

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
МЕЖКАФЕДРАЛЬНАЯ ЛАБОРАТОРИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ХТФ
КАФЕДРА ХИМИИ И ТЕХНОЛОГИИ ПЕРЕРАБОТКИ ЭЛАСТОМЕРОВ

А.Н. Гайдадин, С.А. Ефремова, Н.Н.Бакумова

ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА
В ТЕХНОЛОГИЧЕСКИХ РАСЧЕТАХ

Методические указания



Волгоград
2008

УДК 678.04

Рецензент

профессор кафедры «Промышленная экология и безопасность жизнедеятельности»
А.Б. Голованчиков

Издается по решению редакционно-издательского совета
Волгоградского государственного технического университета

Применение корреляционного анализа в технологических расчетах: метод. указания / сост. А.Н.Гайдадин, С.А.Ефремова, Н.Н.Бакумова ; ВолгГТУ. – Волгоград, 2008. – 16 с.

В методических указаниях описана методика проведения корреляционного анализа при оценке экспериментальных данных. Для студентов по направлениям 240100 «Химическая технология и биотехнология», 260100 «Технология продуктов питания», специальности 240502 «Технология переработки пластических масс и эластомеров», а также для студентов, обучающихся по магистерским программам 240115 «Технология переработки эластомеров» и 240110 «Химическая технология высокомолекулярных соединений»

© Волгоградский государственный
технический университет, 2008

ВВЕДЕНИЕ

В практической деятельности инженера-технолога нередки ситуации, в которых оценку одного из свойств объекта необходимо осуществляется с учетом оценки второго свойства. В этом случае возникает необходимость учета взаимного влияния свойств. Закономерности такого влияния достаточно сложно описать математическими моделями. В подобной ситуации удобно использовать корреляционную оценку показателей, устанавливая элемент качественной, экспертной оценки влияния одного показателя на другой. Целью исследователя при решении указанной задачи является не только нахождение корреляционной зависимости между двумя свойствами объекта, но и получение качественной (экспертной) оценки влияния одного свойства на другое.

1. Цель и задачи лабораторной работы

Целью настоящей работы является ознакомление с процедурой проведения корреляционного анализа по результатам эксперимента и определения значения коэффициента корреляции.

В ходе лабораторной работе студенты должны приобрести навыки по установлению экспертных связей для принятия решения, оцениванию влияния двух функций отклика друг на друга, а также по расчету корреляционного коэффициента и установлению взаимного роста и убывания функций друг от друга в зависимости от изменения аргумента.

Задачей студента является освоение теории по использованию корреляционного анализа. В ходе лабораторной работы студент должен обоснованно принять решение о корреляции исследуемых функций. Задачей лабораторной работы является получение студентами навыков разработки экспертных правил и закономерностей.

2. Теоретическая часть

В технологической практике исследователь часто сталкивается с необходимостью установления факта существования функциональных или иных зависимостей между экспериментальными данными, нередко такая связь может быть случайной. Различные постановки задач статистического исследования можно классифицировать следующим образом [1]: задачи корреляционного анализа (задачи исследования наличия взаимосвязей между отдельными группами переменных); задачи регрессионного анализа; задачи дисперсионного анализа.

Методы корреляционного анализа широко применяются для выявления и описания стохастических зависимостей между случайными величинами по экспериментальным данным. Для экспериментального изучения зависимости между случайными величинами Z и Y производят некоторое количество n независимых опытов. Результат i -го опыта дает пару значений (z_i, x_i) и (y_i, x_i) , $i=1,2,\dots, n$, таким образом можно записать [2]:

$$Z=f(x_1, x_2, x_3, \dots, x_n);$$

$$Y=f(x_1, x_2, x_3, \dots, x_n);$$

Если функции зависят от одного аргумента, то, проведя корреляционный анализ, можно установить взаимный рост или убывание функций друг от друга в зависимости от изменения аргумента.

О наличии или отсутствии корреляции между двумя случайными величинами качественно можно судить по виду поля корреляции, поместив экспериментальные точки на координатную плоскость. Положительная корреляция между случайными величинами представлена на рисунке 1, а на рисунке 2 - обратная корреляция. Сравнительно слабая корреляция представлена на рисунке 3, на рисунке 4 приведен пример отсутствия корреляции.

Для количественной оценки тесноты связи служит выборочный коэффициент корреляции.

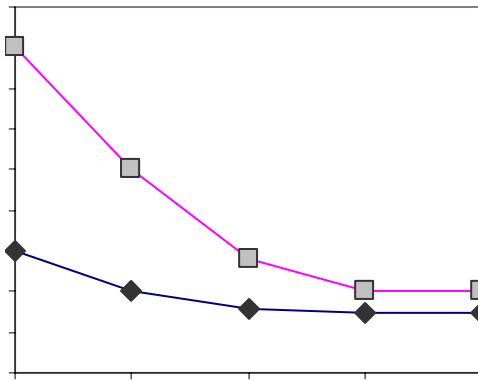


Рисунок 1. Корреляция линейная, рост одной функции приводит к увеличению другой. $r > 0$, $r = 0,98$.

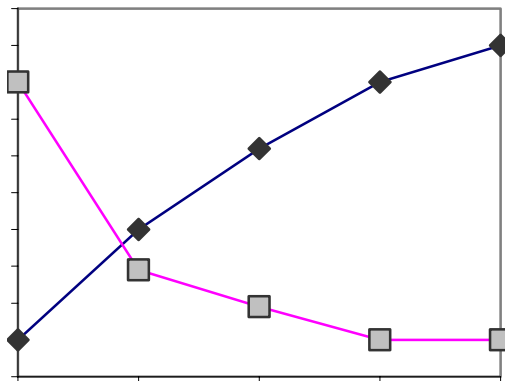


Рисунок 2. Корреляция обратно линейна, рост одной функции приводит к снижению другой. $r < 0$, $r = -0,97$;

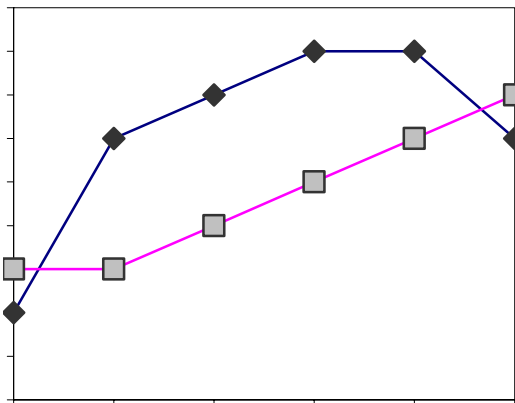


Рисунок 3. Слабая корреляция, $r = 0,51$.

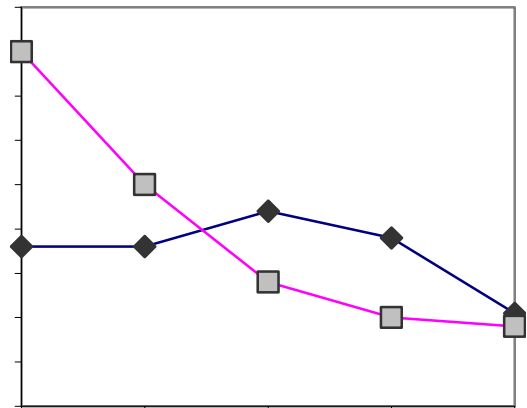


Рисунок 4. Корреляция отсутствует. $r = -0,06$.

Коэффициент корреляции r вычисляется по значениям функций отклика в области эксперимента [2]:

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (z_i - \bar{z})^2}}, \quad (1)$$

где y_i, z_i — каждое текущее значение функции в области эксперимента;
 \bar{y}, \bar{z} — среднее значение функции отклика в области исследования.

Средние значения \bar{y}, \bar{z} вычисляются по следующим формулам:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i \quad (2)$$

где n - число независимых опытов;

Выборочный коэффициент корреляции r по абсолютной величине не превосходит единицы.

Для независимых случайных величин коэффициент корреляции равен нулю, но он может быть равен нулю для некоторых зависимых величин, которые при этом называются некоррелированными. Для случайных величин, имеющих нормальное распределение, отсутствие корреляции означает и отсутствие всякой зависимости.

Выборочный коэффициент корреляции не изменяется при изменении начала отсчета и масштаба величин.

Коэффициент корреляции характеризует не всякую зависимость, а только линейную. Линейная вероятностная зависимость случайных величин заключается в том, что при возрастании одной случайной величины другая имеет тенденцию возрастать (или убывать) по линейному закону. Коэффициент корреляции характеризует степень тесноты линейной зависимости. В общем случае, когда величины Y и Z связаны произвольной стохастической зависимостью, коэффициент корреляции может иметь значение в пределах [3]

$$-1 \leq r \leq +1.$$

В таблице 1 [2] показан характер корреляции в зависимости от величины коэффициента корреляции при исследовании экспериментальных данных в химической и пищевой технологии.

Отметим следующие свойства коэффициента корреляции:

- 1) величина r не меняется от прибавления к Z и Y неслучайных слагаемых;
- 2) величина r не меняется от умножения Z и Y на положительные числа;
- 3) если одну из величин, не меняя другой, умножить на -1 , то на -1 умножится и коэффициент корреляции.

Таблица 1. Зависимость характера корреляции от величины коэффициента корреляции.

Коэффициент корреляции $ r $	Характер корреляции
$ r \geq 0,95$	Отлично коррелирующие функции
$0,9 \leq r \leq 0,95$	Хорошо коррелирующие функции
$0,8 \leq r \leq 0,85$	Допустимо коррелирующие функции
$ r \leq 0,8$	Слабо коррелирующие функции

При интерпретации результатов корреляционного анализа нужно иметь в виду, что коэффициент корреляции – статистический показатель. Он не содержит предположения, что изучаемые величины находятся в причинно- следственной связи. Поэтому любая трактовка корреляционной зависимости должна основываться на информации физико–химического характера.

К достоинствам корреляционного анализа можно отнести возможность создания нового правила взаимодействия функций друг с другом, а также оценку взаимодействия функций полученных неизвестным путем.

Недостатками является то, что все результаты, полученные с помощью этой методики можно использовать только в области исследования или близко к ней.

После обнаружения стохастических связей между изучаемыми переменными величинами исследователь приступает к математическому описанию интересующих его зависимости. Другими словами необходимо перейти от корреляционного анализа к регрессионному анализу.

Пример: Сырье, поступающее из ближайшего карьера, содержат два полезных компонента - минералы А и Б. При этом в партиях сырья с повышенным содержанием А обычно обнаруживается и более высокое содержание Б, так что имеются основания ожидать, что эти величины находятся в связи друг с другом. [4]

Анализы 10 образцов сырья, поступившего в разное время, приведены в таблице 2. Необходимо найти коэффициент корреляции.

Таблица 2. Содержание минералов А и Б (в %)

№ образца	1	2	3	4	5	6	7	8	9	10
% А (Z)	67	54	72	64	39	22	58	43	46	34
%Б (Y)	24	15	23	19	16	11	20	16	17	13

Для использования формулы (1) сведем данные в таблицу 3.

Таблица 3. Вычисление коэффициента корреляции

№	Z	Y	$Z - \bar{Z}$	$Y - \bar{Y}$	$(Z - \bar{Z}) \cdot (Y - \bar{Y})$	$(Z - \bar{Z})^2$	$(Y - \bar{Y})^2$
1	67	24	17,1	6,6	112,86	292,41	43,56
2	54	15	4,1	-2,4	-9,84	16,81	5,76
3	72	23	22,1	5,6	123,76	488,41	31,36
4	64	19	14,1	1,6	22,56	198,81	2,56
5	39	16	-10,9	-1,4	15,26	118,81	1,96
6	22	11	-27,9	-6,4	178,56	778,41	40,96
7	58	20	8,1	2,6	21,06	65,61	6,76
8	43	16	-6,9	-1,4	9,66	47,61	1,96
9	46	17	-3,9	-0,4	1,56	15,21	0,16
10	34	13	-15,9	-4,4	69,96	252,81	19,36
Σ	499	174			545,4	2274,9	154,4

Рассчитаем коэффициент корреляции по формуле (1):

$$r = \frac{545,5}{\sqrt{2274,9 \cdot 154,4}} = 0,92.$$

Полученный коэффициент корреляции $r=0,92$ достаточно высок, исходя из того можно сделать вывод о наличии тесной связи между содержанием минералов А и Б в сырье. Следующим шагом является проведение регрессионного анализа.

3.Использование вычислительной техники при проведении корреляционного анализа

При применении специальных программных продуктов для корреляционного анализа можно воспользоваться встроенными функциями.

1) Расчет в Microsoft Office Excel 2000/XP/2003.

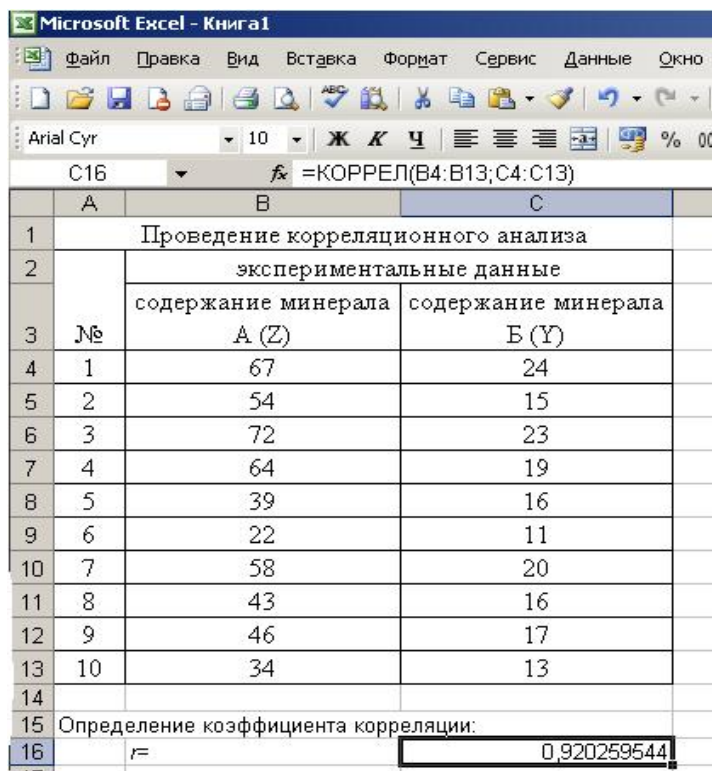
При расчете в Excel применяют встроенную функцию КОРРЕЛ (рис.5). Данная функция имеет следующее описание:

$$\text{КОРРЕЛ}(\text{массив1}; \text{массив2}),$$

где массив1, массив2 – это интервалы значений данных.

При этом надо отметить, что

- если массив1 и массив2 имеют различное значение точек данных, то функция КОРРЕЛ возвращает значение ошибки #Н/Д.
- если массив1 либо массив2 пуст, или если стандартное отклонение их значений равно нулю, то функция КОРРЕЛ возвращает значение ошибки #ДЕЛ/0!.



Проведение корреляционного анализа			
экспериментальные данные			
	содержание минерала		содержание минерала
№	A (Z)	B (Y)	
1	67	24	
2	54	15	
3	72	23	
4	64	19	
5	39	16	
6	22	11	
7	58	20	
8	43	16	
9	46	17	
10	34	13	
Определение коэффициента корреляции:			
r=			0,920259544

Рисунок 5 – Расчет коэффициента корреляции в EXCEL.

2) Расчет в OpenOffice.org Calc 2.3.0. При расчете в Calc применяют встроенную функцию CORREL (рис.6).

Скриншот программы OpenOffice.org Calc. В строке формул введена формула $=CORREL(B4:B13;C4:C13)$. В таблице ниже представлены экспериментальные данные.

	A	B	C
1	Проведение корреляционного анализа		
2	экспериментальные данные		
3	№	содержание минерала A (Z)	содержание минерала B (Y)
4	1	67	24
5	2	54	15
6	3	72	23
7	4	64	19
8	5	39	16
9	6	22	11
10	7	58	20
11	8	43	16
12	9	46	17
13	10	34	13
14			
15	Определение коэффициента корреляции:		
16	r=		0,92025954
17			

Рисунок 6 – Расчет коэффициента корреляции в OpenOffice.org Calc.

Данная функция имеет следующее описание:

$$\text{CORELL}(\text{данные1}; \text{данные2}),$$

где данные1, данные2 – это множества данных.

3) Расчет в MathCAD 14.0.

При расчете в MathCAD применяют встроенную функцию $\text{corr}(A,B)$, где A, B – это массивы данных. Рассмотрим на примере (стр.7).

а) Вводим массивы данных Z и Y (табл.2) (рис.7).

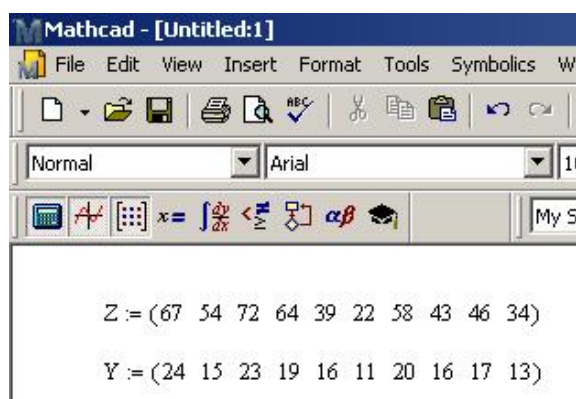
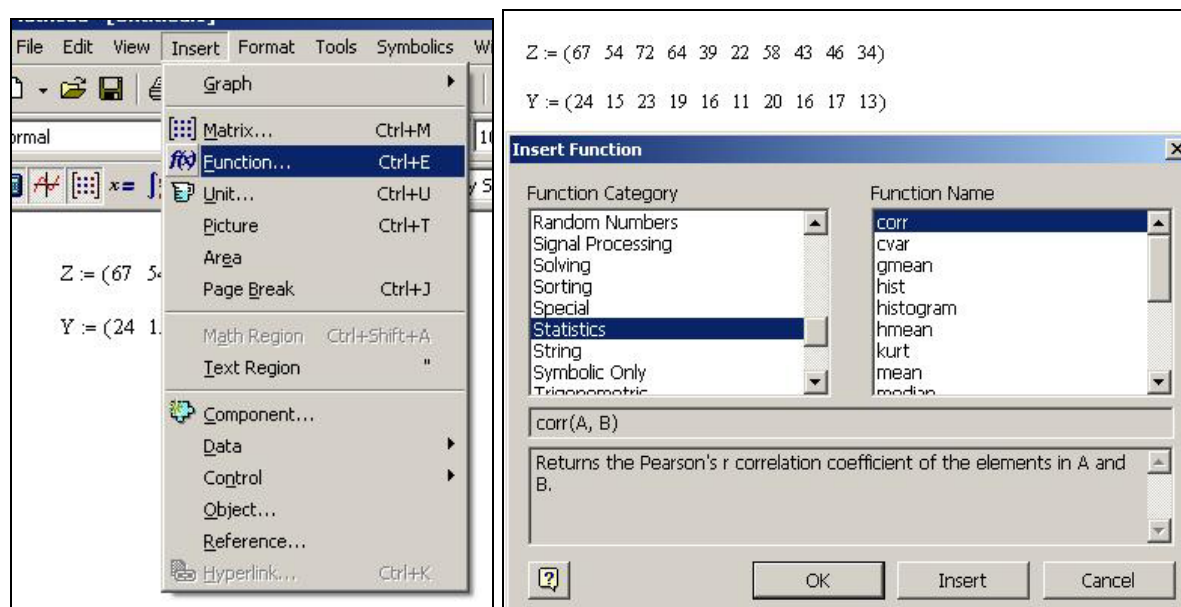


Рисунок 7 - Ввод данных.

б) Добавляем в документ функцию $\text{corr}(A,B)$.

Для этого находим в меню Insert пункт Function (рис. 8, а). После этого находим необходимую функцию (рис.8. b).



a)

b)

Рисунок 8 – Добавление функции corr.

с) Вычисляем коэффициент корреляции (рис.9).

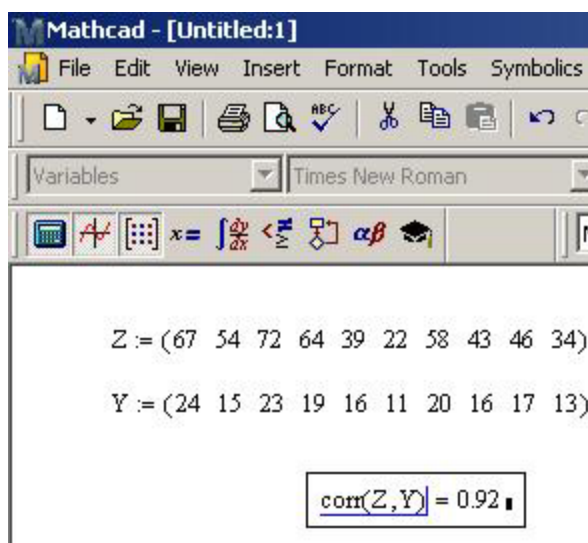


Рисунок – 9 Вычисление коэффициента корреляции.

4) Расчет в программе «Корреляционный анализ».

a) Запустить программу «Корреляционный анализ». На первом шаге предлагается ввести названия исследуемых параметров.

b) Ввод экспериментальных данных X , $Y(X)$, $Z(X)$ (рис.10). В конце ввода

каждого значения нажимается клавиша «ENTER» В качестве разделителя целой и дробной части числа ставится ЗАПЯТАЯ.



КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

ВВЕДИТЕ ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

X **Y(X)** **Z(X)**

10 34 13

ТАБЛИЦА
ЭКСПЕРИМЕНТАЛЬНЫХ ЗНАЧЕНИЙ

X	Y	Z
1	67	24
2	54	15
3	72	23
4	64	19
5	39	16
6	22	11
7	58	20
8	43	16
9	46	17

Рисунок 10– Ввод экспериментальных данных

с) По мере ввода данных можно посмотреть графическую интерпретацию экспериментальных данных, для этого используют кнопку «ПОСТРОИТЬ ЛИНИИ» .

d) Перед расчетом необходимо визуально оценить полученные кривые и ответить на вопрос: коррелируют ли полученные зависимости; после этого программа произведет расчет коэффициента корреляции и делается вывод о существовании корреляционной зависимости для данных кривых (рис.11).

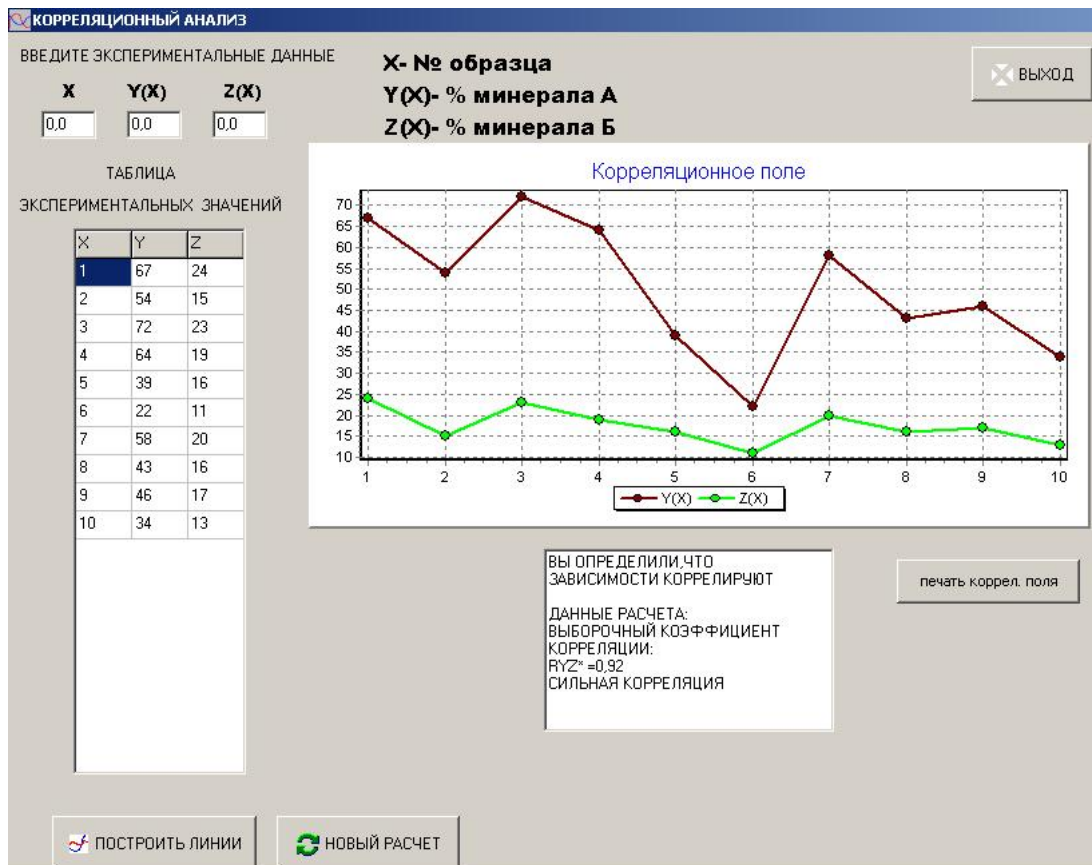


Рисунок 11 – Расчет коэффициента корреляции

е) После анализа полученных значений можно произвести корректировку экспериментальных данных, расположенных в таблице, после чего заново произвести расчет.

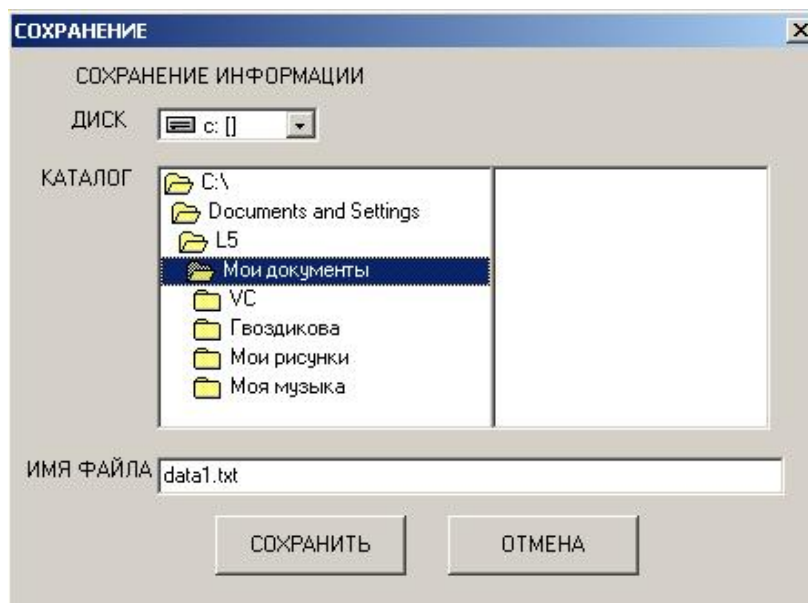


Рисунок 12 – Сохранение данных.

f) При выходе из программы предусмотрена возможность сохранения данных, при этом надо выбрать что именно надо сохранить (диаграмму, экспериментальные данные или коэффициент корреляции). После чего надо ввести имя файла (расширение файла с рисунком - *.bmp, с экспериментальными данными - *.txt), а также выбрать каталог, в котором будет сохранена информация, по умолчанию данные сохраняются в текущей папке (рис.12).

4. Порядок выполнения работы

1.Подготовить экспериментальные данные и получить допуск на проведение лабораторной работы у преподавателя.

3.По экспериментальным данным построить графики исследуемых экспериментальных данных. По взаиморасположению графиков оценить коэффициент корреляции.

4. Занести экспериментальные данные и полученные графики в протокол лабораторной работы.

5. Рассчитать коэффициент корреляции по формуле (1).

6. Провести корреляционный анализ с помощью ЭВМ.

7. Занести результаты анализа в протокол. Сделать выводы о проделанной работе.

4. Контрольные вопросы.

1. Для чего проводится корреляционный анализ?
2. Какую зависимость между величинами характеризует корреляционный анализ?
3. При каком значении коэффициента корреляции можно говорить о существовании корреляции между случайными величинами?
4. По какой формуле можно вычислить коэффициент корреляции?
5. Отметьте достоинства и недостатки корреляционного анализа.

6. В каком случае корреляционная связь прямая, а в каком случае – обратная?
7. Применение корреляционного анализа в инженерной практике.

Список рекомендуемой литературы

1. Математическая статистика: Учеб. для вузов/В.Б.Гориянов, И.В. Павлов, Г.М. Цветкова и др.; Под ред. В.С. Зарубина, А.П. Крищенко. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2001. – 424 с.
2. Ахназарова С.Л., Кафаров В.В. Методы оптимизации эксперимента химической технологии: Учеб. пособие для хим. – технол. спец. вузов. – 2-е изд., перераб. и доп. – М.: Высш. шк., 1985.-327 с.
3. Зельдович Я. Б., Мышкис А. Д. Элементы прикладной математики.- М.:Наука, 1972. –592с.
4. Батунер Л.М., Позин М.Е. Математические методы в химической технике. –Л.: Химия, 1971. – 824 с.

Алексей Николаевич **Гайдадин**
Светлана Анатольевна **Ефремова**
Наталья Николаевна **Бакумова**

ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА
В ТЕХНОЛОГИЧЕСКИХ РАСЧЕТАХ
Методические указания

Редактор

Темплан изданий 2008 г., поз. № .
Формат 60x84 1/16.
Бумага газетная. Печать офсетная. Уч.-изд. л. 1,0.
Подписано в печать Заказ № .

Волгоградский государственный технический университет.
400131, г. Волгоград, пр. им. В. И. Ленина, 28.

РИО РПК «Политехник»
Волгоградского государственного технического университета.
400131, г. Волгоград, ул. Советская, 35.