функциональная зависимость.

Для проведения расчетов данные целесообразно расположить в виде таблицы 2, используя средства табличного процессора Microsoft Excel.

## Расчет сумм.

	A	B	C	D	E	F	G	H	I
1	x	x	x <sup>2</sup>	xy	x <sup>3</sup>	x <sup>4</sup>	x <sup>2</sup> y	lny	xlny
2	12,85	154,77	165,12	1988,79	2121,82	27265,44	25556,01	5,04	64,79
3	12,32	145,59	151,78	1793,67	1869,96	23037,90	22098,00	4,98	61,36
4	11,43	108,37	130,64	1238,67	1493,27	17068,09	14157,99	4,69	53,56
5	10,59	100,76	112,15	1067,05	1187,65	12577,20	11300,04	4,61	48,85
6	10,21	98,32	104,24	1003,85	1064,33	10866,83	10249,28	4,59	46,85
7	9,65	81,43	93,12	785,80	898,63	8671,80	7582,97	4,40	42,46
8	9,63	80,97	92,74	779,74	893,06	8600,13	7508,91	4,39	42,31
9	9,22	79,04	85,01	728,75	783,78	7226,43	6719,06	4,37	40,29
10	8,44	61,76	71,23	521,25	601,21	5074,23	4399,39	4,12	34,80
11	8,07	60,54	65,12	488,56	525,56	4241,25	3942,66	4,10	33,11
12	7,74	55,86	59,91	432,36	463,68	3588,92	3346,44	4,02	31,14
13	7,32	47,63	53,58	348,65	392,22	2871,07	2552,13	3,86	28,28
14	7,08	48,03	50,13	340,05	354,89	2512,66	2407,57	3,87	27,41
15	6,87	36,85	47,20	253,16	324,24	2227,55	1739,21	3,61	24,78
16	5,23	25,65	27,35	134,15	143,06	748,18	701,60	3,24	16,97
17	5,02	24,98	25,20	125,40	126,51	635,06	629,51	3,22	16,15
18	4,65	22,87	21,62	106,35	100,54	467,53	494,51	3,13	14,55
19	4,53	20,32	20,52	92,05	92,96	421,11	416,98	3,01	13,64
20	3,24	9,06	10,50	29,35	34,01	110,20	95,11	2,20	7,14
21	2,55	6,23	6,50	15,89	16,58	42,28	40,51	1,83	4,66
22	1,86	3,91	3,46	7,27	6,43	11,97	13,53	1,36	2,54
23	1,76	3,22	3,10	5,67	5,45	9,60	9,97	1,17	2,06
24	1,11	1,22	1,23	1,35	1,37	1,52	1,50	0,20	0,22
25	0,99	1,10	0,98	1,09	0,97	0,96	1,08	0,10	0,09
26	0,72	0,53	0,52	0,38	0,37	0,27	0,27	-0,63	-0,46
27	163,08	1279,01	1402,97	12289,30	13502,57	138278,17	125964,22	79,49	657,57

Поясним как таблица 2 составляется.

- Шаг 1. В ячейки A2:A26 заносим значения  $x_i$ .
- Шаг 2. В ячейки B2:B26 заносим значения  $y_i$ .
- Шаг 3. В ячейку C2 вводим формулу =A2^2.
- Шаг 4. В ячейки C3:C26 эта формула копируется.
- Шаг 5. В ячейку D2 вводим формулу =A2\*B2.
- Шаг 6. В ячейки D3:D26 эта формула копируется.
- Шаг 7. В ячейку F2 вводим формулу =A2^4.
- Шаг 8. В ячейки F3:F26 эта формула копируется.
- Шаг 9. В ячейку G2 вводим формулу =A2^2\*B2.
- Шаг 10. В ячейки G3:G26 эта формула копируется.
- Шаг 11. В ячейку H2 вводим формулу =LN(B2).
- Шаг 12. В ячейки H3:H26 эта формула копируется.
- Шаг 13. В ячейку I2 вводим формулу =A2\*LN(B2).

Шаг 14. В ячейки I3:I26 эта формула копируется.

Последующие шаги делаем с помощью автосуммирования  $\Sigma$  .

Шаг 15. В ячейку A27 вводим формулу =СУММ(A2:A26).

Шаг 16. В ячейку B27 вводим формулу =СУММ(B2:B26).

Шаг 17. В ячейку C27 вводим формулу =СУММ(C2:C26).

Шаг 18. В ячейку D27 вводим формулу =СУММ(D2:D26).

Шаг 19. В ячейку E27 вводим формулу =СУММ(E2:E26).

Шаг 20. В ячейку F27 вводим формулу =СУММ(F2:F26).

Шаг 21. В ячейку G27 вводим формулу =СУММ(G2:G26).

Шаг 22. В ячейку H27 вводим формулу =СУММ(H2:H26).

Шаг 23. В ячейку I27 вводим формулу =СУММ(I2:I26).



Аппроксимируем функцию  $y = f(x)$  линейной функцией  $y = a_1 + a_2x$ . Для определения коэффициентов  $a_1$  и  $a_2$  воспользуемся системой

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases}$$

Используя итоговые суммы таблицы 2, расположенные в ячейках A27, B27, C27 и D27, запишем систему в виде

$$\begin{cases} 25 \cdot a_1 + 95,93 \cdot a_2 = 2089,99 \\ 95,93 \cdot a_1 + 453,31 \cdot a_2 = 11850,65 \end{cases}$$

решив которую, получим  $a_1 = -24,74$  и  $a_2 = 11,63$ .

Таким образом, линейная аппроксимация имеет вид  $y = -24,74 + 11,63 \cdot x$ .

Решение системы проводили, пользуясь средствами Microsoft Excel.

Результаты представлены в таблице 3.

Таблица 3

**Результаты коэффициентов линейной аппроксимации.**

	A	B	C	D
33	25	163,0800	1279,0100	
34	163,0800	1402,9662	12289,2992	
35				
36	Обратная матрица			
37	0,1654623	-0,019233	a1=	-24,7352
38	-0,019233	0,0029484	a2=	11,6347

В таблице 3 в ячейках A37:B38 записана формула {=МОБР(A33:B34)}.

В ячейках D37:D38 записана формула {=МУМНОЖ(A37:B38;C33:C34)}.

Далее аппроксимируем функцию  $y = f(x)$  квадратичной функцией  $y = a_1 + a_2x + a_3x^2$ . Для определения коэффициентов  $a_1$ ,  $a_2$  и  $a_3$  воспользуемся системой

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i + a_3 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i, \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 + a_3 \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i, \\ a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 + a_3 \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i. \end{cases}$$

Используя итоговые суммы таблицы 2, расположенные в ячейках A27, B27, C27, D27, E27, F27 и G27 запишем систему в виде

$$\begin{cases} 25 \cdot a_1 + 95,93 \cdot a_2 + 453,31 a_3 = 2089,99 \\ 95,93 \cdot a_1 + 453,31 \cdot a_2 + 2417,57 a_3 = 11850,65 \\ 453,31 a_1 + 2417,57 a_2 + 13982,99 a_3 = 71327,34 \end{cases}$$

решив которую, получим  $a_1 = 1,596$ ,  $a_2 = -0,621$  и  $a_3 = 0,955$ .

Таким образом, квадратичная аппроксимация имеет вид

$$y = 1,596 - 0,621 \cdot x + 0,955 x^2.$$

Решение системы проводили, пользуясь средствами Microsoft Excel. Результаты представлены в таблице 4.

Таблица 4

**Результаты коэффициентов квадратичной аппроксимации.**

	E	F	G	H	I
33	25	163,0800	1402,9662	1279,0100	
34	163,0800	1402,9662	13502,5736	12289,2992	
35	1402,9662	13502,5736	138278,1667	125964,2240	
36					
37					
38	0,368843786	-0,113896875	0,007379513	a1=	1,5968
39	-0,113896875	0,047009493	-0,003434784	a2=	-0,6215
40	0,007379513	-0,003434784	0,000267759	a3=	0,9554

В таблице 4 в ячейках E38:G40 записана формула {=МОБР(E33:G35)}.

В ячейках I38:I40 записана формула {=МУМНОЖ(E38:G40;H33:H35)}.

Теперь аппроксимируем функцию  $y = f(x)$  экспоненциальной функцией  $y = a_1 \cdot e^{a_2 x}$ . Для определения коэффициентов  $a_1$  и  $a_2$  прологарифмируем значения  $y_i$  и используя итоговые суммы таблицы 2, расположенные в ячейках A27, C27, H27 и I27 получим систему

$$\begin{cases} 25 \cdot c + 95,93 \cdot a_2 = 90,9771 \\ 95,93 c + 453,3105 \cdot a_2 = 415,0797 \end{cases}$$

где  $c = \ln(a_1)$ .

Решив систему, найдем  $c = 0,5061$ ,  $a_2 = 0,4099$ .

После потенцирования получим  $a_1 = 1,6589$ .

Таким образом, экспоненциальная аппроксимация имеет вид

$$y = 1,6589 \cdot e^{0,4099x}.$$

Решение системы проводили, пользуясь средствами Microsoft Excel.

Результаты представлены в таблице 5.

Таблица 5

**Результаты коэффициентов экспоненциальной аппроксимации.**

	D	E	F	G
42	25	163,0800	79,4941	
43	163,0800	1402,9662	657,5670	
44				
45	0,165462301	-0,019233245	c=	0,5061
46	-0,019233245	0,002948437	a2=	0,4099
47			a1=	1,6589

В таблице 5 в ячейках D45:E46 записана формула {=МОБР(D42:943)}.

В ячейках G45:G46 записана формула {=МУМНОЖ(D45:E46;F42:F43)}.

В ячейке G47 записана формула =EXP(G45).

Вычислим среднее арифметическое  $\bar{x}$  и  $\bar{y}$  по формулам:

$$\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i; \quad \bar{y} = \frac{1}{25} \sum_{i=1}^{25} y_i.$$

Результаты расчета  $\bar{x}$  и  $\bar{y}$  средствами Microsoft Excel представлены в таблице

6.

Таблица 6

**Вычисление средних значений X и Y.**

	E	F
49	X <sub>ср</sub> =	6,5232
50	Y <sub>ср</sub> =	51,1604

В ячейке F49 записана формула =A26/25.

В ячейке F50 записана формула =B26/25.

Для того, чтобы рассчитать коэффициент корреляции и коэффициент детерминированности данные целесообразно расположить в виде таблицы 7, которая является продолжением таблицы 2.

Таблица 7

**Вычисление остаточных сумм.**

	A	B	J	K	L	M	N	O
1	x	y	(X-X <sub>ср</sub> )(Y-Y <sub>ср</sub> )	(X-X <sub>ср</sub> ) <sup>2</sup>	(Y-Y <sub>ср</sub> ) <sup>2</sup>	линейн.	квадр.	эксп.
2	12,85	154,77	655,5172	40,0284	10734,9492	899,9459	11,5313	27788,7125
3	12,32	145,59	547,3895	33,6029	8916,9494	728,2170	43,9828	12794,1401
4	11,43	108,37	280,7161	24,0767	3272,9383	0,0145	119,8089	5077,8309
5	10,59	100,76	201,7117	16,5389	2460,1203	5,2146	1,9750	704,7719
6	10,21	98,32	173,8680	13,5925	2224,0279	18,1880	12,0426	112,9220
7	9,65	81,43	94,6470	9,7769	916,2487	37,3299	9,8713	26,7576
8	9,63	80,97	92,6125	9,6522	888,6123	40,1592	10,5357	24,2632
9	9,22	79,04	75,1857	7,2727	777,2721	12,2283	3,8157	41,3572
10	8,44	61,76	20,3173	3,6741	112,3515	136,9325	7,0251	81,3414
11	8,07	60,54	14,5084	2,3926	87,9769	74,2522	3,0137	231,6548
12	7,74	55,86	5,7185	1,4806	22,0862	89,4447	3,3698	264,8269
13	7,32	47,63	-2,8130	0,6349	12,4637	163,8640	0,3746	204,5982
14	7,08	48,03	-1,7430	0,3100	9,7994	92,3254	8,6483	317,7639
15	6,87	36,85	-4,9628	0,1203	204,7875	336,5507	31,0335	83,4832
16	5,23	25,65	32,9900	1,6724	650,7805	109,5034	1,3680	132,2467
17	5,02	24,98	39,3544	2,2596	685,4133	75,5352	5,8839	143,9242
18	4,65	22,87	52,9936	3,5089	800,3467	42,2013	12,2792	137,2112
19	4,53	20,32	61,4711	3,9728	951,1303	58,5239	3,7330	94,0736
20	3,24	9,06	138,2240	10,7794	1772,4437	15,2202	0,3058	7,8431
21	2,55	6,23	178,5175	15,7863	2018,7408	1,6813	0,0000	2,2876
22	1,86	3,91	220,3381	21,7454	2232,6003	49,0643	0,0268	0,1257
23	1,76	3,22	228,3497	22,6881	2298,2820	55,9215	0,0588	0,0371
24	1,11	1,22	270,3374	29,3027	2494,0436	170,0580	0,7468	1,9447
25	0,99	1,10	276,9942	30,6163	2506,0436	204,9706	0,6691	1,9294
26	0,72	0,53	293,8183	33,6771	2563,4374	285,2102	1,2424	2,8842
27	163,0800	1279,0100	3946,0612	339,1627	49613,8457	3702,5567	293,3420	48278,9311

Поясним как таблица 7 составляется.

Ячейки A2:A27 и B2:B27 уже заполнены (см. табл. 2).

Далее делаем следующие шаги.

Шаг 1. В ячейку J2 вводим формулу =(A2-\$F\$49)\*(B2-\$F\$50).

Шаг 2. В ячейки J3:J26 эта формула копируется.

Шаг 3. В ячейку K2 вводим формулу =(A2-\$F\$49)^2.

Шаг 4. В ячейки K3:K26 эта формула копируется.

Шаг 5. В ячейку L2 вводим формулу  $=(B2-\$F\$50)^2$ .

Шаг 6. В ячейки L3:L26 эта формула копируется.

Шаг 7. В ячейку M2 вводим формулу  $=(D\$37+\$D\$38*A2-B2)^2$ .

Шаг 8. В ячейки M3:M26 эта формула копируется.

Шаг 9. В ячейку N2 вводим формулу

$$=(I\$38+I\$39*A2+I\$40*A2^2-B2)^2.$$

Шаг 10. В ячейки N3:N26 эта формула копируется.

Шаг 11. В ячейку O2 вводим формулу

$$=(G\$47*EXP(G\$46*A2)-B2)^2.$$

Шаг 12. В ячейки O3:O26 эта формула копируется.

Последующие шаги делаем с помощью автосуммирования  $\Sigma$ .

Шаг 13. В ячейку J27 вводим формулу  $=СУММ(J2:J26)$ .

Шаг 14. В ячейку K27 вводим формулу  $=СУММ(K2:K26)$ .

Шаг 15. В ячейку L27 вводим формулу  $=СУММ(L2:L26)$ .

Шаг 16. В ячейку M27 вводим формулу  $=СУММ(M2:M26)$ .

Шаг 17. В ячейку N27 вводим формулу  $=СУММ(N2:N26)$ .

Шаг 18. В ячейку O27 вводим формулу  $=СУММ(O2:O26)$ .

Теперь проведем расчеты коэффициента корреляции по формуле

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{только для линейной аппроксимации})$$

и коэффициента детерминированности по формуле  $r^2 = 1 - \frac{S_{\text{ост}}}{S_{\text{полн}}}$ . Результаты расчетов средствами Microsoft Excel представлены в таблице 8.

Таблица 8

### Результаты расчета.

	A	B	C	D
53	Коэффициент корреляции			0,9620
54	Коэффициент детерминированности (линейная аппроксимация)			0,925373
55	Квадратичная аппроксимация			0,994087
56	Экспоненциальная аппроксимация			0,0269

В таблице 8 в ячейке D53 записана формула  $=J27/(K27*L27)^{(1/2)}$ .

В ячейке D54 записана формула  $=1- M27/L27$ .

В ячейке D55 записана формула  $=1- N27/L27$ .

В ячейке D56 записана формула  $=1- O27/L27$ .

Анализ результатов расчетов показывает, что квадратичная аппроксимация наилучшим образом описывает экспериментальные данные.

#### 4. Построение графиков в Excel и использование функции ЛИНЕЙН.

Рассмотрим результаты эксперимента, приведенные в исследованном выше примере.

Исследуем характер зависимости в три этапа:

- Построим график зависимости.
- Построим линию тренда ( $y = P_1(x) \equiv a_1 + a_2x$ ,  $y = P_2(x) \equiv a_1 + a_2x + a_3x^2$ ,  $y = a_1e^{a_2x}$ ).
- Получим числовые характеристики коэффициентов этого уравнения.

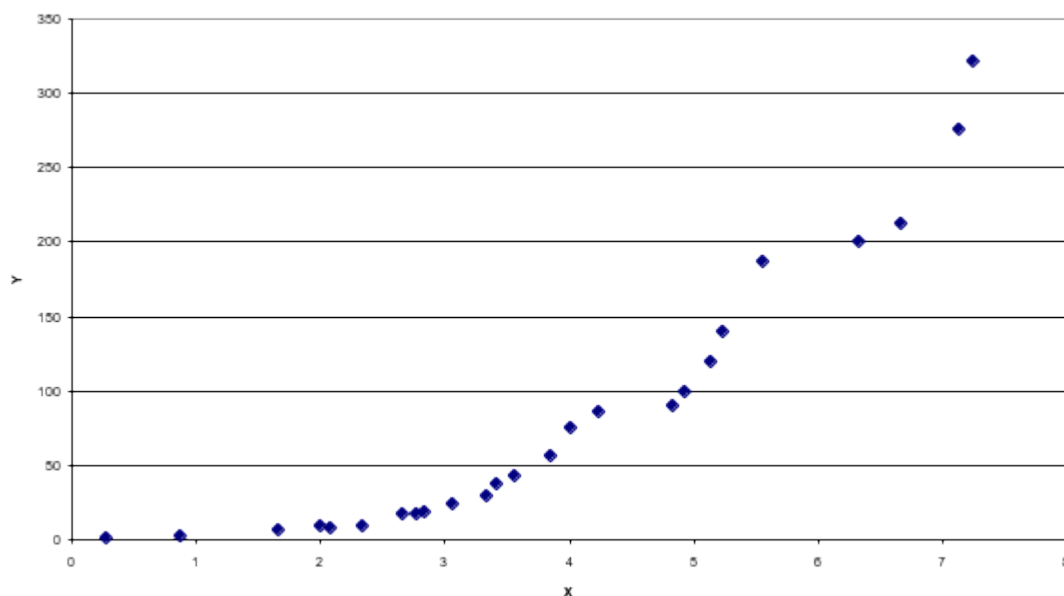
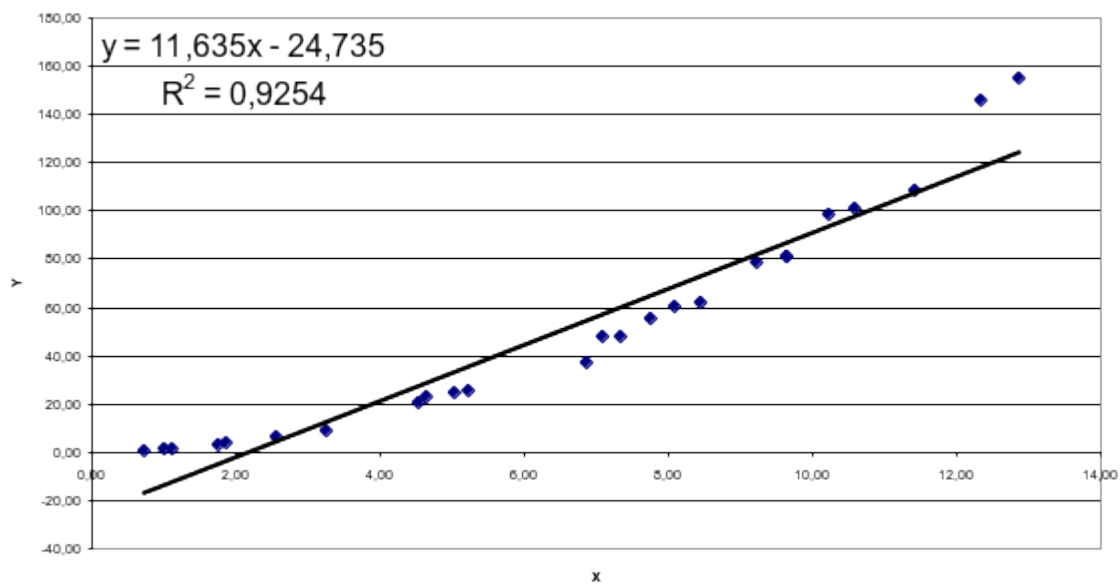
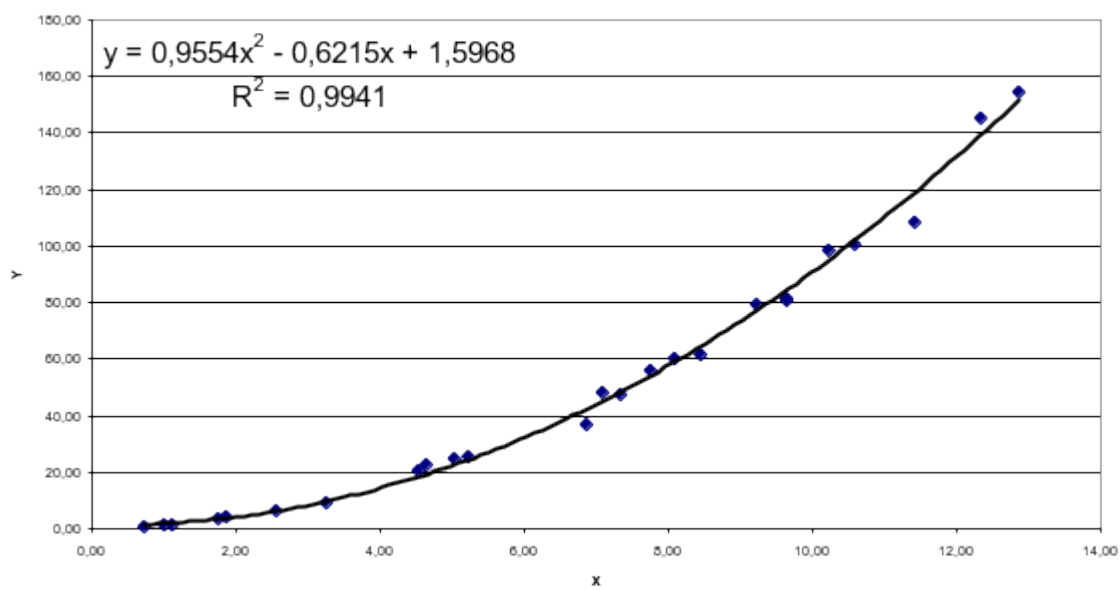


Рис.4.1. График зависимости y от x

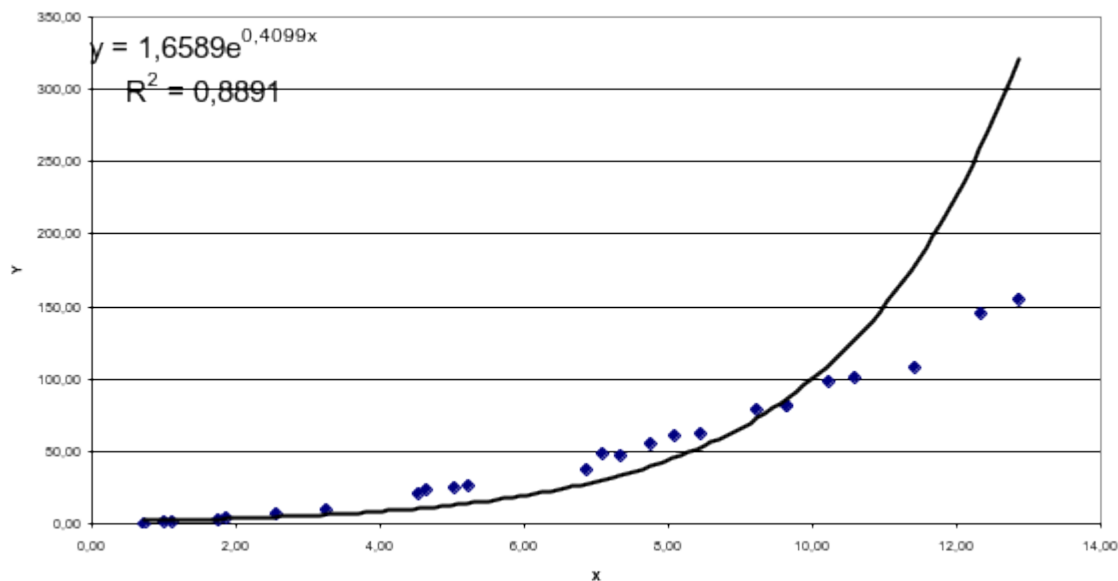


**Рис.4.2. График линейной аппроксимации**



**Рис.4.3. График квадратичной аппроксимации.**





**Рис.4.4. График экспоненциальной аппроксимации.**

Примечание: Полученное при построении линии тренда значение коэффициента детерминированности для экспоненциальной зависимости  $R^2 = 0,889$  не совпадает с истинным значением  $R^2 = 0,027$ , поскольку при вычислении коэффициента детерминированности используются не истинные значения  $y_i$ , а преобразованные значения  $\ln y_i$  с дальнейшей линейризацией.

Таблица 9

	F	G
53	11,635	-24,735
54	0,689	5,161
55	0,925	12,688
56	285,197	23,000
57	45911,289	3702,557

## 5. Программа на языке Pascal.

### 5.1. Схема алгоритма.

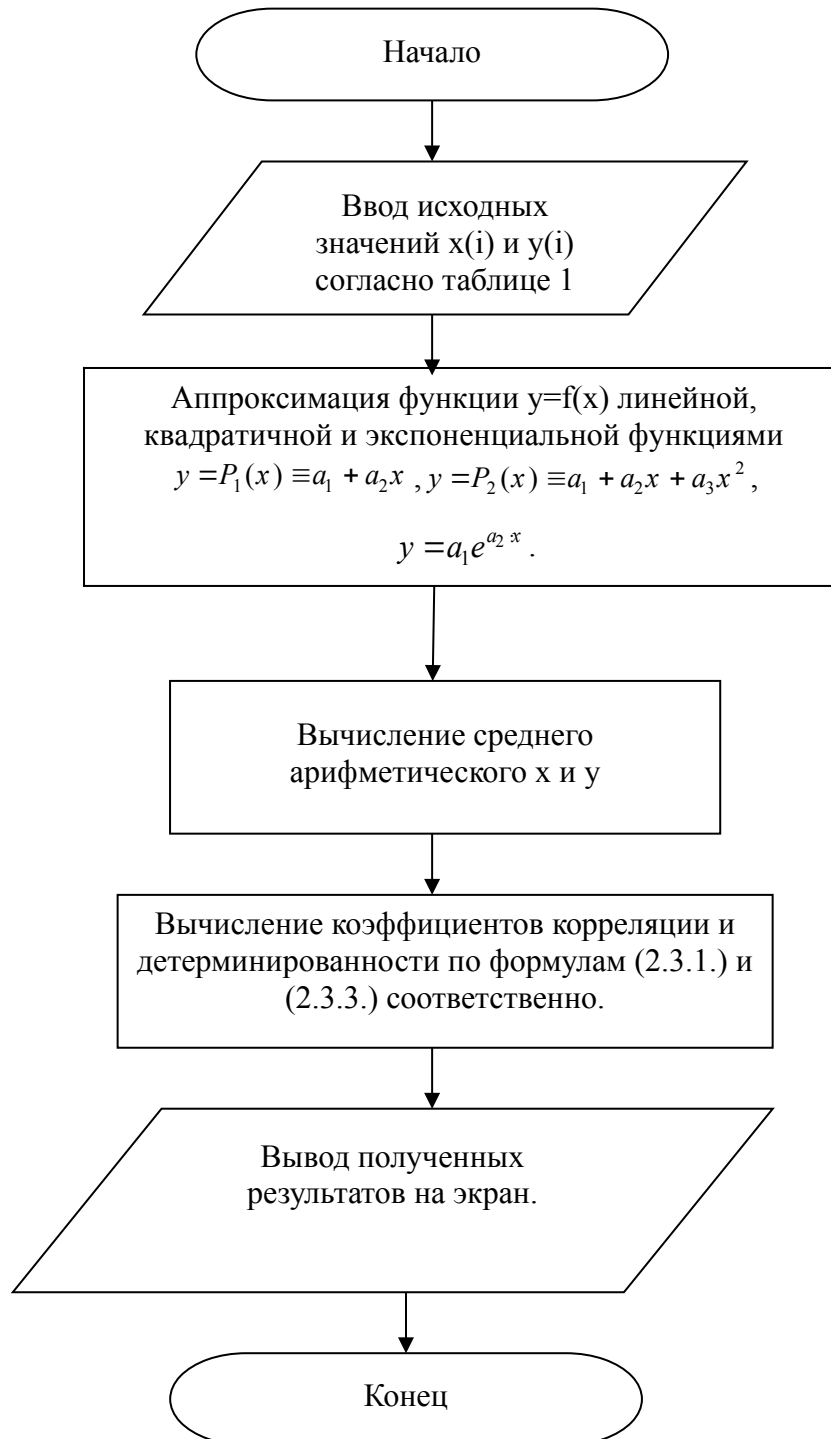


Рис.5.1. Блок-схема

```

program Kramer;
uses CRT;
const
n=25;
type
TArrayXY = array[1..2,1..n] of real;
TArray = array[1..n] of real;
var
SumX,SumY,SumX2,SumXY,SumX3,SumX4,SumX2Y,SumLnY,SumXLnY: real;
OPRlin,OPRkvadr,OPRa1,OPRa2,OPRa3:real;
a1lin,a2lin,a1kvadr,a2kvadr,a3kvadr,a1exp,a2exp,cexp:real;
Xsr,Ysr,S1,S2,S3,Slin,Skvadr,Sexp:real;
Kkor,KdetLin,KdetKvadr,KdetExp:real;
i:byte;
const
ArrayXY:TArrayXY=((12.85,12.32,11.43,10.59,10.21,9.65,9.63,9.22,8.44,8.07,7.74,7.32,
7.08,6.87,5.23,5.02,4.65,4.53,3.24,2.55,1.86,1.76,1.11,0.99,0.72) , (154.77
145.59,108.37,100.76,98.32,81.43,80.97,79.04,61.76,60.54,55.86,47.63,48.03,36.85,25.65,
24.98,22.87,20.32,9.06,6.23,3.91,3.22,1.22,1.10,0.53));
begin
ClrScr;
SumX:=0.0;
SumY:=0.0;
SumXY:=0.0;
SumX2:=0.0;
SumX3:=0.0;
SumX4:=0.0;
SumX2Y:=0.0;
SumLnY:=0.0;
SumXLnY:=0.0;
{ Вычисление сумм x, y, x*y, x^2, x^3, x^4, (x^2)*y, Ln(y), x*Ln(y) }
for i:=1 to n do
begin

```

```

SumX:=SumX+ArrayXY[1,i];
SumY:=SumY+ArrayXY[2,i];
SumXY:=SumXY+ArrayXY[1,i]*ArrayXY[2,i];
SumX2:=SumX2+sqr(ArrayXY[1,i]);
SumX3:=SumX3+ArrayXY[1,i]*ArrayXY[1,i]*ArrayXY[1,i];
SumX4:=SumX4+sqr(ArrayXY[1,i])*sqr(ArrayXY[1,i]);
SumX2Y:=SumX2Y+sqr(ArrayXY[1,i])*ArrayXY[2,i];
SumLnY:=SumLnY+ln(ArrayXY[2,i]);
SumXlnY:=SumXlnY+ArrayXY[1,i]*ln(ArrayXY[2,i])
end;
{ Вычисление коэффициентов }
OPRlin:=0.0;
a1lin:=0.0;
a2lin:=0.0;
a1kvadr:=0.0;
OPRkvadr:=0.0;
a2kvadr:=0.0;
a2kvadr:=0.0;
a1exp:=0.0;
a2exp:=0.0;
OPRlin:=n*SumX2-SumX*SumX;
a1lin:=(SumX2*SumY-SumX*SumXY)/OPRlin;
a2lin:=(n*SumXY-SumX*SumY)/OPRlin;
OPRkvadr:=n*SumX2*SumX4+SumX*SumX3*SumX2+SumX2*SumX*SumX3-
SumX2*SumX2*SumX2-n*SumX3*SumX3-SumX*SumX*SumX4;
a1kvadr:=(SumY*SumX2*SumX4+SumX*SumX2Y*SumX3+SumX2*SumXY*SumX3-
SumX2*SumX2*SumX2Y-SumY*SumX3*SumX3-SumX*SumXY*SumX4)/OPRkvadr;
a2kvadr:=(n*SumXY*SumX4+SumY*SumX3*SumX2+SumX2*SumX*SumX2Y-
SumX2*SumX2*SumXY-n*SumX3*SumX2Y-SumY*SumX*SumX4)/OPRkvadr;
a3kvadr:=(n*SumX2*SumX2Y+SumX*SumXY*SumX2+SumY*SumX*SumX3-
SumY*SumX2*SumX2-n*SumXY*SumX3-SumX*SumX*SumX2Y)/OPRkvadr;
a2exp:=(n*SumXlnY-SumX*SumLnY)/OPRlin;
cexp:=(SumX2*SumLnY-SumX*SumXlnY)/OPRlin;
a1exp:=exp(cexp);

```

```

{ Вычисление средних арифметических x и y }
Xsr:=SumX/n;
Ysr:=SumY/n;
S1:=0.0;
S2:=0.0;
S3:=0.0;
Slin:=0.0;
Skvadr:=0.0;
Sexp:=0.0;
Kkor:=0.0;
KdetLin:=0.0;
KdetKvadr:=0.0;
KdetExp:=0.0;

for i:=1 to n do
begin
  S1:=S1+(ArrayXY[1,i]-Xsr)*(ArrayXY[2,i]-Ysr);
  S2:=S2+sqr(ArrayXY[1,i]-Xsr);
  S3:=S3+sqr(ArrayXY[2,i]-Ysr);
  Slin:=Slin+sqr(a1lin+a2lin*ArrayXY[1,i]-ArrayXY[2,i]);
  Skvadr:=Skvadr+sqr(a1kvadr+a2kvadr*ArrayXY[1,i]
+a3kvadr*ArrayXY[1,i]*ArrayXY[1,i]-ArrayXY[2,i]);
  Sexp:=Sexp+sqr(a1exp*exp(a2exp*ArrayXY[1,i])-ArrayXY[2,i]);
end;
{ Вычисление коэффициентов корреляции и детерминированности }
Kkor:=S1/sqrt(S2*S3);
KdetLin:=1-Slin/S3;
KdetKvadr:=1-Skvadr/S3;
KdetExp:=1-Sexp/S3;
{ Вывод результатов }
WriteLn('Линейная функция');
WriteLn('a1=',a1lin:8:5);
WriteLn('a2=',a2lin:8:5);
WriteLn('Квадратичная функция');

```

```
WriteLn('a1=',a1kvadr:8:5);
WriteLn('a2=',a2kvadr:8:5);
WriteLn('a3=',a3kvadr:8:5);
WriteLn('Экспоненциальная функция');
WriteLn('a1=',a1exp:8:5);
WriteLn('a2=',a2exp:8:5);
WriteLn('c=',cexp:8:5);
WriteLn('Xcp=',Xsr:8:5);
WriteLn('Ycp=',Ysr:8:5);
WriteLn('Коэффициент корреляции ',Kkor:8:5);
WriteLn('Коэффициент детерминированности (линейная аппроксимация)
',KdetLin:2:5);
WriteLn('Коэффициент детерминированности (квадратическая аппроксимация)
',KdetKvadr:2:5);
WriteLn('Коэффициент детерминированности (экспоненциальная аппроксимация)
',KdetExp:2:5);
end.
```

## 5.2. Результаты расчета Pascal.

Коэффициенты линейной функции

$$a1 = -24.73516$$

$$a2 = 11.63471$$

Коэффициенты квадратичной функции

$$a1 = 1.59678$$

$$a2 = -0.62145$$

$$a3 = 0.95543$$

Коэффициенты экспоненциальной функции

$$a1 = 1.65885$$

$$a2 = 0.40987$$

$$c = 0.50613$$

$$X_{\text{ср}} = 6.52320$$

$$Y_{\text{ср}} = 51.16040$$

Коэффициент корреляции 0.96196

Коэффициент детерминированности (линейная аппроксимация) 0.92537

Коэффициент детерминированности (квадратическая аппроксимация) 0.99409

Коэффициент детерминированности (экспоненциальная аппроксимация) 0.02691

## Заключение.

Сделаем заключение по результатам полученных данных:

1. Анализ результатов расчетов показывает, что квадратичная аппроксимация наилучшим образом описывает экспериментальные данные т.к. согласно таблице 8 коэффициент корреляции - 0,9620; Коэффициенты детерминированности линейной аппроксимации - 0,9253; квадратической аппроксимации – 0,994; экспоненциальной аппроксимации – 0,0269.
2. Сравнивая результаты, полученные при помощи функции ЛИНЕЙН видим что они полностью совпадают с вычислениями, проведенными выше. Это указывает на то, что вычисления верны.
3. Полученное при построении линии тренда значение коэффициента детерминированности для экспоненциальной зависимости не совпадает с истинным значением поскольку при вычислении коэффициента детерминированности используются не истинные значения  $y$ , а преобразованные значения  $\ln(y)$  с дальнейшей линеаризацией.
4. Результаты полученные с помощью программы на языке PASCAL полностью совпадают со значениями приведенными выше. Это говорит о верности вычислений.



### **Список литературы.**

1. Ахметов К.С. Windows 95 для всех. - М.:ТОО "КомпьютерПресс", 1995.
2. Вычислительная техника и программирование. Под ред. А.В. Петрова. М.: Высшая школа, 1991.
3. Гончаров А., Excel 97 в примерах. — СПб: Питер, 1997.
4. Левин А., Самоучитель работы на компьютере. - М.: Международное агентство А.Д.Т., 1996.
5. Информатика: Методические указания к курсовой работе. Санкт-Петербургский горный институт. Сост. Д.Е. Гусев, Г.Н. Журов. СПб, 1999