

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ВОЛЖСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ (ФИЛИАЛ)
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Л.И. Медведева, В.И. Капля

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Электронное учебно-методическое пособие



Волжский

2018

УДК 681.5(07)

ББК 32я73

М 42

Рецензенты:

канд. тех. наук, доцент, и.о. зав. кафедрой
«Автоматизация производственных процессов» ВолгГТУ

А.М. Макаров,

канд. тех. наук, начальник Волжского
филиала ФБУ "Волгоградский ЦСМ"

Е.Н. Бельчанская

Издается по решению редакционно-издательского совета
Волгоградского государственного технического университета

Медведева, Л.И.

Статистическая обработка экспериментальных данных
[Электронный ресурс] : учебно-методическое пособие / Л.И.
Медведева, В.И. Капля ; ВПИ (филиал) ВолгГТУ. – Электрон.
текстовые дан. (1 файл: 8,76 МБ). – Волжский, 2018. – Режим доступа:
<http://lib.volpi.ru>. – Загл. с титул. экрана.

ISBN 978-5-9948-3096-3

Учебно-методическое пособие содержит теоретический материал,
необходимый для выполнения и защиты лабораторных работ, задания для
лабораторных работ и примеры их выполнения.

Предназначено для студентов, обучающихся дисциплине «Статистическая
обработка экспериментальных данных» по направлениям бакалавриата 15.03.04 и
магистратуры 15.04.04 «Автоматизация технологических процессов и производств»
всех форм обучения.

Ил. 18, табл.1, библиограф. : 8 назв.

ISBN 978-5-9948-3096-3

© Волгоградский
государственный
технический
университет, 2018
© Волжский
политехнический
институт, 2018

СОДЕРЖАНИЕ

1. Введение	4
2. Теоретический материал для выполнения и отчета лабораторных работ по дисциплине «Статистическая обработка экспериментальных данных»	6
3. Протоколы лабораторных работ по дисциплине «Статистическая обработка экспериментальных данных»	50
4. Литература	66

ВВЕДЕНИЕ

Изучение дисциплины «Статистическая обработка экспериментальных данных» основано на сочетании теоретического обучения с производственной и научно-исследовательской практикой и взаимосвязи учебного процесса с научной работой через тематику курсовых и дипломных работ. Практические занятия на младших курсах, учебно-исследовательские лабораторные практикумы для общепрофессиональных и специальных дисциплин, научно-исследовательская работа студентов на старших курсах позволяют учащимся приобрести навыки научно-исследовательской работы, обработки и оформления результатов эксперимента, представления результатов работы в форме научного доклада и отчета (научной статьи).

Одним из основных этапов изучения дисциплины и закрепления полученных навыков, является проведение лабораторных работ, которые представляют собой миниатюрную модель обычного цикла научной работы, хотя в ней отсутствует принципиальная новизна полученного результата. Учащийся при выполнении лабораторной работы, практического задания учебно-исследовательского характера, курсовой научно-исследовательской работы, наконец, дипломной работы в меньшей или большей степени должен освоить ряд элементов аналогичной научной работы.

Лабораторная работа, как часть экспериментального исследования, представляет собой небольшой научный отчет, обобщающий проведенную студентом работу, которую представляют для защиты преподавателю. К лабораторным работам предъявляется ряд требований, основным из которых является полное, исчерпывающее описание всей проделанной работы, позволяющее судить о полученных результатах, степени выполнения заданий и профессиональной подготовке студентов [5].

Основными разделами в отчете по лабораторной работе являются следующие пункты.

Титульный лист – это первая страница любой научной работы и для конкретного вида работы заполняется по определенным правилам. Для лабораторной работы титульный лист оформляется следующим образом.

В верхнем поле листа указывают полное наименование учебного заведения и кафедры, на которой выполнялась данная работа.

В среднем поле указывается вид работы, в данном случае лабораторная работа с указанием курса, по которому она выполнена, и ниже ее название. Название лабораторной работы приводится без слова *тема* и в кавычки не заключается.

Далее ближе к правому краю титульного листа указывают фамилию, инициалы, курс и группу учащегося, выполнившего работу, а также фамилию, инициалы, ученую степень и должность преподавателя, принявшего работу.

В нижнем поле листа указывается место выполнения работы и год ее написания (без слова *год*).

Цель работы должна отражать тему лабораторной работы, а также конкретные задачи, поставленные студенту на период выполнения работы. По объему цель работы в зависимости от сложности и многозадачности работы составляет от нескольких строк до 0,5 страницы.

Описание экспериментальной установки и методики эксперимента. В данном разделе приводится схема экспериментальной установки с описанием ее работы и подробно излагается методика проведения эксперимента, процесс получения данных и способ их обработки. Если используются стандартные пакеты компьютерных программ для обработки экспериментальных результатов, то необходимо обосновать возможность и целесообразность их применения, а также подробности обработки данных с их помощью. Для лабораторных работ, связанных с компьютерным моделированием динамических процессов, необходимо в этом разделе описать математическую модель и компьютерные программы, моделирующие данные явления.

Экспериментальные результаты. В этом разделе приводятся непосредственно результаты, полученные в ходе проведения лабораторных работ: экспериментально или в результате компьютерного моделирования определенные значения величин, графики, таблицы, диаграммы.

Анализ результатов работы. Раздел отчета должен содержать подробный анализ полученных результатов, интерпретацию этих результатов на основе изученного теоретического материала. Следует сравнить полученные результаты с известными литературными данными, обсудить их соответствие существующим моделям. Если обнаружено несоответствие полученных результатов и теоретических расчетов или литературных данных, необходимо обсудить возможные причины этих несоответствий.

Выводы. В выводах кратко излагаются результаты работы: полученные экспериментально или теоретически значения физических величин, их зависимости от условий эксперимента или выбранной расчетной модели, указывается их соответствие или несоответствие физическим законам и теоретическим моделям, возможные причины несоответствия.

Отчет по лабораторной работе оформляется на писчей бумаге стандартного формата А4 на одной стороне листа, которые сшиваются в скоросшивателе или переплетаются. Если по специальному лабораторному практикуму требуется оформить в конце семестра общий отчет по всему циклу лабораторных работ, посвященных исследованию одного и того материала разными методами, оформляются также и отдельные отчеты по каждой работе цикла по мере их выполнения. На основе отчетов по каждой работе в конце семестра оформляется итоговый отчет, в котором основное внимание должно быть уделено анализу результатов, полученных в разных лабораторных работах.

ТЕОРЕТИЧЕСКИЙ МАТЕРИАЛ ДЛЯ ВЫПОЛНЕНИЯ И ОТЧЕТА ЛАБОРАТОРНЫХ РАБОТ ПО ДИСЦИПЛИНЕ «СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Глава 1. Основные понятия и определения дисциплины «Статистическая обработка экспериментальных данных»

Эксперимент – система операций, воздействий и (или) наблюдений, направленных на получение информации об объекте при исследовательских испытаниях.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям.

План эксперимента – совокупность данных, определяющих число, условия и порядок реализации опытов.

Опыт – воспроизведение исследуемого явления в определенных условиях проведения эксперимента при возможности регистрации его результатов [1].

Фактор – переменная величина, по предположению влияющая на проведение эксперимента.

Факторное пространство – пространство, координатные оси которого соответствуют значениям факторов.

Рассматривая в качестве исследуемой системы баллон, заполненный газом, факторами можно считать основные термодинамические параметры: давление P , объем газа V и температуру T . Каждому состоянию такой системы соответствует точка в факторном пространстве $\{P, V, T\}$. Каждому термодинамическому процессу соответствует линия в факторном пространстве $\{P, V, T\}$ (рис. 1, а, б, в).

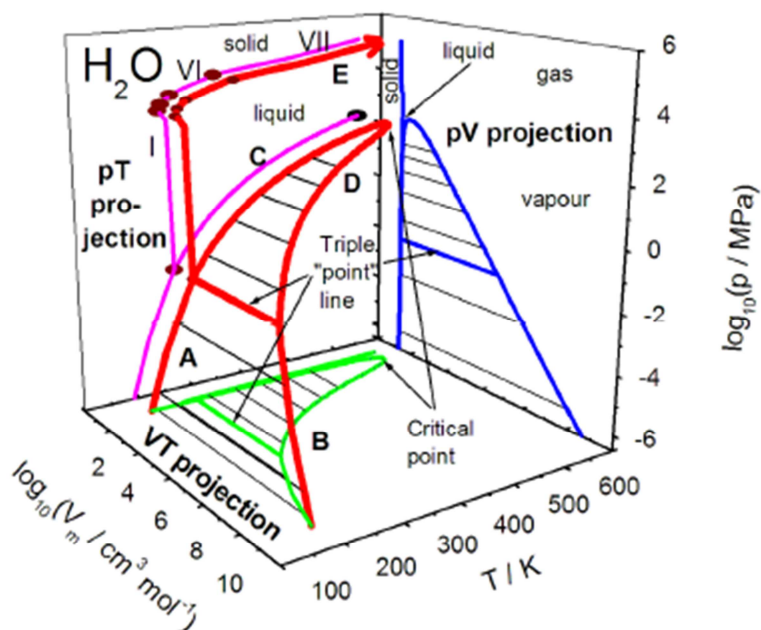


Рисунок – 1, а. Факторное пространство для баллона, заполненного водой.

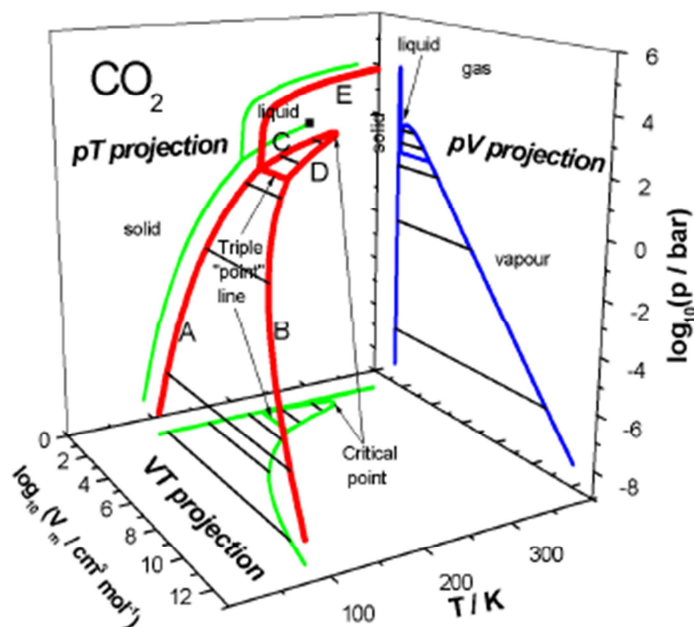


Рисунок – 1, б. Факторное пространство для баллона, заполненного углекислым газом.

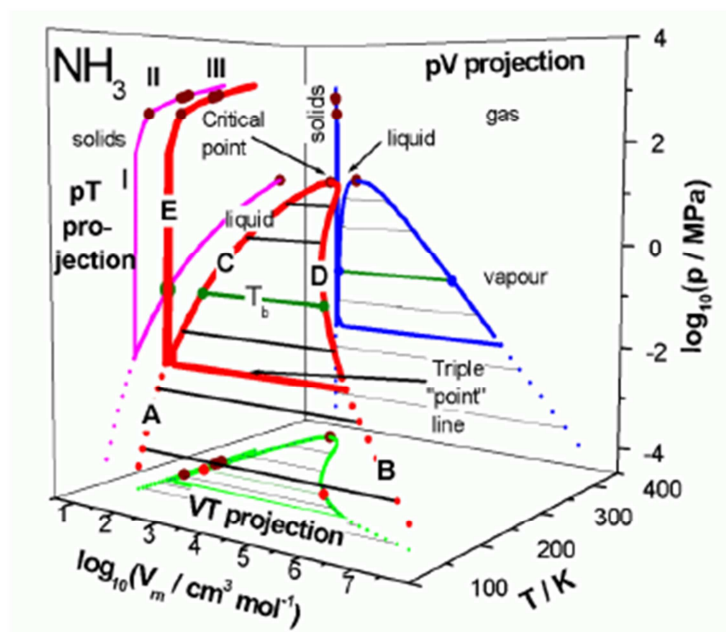


Рисунок – 1, в. Факторное пространство для баллона, заполненного метаном.

Диапазон экспериментальных значений факторов ограничен физически и может дополнительно ограничиваться экспериментатором. В связи с этим используют следующие термины.

Область экспериментирования (область планирования) – область факторного пространства, где могут размещаться точки, отвечающие условиям проведения опытов.

Размах варьирования фактора – разность между максимальными и минимальными натуральными значениями фактора в данном экспериментальном плане.

Интервал варьирования фактора – половина размаха варьирования фактора.

Цель планирования эксперимента – минимизация общего объема испытаний при соблюдении требований к достоверности и точности получаемых результатов.

План эксперимента представляет собой множество точек факторного пространства (множества допустимых значений факторов), заданных по определенному правилу, соответствующему цели исследований.

Точка плана – упорядоченная совокупность численных значений факторов, соответствующая условиям проведения опыта.

Цель плана (центральная точка плана) – точка факторного пространства, соответствующая средним значениям факторов.

Уровень фактора – фиксированное значение фактора относительно начала отсчета.

Основной уровень фактора – натуральное значение фактора, соответствующее нулю в безразмерной шкале.

Активный эксперимент – эксперимент, в котором уровни факторов в каждом опыте задаются исследователем.

Пассивный эксперимент – эксперимент, при котором уровни факторов в каждом опыте регистрируются исследователем, но не задаются [1].

Последовательный эксперимент – эксперимент, реализуемый в виде серий, в котором условия проведения каждой последующей серии определяются результатами предыдущей.

При планировании эксперимента выполняют *нормализацию факторов* – преобразование натуральных значений факторов в безразмерные.

Компоненты вектора управляющих факторов обычно формируют так, чтобы диапазон их варьирования соответствовал интервалу $[-1; +1]$. Формула такого нормирования:

$$z_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \cdot 2 - 1,$$

где i – номер фактора ($i=1, 2, \dots k$);

k – количество факторов.

Матрица плана – стандартная форма записи условий проведения эксперимента в виде прямоугольной таблицы, строки которой отвечают опытам, столбцы – факторам (рис. 2).

Обычно формируют матрицу спектр плана.

Спектр плана – совокупность всех точек плана, отличающихся уровнями хотя бы одного фактора.

Матрица спектр плана – матрица, составленная из всех строк матрицы плана, отличающихся уровнями хотя бы одного фактора.

n	z_1	z_2	z_3	y
1	-1	-1	-1	y_1
2	+1	-1	-1	y_2
3	-1	+1	-1	y_3
4	+1	+1	-1	y_4
5	-1	-1	+1	y_5
6	+1	-1	+1	y_6
7	-1	+1	+1	y_7
8	+1	+1	+1	y_8

Рисунок – 2. Матрица плана.

Количество параллельных опытов для каждого сочетания факторов заносят в *матрицу дублирования* – квадратную диагональную матрицу, диагональные элементы которой равны числам параллельных опытов в соответствующих точках спектра плана.

Полный факторный план (ПФП) – план, содержащий все возможные комбинации всех факторов на определенном числе уровней равное количество раз.

Если уровней два, то ПФП состоит из $2^k m$ опытов (m – число параллельных опытов в каждой точке плана). В этом случае точки факторного пространства, в которых проводятся опыты, являются вершинами k -мерного куба (рис. 3).

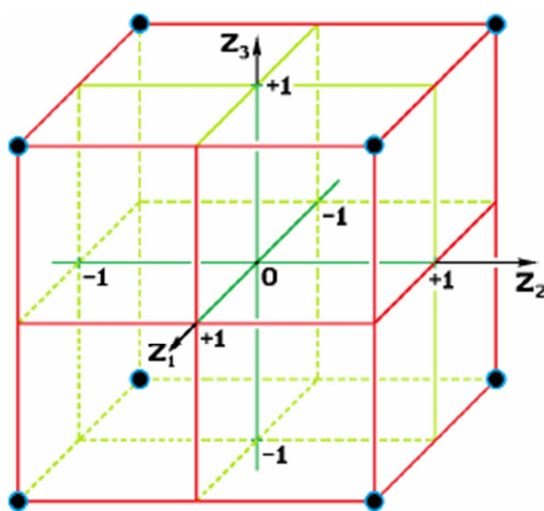


Рисунок – 3. Геометрическая интерпретация ПФП при $k=3$, $u=2$.

При $k \leq 3$ ПФП представляется набором проекций k -мерного куба на координатные плоскости $z_i 0 z_j$, где $i, j=1, 2, \dots, k$ и $i \neq j$ (рис. 4).

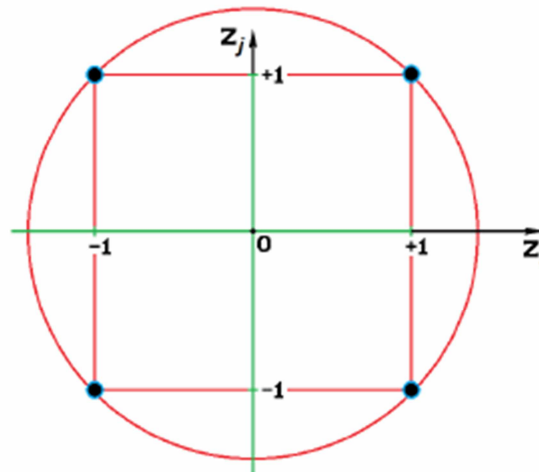


Рисунок – 4. Проекция ПФП на координатную плоскость z_iOz_j .

План полного факторного эксперимента представляет собой множество всех возможных сочетаний уровней факторов. В общем случае, когда каждый фактор ПФП имеет u уровней градации, каждый ПФП содержит u^k опытов.

Расширенная матрица ПФП позволяет определить коэффициенты нелинейного уравнения регрессии с учетом эффекта парного влияния факторов на результат эксперимента (Рис. 5).

n	z_1	z_2	z_3	$z_1 z_2$	$z_1 z_3$	$z_2 z_3$	y
1	-1	-1	-1	+1	+1	+1	y_1
2	+1	-1	-1	-1	-1	+1	y_2
3	-1	+1	-1	-1	+1	-1	y_3
4	+1	+1	-1	+1	-1	-1	y_4
5	-1	-1	+1	+1	-1	-1	y_5
6	+1	-1	+1	-1	+1	-1	y_6
7	-1	+1	+1	-1	-1	+1	y_7
8	+1	+1	+1	+1	+1	+1	y_8

Рисунок – 5. Расширенная матрица ПФП.

Центральный композиционный план представляет собой комбинацию ПФП для двух уровней факторов (± 1), дополненный центральной точкой факторного пространства и вариантами изменения факторов по одному на уровнях ($\pm \alpha$), где α определяется из уравнения:

$$4 \cdot \alpha^4 + 4 \cdot 2^k \cdot \alpha^2 - 2^k \cdot (2k + 1) = 0.$$

Аналитической формой плана экспериментов служит матрица планирования, которая образуется из транспонированных векторов регрессоров, вычисленных для различных запланированных N точек факторного пространства:

$$F = (r(X_1)^T, r(X_2)^T, \dots, r(X_N)^T)^T = \begin{pmatrix} r(X_1)_1 & \dots & r(X_1)_p \\ \dots & \dots & \dots \\ r(X_N)_1 & \dots & r(X_N)_p \end{pmatrix},$$

где $r(X_j)_i$ – компонента с номером i вектора в j -той точке факторного пространства;

$X_1, X_2, \dots, X_N$ – множество точек факторного пространства, образующее выбранный план эксперимента.

Матрица центрального композиционного плана второго порядка для трех факторов (строки соответствуют различным точкам факторного пространства, столбцы – искомым параметрам модели) имеет вид:

$$F = \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -\alpha & 0 & 0 & 0 & 0 & 0 & \alpha^2 & 0 & 0 \\ 1 & \alpha & 0 & 0 & 0 & 0 & 0 & \alpha^2 & 0 & 0 \\ 1 & 0 & -\alpha & 0 & 0 & 0 & 0 & 0 & \alpha^2 & 0 \\ 1 & 0 & \alpha & 0 & 0 & 0 & 0 & 0 & \alpha^2 & 0 \\ 1 & 0 & 0 & -\alpha & 0 & 0 & 0 & 0 & 0 & \alpha^2 \\ 1 & 0 & 0 & \alpha & 0 & 0 & 0 & 0 & 0 & \alpha^2 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Независимыми являются 1-4 столбики, а остальные сформированы как линейные комбинации первых четырех столбиков.

При большом количестве многоуровневых факторов ПФП не осуществим. В таких случаях составляют *дробный факторный план* – план, содержащий часть комбинаций полного факторного плана. Выбор комбинаций факторов осуществляют с помощью *генератора плана* – алгебраического выражения, используемого при построении дробного факторного плана.

Латинский план – план проведения эксперимента с одним первичным фактором и несколькими вторичными.

Латинский квадрат – план дисперсионного анализа, задаваемый расположением некоторого числа символов в ячейках, сгруппированных в строки и столбцы так, что каждый символ встречается один раз в каждой строке и каждом столбце.

Введем обозначения: А – первичный фактор, В и С – вторичные факторы. Пусть каждый из рассматриваемых факторов имеет четыре уровня градации, тогда латинский план представляется таблицей:

	C_1	C_2	C_3	C_4
B_1	A_1	A_2	A_3	A_4
B_2	A_2	A_3	A_4	A_1
B_3	A_3	A_4	A_1	A_2
B_4	A_4	A_1	A_2	A_3

Если каждый из трех факторов латинского квадрата имеет u уровней градации, то латинский план содержит u^2m опытов. В общем случае, когда каждый из k факторов латинского плана имеет u уровней, латинский план содержит $u^{k-1}m$ опытов. Латинский квадрат позволяет сократить количество опытов в u раз по сравнению с ПФП.

При варьировании более трех факторов составляют *латинский куб* – план дисперсионного анализа, задаваемый расположением некоторого числа символов в квадратах из строк и столбцов так, что каждый символ встречается одинаковое число раз в каждом квадрате.

Существенное сокращение объема испытаний возможно в результате *рандомизации плана эксперимента* – одного из приемов планирования эксперимента, имеющего целью свести эффект некоторого неслучайного фактора к случайной ошибке. Обычно план эксперимента соответствует равномерному распределению выбранных точек в факторном пространстве.

Свойства плана:

- *ортogonalность плана* – свойство плана, при котором матрица моментов для заданной модели является диагональной;
- *ротатабельность плана* – свойство плана, при котором дисперсия оценки функции отклика зависит только от расстояния от центра плана;
- *композиционность плана* – свойство плана, позволяющее выполнять эксперимент последовательно, переходя от более простых моделей к более сложным;
- *насыщенность плана* – свойство плана, задающееся разностью между числом точек спектра плана и числом оцениваемых параметров модели.

Глава 2. Метод случайного баланса

Метод случайного баланса (МСБ) – метод отсеивания факторов, основанный на использовании сверх насыщенных планов со случайным уровнем выборов сочетаний уровней факторов.

Метод случайного баланса использует перенасыщенные планы, при которых выполняется условие $u+1 > N$, где u – общее количество точек используемых переменных, N – число строк матрицы планирования, соответствующее числу точек факторного пространства, в которых ставятся

опыты. Применение МСБ основано на априорном знании того, что из всей совокупности рассматриваемых факторов только небольшое их число являются действительно существенными, остальные же могут быть отнесены к шумовому полю.

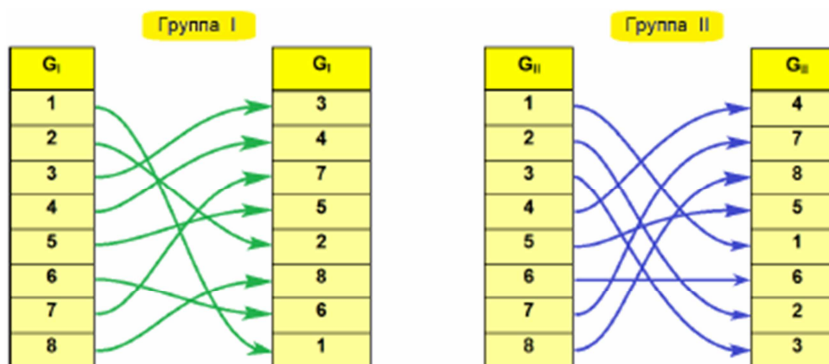
Априори – знание, полученное до опыта и независимо от него.

Этапы МСБ:

- 1) Факторы z_i ($i=1, 2, \dots, k$) разбивают на группы. Заведомо взаимодействующие факторы стремятся включить в одну группу. Если нет априорных сведений о физике процесса, то разбивку факторов по группам можно производить формально, с использованием таблицы случайных чисел. В этом случае предварительно составляют пронумерованный список факторов, а затем с помощью таблицы случайных чисел каждому фактору присваивают свой случайный порядковый номер, после чего факторы случайным образом объединяют в группы.
- 2) Для каждой группы составляют матрицу планирования. Все групповые матрицы планирования должны иметь одинаковое количество строк, чтобы их можно было состыковать.

Группа I					Группа II			
G_I	z_1	z_2	z_3		G_{II}	z_4	z_5	z_6
1	-1	-1	-1		1	-1	-1	-1
2	+1	-1	-1		2	+1	-1	-1
3	-1	+1	-1		3	-1	+1	-1
4	+1	+1	-1		4	+1	+1	-1
5	-1	-1	+1		5	-1	-1	+1
6	+1	-1	+1		6	+1	-1	+1
7	-1	+1	+1		7	-1	+1	+1
8	+1	+1	+1		8	+1	+1	+1

- 3) Выполняют рандомизацию строк матриц полного факторного эксперимента с помощью таблицы случайных чисел. Вводят нумерацию строк общей матрицы планирования.



- 4) Составляют общую матрицу планирования путем построчной стыковки матриц планирования всех групп после рандомизации.

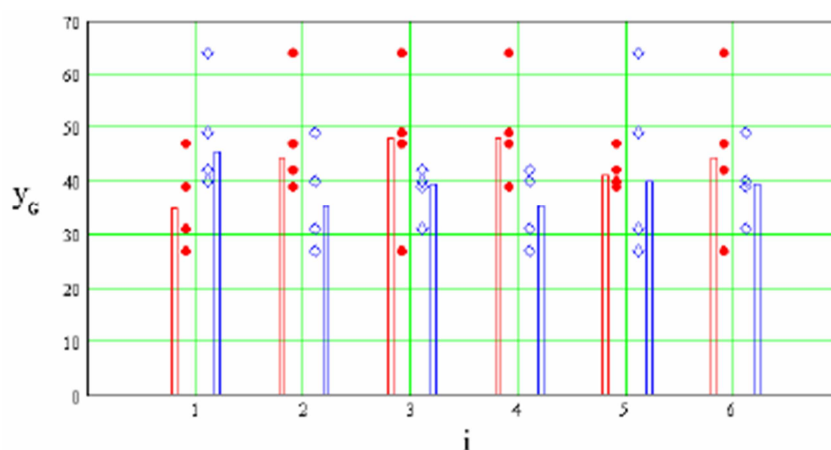
	Группа I				Группа II			
G	G _I	z ₁	z ₂	z ₃	G _{II}	z ₄	z ₅	z ₆
1	3	-1	+1	-1	4	+1	+1	-1
2	4	+1	+1	-1	7	-1	+1	+1
3	7	-1	+1	+1	8	+1	+1	+1
4	5	-1	-1	+1	5	-1	-1	+1
5	2	+1	-1	-1	1	-1	-1	-1
6	8	+1	+1	+1	6	+1	-1	+1
7	6	+1	-1	+1	2	+1	-1	-1
8	1	-1	-1	-1	3	-1	+1	-1

5) Вычисляют «медиану слева» и «медиану справа» для каждого фактора:

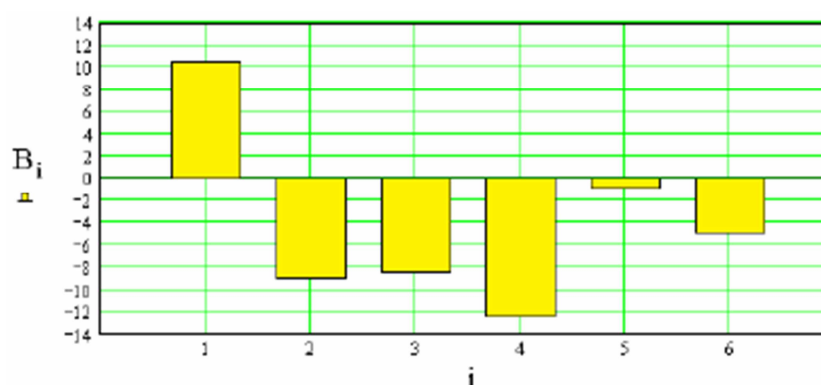
$$M_i^- = Me\{y_n\}_{z_{n,i}=-1}$$

$$M_i^+ = Me\{y_n\}_{z_{n,i}=+1}$$

6) Строят диаграмму рассеяния результатов эксперимента.



7) Определяют вклад каждого фактора: $B_i = M_i^+ - M_i^-$.



8) По вкладу факторов выявляют первичные и вторичные факторы.

Визуальное и численное сравнение вкладов дает возможность обнаружить наиболее существенные факторы: их вклады являются наибольшими по модулю.

Но в ряде случаев абсолютная величина вкладов не является достаточным критерием наибольшей существенности факторов. Поэтому используют дополнительные критерии: коррелированность вектор-столбцов, наличие, число и характер выделяющихся точек.

Глава 3. Метод наименьших квадратов

Метод наименьших квадратов (МНК) – математический метод, основанный на минимизации сумм квадратов отклонений некоторый функций от искомым переменных [2, 7].

МНК применяется для решения:

- 1) переопределенных систем уравнений, в которых количество уравнений превышает количество неизвестных;
- 2) нелинейных систем уравнений.

МНК используют при аппроксимации точечных экспериментальных значений некоторой функцией (рис. 6, а, б).

Аппроксимация – замена одних математических объектов другими, близкими к исходным.

Аппроксимация позволяет исследовать числовые характеристики и качественные свойства объекта, сводя задачу к изучению более простых или более удобных объектов, характеристики которых легко вычисляются и свойства которых хорошо известны.

Аппроксимационная функция – основа регрессионной модели объекта или процесса.

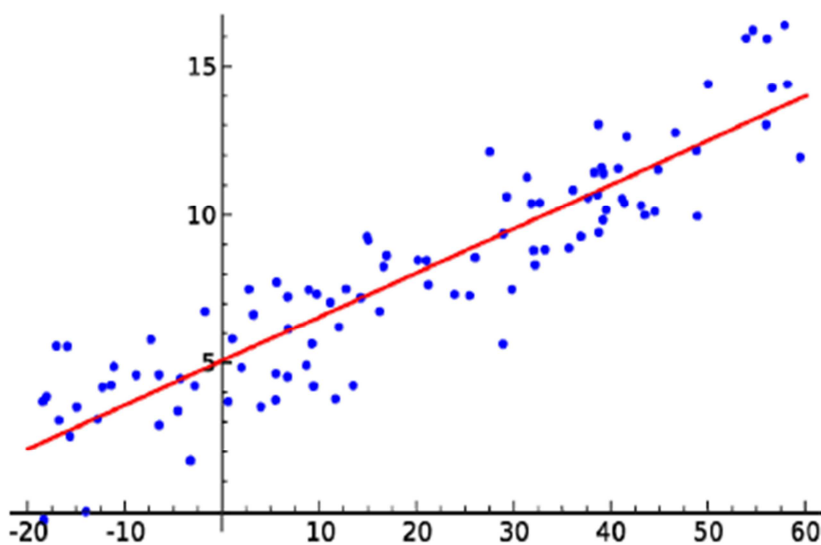


Рисунок – 6, а. Аппроксимация линейной функции.

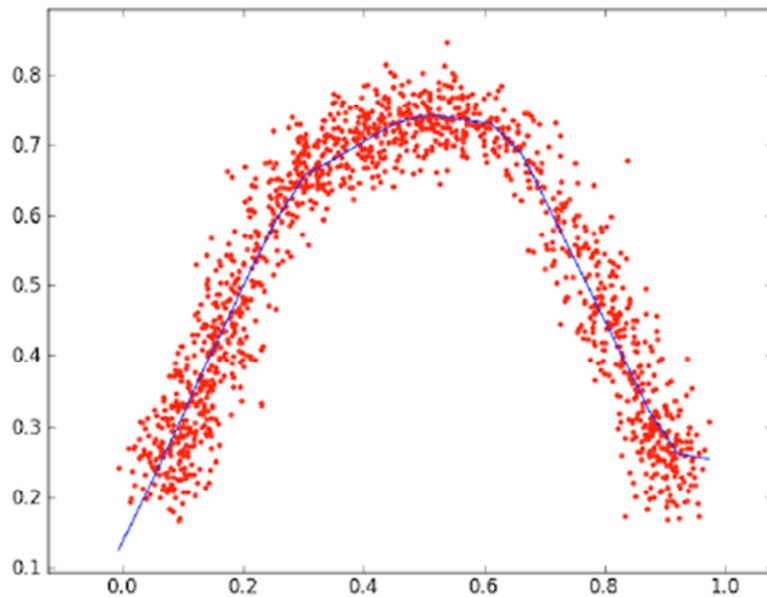


Рисунок – 6, б. Аппроксимация нелинейной функции.

МНК один из базовых методов регрессионного анализа результатов эксперимента для оценки неизвестных параметров регрессионных моделей объектов по выборочным данным.

Поиск комбинации значений параметров аппроксимационной функции приводит к минимизации целевой функции:

$$F(a_1, a_2, \dots, a_k) = \sum_{n=1}^N (\tilde{Y}(x_n; a_1, a_2, \dots, a_k) - y_n)^2, \quad (1)$$

где n – номер опыта (номер пары результатов измерения величин X и Y);

$\{x_n, y_n\}$ – результат n -го опыта ($n=1, 2, \dots, N$);

$\tilde{Y}(x; a_1, a_2, \dots, a_k)$ – аппроксимационная функция;

$A = [a_1, a_2, \dots, a_k]^T$ – вектор неизвестных параметров аппроксимационной функции.

Мерой точности можно считать значение целевой функции в найденной точке $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k)$ пространства параметров. Обычно точность аппроксимации характеризуют среднеквадратическим отклонением аппроксимационной функции от экспериментальных точек:

$$\sigma_Y = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^N (\tilde{Y}(x_n; \hat{a}_1, \hat{a}_2, \dots, \hat{a}_k) - y_n)^2}. \quad (2)$$

Среднеквадратичное отклонение (2) измеряется в единицах измерения случайной величины, для которой оно определяется. Функциональная зависимость среднеквадратического отклонения $\sigma_Y(a_1, a_2, \dots, a_k)$ от

параметров аппроксимационной функции тоже может использоваться в качестве целевой функции. Целевая функция (1) имеет более простой вид, но имеет квадратичную размерность.

Глава 4. Интерполяция результатов эксперимента.

Интерполяция – приближенное или точное нахождение какой-либо величины по известным отдельным значениям этой величины или других, связанных с ней.

Интерполяционная формула – формула для приближенного вычисления значений таблично-заданной функции.

Интерполяция функции $y(x)$ в отрезке $[a, b]$ по значениям в узлах x_n сетки сводится к нахождению функции $L_N(x)$, такой, что $L_N(x_n) = y(x_n)$, где $n=0, 1, 2 \dots N$. Обычно функция $L_N(x)$ строится в виде:

$$L_N(x) = \sum_{k=0}^N \alpha_k \cdot \varphi_k(x), \quad (3)$$

где $L_N(x)$ – интерполяционный полином;

$\varphi_k(x)$ – некоторая заранее выбранная система линейно независимых функций.

Чаще всего используется алгебраическая интерполяция. В этом случае $\varphi_k(x) = x^k$, а интерполяционный полином имеет вид алгебраического многочлена:

$$L_N(x) = \sum_{k=0}^N \alpha_k \cdot x^k.$$

Простейший вариант интерполяции – линейная интерполяция с двумя узлами x_n и x_{n+1} . В этом случае интерполяционный полином определяется формулой:

$$L_1(x) = \frac{x - x_n}{x_{n+1} - x_n} \cdot [f(x_{n+1}) - f(x_n)] + f(x_n), \quad (4)$$

где $x_n \leq x \leq x_{n+1}$.

В случае квадратичной интерполяции с тремя узлами интерполяционный полином имеет вид:

$$\begin{aligned}
L_2(x) = & \frac{(x - x_n) \cdot (x - x_{n+1})}{(x_{n-1} - x_n) \cdot (x_{n-1} - x_{n+1})} \cdot f(x_{n-1}) + \\
& + \frac{(x - x_{n-1}) \cdot (x - x_{n+1})}{(x_n - x_{n-1}) \cdot (x_n - x_{n+1})} \cdot f(x_n) + \\
& + \frac{(x - x_{n-1}) \cdot (x - x_n)}{(x_{n+1} - x_{n-1}) \cdot (x_{n+1} - x_n)} \cdot f(x_{n+1})
\end{aligned}
\tag{5}$$

где $x_{n-1} \leq x \leq x_{n+1}$.

Порядок интерполяционной функции в общем случае на единицу меньше количества экспериментальных точек, по которым она строится.

Интерполяционный многочлен Лагранжа – многочлен минимальной степени, принимающий заданные значения в заданном наборе точек: $L(x_n) = y_n$. Многочлен Лагранжа представляют в виде:

$$L_N(x) = \sum_{n=0}^N y_n \cdot \Lambda_n(x) \tag{6}$$

$$\Lambda_n(x) = \prod_{\substack{s=0 \\ s \neq n}}^N \frac{x - x_s}{x_n - x_s}$$

где – базисные полиномы.

Многочлен $L(x_n)$ также как и $\Lambda_n(x)$ может иметь степень не больше N .

Наиболее современной является интерполяция сплайнами. *Сплайн* – это функция, область определения которой разбита на конечное число отрезков, на каждом из которых сплайн совпадает с некоторым алгебраическим полиномом. Максимальная степень из использованных полиномов называется *степенью сплайна*. Разность между степенью сплайна и порядком гладкости называется *дефектом сплайна*. Например, ломаная есть сплайн степени 1 и дефекта 1.

Интерполяционный сплайн в общем случае имеет вид:

$$S_n(x) = a_0 + a_1 \cdot x + \dots + a_{n-1} \cdot x^{n-1} + \sum_{k=0}^{N-1} c_k \cdot (x - x_k)_+^n, \tag{7}$$

$$t_+ = \begin{cases} t, & \text{если } t \geq 0, \\ 0, & \text{если } t < 0, \end{cases} \quad x_0 < x_1 < \dots < x_{N-1} < x_N$$

где

c_k – действительные числа.

Чаще всего используются параболические или кубические сплайны. Интерполяционным кубическим сплайном дефекта 1 для функции $y(x)$ относительно сетки Δ_N называют функцию $S_3(x)$ – многочлен третьей степени на каждом из отрезков $[x_{k-1}; x_k]$, принадлежащую классу дважды непрерывно дифференцируемых функций и удовлетворяющую условиям:
 $S_3(x_k) = y(x_k)$, для $k = 0, 1, \dots, N$; $N \geq 2$.

В качестве аргумента x и функции y могут анализироваться любые физические величины. При интерполяции вольтамперной характеристики (ВАХ) туннельного диода $I_{пр}$ – прямой ток, протекающий через туннельный диод при напряжении $U_{пр}$ (рис. 7).

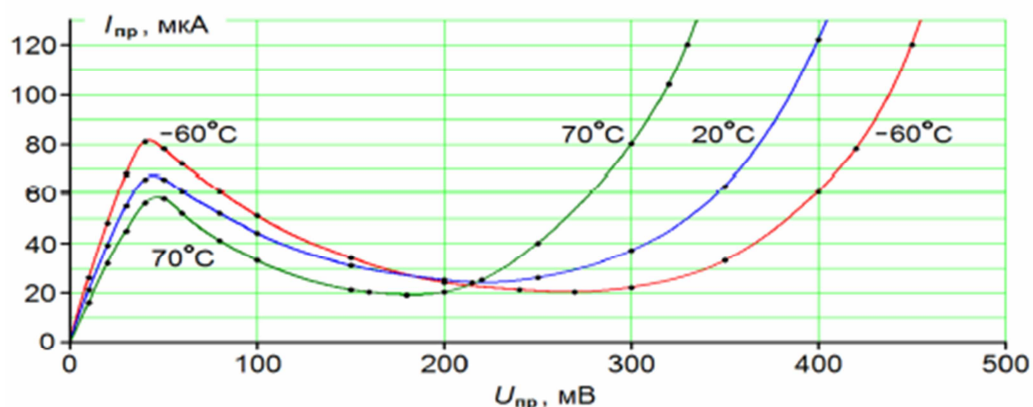


Рисунок – 7. Результат интерполяции кубическими сплайнами ВАХ туннельного диода ГИ403А при различных температурах.

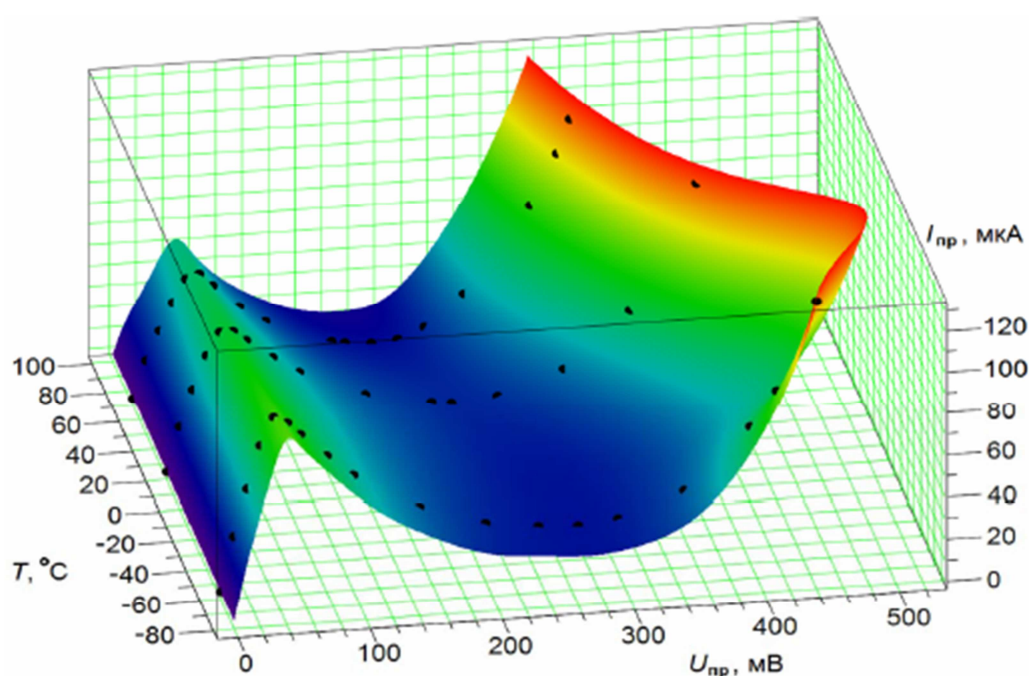


Рисунок – 8. Интерполяционная поверхность $I_{пр}(U_{пр}, T)$ для туннельного диода ГИ403А.

При обработке результатов эксперимента с участием нескольких варьируемых факторов применяют многомерную аппроксимацию или интерполяцию. На рисунке 8 показана интерполяционная поверхность $I_{пр}(U_{пр}, T)$, построенная по трем наборам экспериментальных точек, соответствующим трем уровням температуры.

При построении рассмотренных графиков кроме интерполяции выполнена и *экстраполяция* результатов эксперимента – то есть продолжение функции за пределы ее области определения. Экстраполяция в широких диапазонах, значительно превышающих область экспериментирования, может привести к ошибочным результатам. Характеристики любого объекта имеют физические ограничения, которые должны учитываться при экстраполяции.

Глава 5. Регрессионный анализ результатов эксперимента.

Регрессионный анализ – статистический метод анализа и обработки экспериментальных данных при воздействии на отклик только количественных факторов, основанный на сочетании метода наименьших квадратов и техники статистической проверки гипотез.

Отклик – наблюдаемая случайная переменная, по предположению, зависящая от факторов.

Функция отклика – зависимость математического ожидания отклика от факторов.

Оценка функции отклика – зависимость, получаемая при подстановке в функцию отклика оценок значений ее параметров.

Поверхность отклика – геометрическое представление функции отклика.

Цель регрессионного анализа – планирование экспериментов для получения линейной или неполной степенной математической модели статистики сложного объекта (системы). Вариация входных воздействий на объект и регистрация соответствующих состояний объекта позволяет получить систему уравнений для определения параметров математической модели объекта.

Регрессионная модель – зависимость отклика от количественных факторов и ошибок наблюдения отклика. Регрессионная модель объекта задается уравнением:

$$Y(X) = \eta(X, B) + \varepsilon,$$

где $Y(X)$ – функция отклика (измеряемая реакция объекта),

$X = (x_1, x_2, \dots, x_k)^T$ – вектор факторов (входных переменных),

$\eta(X, B)$ – параметрическая функция регрессии,

$B = (\beta_1, \beta_2, \dots, \beta_\Psi)^T$ – вектор параметров, определяемых по результатам

- измерений,
 ψ – количество параметров,
 k – количество факторов,
 $\square$ – случайные воздействия на исследуемый объект.

Модель называется линейной регрессионной моделью, если она линейна относительно компонент вектора параметров. Уравнение линейной регрессионной модели имеет вид:

$$Y(X) = r(X)^T \cdot B + \varepsilon = \sum_{i=1}^{\Psi} \beta_i \cdot r(X)_i + \varepsilon$$

где $r(X)$ – вектор регрессоров.

Уравнение линейной регрессионной модели для трех факторов:

$$\hat{y}(x_1, x_2, x_3) = a_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 \quad (8)$$

Вектор параметров уравнения (8):

$$B = (a_0, b_1, b_2, b_3)^T.$$

Вектор регрессоров уравнения первого порядка:

$$r(X) = (1, x_1, x_2, x_3)^T.$$

Первый элемент вектора регрессоров равен единице, так как отсутствие этого элемента влечет за собой равенство $\hat{y}(0) = 0$ для любых значений параметров, что недопустимо в общем случае. Количество элементов вектора регрессоров равно количеству элементов вектора параметров.

Модель регрессионного анализа второго порядка (*квадратичная модель*) – регрессионная модель, задаваемая полиномом второго порядка по факторам. Уравнение регрессии второго порядка:

$$\hat{y}(X) = a_0 + \sum_{i=1}^k b_i \cdot x_i + \sum_{i < j} c_{i,j} \cdot x_i \cdot x_j + \sum_{i=1}^k d_i \cdot x_i^2,$$

где $\hat{y}$ – расчетное значение отклика системы,

$\{a, b, c, d\}$ – коэффициенты уравнения регрессии, которые требуется определить.

Уравнение регрессии второго порядка для трех факторов:

$$\begin{aligned} \hat{y}(x_1, x_2, x_3) = & a_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \\ & + c_{1,2} \cdot x_1 \cdot x_2 + c_{1,3} \cdot x_1 \cdot x_3 + c_{2,3} \cdot x_2 \cdot x_3 + \\ & + d_1 \cdot x_1^2 + d_2 \cdot x_2^2 + d_3 \cdot x_3^2. \end{aligned}$$

Вектор параметров уравнения регрессии имеет вид:

$$B = (a_0, b_1, b_2, b_3, c_{1,2}, c_{1,3}, c_{2,3}, d_1, d_2, d_3)^T.$$

Вектор регрессоров уравнения второго порядка:

$$r(X) = (1, x_1, x_2, x_3, x_1 \cdot x_2, x_1 \cdot x_3, x_2 \cdot x_3, x_1^2, x_2^2, x_3^2)^T.$$

Форма задания уравнения регрессии в общем случае произвольна. Полиномиальная форма соответствует разложению в ряд Тейлора уравнения регрессии в окрестностях точки a_0 .

Условия задачи множественного регрессионного анализа:

1. Результаты наблюдений $y = (y_1, y_2, \dots, y_N)^T$ должны представлять собой независимые, нормально распределенные случайные величины.
2. Выборочные оценки дисперсии S_y^2 однородны.
3. Погрешность измерения факторов $X = (x_1, x_2, \dots, x_k)^T$ пренебрежительно мала относительно их диапазона измерения.

Этапы реализации регрессионного анализа:

1. Планирование экспериментов.
2. Проведение экспериментов на исследуемой схеме.
3. Получение математической модели системы.
4. Проверка воспроизводимости экспериментов.
5. Проверка адекватности математического описания.

Глава 6. Дисперсионный анализ результатов эксперимента.

В практической деятельности встречаются объекты исследования, состояние которых описывается входными факторами, не имеющими количественного описания. Такими факторами могут быть переменные, которые не доступны для измерения, однако, их уровни варьирования можно произвольно выбирать и фиксировать во времени. Например, на качество изготавливаемой продукции влияют комплектующие, которые могут поступать от разных поставщиков, производственное оборудование однотипных производственных участков, технология изготовления, конкретные работники различных участков и смен.

Постановка задач дисперсионного анализа предполагает известным:

- отклик Y зависит от k независимых управляемых факторов $X = (x_1, x_2, \dots, x_k)^T$, не имеющих количественного описания;
- каждый фактор x_j может варьироваться на u_j уровнях;
- полный факторный эксперимент состоит из $N = u_1 \cdot u_2 \cdot \dots \cdot u_k$ серий независимых наблюдений;
- каждая j -тая серия содержит m_j наблюдений параллельных опытов.

Требуется определить меру влияния каждого фактора на отклик Y системы и выделить наиболее существенные факторы.

Дисперсионный анализ базируется на следующих допущениях:

- результаты измерений выходного сигнала Y характеризуются нормальным распределением;
- дисперсия единичного наблюдения σ_{ε}^2 обусловлена случайными ошибками $\square$, постоянна во всех опытах и не зависит от сочетания входных факторов.

В качестве показателя влияния фактора X принимают *дисперсию фактора* – то есть дисперсию выходной величины, обусловленную вариацией данного входного фактора:

$$\sigma_X^2 = \frac{1}{u} \sum_{i=1}^u (y_i - \bar{y})^2$$

$$\bar{y} = \frac{1}{u} \sum_{i=1}^u y_i .$$

где

Сравнение дисперсий σ_X^2 и σ_{ε}^2 соответствует сравнению влияния фактора x и фактора влияния $\square$. Результаты экспериментов позволяют вычислить оценку выборочной дисперсии:

$$S_0^2 = \frac{1}{u-1} \cdot \sum_{i=1}^u (y_i - \bar{y})^2$$

Если отличие S_0^2 от σ_{ε}^2 незначительно, то влияние фактора не существенно, иначе – существенно. Дисперсия фактора может быть оценена по следующей формуле:

$$\sigma_X^2 = S_0^2 - \sigma_{\varepsilon}^2, \text{ так как } S_0^2 = \sigma_X^2 + \sigma_{\varepsilon}^2 .$$

Дисперсионный анализ основан на разложении оценки общего рассеивания выходного сигнала на составляющие, которые зависят от факторов и от случайных величин.

При однофакторном дисперсионном анализе исследуется влияние только одного фактора X и результаты эксперимента сводятся в таблицу, которая содержит результаты равночисленных параллельных опытов (рис. 9).

Уровень варьирования	Номер параллельного опыта				Результаты расчета
	1	2	...	m	
1	$y_{1,1}$	$y_{1,2}$	...	$y_{1,m}$	$\bar{y}_1$
2	$y_{2,1}$	$y_{2,2}$	...	$y_{2,m}$	$\bar{y}_2$
...	...	...	...	...	...
u	$y_{u,1}$	$y_{u,2}$	...	$y_{u,m}$	$\bar{y}_u$

Рисунок – 9. Таблица результатов эксперимента.

Вычисляются средние значения отклика $\bar{y}_g$ по множеству параллельных опытов для каждого уровня фактора и общее среднее арифметическое по всем уровням:

$$\bar{y}_g = \frac{1}{m} \sum_{i=1}^m y_{g,i} \quad \bar{\bar{y}} = \frac{1}{u} \sum_{g=1}^N \bar{y}_g$$

Фактор случайности σ_{ε}^2 проявляется в рассеивании наблюдений серий параллельных опытов на каждом уровне X_j вокруг $\bar{y}_g$ своей серии изменчивости. Входной фактор X вызывает повышенное рассеивание средних арифметических $\bar{y}_g$ относительно общего среднего $\bar{\bar{y}}$.

Оценка дисперсии фактора случайности вычисляется по формуле:

$$S_{\varepsilon}^2 = \frac{1}{u(m-1)} \sum_{g=1}^N \sum_{i=1}^m (y_{g,i} - \bar{y}_g)^2$$

Оценка дисперсии фактора X вычисляется по формуле:

$$S_X^2 = \frac{1}{u-1} \cdot \sum_{g=1}^N (\bar{y}_g - \bar{\bar{y}})^2$$

При проверке гипотезы о существенности фактора X используется следующая функция статистики:

$$F = \frac{S_X^2}{S_{\varepsilon}^2}$$

Критическое значение статистики: $F_K = qF(q, v_1, v_2)$, где $v_1 = u-1$, $v_2 = u \cdot (m-1)$. Гипотеза о существенности фактора считается подтвержденной, если выполняется неравенство: $F \leq F_K$.

Если серии наблюдений состоят из разного числа опытов, то степени свободы распределения Фишера рассчитывают следующим образом:

$$v_1 = u - 1, \quad v_2 = \sum_{j=1}^u m_j - u$$

При двухфакторном дисперсионном анализе изучается влияние двух одновременно действующих входных факторов X_1 и X_2 . При этом вводятся следующие обозначения:

- $\{Y_{j,g,k}\}$ – множество результатов измерений;
- $j = 1, \dots, u_1$ – номер уровня X_1 ;
- $g = 1, \dots, u_2$ – номер уровня X_2 ;
- $k = 1, \dots, m$ – номер параллельного опыта;
- $u_1 \cdot u_2 \cdot m$ – общее число наблюдений.

j	g				Результаты расчета
	1	2	...	u_2	
1	$y_{1,1,1}$	$y_{1,2,1}$	...	$y_{1,u_2,1}$	$\bar{y}_1$
	$y_{1,1,2}$	$y_{1,2,2}$	...	$y_{1,u_2,2}$	
	...	...	...	...	
	$y_{1,1,m}$	$y_{1,2,m}$	...	$y_{1,u_2,m}$	
...	...	...	...	...	...
u_1	$y_{u_1,1,1}$	$y_{u_1,2,1}$	...	$y_{u_1,u_2,1}$	$\bar{y}_{u_1}$
	$y_{u_1,1,2}$	$y_{u_1,2,2}$	...	$y_{u_1,u_2,2}$	
	...	...	...	...	
	$y_{u_1,1,m}$	$y_{u_1,2,m}$	...	$y_{u_1,u_2,m}$	
Результаты расчета	$\bar{y}_1$	$\bar{y}_2$	...	$\bar{y}_{u_2}$	$\bar{y}$

Рисунок – 10. Таблица результатов эксперимента при двухфакторном анализе.

Вычисляют средние значения выходного сигнала по уровням каждого фактора и общее среднее по следующим формулам:

$$\bar{y}_j = \frac{1}{m \cdot u_2} \sum_{g=1}^{u_2} \sum_{k=1}^m y_{j,g,k}$$

$$\bar{y}_g = \frac{1}{m \cdot u_1} \sum_{j=1}^{u_1} \sum_{k=1}^m y_{j,g,k}$$

$$\bar{\bar{y}} = \frac{1}{u_2} \sum_{g=1}^{u_2} \bar{y}'_g = \frac{1}{u_1} \sum_{j=1}^{u_1} \bar{y}_j$$

На рассеивание средних по строкам таблицы (рис. 10) оказывает влияние фактор X_1 , так как значения другого фактора усреднены. На рассеивание средних по столбцам оказывает влияние фактор X_2 . Оценки дисперсий факторов X_1 и X_2 вычисляются по формулам:

$$S_{X_1}^2 = \frac{1}{u_1 - 1} \cdot \sum_{j=1}^{u_1} (\bar{y}_j - \bar{\bar{y}})^2$$

$$S_{X_2}^2 = \frac{1}{u_2 - 1} \cdot \sum_{g=1}^{u_2} (\bar{y}'_g - \bar{\bar{y}})^2$$

Общая выборочная дисперсия определяется следующим соотношением:

$$S_{X_1, X_2}^2 = \frac{1}{\sqrt{(u_1 - 1)(u_2 - 1)}} \sum_{j=1}^{u_1} \sum_{g=1}^{u_2} (\bar{y}_j - \bar{\bar{y}})(\bar{y}'_g - \bar{\bar{y}})$$

Существенность влияния двух и более факторов определяется как их относительное влияние. Фактор случайности $\square$ считается малозначительным. Оценивание существенности влияния факторов реализуется с помощью функций статистик:

$$F_{X_1} = \frac{S_{X_1}^2}{S_{X_1, X_2}^2} \quad F_{X_2} = \frac{S_{X_2}^2}{S_{X_1, X_2}^2}$$

Критические значения статистик определяются с помощью квантилей:

$$F_{K_1} = qF(q_1, v_1, v)$$

$$F_{K_2} = qF(q_2, v_2, v)$$

где $v_1 = u_1 - 1$, $v_2 = u_2 - 1$, $v = (u_1 - 1) \cdot (u_2 - 1)$.

Гипотезы о существенности факторов считаются подтвержденными,

если выполняются неравенства: $F_{X_1} > F_{K_1}$ $F_{X_2} > F_{K_2}$.

При многофакторном анализе последовательность операций аналогична, но значительно усложняются таблицы наблюдений и расчетные формулы.

Глава 7. Корреляционный анализ результатов эксперимента.

Корреляция – статистическая зависимость между величинами, не имеющая в общем случае строгого функционального характера.

Корреляционная зависимость – статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). Корреляционная зависимость строится в тех случаях, когда одна из величин зависит не только от заданного набора двух величин, но и от ряда случайных факторов.

Корреляционный анализ – совокупность методов обнаружения корреляционной зависимости между случайными величинами или признаками.

Коэффициент корреляции – мера корреляции двух случайных величин. Коэффициент корреляции изменяется в пределах $-1 \leq K \leq 1$.

Положительная корреляция – связь, при которой увеличение одной переменной связано с увеличением другой переменной. При этом $K \geq 0$.

Отрицательная корреляция – связь, при которой увеличение одной переменной связано с уменьшением другой переменной. При этом $K \leq 0$.

Возможна ситуация отсутствия статистической взаимосвязи. Например, в случае независимых случайных величин. Тогда $K \sim 0$.

Существует несколько методик расчета коэффициента корреляции [6]. При корреляционном анализе ранжированных величин (например, экспертных решений) применяют методики Кендалла и Спирмена. При анализе метрических величин применяют методику Пирсона.

Показатели корреляции:

1. Параметрические показатели корреляции:
 - 1.1. Ковариация;
 - 1.2. Линейный коэффициент корреляции.
2. Непараметрические показатели корреляции:
 - 2.1. Коэффициент ранговой корреляции Кендалла;
 - 2.2. Коэффициент ранговой корреляции Спирмена;
 - 2.3. Коэффициент корреляции знаков Фехнера;
 - 2.4. Коэффициент множественной ранговой корреляции (конкордации).

Ковариация (корреляционный момент) – характеристика совместного распределения двух случайных величин. Ковариация определяется математическим ожиданием произведения отклонений случайных величин:

$$\text{cov}(X, Y) = M[(X - M(X)) \cdot (Y - M(Y))]$$

$$\text{cov}(X, Y) = M(X \cdot Y) - M(X) \cdot M(Y)$$

где M – математическое ожидание;

X и Y – две случайные величины, определенные на одном вероятностном пространстве.

Свойства ковариации:

1. Ковариация двух независимых случайных величин X и Y равна нулю.
2. Абсолютная величина ковариации двух случайных величин X и Y не превышает среднего арифметического их дисперсий:

$$|\text{cov}(X, Y)| \leq \sqrt{D_X \cdot D_Y} = \sigma_X \cdot \sigma_Y$$

3. Ковариация имеет размерность, равную произведению размерности случайных величин.

Линейный коэффициент корреляции (коэффициент корреляции Пирсона) – рассчитывается по формуле:

$$K_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{D_X} \cdot \sqrt{D_Y}}$$

где cov – ковариация;

D – дисперсия.

$$K_{XY} = \frac{M(X \cdot Y) - M(X) \cdot M(Y)}{\sqrt{M(X^2) - (M(X))^2} \cdot \sqrt{M(Y^2) - (M(Y))^2}}$$

Коэффициент корреляции двух непрерывных случайных величин X и Y вычисляют по формуле:

$$K_{X,Y} = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - \mu_X] \cdot [y - \mu_Y] \cdot f(x, y) \, dx \, dy}{\sigma_X \cdot \sigma_Y}$$

Выборочный линейный коэффициент корреляции Пирсона вычисляют для дискретных метрических случайных величин по формуле:

$$K_{XY} = \frac{\sum_{i=1}^p \sum_{j=1}^q (x_i - \bar{x}) \cdot (y_j - \bar{y}) \cdot n_{i,j}}{\sqrt{\sum_{i=1}^p (x_i - \bar{x})^2 \cdot n_{i,*}} \cdot \sqrt{\sum_{j=1}^q (y_j - \bar{y})^2 \cdot n_{*,j}}}$$

где $n_{i,*} = \sum_{j=1}^q n_{i,j}$, $n_{*,j} = \sum_{i=1}^p n_{i,j}$, $N = \sum_{i=1}^p \sum_{j=1}^q n_{i,j}$,

$$\bar{x} = \frac{1}{N} \cdot \sum_{i=1}^p n_{i,*} \cdot x_i, \quad \bar{y} = \frac{1}{N} \cdot \sum_{j=1}^q n_{*,j} \cdot y_j$$

Для графического представления корреляционной связи можно использовать прямоугольную систему координат с осями, которые соответствуют обеим переменным. Каждая пара значений маркируется при помощи определенного символа и называется *диаграммой рассеяния* (рис. 11).

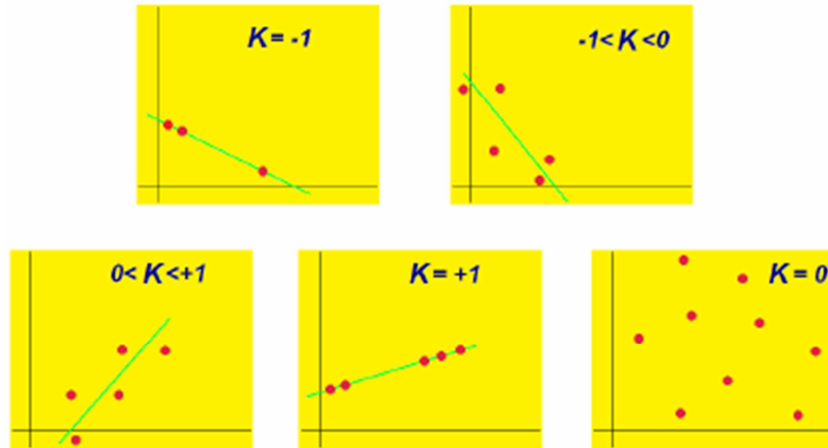


Рисунок – 11. Диаграммы рассеяния.

Корреляционная матрица – матрица коэффициентов корреляции нескольких случайных величин (МЭС). Пусть исследуется взаимосвязь четырех случайных величин V, X, Y, Z., принимающих дискретные значения. Тогда корреляционная матрица имеет вид:

$$K = \begin{bmatrix} K_{VV} & K_{VX} & K_{VY} & K_{VZ} \\ K_{XV} & K_{XX} & K_{XY} & K_{XZ} \\ K_{YV} & K_{YX} & K_{YY} & K_{YZ} \\ K_{ZV} & K_{ZX} & K_{ZY} & K_{ZZ} \end{bmatrix},$$

где $K_{VV} = K_{XX} = K_{YY} = K_{ZZ} = 1$, $K_{VX} = K_{XV}$, $K_{VY} = K_{YV}$, $K_{VZ} = K_{ZV}$, $K_{XY} = K_{YX}$, $K_{XZ} = K_{ZX}$, $K_{YZ} = K_{ZY}$.

При вычислении коэффициента корреляции Спирмена K_c каждому показателю X и Y присваивается ранг:

$$K_c = 1 - \frac{6 \cdot \left[\sum_{j=1}^n (x_{j,j} - y_{g,j})^2 \right]}{n \cdot (n^2 - 1)},$$

где $x_{j,j} - y_{g,j}$ – разность рангов.

Ранжированием множества V вариантов решения $V = \{v_1, v_2, \dots, v_n\}$ называется нумерация вариантов V_i в соответствии с возрастанием (или

убыванием) некоторого критерия g . Ранг $x_{j,i}$ варианта V_i указывает место, которое занимает i -тый вариант среди других вариантов, расположенных в соответствии с данным критерием. Каждый j -тый эксперт может указать варианту V_i особый ранг $x_{j,i}$. Результатом ранжирования n вариантов решения j -тым экспертом является нормированная последовательность (вектор):

$$X_j = \{x_{j,1}, x_{j,2}, \dots, x_{j,n}\}.$$

Ранги $x_{j,i}$ должны удовлетворять условию нормировки:

$$\sum_{i=1}^n x_{j,i} = \frac{1}{2} \cdot n \cdot (n+1).$$

Ранжирование может применяться даже, когда значения критерия $q(v_i)$ для вариантов нельзя измерить или рассчитать. Величина $q(v_i)$ оценивается экспертами. Результаты ранжирования вариантов решения, полученные двумя экспертами « j » и « g »:

$$X_j = \{x_{j,1}, x_{j,2}, \dots, x_{j,n}\}, \quad X_g = \{x_{g,1}, x_{g,2}, \dots, x_{g,n}\}$$

можно сравнить, применяя корреляционный анализ.

Степень связи между последовательностями рангов X_j и X_g оценивается с помощью коэффициентов ранговой корреляции по Спирмену или по Кендаллу. Значения коэффициентов ранговой корреляции K принадлежат интервалу $[-1;1]$. Если вектора X_j и X_g равны, то есть мнения экспертов « j » и « g » совпадают, то $K=1$. Если ранжирование вариантов двумя экспертами полностью противоположно, то $K=-1$. Если ранги в последовательностях X_j и X_g независимы, то в этом случае $K \rightarrow 0$.

Коэффициент ранговой корреляции по Спирмену рассчитывается по формуле:

$$K_c = 1 - \frac{6 \cdot \left[\sum_{i=1}^n (x_{j,i} - x_{g,i})^2 \right]}{n \cdot (n^2 - 1)}.$$

Экспертная система – компьютерная система, способная частично заменить специалиста эксперта в разрешении проблемной ситуации. В информатике экспертные системы рассматриваются совместно с базами знаний как модели поведения экспертов в определенной области знаний с использованием процедур логического вывода и принятия решений, а базы знаний – как совокупность факторов и правил логического вывода в выбранной предметной области деятельности.

Автокорреляционная функция – функция, характеризующая взаимосвязь между значениями случайной величины из одного ряда, взятых со сдвигом. Обычно рассматривается характеристика случайного процесса со

сдвигом по времени. Автокорреляционная функция связывает между собой отклонения случайной функции от ее математического ожидания при двух значениях аргумента t_1 и t_2 и равна математическому ожиданию этих отклонений:

$$K_x(t_1, t_2) = M \{ [x(t_1) - M\{x(t_1)\}] \cdot [x(t_2) - M\{x(t_2)\}] \}, \quad (9)$$

где $M\{\alpha(t)\}$ – математическое ожидание функции $\alpha(t)$.

При обработке сигналов автокорреляционная функция определяется интегралом:

$$K_x(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t_1) - M\{x(t_1)\}] \cdot [x(t_2) - M\{x(t_2)\}] \cdot f(x(t_1), x(t_2)) dx(t_1) dx(t_2), \quad (10)$$

Если рассматриваемые процессы стационарные или стационарно-связанные, то средние значения и математические ожидания $M\{x(t_1)\}$, $M\{x(t_2)\}$ не зависят от времени, а автокорреляционная функция будет зависеть лишь от величины $\tau = t_1 - t_2$.

$$K_x(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t) - M\{x\}] \cdot [x(t - \tau) - M\{x\}] \cdot f(x(t), x(t - \tau)) dx(t) dx(t - \tau),$$

где τ – смещение во времени.

График автокорреляционной функции можно построить, отложив по оси ординат коэффициент корреляции двух функций (базовой и функции сдвинутой на величину τ), а по оси абсцисс величину τ .

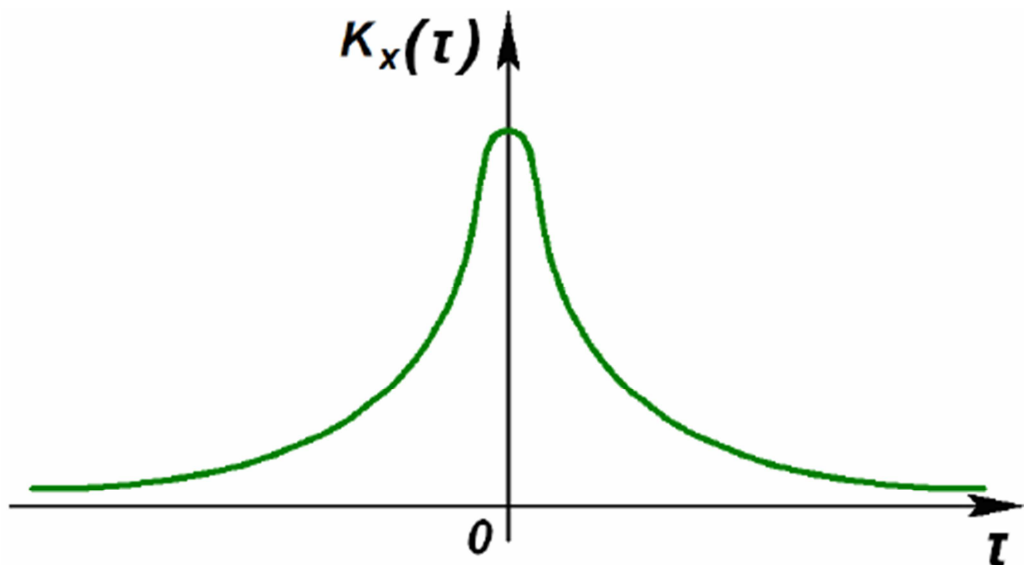


Рисунок – 12. График автокорреляционной функции для случайного процесса без периодической составляющей.

Если исследуемая функция является периодической, тогда автокорреляционная функция тоже будет периодической (Рис. 12).

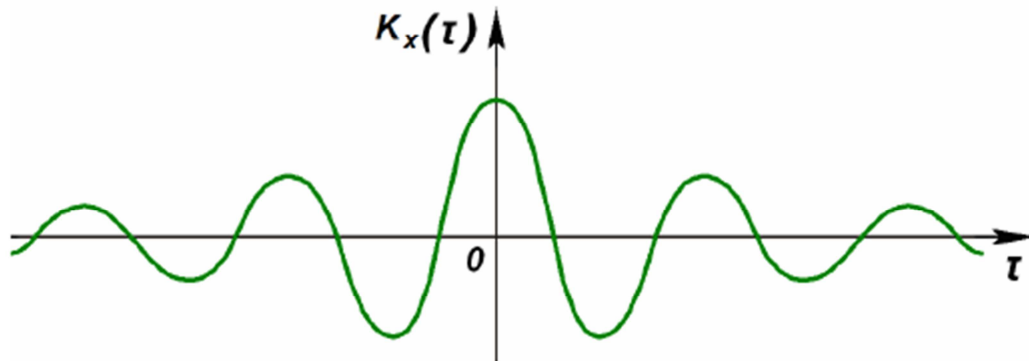


Рисунок – 12. График автокорреляционной функции для случайного процесса с периодической составляющей.

Основное свойство автокорреляционной функции:

$$\lim_{\tau \rightarrow \pm\infty} K_x(\tau) = 0$$

Свертка функций – операция, показывающая сходство одной функции к отраженной и сдвинутой копии другой (особый вид интегрального преобразования функций). Свертка двух функций $x(t)$ и $y(t)$ равна:

$$(x * y)(t) \stackrel{\text{def}}{=} \int_{R^d} x(\tau) \cdot y(t - \tau) d\tau = \int_{R^d} y(\tau) \cdot x(t - \tau) d\tau$$

В частности, при $d=1$ формула принимает вид:

$$(x * y)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} x(\tau) \cdot y(t - \tau) d\tau = \int_{-\infty}^{+\infty} y(\tau) \cdot x(t - \tau) d\tau$$

Статистическая связь между двумя процессами $x(t)$ и $y(t)$ в два момента времени t_1 и t_2 характеризуется функцией *взаимной корреляции*, которая определяется интегралами:

$$K_{x,y}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t_1) - M\{x(t_1)\}] \cdot [y(t_2) - M\{y(t_2)\}] \cdot f(x(t_1), y(t_2)) dx(t_1) dy(t_2),$$

$$K_{x,y}(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x(t) - M\{x\}] \cdot [y(t - \tau) - M\{y\}] \cdot f(x(t), y(t - \tau)) dx(t) dy(t - \tau),$$

Автокорреляционная функция стационарного эргодического случайного процесса $x(t)$ с нулевым математическим ожиданием вычисляется по формуле:

$$K_x(\tau) = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \cdot \int_0^T x(t) \cdot x(t + \tau) dt \right), \quad (11)$$

где T – длина реализации;

τ – корреляционный сдвиг.

Аналогично определяется взаимная корреляционная функция стационарных и стационарно связанных эргодических случайных процессов $x(t)$ и $y(t)$ с нулевым математическим ожиданием:

$$K_{xy}(\tau) = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \cdot \int_0^T x(t) \cdot y(t + \tau) dt \right). \quad (12)$$

Корреляционные функции (11) и (12) представляют собой неслучайные функции. На практике длина реализации T ограничена, поэтому вычисляемые корреляционные функции являются случайными и называются *оценками корреляционных функций* $K_x^*(\tau, T)$ и $K_{xy}^*(\tau, T)$. Оценка корреляционной функции, вычисленная по экспериментальным данным, обычно представляет собой ряд дискретных отсчетов в фиксированных точках τ_i .

Оценка корреляционной функции может быть получена по непрерывно-шаговому или выборочно-шаговому алгоритмам. *Непрерывно-шаговый алгоритм* основан на интегральной формуле. А *выборочно-шаговый алгоритм* основан на дискретной формуле, в которой интеграл заменен суммой.

Оценки взаимной корреляционной функции:

$$K_{xy}^*(k \cdot \Delta\tau, T) = \frac{1}{T} \cdot \int_0^T x(t) \cdot y(t + k \cdot \Delta\tau) dt$$

$$K_{xy}^*(k \cdot \Delta\tau, N) = \frac{1}{N} \cdot \sum_{n=1}^N x(n \cdot \Delta t) \cdot y(n \cdot \Delta t + k \cdot \Delta\tau)$$

где $\Delta\tau$ – шаг дискретности корреляционного сдвига;

Δt_b – шаг отбора выборочных пар (x_n, y_n) ;

$N = T / \Delta t_b$ – количество выборочных пар (x_n, y_n) ;

n – номер пары (x_n, y_n) ;

$n \cdot \Delta\tau = \tau$ – корреляционный сдвиг.

Глава 8. Законы распределения результатов экспериментов.

1 закон. Дискретное равномерное распределение. Случайная величина имеет равномерное дискретное распределение, если она принимает конечное число значений с равными вероятностями.

2 закон. Непрерывное равномерное распределение – распределение случайной величины, принимающей значения, принадлежащие интервалу (а, b), характеризующееся тем, что плотность вероятности на этом интервале постоянна.

Плотность равномерного распределения:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a; b] \\ 0, & x \notin [a; b] \end{cases}$$

Функция распределения имеет вид:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

3 закон. Нормальное распределение (распределение Гаусса). Формула Гаусса (Рис. 13):

$$f(x) = \frac{1}{\sigma_x \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(x - \mu_x)^2}{2 \cdot \sigma_x^2}\right)$$

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot D_x}} \cdot \exp\left(-\frac{(x - \mu_x)^2}{2 \cdot D_x}\right)$$

$$f(x) = \text{dnorm}(x, \mu_x, \sigma_x)$$

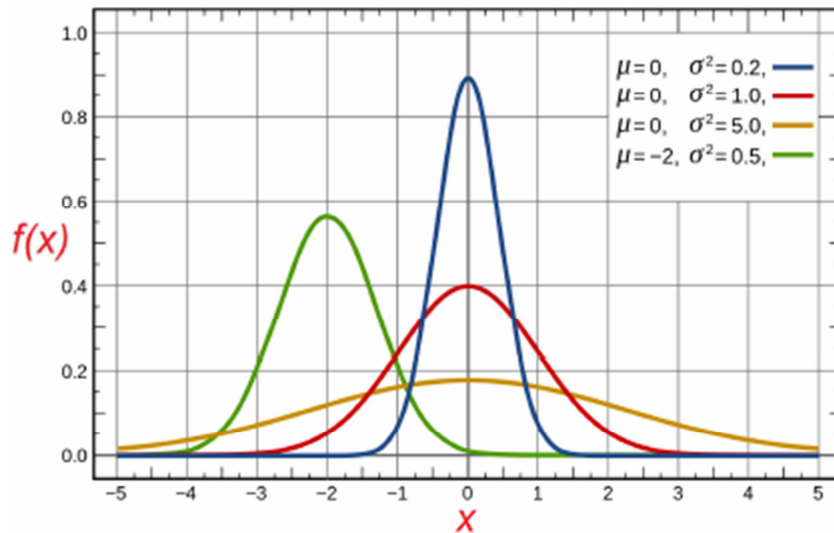


Рисунок13. – График нормального распределения Гаусса.

Функция распределения Гаусса:

$$F(x) = \frac{1}{2} \cdot \left(1 + \operatorname{erf}\left(\frac{x - \mu_x}{\sigma_x \sqrt{2}}\right) \right)$$

$$F(x) = \frac{1}{2} \cdot \left(1 + \operatorname{erf}\left(\frac{x - \mu_x}{\sqrt{2 \cdot D_x}}\right) \right)$$

$$F(x) = \text{pnorm}(x, \mu_x, \sigma_x)$$

Функция ошибок:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \cdot \int_0^x \exp(-t^2) dt$$

Нормальный закон наблюдается в тех многочисленных случаях, когда на измеряемую случайную величину действуют факторы незначительно связанные между собой и равнозначно действующие на случайную величину [3].

4 закон. Распределение χ^2 (хи-квадрат) – распределение суммы квадратов независимых случайных величин, с нормальным законом распределения [7]:

$$f(x) = \Gamma\left(2, \frac{k}{2}\right) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \cdot x^{\frac{k}{2}-1} \cdot \exp\left(-\frac{x}{2}\right)$$

где $\Gamma(\bullet, \bullet)$ – гамма-распределение;

$\Gamma(\bullet)$ – гамма-функция.

5 закон. Распределение Максвелла. Формула Максвелла-Больцмана – плотность распределение Максвелла для модуля скорости частиц массой m :

$$f(v) = 4 \cdot \pi \cdot v^2 \cdot \left(\frac{m}{2 \cdot \pi \cdot k \cdot T} \right)^{3/2} \cdot \exp \left(- \frac{m \cdot v^2}{2 \cdot k \cdot T} \right)$$

где $k \approx 1.381 \cdot 10^{-23} \frac{\text{Дж}}{\text{К}}$ – постоянная Больцмана.

6 закон. Распределение Вейбулла – функция от экспоненциального распределения:

$$f(x) = \frac{k}{\lambda} \cdot \left(\frac{x - \gamma}{\lambda} \right)^{k-1} \cdot \exp \left(- \left(\frac{x - \gamma}{\lambda} \right)^k \right)$$

где k – параметр формы;
 λ – параметр масштаба;
 γ – смещение.

Функция распределения Вейбулла:

$$F(x) = 1 - \exp \left(- \left(\frac{x - \gamma}{\lambda} \right)^k \right)$$

Экспоненциальный закон характерен для распределения случайных величин, изменение которых обусловлено влиянием какого-то доминирующего фактора. Распределение скорости ветра:

$$f(v) = \frac{k}{\lambda} \cdot \left(\frac{v - \gamma}{\lambda} \right)^{k-1} \cdot \exp \left(- \left(\frac{v - \gamma}{\lambda} \right)^k \right)$$

7 закон. Логистическое распределение. Плотность логистического распределения:

$$f(x) = \frac{1}{2 \cdot |E|} \cdot \text{sech}^2 \left(\frac{x - \mu}{E} \right)$$

где $\text{sech}(\ast) = \frac{1}{\cosh(\ast)} = \frac{2}{\exp(\ast) + \exp(-\ast)}$ – гиперболический секанс.

Функция распределения:

$$F(x) = \frac{1}{2} \cdot \left(1 + \tanh\left(\frac{x - \mu}{E}\right) \right)$$

8 закон. Распределение Стюдента. Пусть $y_0, y_1, \dots, y_n$ – независимые случайные величины с нормальным законом распределения, тогда случайная

$$t = \frac{y_0}{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n y_i^2}}$$

величина имеет распределение Стюдента с n степенями свободы. Плотность распределения Стюдента [4]:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \cdot \sqrt{\pi \cdot n}} \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

где $\Gamma(\bullet)$ – гамма-функция Эйлера.

9 закон. Распределение Фишера (F-распределение) – это двухпараметрическое семейство абсолютно непрерывных распределений.

Случайную величину с распределением Фишера обозначают F_{v_1, v_2} – это функция от $v_1 + v_2$ взаимно простых независимых случайных величин с нормальным распределением и нулевым математическим ожиданием:

$$F_{v_1, v_2} = \frac{\frac{1}{v_1} \sum_{i=1}^{v_1} x_i^2}{\frac{1}{v_2} \sum_{i=v_1+1}^{v_1+v_2} x_i^2}$$

где v_1, v_2 – степени свободы;

F_{v_1, v_2} – случайная величина, не зависит от дисперсии σ_x^2 .

Плотность распределения Фишера имеет вид:

$$f_{v_1, v_2}(x) = k \cdot x^{\frac{v_1-2}{2}} \cdot \left(1 + \frac{v_1}{v_2} x\right)^{-(v_1+v_2)/2}$$

$$k = \left(\frac{v_1}{v_2}\right)^{v_1/2} \Gamma\left(\frac{v_1+v_2}{2}\right) \cdot \Gamma\left(\frac{v_1}{2}\right)^{-1} \cdot \Gamma\left(\frac{v_2}{2}\right)^{-1}$$

где – нормирующий множитель.

10 закон. Мультимодальное распределение – распределение вероятностей с несколькими модами, то есть с несколькими относительными максимумами, соответствующими этим модам. Мультимодальные распределения, как правило, представляют собой смеси стандартных унимодальных распределений вероятностей.

Глава 9. Экспериментальная проверка статистических гипотез.

Статистическая гипотеза – это некоторое предположение относительно свойств генеральной совокупности, из которой извлекается выборка.

Простая гипотеза соответствует одному распределению и одной точке пространства параметров. *Сложная гипотеза* – выбор распределения из множества распределений или точки пространства параметров из интервала.

Критерий статистической гипотезы – правило, позволяющее принять или отвергнуть данную гипотезу на основании выборки.

Статистика – функция от результатов повторяющихся опытов $g(x_1, x_2, \dots, x_N)$, которая обладает свойством роста предсказуемости с увеличением количества опытов. Это свойство называется *статистической устойчивостью*, которая может быть проверена только экспериментальным путем.

Область возможных значений статистик делят на две части (Рис. 14):

- область принятия гипотезы;
- критическую область.

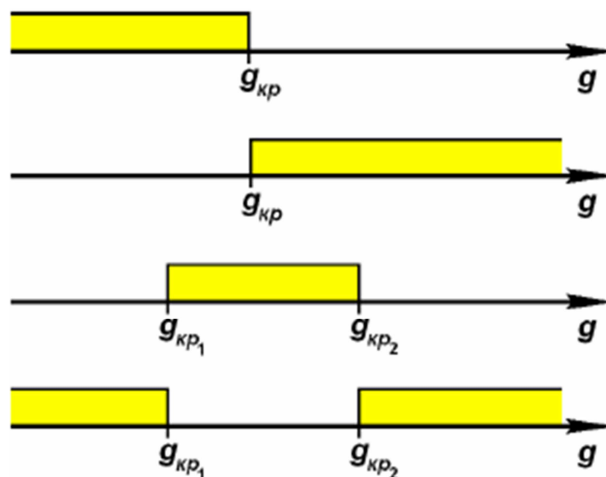


Рисунок 14. – Область принятия гипотезы.

Этапы проверки гипотез:

- 1) формулируется основная гипотеза;
- 2) задается уровень значимости гипотезы g ;
- 3) выбирается статистика $g(x_1, x_2, \dots, x_N)$ и определяется ее закон распределения;

- 4) вычисляются критические значения статистики $g_{кр}(q, N)$ на основе данных о количестве опытов N ;
- 5) по данным выборки вычисляют экспериментальное значение статистики;
- 6) сравнивают критические значения и экспериментальное значение статистики;
- 7) если экспериментальное значение попало в область принятия гипотезы, то гипотезу считают подтвержденной по результатам опытной выборки.

В процессе проверки гипотезы возможны правильные и ложные решения. Возможные варианты приведены в таблице 1:

Таблица 1.

Гипотеза	Объективно верна	Объективно не верна
Принимается	Правильное решение	Ошибка 2-го рода
Отвергается	Ошибка 1-го рода	Правильное решение

Уровень значимости критерия – вероятность ошибочно отвергнуть проверяемую гипотезу, когда она верна. Уровень значимости критерия соответствует вероятности ошибки первого рода и обозначается q ($0 < q < 1$). *Надежность критерия*: $p = 1 - q$. *Мощность критерия* – вероятность ошибки второго рода.

Критерии проверки согласия законов распределения.

1. Критерий согласия Пирсона – критерий проверки гипотезы, состоящей в том, что генеральная совокупность имеет распределение заданного типа. Статистика критерия Пирсона имеет χ^2 -распределение. Критерий реализуется в соответствии со следующим алгоритмом:

- 1) По выборке строят гистограмму n_m , где m – номер интервала. Если в каком-либо интервале число наблюдений меньше пяти, то его объединяют с соседним интервалом.
- 2) Задается вид гипотетической функции распределения (Рис. 15) и по вектору выборки (или по гистограмме) вычисляют точечные оценки r параметров распределения (например, математическое ожидание и дисперсию, тогда $r=2$).
- 3) Определяется вектор p_m теоретической вероятности попадания случайной величины в каждый интервал гистограммы:

$$p_m = \frac{1}{x_{m+1} - x_m} \cdot \int_{x_m}^{x_{m+1}} f(x) dx$$

- 4) Вычисляется значение статистики по вектору выборки:

$$G = \sum_{m=1}^K \frac{(n_m - N \cdot p_m)^2}{N \cdot p_m}$$

где K – число интервалов группирования;

p_m – теоретическая вероятность попадания в интервал гистограммы с номером m ;

n_m – экспериментальная частота попадания в интервал гистограммы.

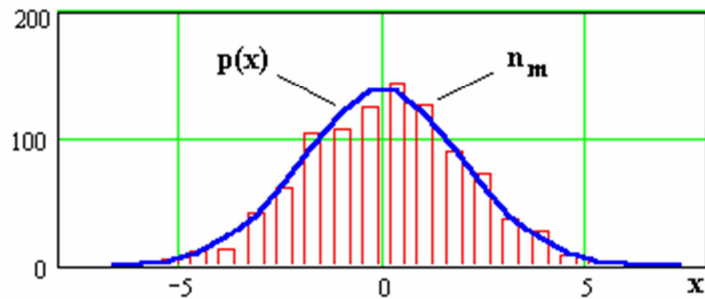


Рисунок 15. – Гипотетическая функция распределения (нормированная) и экспериментальная гистограмма.

- 5) Вычисляется критическое значение статистики, которое равно квантили распределения χ^2 :

$$G_{kr} = \chi_{q,v}^2$$

где g – заданный уровень значимости;

$$v = K - r - 1;$$

r – число проверяемых параметров распределения.

- 6) Гипотеза о согласии экспериментальных данных с распределением заданного типа считается подтвержденной, если $G \leq G_{kr}$.

Критерий Пирсона позволяет контролировать согласованность гипотетических вероятностей с их опытными относительными частотами [8].

2. Критерий согласия Колмогорова предназначен для проверки гипотезы о соответствии выборки экспериментальных данных заданному закону распределения, то есть проверки соответствия эмпирического распределения предполагаемой модели. Статистика критерия Колмогорова определяется формулой:

$$G(X) = \sqrt{N} \cdot \sup_{x_i} |F_{\text{эсп}}(x_i) - F_{\text{теор}}(x_i; a, b, \dots)|$$

где $F_{\text{эсп}}(\cdot)$ – эмпирическая функция распределения x , построенная по экспериментальным данным;

$F_{теор}(\cdot)$ – теоретическая функция распределения.

Критическое значение $G_{кр}$ статистики $G(x)$ критерия Колмогорова равно квантили распределения Стьюдента с заданной доверительной вероятностью и известным числом степеней свободы. Гипотеза считается подтвержденной, если $G \leq G_{кр}$.

3. Непараметрический критерий согласия Купера является развитием критерия согласия Колмогорова. Критерий предложен для проверки простых гипотез о принадлежности экспериментальной выборки полностью известному закону с известным вектором параметров. В критерии Купера используется статистика вида:

$$V_N = D_N^+ + D_N^- ,$$

$$D_N^+ = \max \left(\frac{i}{N} - F(x_i, A) \right) ,$$

$$D_N^- = \max \left(F(x_i, A) - \frac{i-1}{N} \right) ,$$

где N – объем выборки;

$[x_1; x_2; \dots; x_N]$ – упорядоченные по возрастанию элементы выборки;

i – номер измерения ($i = 1, 2, \dots, N$);

A – вектор параметров.

При справедливости простой проверяемой гипотезы статистика $V_N \cdot \sqrt{N}$ в пределе подчиняется распределению:

$$G(u) = 1 - \sum_{m=1}^{\infty} \left(2 \cdot (4 \cdot m^2 \cdot u^2 - 1) \cdot \exp(-2 \cdot m^2 \cdot u^2) \right)$$

Проверяемая гипотеза отклоняется при больших значениях статистики. При проверке простых гипотез критерий Купера является свободным от распределения, то есть не зависит от вида закона, с которым проверяется согласие.

4. Критерий согласия Крамера-Мизеса-Смирнова – непараметрический критерий согласия, предназначенный для проверки простых гипотез о принадлежности экспериментальной выборки полностью известному закону с известным вектором параметров. В критерии Крамера-Мизеса-Смирнова используется статистика вида:

$$U_N^2 = \sum_{i=1}^N \left(F(x_i, A) - \frac{2 \cdot i - 1}{2 \cdot N} \right)^2 + \frac{1}{12 \cdot N} ,$$

где N – объем выборки;

$[x_1; x_2; \dots; x_N]$ – упорядоченные по возрастанию элементы выборки.

Проверяемая гипотеза отклоняется при больших значениях статистики.
 5. Критерий согласия Ватсона – непараметрический критерий согласия, полученный в результате развития критерия согласия Крамера-Мизеса-Смирнова. Критерий предложен для проверки простых гипотез о принадлежности экспериментальной выборки полностью известному закону с известным вектором параметров. В критерии Ватсона используется статистика вида:

$$U_N^2 = \sum_{i=1}^N \left(F(x_i, A) - \frac{2 \cdot i - 1}{2 \cdot N} \right)^2 - N \cdot \left(\frac{1}{N} \cdot \sum_{i=1}^N F(x_i, A) - 0.5 \right)^2 + \frac{1}{12 \cdot N}$$

При справедливости простой проверяемой гипотезы статистика U_n^2 в пределе подчиняется распределению:

$$G(u) = 1 - 2 \cdot \sum_{m=1}^{\infty} (-1)^{m-1} \cdot \exp(-2 \cdot m^2 \cdot \pi^2 \cdot u)$$

Чтобы уменьшить зависимость распределения статистики от объема выборки, можно использовать модификацию статистики вида:

$$U_N^{2*} = \left(U_N^2 - \frac{0.1}{N} + \frac{0.1}{N^2} \right) \cdot \left(1 + \frac{0.8}{N} \right)$$

При $n \leq 20$ отличием распределения статистики U_n^2 от предельного распределения $G(u)$ можно пренебречь. Проверяемая гипотеза отклоняется при больших значениях статистики.

Глава 10. Проверка воспроизводимости эксперимента.

Воспроизводимость – характеристика результатов испытаний, определяемая взаимной близостью результатов повторных испытаний объекта. Воспроизводимость результатов испытаний при использовании одной и той же методики испытаний может зависеть от свойств объекта испытаний. Методика испытаний включает в себя: метод, средства, алгоритм проведения и т.д. Если объектом является партия изделий, подвергаемая выборочным испытаниям, то такие испытания у поставщика и потребителя могут проводиться на разных образцах, выбранных из данной партии, и в этом случае неоднородность изделий может существенно, иногда решающим образом, влиять на воспроизводимость результатов испытаний.

В качестве меры воспроизводимости используются:

- выборочное стандартное отклонение: $S = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$;

- относительное стандартное отклонение: $S_r = \frac{S}{\bar{x}}$;

- дисперсия: S^2 .

Для повышения достоверности исследований могут проводиться *параллельные опыты* для каждого запланированного сочетания факторов. Обычно, число параллельных опытов для всех сочетаний факторов назначается одинаковым.

Введем следующие обозначения:

- m – количество параллельных опытов;

- $Y = (y_1, y_2, \dots, y_N)^T$ – вектор результатов наблюдений отклика системы в различных запланированных точках факторного пространства при отсутствии параллельных опытов ($m=1$);

- $YM = \begin{pmatrix} y_{1,1} & \dots & y_{1,m} \\ \dots & \dots & \dots \\ y_{N,1} & \dots & y_{N,m} \end{pmatrix}$ – матрица наблюдений отклика системы в различных запланированных точках факторного пространства при проведении параллельных опытов ($m \geq 1$).

Проверка воспроизводимости эксперимента состоит в проверке гипотезы о равенстве генеральных дисперсий

$$\sigma^2(YM^{T<1>}) = \sigma^2(YM^{T<2>}) = \dots = \sigma^2(YM^{T<N>})$$

результатов параллельных опытов при экспериментах в соответствующих точках факторного пространства ($x_1, x_2, \dots, x_N$). Обозначение $YM^{<i>}$ используется для выделения i -го столбика матрицы YM , а обозначение $YM^{T<i>}$ – для обозначения i -той строки. Усреднение результатов m параллельных опытов (для одной точки факторного пространства) проводится по следующей формуле:

$$\bar{y}_g = \frac{1}{m} \sum_{i=1}^m y_{g,i}$$

Дисперсия воспроизводимости определяется следующим образом:

$$s_{\text{Вос}}^2 = \frac{1}{N} \cdot \sum_{g=1}^N s_g^2,$$

$$s_g^2 = \frac{1}{m-1} \cdot \sum_{i=1}^m (y_{g,i} - \bar{y}_g)^2,$$

$$\bar{y}_g = \frac{1}{m} \cdot \sum_{i=1}^m y_{g,i}$$

Проверка воспроизводимости эксперимента осуществляется с помощью критерия Кокрена, который использует следующее определение статистики:

$$G = \frac{\max \{S_g^2\}}{N \cdot S_{\text{вс}}^2}$$

Критерий Кокрена звучит: гипотеза о воспроизводимости эксперимента считается подтвержденной, если $G \leq G_{\text{кр}}$.

Критическое значение статистики вычисляется с использованием квантилей распределения Фишера по следующей формуле:

$$G_{\text{кр}} = \frac{qF(\alpha, v_1, v_2)}{(k-1) + qF(\alpha, v_1, v_2)}$$

где g – заданный уровень значимости;

m – количество параллельных опытов;

k – количество факторов;

N – количество опытов с разными сочетаниями факторов;

$$\alpha = 1 - \frac{g}{N}, \quad v_1 = m - 1, \quad v_2 = (N - 1) \cdot (m - 1).$$

Если гипотеза не подтверждена, то рекомендуется увеличить число параллельных опытов или использовать другую модификацию метода наименьших квадратов.

Проверка адекватности математического описания объекта по экспериментальным данным представляет собой оценку отклонения предсказанной по уравнению регрессии величины отклика системы $\hat{Y}$ от результатов наблюдений Y в одних и тех же точках факторного пространства. Рассеяние результатов наблюдений вблизи уравнения регрессии, оценивающего истинную функцию отклика, можно охарактеризовать с помощью дисперсии адекватности:

$$D_{\text{ад}} = S_{\text{ад}}^2 = \frac{m}{N - d} \cdot \sum_{g=1}^N (\hat{y}_g - \bar{y}_g)^2$$

где m – количество параллельных опытов;

N – количество опытов с разными сочетаниями факторов;

g – номер сочетания управляющих факторов;

d – количество параметров математической модели (параметры определяются по результатам эксперимента);

$\hat{y}_g$ – значение отклика, вычисленное по полученной функции регрессии;

$\bar{y}_g$ – значение отклика системы, усредненное по множеству параллельных опытов.

Функция статистики для оценки адекватности модели определяется как отношение оценки дисперсии адекватности к оценке дисперсии воспроизводимости:

$$F_{ад} = \frac{s_{ад}^2}{s_{вос}^2} = \frac{D_{ад}}{D_{вос}}$$

Критическое значение статистики определяется с помощью следующего соотношения:

$$(F_{ад})_{кр} = qF(q_{ад}, v_1, v_2)$$

где $v_1 = N - d > 0$;

$v_2 = N \cdot (m - 1)$;

$q_{ад}$ – заданный уровень значимости критерия.

Гипотеза об адекватности регрессионной математической модели экспериментальным данным считается подтвержденной, если выполняется

неравенство $F_{ад} \leq (F_{ад})_{кр}$. Если гипотеза об адекватности отвергнута, то рекомендуется перейти к более сложной форме математического описания или уменьшить интервал варьирования факторов до 10% - 60% от первоначального. Однако, уменьшение интервала варьирования приводит к снижению отношения сигнал/шум, что влечет за собой увеличение числа параллельных опытов.

Глава 11. Метод максимального правдоподобия.

Ограниченная выборка экспериментальных данных не дает возможности найти точные значения параметров распределения. На основе выборки можно дать оценки этих параметров.

Оценка параметра – функция наблюдаемых величин, которая приближенно равна искомому параметру. Чем больше число наблюдений, тем ближе оценка параметра к его истинному значению.

Метод максимального правдоподобия (ММП) – метод получения статистических точечных оценок неизвестных параметров функции распределения, для которых функция правдоподобия при случайной выборке заданного объема будет иметь максимальное значение.

Введем обозначения:

- n – объем выборки (количество опытов);

- $f(x, a)$ – плотность распределения случайной величины X ;
- a – неизвестный параметр, входящий в выражение закона распределения;
- $[x_1, x_2, \dots, x_n]$ – выборка экспериментальных значений величины X ;
- i – номер измерения величины ($i=1, 2, \dots, n$);
- δ – длина окрестности вокруг точки X_i ;
- $f(x, a) \cdot \delta$ – вероятность попадания результата одного измерения величины X в интервал $\left(x_i - \frac{\delta}{2}; x_i + \frac{\delta}{2}\right)$.

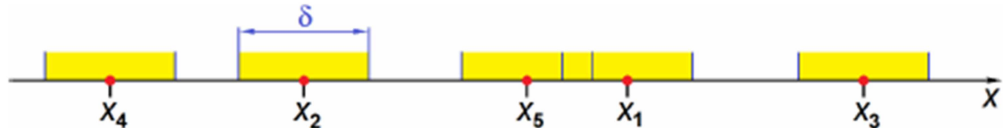


Рисунок 16. – Выборка экспериментальных значений величины X с окрестностями.

Вероятность $P(x, a)$ попадания результатов n опытов в соответствующие интервалы равна произведению:

$$P(x, a) = \delta^n \cdot \prod_{i=1}^n f(x_i, a) \quad (13)$$

Функция $P(x, a)$ имеет максимум $P_{\max} = P(x, \hat{a})$, который достигается в точке $\hat{a}$. Величина $\hat{a} = \hat{a}(x_1, x_2, \dots, x_n)$ – оценка максимального правдоподобия параметра a , соответствующая заданной экспериментальной выборке $[x_1, x_2, \dots, x_n]$. Значение $\hat{a}$ можно определить в результате решения уравнения:

$$\left. \frac{\partial P(x, a)}{\partial a} \right|_{a=\hat{a}} = 0 \quad (14)$$

При увеличении объема выборки функция (13) и ее производная принимают значения, стремящиеся к нулю.

С целью быстрого и достаточно точного определения оценки параметра закона распределения используют *функцию правдоподобия*, под которой понимают распределение выборки из параметрического распределения как функции параметра:

$$\tilde{L}(a | x) = \frac{P(x, a)}{\delta^n} = \prod_{i=1}^n f(x_i, a) \quad (15)$$

Во многих приложениях необходимо найти максимум функции правдоподобия, что связано с вычислением производной. *Логарифм* –

монотонно возрастающая функция, поэтому логарифм от функции достигнет максимума в той же точке, что и сама функция. Логарифм произведения равен сумме логарифмов, что упрощает дифференцирование. Поэтому предпочитают использовать логарифмическую функцию правдоподобия:

$$L(a | x) = \ln \left(\frac{P(x, a)}{\delta^n} \right). \quad (16)$$

С учетом (13) функция (16) преобразуется к виду:

$$L(a | x) = \ln \left(\prod_{i=1}^n f(x_i, a) \right). \quad (17)$$

На основе свойства логарифма произведение (17) приводится к сумме:

$$L(a | x) = \sum_{i=1}^n \ln(f(x_i, a)). \quad (18)$$

Логарифмическая функция правдоподобия также имеет максимум в точке $\hat{a}$, величину которой можно найти, решив уравнение:

$$\left. \frac{dL(a | x)}{da} \right|_{a = \hat{a}} = 0. \quad (19)$$

С целью получения статических точечных оценок нескольких неизвестных параметров $a_1, a_2, \dots, a_k$ закона распределения составляют логарифмическую функцию правдоподобия $L(a_1, a_2, \dots, a_k | x)$. Приравнявая частные производные функции правдоподобия по каждому из параметров в точке максимума к нулю, составляют систему уравнений:

$$\begin{cases} \left. \frac{\partial L(a_1, a_2, \dots, a_k | x)}{\partial a_1} \right|_{a_1 = \hat{a}_1, a_2 = \hat{a}_2, \dots, a_k = \hat{a}_k} = 0, \\ \left. \frac{\partial L(a_1, a_2, \dots, a_k | x)}{\partial a_2} \right|_{a_1 = \hat{a}_1, a_2 = \hat{a}_2, \dots, a_k = \hat{a}_k} = 0, \\ \dots \dots \dots \\ \left. \frac{\partial L(a_1, a_2, \dots, a_k | x)}{\partial a_k} \right|_{a_1 = \hat{a}_1, a_2 = \hat{a}_2, \dots, a_k = \hat{a}_k} = 0. \end{cases} \quad (20)$$

Полученную систему уравнений решают относительно $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k$. Корни системы (20) могут соответствовать максимуму или минимуму функции правдоподобия, а также могут являться точками перегиба. Поэтому необходимо проверить, соответствует ли полученное решение $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k$ максимуму функции $L(a_1, a_2, \dots, a_k)$.

Применяя метод максимального правдоподобия можно определить оценки неизвестных параметров любого другого закона распределения, соответствующего исследуемой экспериментальной выборке. Метод

максимального правдоподобия является популярным статистическим методом, который используется для создания статистической модели на основе экспериментальных данных и обеспечения оценки параметров математической модели объекта или процесса.

Глава 12. Метод главных компонент.

Метод главных компонент – один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации, и применяется к экспериментальным данным, записанным в виде матрицы X .

Строки этой матрицы называются образцами и соответствуют опытам. Столбцы называются переменными и соответствуют факторам. Каждой строке (опыту) соответствует точка на плоскости. Через экспериментальные точки проводится прямая так, чтобы вдоль нее происходило максимальное изменение данных. Эта прямая называется *главной компонентой* – PC1. Экспериментальные точки проецируются на прямую PC1, а отклонения от новой оси называются *остатками*.

Проверить шум ли это или нужная часть данных можно, поступив с этими остатками так же, как мы поступили с исходными данными – найти в них ось максимальных изменений. Она называется второй главной компонентой (PC2). Эти действия повторяются до тех пор, пока остатки не станут явным шумом, то есть случайным хаотическим набором величин, которыми можно пренебречь.

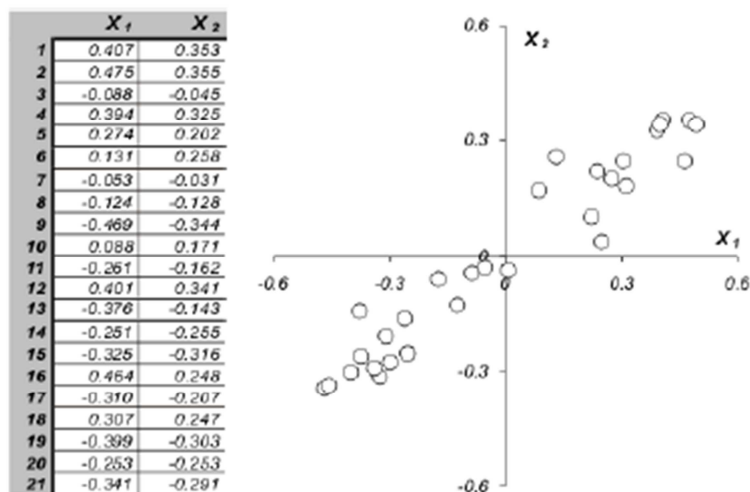


Рисунок 17. – Графическое представление двумерных данных.

В общем многомерном случае процесс выделения главных компонент осуществляется в следующем порядке:

- 1) Определяется центр облака данных и в него переносится новое начало координат – это нулевая главная компонента (PC0).
- 2) Выбирается направление максимального изменения данных – это первая главная компонента (PC1).

- 3) Если величина шума существенна, то выбирается новое направление (РС 2) – перпендикулярное к первому, так чтобы описать оставшееся изменение в данных и так далее (Рис. 18).

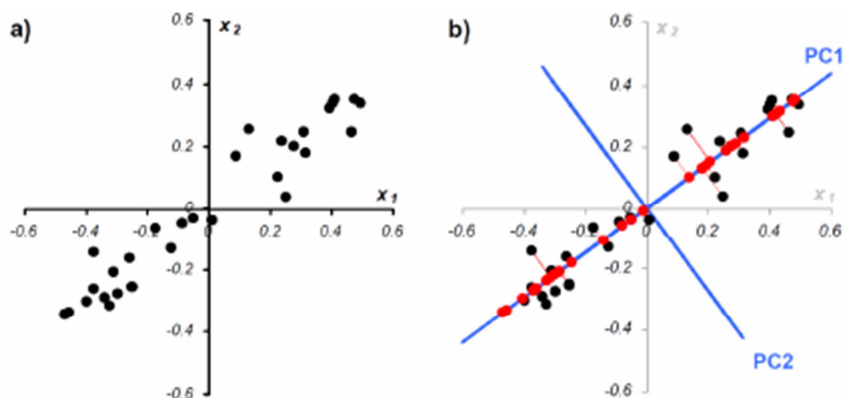


Рисунок 18. – графическая иллюстрация метода главных компонент (а – данные в исходных координатах; б – данные в координатах главных компонент).

Метод главных компонент осуществляется последовательно, шагами. На каждом шаге исследуются остатки, и среди остатков выбирается направление наибольшего изменения. Данные проецируются на эту ось и вычисляются новые остатки. Этот алгоритм называется NIPALS.

ПРОТОКОЛЫ ЛАБОРАТОРНЫХ РАБОТ ПО ДИСЦИПЛИНЕ «СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Лабораторная работа № 1.

«Рандомизированный план эксперимента»

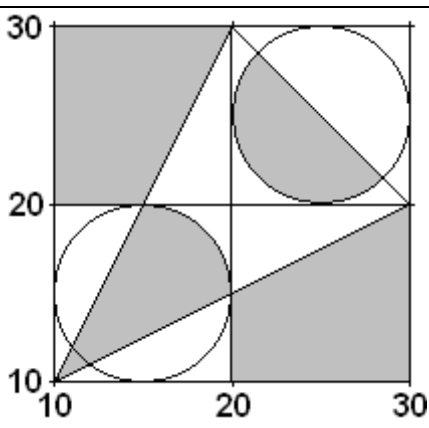
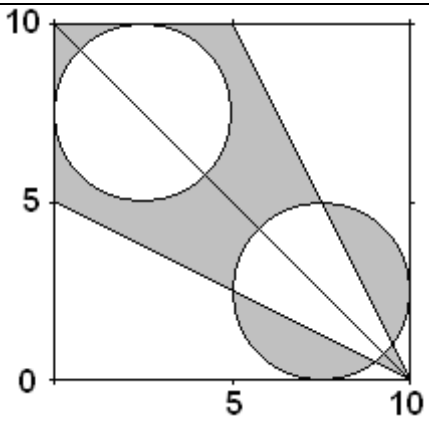
Цель работы: составление рандомизированного плана эксперимента.

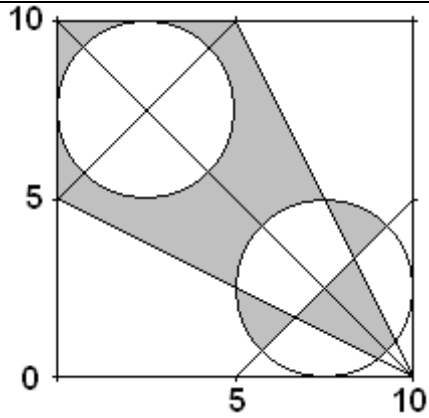
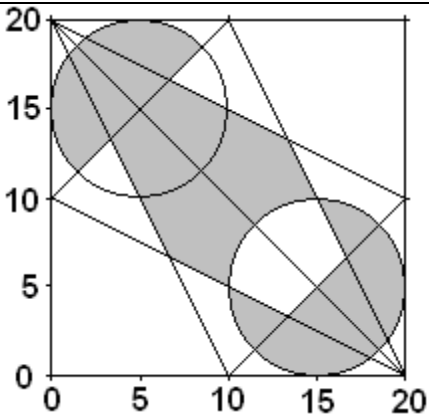
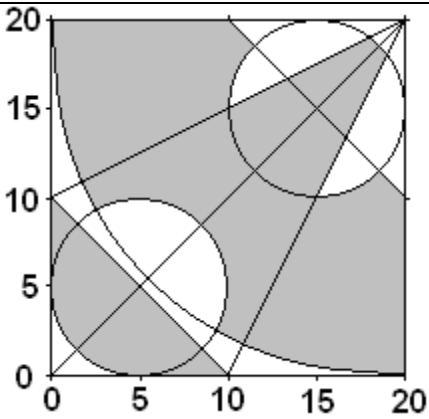
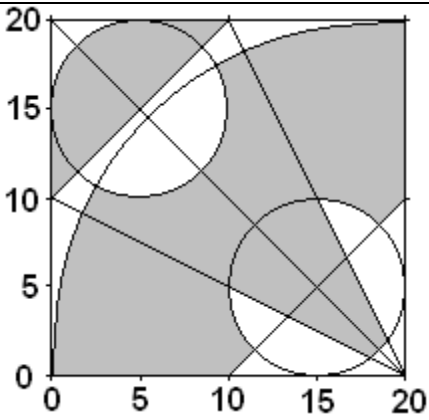
Работа выполняется средствами математической среды "MathCad".

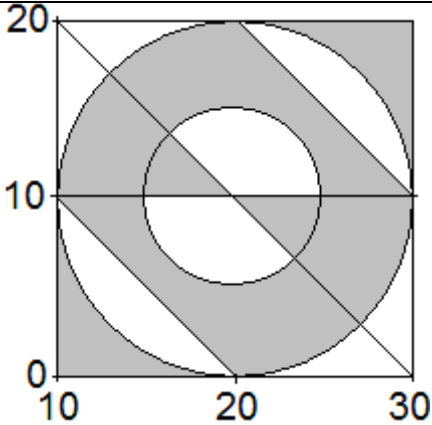
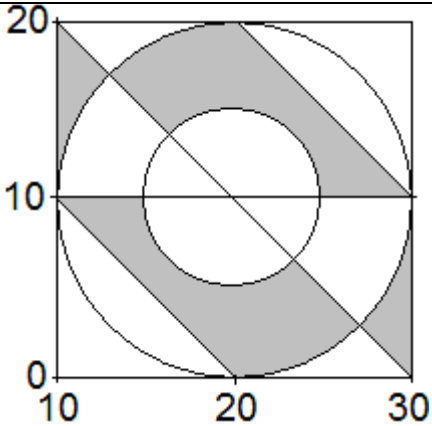
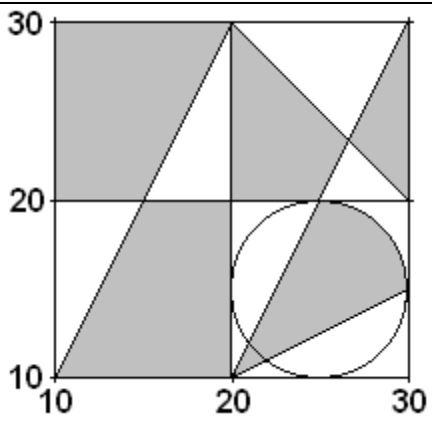
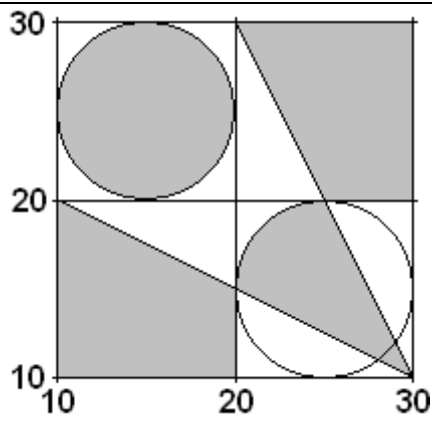
Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

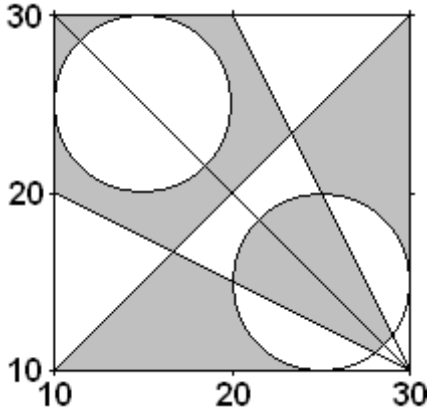
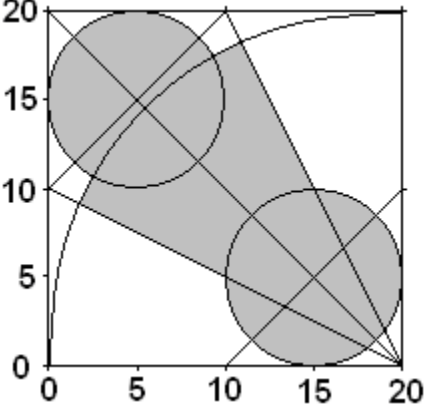
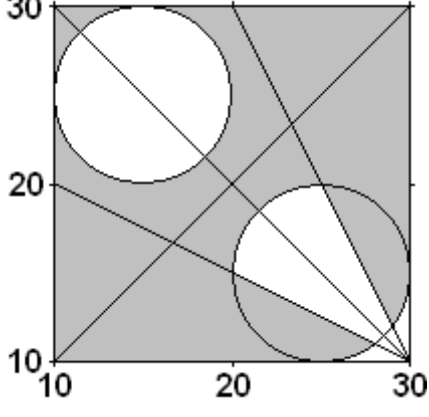
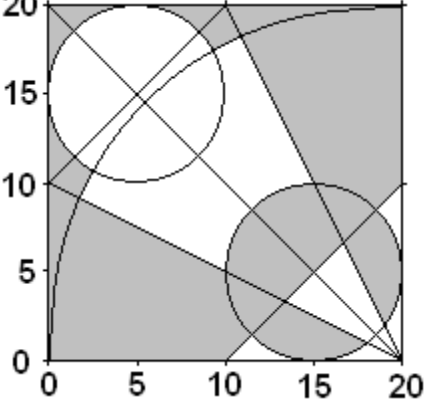
Содержание задания:

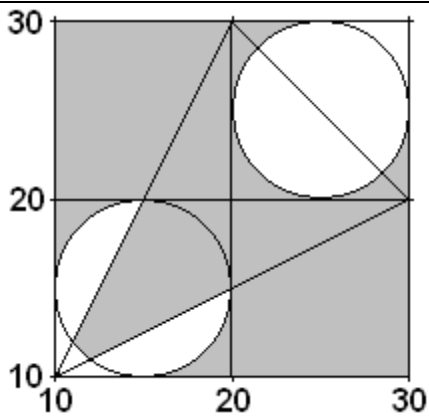
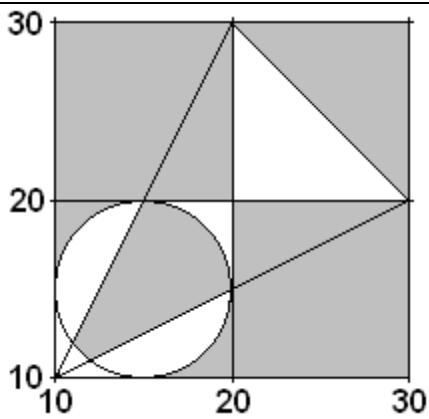
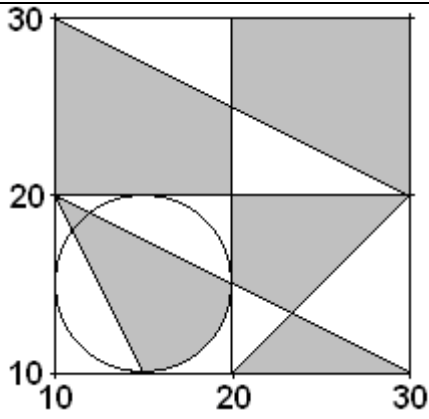
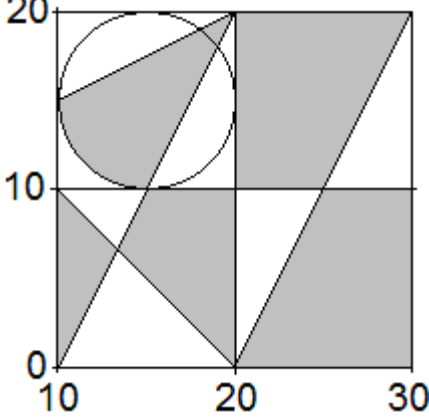
Сгенерировать две последовательности случайных чисел X и Y , содержащих по 1000 элементов по равномерному закону в заданном диапазоне. Выполнить сортировку пар элементов, соответствующих выделенным областям. Отметить точки, попавшие в выделенные области символом «X», а не попавшие символом «•».

Вариант	Визуализация плана	ФИО исполнителя
1		
2		

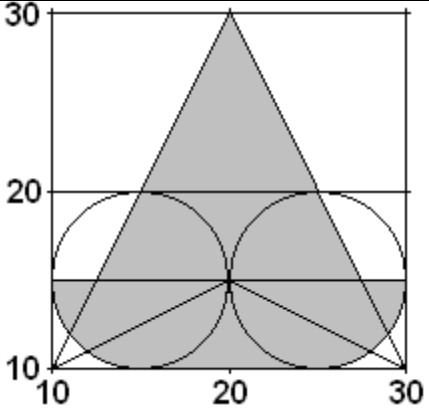
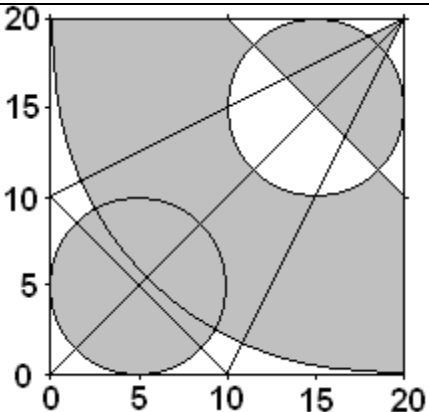
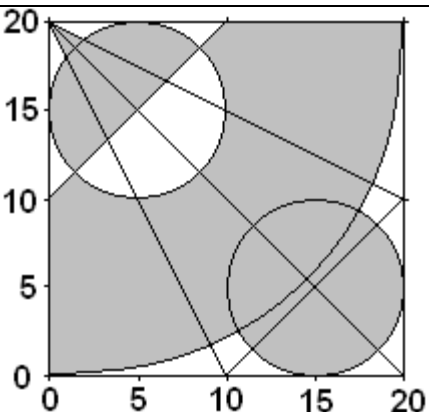
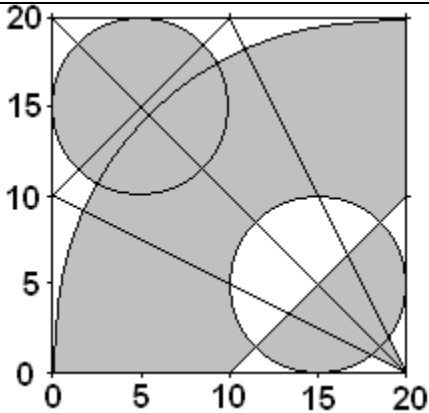
3		
4		
5		
6		

7		
8		
9		
10		

11		
12		
13		
14		

15		
16		
17		
18		

19		
20		
21		
22		

23		
24		
25		
26		

Лабораторная работа № 2.
«Метод случайного баланса»

Цель работы: оценка значимости факторов по экспериментальным данным.

Работа выполняется средствами математической среды "MathCad".

Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

Содержание задания:

1. Загрузите вектор результатов эксперимента из файла "Results.prn"
2. Загрузите матрицу значений факторов, варьируемых в ходе эксперимента из файла "Factors.prn". Стоки матрицы соответствуют опытам, столбцы – факторам.
3. Разбейте факторы на группы в соответствии с вариантом задания. Номера факторов в группах указаны в таблице.
4. Приведите значения факторов к двухуровневому виду.
5. Сформируйте групповые матрицы планирования эксперимента.
6. Выполните рандомизацию строк групповых матриц планирования.
7. Составьте общую матрицу путём построчной стыковки групповых матриц.
8. Определите медианы для каждого фактора.
9. Постройте диаграмму рассеяния результатов эксперимента.
10. Вычислите вклад каждого фактора в результаты эксперимента. Постройте гистограмму вкладов факторов.
11. Создайте подпрограмму, осуществляющую ранжировку факторов по величине вклада, т.е. формирующую вектор «Numbers» из номеров факторов в порядке убывания абсолютной величины вклада.
12. Создайте подпрограмму, формирующую вектор «First» из номеров первичных факторов и вектор «Second» из номеров вторичных факторов так, чтобы абсолютная величина вклада каждого первичного фактора была не меньше 60% от максимальной абсолютной величины вклада.
13. Подготовьте отчет и предоставьте его на проверку преподавателю. При построении графиков необходимо эффективно использовать возможности выбора пределов и масштабов.

Вариант	Номера факторов в общем списке			ФИО исполнителя
	Группа I	Группа II	Группа III	
1	2,7,9	3,4,10	1,5,6,8	
2	1,4,10	2,5,8	3,6,7,9	
3	3,5,8	1,6,7	2,4,9,10	
4	4,6,9	2,5,8	1,3,7,10	
5	2,3,8,10	1,5,7	4,6,9	
6	3,5,8,9	4,6,7	1,2,10	
7	1,5,6	2,4,9,10	3,7,8	
8	2,4,5	1,7,8,9	3,6,10	
9	1,3,6	4,5,10	2,7,8,9	
10	2,6,8	1,3,4	5,7,9,10	
11	3,4,10	1,7,8,9	2,5,6	
12	2,3,7,9	1,4,10	5,6,8	

13	1,3,6,8	4,7,10	2,5,9	
14	2,7,9	4,8,10	1,3,5,6	
15	1,4,10	2,6,8	3,5,7,9	
16	3,6,8	1,5,7	2,4,9,10	
17	4,6,9	2,5,7,8	1,3,10	
18	2,3,8,10	1,6,9	4,5,7	
19	3,5,8	4,6,7,9	1,2,10	
20	1,5,6	4,9,10	2,3,7,8	
21	2,4,5	1,6,8,9	3,7,10	
22	1,3,5	4,6,10	2,7,8,9	
23	2,6,8,10	1,3,4	5,7,9	

Лабораторная работа № 3.

«Интерполяция экспериментальных данных»

Цель работы: Интерполяция экспериментальных данных.

Работа выполняется средствами математической среды "MathCad".

Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

Содержание задания:

В таблице 1 приведены результаты эксперимента:

Таблица 1

X	-2	4	22	25	8	21	15	13	7	-5	24	12
Y	13	7	11	7	10	13	14	14	5	22	8	11

1. Выполните интерполяцию данных, используя полином Лагранжа.
2. Постройте графики интерполяционных функций $f(x)$, применяя линейную, квадратичную и кубическую интерполяцию.
3. Постройте все четыре графика на одном поле в среде MathCad.
4. Выделите вертикальными маркерами область экспериментирования, а экспериментальные точки символом «•».
5. Подготовьте отчет и предоставьте его на проверку преподавателю. При построении графиков необходимо эффективно использовать возможности выбора пределов и масштабов.

Лабораторная работа № 4.

«Интерполяция температурного поля»

Цель работы: Восстановление температурного поля по экспериментальным значениям температуры в узловых точках.

Работа выполняется средствами математической среды "MathCad". Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

Содержание задания:

В таблице указаны значения температуры T в некоторых точках плоскости с известными координатами X и Y :

$Y, мм$	$X, мм$				
	20	40	50	80	120
10	13	72	124	112	72
20	41	56	109	82	56
40	53	38	88	71	42
70	39	29	62	59	39
100	26	43	30	51	72

1. Сформируйте вектора значений факторов и матрицу значений температуры.

2. Постройте интерполяционную функцию $T(X,Y)$ с шагом $\Delta X = \Delta Y = 1\text{мм}$, используя кубическую сплайновую интерполяцию.
3. Постройте поверхность, соответствующую функции $T(X,Y)$. Область визуализации температурного поля должна совпадать с областью экспериментирования. При заливке поверхности использовать палитру «Радуга». Области с наибольшей температурой должны иметь красный цвет, а с наименьшей – синий.
4. Визуализируйте на поверхности линии уровня, соответствующие температурам, кратным 10°C .
5. Включите визуализацию экспериментальных точек в виде красных шариков.
6. Постройте горизонтальную плоскость на уровне 70°C . Коэффициент прозрачности плоскости задать равным 40%.
7. Подготовьте отчет и предоставьте его на проверку преподавателю. При построении графиков необходимо эффективно использовать возможности выбора пределов и масштабов.

Лабораторная работа № 5.

«Аппроксимация и регрессия»

Цель работы: Восстановление температурного поля по экспериментальным значениям температуры в узловых точках методом наименьших квадратов.

Работа выполняется средствами математической среды "MathCad".

Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

Содержание задания:

1. Загрузите экспериментальные данные из файла “*data.txt*”. Массив данных содержит 3 столбца: две координаты X , Y и соответствующее значение температуры $T(X,Y)$. Сформируйте вектора координат X , Y и вектор значений температуры T .
2. Сформируйте линейную аппроксимационную функцию $F1(x,y)$.
3. Вычислите параметры аппроксимационной функции $F1(x,y)$, используя метод наименьших квадратов.
4. Вычислите среднее квадратичное отклонение расчётных точек от экспериментальных.
5. Сформируйте квадратичную аппроксимационную функцию $F2(x,y)$.
6. Вычислите параметры аппроксимационной функции $F2(x,y)$, используя метод наименьших квадратов.
7. Вычислите среднее квадратичное отклонение расчётных точек от экспериментальных.
8. Сформируйте аппроксимационную функцию девятого порядка $F9(x,y)$ с помощью стандартной функции `regress(***)`.

9. Создайте расчётные матрицы температурного поля на базе аппроксимационных функций первого и третьего порядков на пространственной сетке с координатным шагом $\Delta X = \Delta Y = 1\text{мм}$ в пределах: от 2 до 11 по оси Ox и от 4 до 15 по оси Oy .
10. Постройте поверхность, соответствующую функции $F1(x,y)$. Заливку поверхности выполнить синим цветом.
11. Постройте поверхность, соответствующую функции $F9(x,y)$. При визуализации поверхности использовать палитру «Fire».
12. Область визуализации по осям Ox, Oy, Oz ограничить диапазонами [2;11], [4;15], [0;800] (соответственно). Коэффициент прозрачности поверхностей задать равным 40%.
13. Включите визуализацию экспериментальных точек в виде шариков.
14. Подготовьте отчет и предоставьте его на проверку преподавателю. При построении графиков необходимо эффективно использовать возможности выбора пределов и масштабов.

Лабораторная работа № 6.

«Экспериментальная проверка статистической гипотезы»

Цель работы: Экспериментальная проверка статистической гипотезы о соответствии экспериментальных выборок скорости ветра закону Вейбулла.

Работа выполняется средствами математической среды "MathCad".

Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

Содержание задания:

1. Загрузите метеоданные из файла ВММ-2015.
2. Постройте графики зависимостей температуры на высотах 2, 121 и 301м от номера измерения (по вертикали – температура, по горизонтали – номер измерения).
3. Постройте графики зависимостей скорости ветра на высотах 8, 121 и 301м от номера измерения (по вертикали – скорость, по горизонтали – номер измерения).
4. Выполните группировку результатов измерений по величине скорости ветра на заданной в таблице вариантов высоте с шириной интервала $\Delta = 0.1\text{м/с}$.
5. Постройте распределение скорости ветра, полученное в результате группировки, в виде столбчатой диаграммы для годовой выборки (по горизонтали – скорость ветра, по вертикали – соответствующее количество измерений скорости ветра).
6. Выполните нормировку высоты столбцов экспериментальной гистограммы. Проверьте выполнение условия нормировки.
7. Вычислите параметры трёхпараметрического закона Вейбулла* применительно к распределению скорости ветра, используя метод максимального правдоподобия.

8. Постройте график теоретической плотности вероятности скорости ветра совместно с экспериментальной гистограммой.
9. Выполните проверку гипотезы о согласии закона Вейбулла* с экспериментальным распределением, используя заданный в таблице вариантов критерий согласия.
10. Повторите пункты 4 – 9 для экспериментальной выборки заданного в таблице месяца.
11. Подготовьте отчет и предоставьте его на проверку преподавателю. При построении графиков необходимо эффективно использовать возможности выбора пределов и масштабов.

Номер варианта	Фамилия исполнителя	Высота установки датчика, м	Критерий согласия	Месяц
1		8	Пирсона	Январь
2		121	Колмогорова	Февраль
3		301	Купера	Март
4		8	Пирсона	Апрель
5		8	Колмогорова	Май
6		121	Купера	Июнь
7		301	Пирсона	Июль
8		8	Колмогорова	Август
9		8	Купера	Сентябрь
10		121	Пирсона	Октябрь
11		301	Колмогорова	Ноябрь
12		8	Купера	Декабрь
13		121	Пирсона	Январь
14		301	Колмогорова	Февраль
15		8	Купера	Март
16		121	Пирсона	Апрель
17		301	Колмогорова	Май
18		8	Купера	Июнь
19		8	Пирсона	Июль
20		121	Колмогорова	Август
21		301	Купера	Сентябрь
22		8	Пирсона	Октябрь
23		8	Колмогорова	Ноябрь
24		121	Купера	Декабрь

25		301	Пирсона	Январь
26		8	Колмогорова	Февраль
27		121	Купера	Март
28		301	Пирсона	Апрель
29		8	Колмогорова	Май
30		121	Купера	Июнь
31		301	Пирсона	Июль
32		8	Колмогорова	Август

Лабораторная работа № 7.
«Метод главных компонент»

Цель работы: Определение главных компонент выборки экспериментальных данных.

Работа выполняется средствами математической среды "MathCad". Текст программы должен начинаться с указания номера группы, фамилии и инициалов исполнителя.

Содержание задания:

1. Загрузите метеоданные из файлов BMM_2008, 2009, ..., 2015.
2. Сформируйте последовательности экспериментальных значений величин, указанных в таблице вариантов. Исключите элементы последовательностей, соответствующие пропущенным измерениям.
3. Объедините годовые выборки заданных в таблице вариантов величин в многолетние последовательности.
4. Постройте облако экспериментальных данных (указанных в таблице вариантов величин) в соответствующей системе координат. Экспериментальные точки отобразите чёрными шариками.
5. Вычислите координаты центра облака, учитывая все экспериментальные точки. Визуализируйте центр облака в виде синего шарика.
6. Вычислите две главные компоненты $PC1$ и $PC2$.
7. Постройте на плоскости вектора $\overrightarrow{PC1}$ и $\overrightarrow{PC2}$ поверх облака экспериментальных данных. Вектора выделить красным цветом.
8. Проверьте ортогональность найденных главных компонент $\overrightarrow{PC1}$ и $\overrightarrow{PC2}$, используя скалярное произведение векторов: $\overrightarrow{PC1} \cdot \overrightarrow{PC2} = 0$.
9. Выполните аппроксимацию экспериментальных данных главными компонентами и определите средне-квадратическое отклонение аппроксимационной функции от экспериментальных данных.
10. Подготовьте отчет и предоставьте его на проверку преподавателю. При построении графиков необходимо эффективно использовать возможности выбора пределов и масштабов.

Номер варианта	Фамилия исполнителя	Переменная № 1	Переменная № 2
1		V8	T2
2		V121	T121
3		V301	T301
4		V8	V121
5		V121	V301
6		V301	V8
7		T2	F8
8		T121	F121
9		T301	F301
10		T2	T121
11		T121	T301
12		T301	T2
13		R2	P2
14		R2	T2
15		R2	V8
16		R2	F8
17		P2	T2
18		P2	V8
19		P2	F8
20		V8	F8
21		V121	F121
22		V301	F301
23		F8	F121
24		F121	F301
25		F301	F8
26		T2	V8
27		T121	V121
28		T301	V301
29		V121	V8
30		V301	V121
31		V8	V301
32		F8	T2

ЛИТЕРАТУРА

1. Бакеев Д.А., Ильина И.В., Ильин И.А. Основы научных исследований. Экспериментальное исследование технических устройств: Учебное пособие для студентов специальности 201300 «Техническая эксплуатация транспортного радиооборудования». – Петропавловск-Камчатский: КамчатГТУ, 2003. – 94 с.
2. Грановский В.А. Сирая Т.Н. Методы обработки экспериментальных данных при измерениях Л.: ЭНЕРГОАТОМИЗДАТ 1990г. - 288 с.
3. Князев Б.А., Черкасский В.С. Начала обработки экспериментальных данных. Электронный учебник и программа обработки данных для начинающих: Учебное пособие // Новосиб. ун-т. Новосибирск, 1996. 93 с.
4. Никитин В.И. Первичная статистическая обработка экспериментальных данных: мет. кк. по вып. к.р. / В.И. Никитин. – Самара: Самар. гос. техн. ун-т, 2017- 80 с.
5. Подготовка и оформление лабораторной, курсовой и дипломной работ [текст]:метод. указания к выполнению практических работ для студентов специальности 010400 МИФО / сост.: Г. А. Александрова, А. С. Паршин; СибГАУ. Красноярск, 2004. – 29 с.
6. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений М.: Наука, 1968. - 288 с.
7. Румшицкий Л.З. Математическая обработка результатов эксперимента М.: Наука. ФИЗМАТЛИТ, 1971 г. - 192 с.
8. Статистическая обработка и анализ экспериментальных данных массовых случайных явлений: Методические к расчетно-графической работе для студентов II курса всех специальностей / Сост. О.А. Романчук, В. И. Чеботарев, Г. И. Целоусова, В. В. Мухранова. — Хабаровск: Изд-во Хабар. гос. техн. ун-та, 2004. — 21 с.

Электронное учебное издание

Людмила Ивановна **Медведева**

Виктор Иванович **Капля**

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Учебно-методическое пособие

Электронное издание сетевого распространения

Редактор Матвеева Н.И.

Темплан 2018 г. Поз. № 6В.

Подписано к использованию 23.10.2018. Формат 60x84 1/16.

Гарнитура Times. Усл. печ. л. 4,19.

Волгоградский государственный технический университет.

400005, г. Волгоград, пр. Ленина, 28, корп. 1.

ВПИ (филиал) ВолгГТУ.

404121, г. Волжский, ул. Энгельса, 42а.