МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
________ О.Г. Локтионова
« 1 » 02 2018 г.

МЕТА-АНАЛИЗ В МЕДИЦИНСКИХ И ЭКОЛОГИЧЕСКИХ СИСТЕМАХ

Методические рекомендации по организации и выполнению самостоятельной работы для аспирантов направления подготовки 09.06.01 и 12.06.01

Курск 2018

УДК 004.93:61

Составитель: С.А. Филист

Рецензент
Доктор технических наук, профессор А.Ф. Рыбочкин

Мета-анализ в медицинских и экологических системах: методические рекомендации по организации и выполнению самостоятельной работы / Юго-Зап. гос. ун-т; сост.: С.А. Филист. - Курск, 2018. - 96 с.

Методические указания по структуре, содержанию и стилю изложения материала соответствуют методическим и научным требованиям, предъявляемым к учебным и методическим пособиям.

Предназначены для аспирантов направления подготовки 09.06.01 «Информатика и вычислительная техника (Системный анализ, управление и обработка информации (технические и медицинские системы)) и 12.06.01 «Фотоника, приборостроение, оптические и биотехнические системы и технологии (Приборы, системы и изделия медицинского назначения)»

Текст печатается в авторской редакции

Подписано в печать 1.03.18. Формат 60x84 1/16.
Усл.печ.л. 5,4. Уч.-изд.л. 4,9 Тираж 100 экз. Заказ: 1435. Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Самостоятельная работа №1
«Элементарные задачи мета-анализа»

Цель работы: Рассмотреть возможности MathCAD для решения элементарных задач математической статистики. Научиться использовать возможности MathCAD для ввода и вывода файловых данных. Познакомиться с расчетом основных выборочных характеристик в среде MathCAD. Научиться представлять графически выборку случайных величин в виде гистограмм и полигонов.

Краткие теоретические сведения

В большинстве статистических расчетов приходится иметь дело либо со случайными данными, полученными в ходе какого-либо эксперимента (которые выводятся из файла или печатаются непосредственно в документе), либо с результатами генерации случайных чисел.

Случайной выборкой называется случайный вектор, элементы которого независимы и одинаково распределены. Обычно под *выборкой* подразумевают результаты независимых измерений, которые проводятся в одинаковых условиях.

Ввод и вывод файлов данных

Важный компонент ввода-вывода — это ввод-вывод во внешние файлы. Ввод внешних данных в документы Mathcad применяется чаще вывода, поскольку Mathcad имеет гораздо лучшие возможности представления результатов расчетов, чем многие пользовательские программы. Для общения с внешними файлами данных в Mathcad имеется несколько разных способов.

Самый простой из них — использовать имеющееся семейство встроенных функций.

- READPRN (“file”) - чтение данных в матрицу из текстового файла;
- WRITEPRN (“file”) - запись данных в текстовый файл;
- APPENDPRN (“file”) - дозапись данных в существующий текстовый файл;

- file — путь к файлу.

Встроенная функция APPENDPRN может применяться и для создания нового файла. Иными словами, если файла с заданным именем не существовало, то он, после применения, будет создан и наполнен теми данными, которые Вами определены в документе.

Для удобства можно использовать функцию CWD - указания полигона, где необходимо создать файл или где находиться считываемый файл.

Можно задавать как полный путь к файлу, например, C:\Мои документы, так и относительный, имея в виду, что он будет отсчитываться от папки, в которой находится файл с документом Mathcad. В качестве имени файла можно использовать русские буквы.

Пример 1: Запись данных в файл "da.ta.txt"

x:=rpost(N, a, σ) создание выборки случайных величин распределенных по нормальном закону;

CWD:= "D:\tmp\" устанавливается текущий рабочий каталог.

WRITEPRN(data.txt") :=x запись в файл «data»созданной ранее выборки x.

Моделирование выборок из стандартных распределений

Mathcad обладает богатой библиотекой встроенных функций, предназначенных для генерации выборок из генеральных совокупностей с наиболее распространенными стандартными распределениями.

Вставку рассмотренных ранее статистических функций в программы удобно осуществлять с помощью диалогового окна *Insert Function* (Вставка функции).

Для этого необходимо выполнить следующие действия:

1. Установить курсор на место вставки функции в документе.
2. Вызвать диалоговое окно *Insert Functio*n нажатием кнопки f(x) на стандартной панели инструментов или командой меню *Insert/Function* (Вставка/Функция), или нажатием клавиш <Ctrl>+<E>.
3. Выбрать в списке *Function Category* (Категория функции) выберите одну из категорий статистических функций. Категория *Probability Density* (Плотность вероятности) содержит встроенные функции для плотности вероятности, Категория *Probability Distribution* (Функция распределения) — для вставки функций или

квантилей распределения, Категория *Random Numbers* (Случайные числа) — для вставки функции генерации случайных чисел.

4. Выбрать в списке *Function Name* (Имя функции) функцию, соответствующую требующемуся закону распределения. При выборе элемента списка в текстовом поле в нижней части окна будет появляться информация о назначении выбранной функции и ее параметрах.

5. Вставить выбранную функцию в документ нажатием кнопки "Ок".

Функции Mathcad для расчета численных характеристик

В Mathcad имеется ряд встроенных функций для расчетов числовых статистических характеристик рядов случайных данных.

- mean(x) — выборочное среднее значение, оценка математического ожидания выборки;
- median (x) — выборочная медиана (median) — значение аргумента, которое делит гистограмму плотности вероятностей на две равные части;
- var(x) — выборочная дисперсия выборки (variance);
- stdev(x) — среднеквадратичное (или "стандартное") отклонение выборки (standard deviation);
- max(x), min(x) — максимальное и минимальное значения выборки;
- mode(x) — наиболее часто встречающееся значение выборки.

Построение гистограмм

Гистограммой называется график, аппроксимирующий по случайным данным плотность их распределения. При построении гистограммы область значений случайной величины (a, b) разбивается на некоторое количество *bin* сегментов, а затем подсчитывается процент попадания данных в каждый сегмент. Для построения гистограмм в Mathcad имеется несколько встроенных функций.

Гистограмма с произвольными сегментами разбиения

- hist,(intvls,x) - вектор частоты попадания данных в интервалы гистограммы;
- intvls - вектор, элементы которого задают сегменты построения гистограммы в порядке возрастания $a < \mathrm{int\ vis}_i < b$;
- x - вектор случайных данных.

Если вектор *intvls* имеет *bin* элементов, то и результат *hist* имеет столько же элементов.

1. Для построения гистограмм созданную случайную величину предварительно необходимо упорядочить. Для этого в Mathcad имеется встроенная функция.

sort(x) - сортировка выборки в порядке возрастания;

Для того, чтобы построить гистограмму, нужно сначала сгруппировать выборочные данные, записанные в массиве x, и сохранить граничные очки интервалов группировки в векторе *intvls, размерность которого равна числу интервалов.*

2. Сформировать вектор intvls границ интервалов.
3. Определить процент попадания данных в каждый сегмент.
4. Построить гистограмму.

Пример 2. Построение гистограммы

N:=1000

x:=binom (N, Δ, 0,5)

bin:=30; кол-во равных сегментов, на кот. разбивается весь диапазон

; определение границы интервала построения гистограммы

lower:=floor (min(x)): наибольшее целое число $\le \min(x)$

upper:=ceil (max(x)); наименьшее целое число $\ge \max(x)$

$h := \dfrac{upper - lower}{bin}$; размер сегмента

j:=0…bin; счетчик сегментов

$int_j := lower + h \cdot j$

$f := \dfrac{1}{N \cdot h} \cdot hist(int, x)$; массив начальных точек каждого сегмента int:=int+0.5h; от левой границы каждого сегмента к его центру; нормирование гистограммы для удобства отображения на одном графике вместе с плотностью распределения

В векторе *int* можно задать произвольные границы сегментов разбиения так, чтобы они имели разную ширину.

Недостаток упрощенной формы функции *hist* состоит в том, что необходимо дополнительно определять вектор сегментов построения гистограммы. От этого недостатка свободна функция *histogram*.

Гистограмма с разбиением на равные сегменты

histogram (bin, x) — матрица гистограммы размера bin*2, состоящая из столбца сегментов разбиения и столбца частоты попадания в них данных;

- bin — количество сегментов построения гистограммы;
- x — вектор случайных данных.

Пример 3. Построение гистограммы (упрощенный вариант)

N:=100; созданы выборки сл. величины

x:= exp (N, 1)

bin:=30; кол-во равных сегментов, на кот. разбивается весь диапазон

f := histogram(bin, x)

Аргументы у обоих процедур hist() и histogram() одинаковы: первый определяет интервалы для создания гистограммы, а второй - это выборка, на основе которой строится гистограмма. Первый аргумент может быть либо вектором конечных точек интервалов для группировки данных выборки, либо целым числом, задающим число интервалов. В последнем случаи весь диапазон значений в выборке разбивается на равные интервалы.

Создание графика гистограммы

Для того чтобы создать график в виде гистограммы необходимо:

1. Построить двумерный график, по оси x откладываются границы интегралов, по оси y частота(процент) попадания значения сл. величины в заданные интервалы.

2. Перейти в диалоговом окне *Formatting Currently Selected Graph* (Форматирование) выбранного графика (например, двойным щелчком мыши) в раздел *Traces* (Графики). Установить в поле *Type* (Тип) элемент списка *bar* (столбцы) или *solidbar* (гистограмма). Тип solidbar специально предназначен для гистограмм.

3. Нажать кнопку ОК.

Процедура *histogram()* инициализирует объект (матрицу), содержащий срединные точки интервалов гистограммы (первый столбец) и столбец частот, попадания в заданные интервалы.

Для построения таких графиков по оси x откладывается столбец срединные точки интервалов (столбец матрицы с нулевым индексом), а по оси y - столбец с частотами распределения данных по интервалам гистограммы (столбец с первым индексом). Индекс $M^{\langle\rangle}$ столбца вводится с помощью соответствующей пиктограммы п панели Matrix или комбинации клавиш <Ctrl>+<6>.

Полигон частот

Иная форма графического представления группированных данных - полигон частот. *Полигон частот* - это ломанная линия, соединяющая точки с координатами $(\bar{x}_i, h_i)$, т.е. с абсциссами, равными серединам интервалов группировки, и ординатами, равными соответствующим частотам. Если соединить центры элементарных сегментов гистограммы ломанной линией, то получится график полигона.

Задание 1. Создайте выборку из 100 случайных величин с нормальным распределением, среднее значение m=0,1*k (k - номер варианта) и стандартное отклонение σ=0,5. Запишите данную выборку в файл с произвольным именем. Рассчитайте с помощью встроенных функций MathCad числовые статистические характеристики созданной выборки. Постройте гистограмму двумя способами и полигон частот

Порядок выполнения работы:

1. С помощью встроенной функции из категории Random Numbers (Случайные числа) получить заданную выборку.
2. Записать полученную величину в файл с произвольным названием. В отчет вставьте фрагмент этого документа.
3. Упорядочить значения в выборке случайной величины по возрастанию.
4. С помощью стандартных функций Mathcad, получить числовые характеристики: min и max значения выборки, выборочное среднее, выборочную дисперсию, среднеквадратическое отклонение, выборочную медиану.
5. Используя функцию дозаписи, добавить в созданный ранее файл числовые характеристики выборки. Фрагмент вновь созданного файла привести в отчете.
6. Считать полученный файл.
7. Выполните расчет гистограммы с помощью функцию hist (int, x) Отобразить на графиках гистограмму и плотность распределения на одном и полигон частот на другом.
8. Выполните расчет гистограммы, используя функцию histogram (int, x).
9. Выведите на экран результаты процедур hist() и histogram(). Сравните их.
10. Понаблюдайте, как изменится внешний вид гистограммы, если изменить количество интервалов разбиения выборки. Сделать выводы.

В отчете представить все необходимые фрагменты, сделанные в Mathcad, и требуемые выводы.

Контрольные вопросы

1. Что такое случайная выборка?
2. Что такое выборка?
3. Как происходит ввод и вывод данных в MathCad?
4. Как в MathCad произвести моделирование выборок из стандартных распределений?
5. Что такое функция MathCad для расчета численных характеристик?
6. Как в MathCad построить гистограммы?
7. Что такое гистограмма с произвольным сегментом разбиения?

8. Что такое гистограмма с разбиением на равные сегменты?
9. Приведите алгоритм создания графика гистограммы.
10. Что такое полигон частот?

Самостоятельная работа №2
«Структура экспертных систем»

1.1 Краткие теоретические сведения

Экспе́ртная систе́ма (ЭС, англ. expert system) — компьютерная система, способная частично заменить специалиста-эксперта в разрешении проблемной ситуации. Современные ЭС начали разрабатываться исследователями искусственного интеллекта в 1970-х годах, а в 1980-х получили коммерческое подкрепление. Предшественниками экспертных систем были предложены в 1832 году С. Н. Корсаковым, создавшим механические устройства, так называемые «интеллектуальные машины», позволявшие находить решения по заданным условиям, например, определять наиболее подходящие лекарства по наблюдаемым у пациента симптомам заболевания.

В информатике экспертные системы рассматриваются совместно с базами знаний (база знаний (БЗ; англ. knowledge base) в информатике и исследованиях искусственного интеллекта — это особого рода база данных, разработанная для оперирования знаниями. База знаний содержит структурированную информацию, покрывающую некоторую область знаний, для использования кибернетическим устройством (или человеком) с конкретной целью. Современные базы знаний работают совместно с системами поиска информации, имеют классификационную структуру (классификация (классифицирование) (от лат. classis — разряд и лат. facere — делать) — особый случай применения логической операции деления объема понятия, представляющий собой некоторую совокупность делений (деление некоторого класса на виды, деление этих видов и т.д.) и формат представления знаний (представление знаний — вопрос, возникающий в информатике — с подбором представления конкретных и обобщённых знаний, сведений и фактов для накопления и обработки информации в ЭВМ. Главная задача в искусственном интеллекте (ИИ) — научиться хранить знания таким образом, чтобы программы могли осмысленно обрабатывать их и достигнуть тем подобия человеческого интеллекта).

Полноценные базы знаний содержат в себе не только фактическую информацию, но и правила вывода, допускающие автоматические умозаключения о вновь вводимых фактах и, как следствие, осмысленную обработку информации. Область наук об искусственном интеллекте, изучающая базы знаний и методы работы со знаниями, называется инженерией знаний) как модели поведения экспертов в определенной области знаний с использованием процедур логического вывода и принятия решений, а базы знаний — как совокупность фактов и правил логического вывода в выбранной предметной области деятельности. В инструментальную систему входят помимо описанной выше экспертной оболочки программа-редактор баз знаний и программа логического вывода.

1.2 Задание

Знакомство с инструментальным программным обеспечением (ANIES) для построения экспертных систем.

Выбор задачи и предметной области для реализации учебной экспертной системы. Согласование с преподавателем задание на разработку учебной экспертной системы

Реализация базы знаний. Консультации у преподавателя о пути развития базы знаний.

Реализация и тестирование базы знаний. Отладка экспертной системы. Тестирование базы знаний учебной экспертной системы (УЭС);

Демонстрация работы учебной экспертной системы преподавателю.

1.3 Цель работы

Целью лабораторной работы является освоение технологии и методики построения экспертных систем на примере разработки учебной экспертной системы. Студент выступает в роли одновременно эксперта и инженера по знаниям.

1.4 Порядок выполнения самостоятельной работы

Составить в текстовом редакторе описание учебной экспертной системы (файл с расширением *.ies). Пример приведен в приложении А. Количество гипотез – не менее 7, количество параметров – не менее 7, число переменных – не менее 2, количество правил определяется студентом из расчета количества используемых ключевых словIF(не менее 20). Рекомендуется составить не менее 15 правил.

Запустить инструментальную систему ANIESв различных режимах логического вывода (прямой и обратный в глубину и в ширину).

От выбранного метода поиска, то есть стратегии вывода, будет зависеть порядок применения и срабатывания правил. Процедура выбора сводится к определению направления поиска и способа его осуществления.

При разработке стратегии управления выводом важно определить два вопроса:

1. Какую точку в пространстве состояний принять в качестве исходной? От выбора этой точки зависит и метод осуществления поиска — в прямом или обратном направлении.

2. перебора — глубину, в ширину, по подзадачам или иначе.

Какими методами можно повысить эффективность поиска решения? Эти методы определяются выбранной стратегией.

При обратном порядке вывода вначале выдвигается некоторая гипотеза, а затем механизм вывода как бы возвращается назад, переходя к фактам, пытаясь найти те, которые подтверждают гипотезу. Если она оказалась правильной, то выбирается следующая гипотеза, детализирующая первую и являющаяся по отношению к ней подцелью. Далее отыскиваются факты, подтверждающие истинность подчиненной гипотезы. Вывод такого типа называется управляемым целям. Обратный поиск применяется в тех случаях, когда цели известны и их сравнительно немного.

В системах с прямым выводом по известным фактам отыскивается заключение, которое из этих фактов следует. Если такое заключение удается найти, то оно заносится в рабочую память. Прямой вывод часто называются выводом, управляемым данными.

Демонстрационный прототип экспертной системы предъявить преподавателю (файл с расширением *.ies).

Инструментальная экспертная система «ANIES» является обучающей программой, предназначенной для демонстрации возможностей, которые предоставляют продукционные правила (продукционная модель знания — модель, основанная на правилах, позволяет представить знание в виде предложений типа «Если (условие), то (действие)») при логическом выводе (логический вывод — рассуждение, в котором осуществляется переход по правилам от высказывания или системы высказываний к высказыванию или системе высказываний). К логическому выводу обычно предъявляются (совместно или по отдельности) следующие требования: 1) правила перехода должны воспроизводить отношение следования логического (ту или иную его разновидность); 2) переходы в логическом выводе должны осуществляться на основе учета только синтаксических характеристик высказываний или систем высказываний.

Для работы программы необходим процессор Pentium 166 Mz, ОЗУ 16 Mb, HDD 1 Gb, русифицированная версия Windows 95 (98) или выше. Программное обеспечение включает в себя выполняемый файл ANIES.EXE.

В процессе работы программы образуются файлы баз знаний *.ies, хранящие ЭС пользователя. Все файлы хранятся в текстовом формате.

Взаимодействие пользователя с инструментальной экспертной системой осуществляется посредствам интерфейса пользователя. Одним из основных управляющих элементов интерфейса является главное меню программы, которое состоит из горизонтального меню, содержащего имена основных групп команд, и выпадающих подменю, позволяющих выбрать конкретную команду или режим работы. Такие пункты горизонтального меню, как “Файл”, ”Правка”, являются стандартными для программ. Они содержат набор команд для работы с файловой системой, облегчения редактирования текста. При помощи текстового редактора либо используя режим вставки при помощи пункта меню “Ввод данных” и панели ключевых слов, специалист по ИИ создает структуру БЗ, с использованием продукционных правил “IF-THEN-ELSE” и “CASE”, которая в последствии будет участвовать в обработке данных в режиме интерпретации. После запуска ЭС пользователь вводит ответы на запрашиваемые системой вопросы с указанием коэффициента уверенности в диапазоне [-1;1]. Отвечая, на один вопрос пользователь может указать несколько ответов или ни одного. Отсутствие ответа интерпретируется как коэффициент равный нулю. Также возможно остановить процесс обучения в любой момент. Система, используя машину логического вывода, производит подсчет всех заключений и отображает перечень гипотез. При желании пользователь может просмотреть ход срабатывания правил экспертной системы в виде протокола решения.

Знакомство с инструментальным программным обеспечением (ANIES) для построения экспертных систем.

Выбор задачи и предметной области для реализации учебной экспертной системы. Согласование с преподавателем задание на разработку учебной экспертной системы

Реализация базы знаний. Консультации у преподавателя о пути развития базы знаний.

Реализация и тестирование базы знаний. Отладка экспертной системы. Тестирование базы знаний учебной экспертной системы (УЭС);

Демонстрация работы учебной экспертной системы преподавателю

Контрольные вопросы

1. Назначение Экспертных Систем
2. На какие типы можно разделить биотехнические системы?
3. Назовите роли технического компонента в процессе управления БТС.
4. Назовите свойства БТС.
5. Назовите основные этапы синтеза БТС.
6. Как происходит информационное согласование.
7. Что необходимо для установления связи между режимами функционирования воспринимающих систем и состоянием организма оператора?

Порядок защиты работы

Работа может быть зачтена, если студент представил отчет согласно п. 1.5, исследуемые в работе сигналы соответствуют индивидуальному варианту, электронная форма соответствует представленному тексту, и студент дал исчерпывающие ответы на 10 произвольных вопросов из п. 1.6.

Список использованных источников

1. Сергиенко А.Б. Технические и медицинские системы. – СПб.: Питер. 2002. – 608 с.
2. Зубов В.С. Биотехнические системы и технологии (версии 6.0 и 7.0). – М.: Информационно-издательский дом «Филинъ», 1997. – 304 с.
3. Циммерман Франклин. Разработка систем и технологий. – М.: «Издательство БИНОМ», 1997. – 448 с.

Самостоятельная работа №3
«Моделирование случайных чисел с заданным законом распределения»

Целью работы является 1) практическое ознакомление с алгоритмами моделирования случайных чисел с заданным законом распределения; 2) изучение основных способов статистической оценки характеристик случайных чисел.

Краткие теоретические сведения

Дискретные случайные величины

Слова "случайная величина" в обыденном смысле употребляют тогда, когда хотят подчеркнуть, что неизвестно, каким будет конкретное значение этой величины. Причем иногда за этими словами скрывается просто незнание, какова эта величина.

Математик употребляет эти же слова "случайная величина", вкладывая в них определенное содержание.

«Действительно, - говорит он, - мы не знаем, какое значение примет эта величина в данном конкретном случае, но мы знаем, какие значения она может принимать, и знаем, каковы вероятности тех или иных значений. На основании этих данных мы не можем точно предсказать результат одного испытания, связанного с этой случайной величиной, но можем весьма надежно предсказать совокупность результатов большого числа испытаний. Чем больше испытаний, тем точнее будут наши предсказания».

Итак, чтобы задать случайную величину, надо указать, какие значения она может принимать, и каковы вероятности этих значений.

Случайная величина X называется дискретной, если она может принимать дискретное множество значений $x_1, x_2, ..., x_n$.

Формально случайная дискретная величина X определяется таблицей

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}. \qquad (1.1)$$

где $x_1, x_2, \dots, x_n$ - возможные значения величины X;

$p_1, p_2, \dots, p_n$ - соответствующие вероятности.

Точнее говоря, вероятность $P\{X = x_i\}$ того, что случайная величина X примет значение x_i, равна:

$$P\{X = x_i\} = p_i. \qquad (1.2)$$

Таблица (1) называется распределением случайной дискретной величины.

Числа $x_1, x_2, \dots, x_n$ могут быть вообще говоря, любыми. Однако вероятности $p_1, p_2, \dots, p_n$ должны удовлетворять двум условиям:

$$p_i > 0 \qquad (1.3)$$

и

$$p_1 + p_2 + \dots + p_n = 1. \qquad (1.4)$$

Последнее условие означает, что X обязана в каждом случае принять одно из значений $x_1, x_2, \dots, x_n$.

Кроме распределения случайной величины, которая является исчерпывающей характеристикой, вводятся числовые характеристики, основными среди которых являются математическое ожидание и дисперсия.

Получение случайных величин на ЭВМ

Сама постановка вопроса "получение случайных чисел на ЭВМ" иногда вызывает недоумение: ведь все, что делает компьютер, должно быть заранее запрограммировано; откуда же может появиться случайность?

Специалисты считают, что в этом вопросе есть определенные трудности, но они относятся скорее к философии, так что мы на них останавливаться не будем. Отметим лишь, что случайные величины, о которых шла речь в предыдущем разделе это идеальные математические понятия.

Вопрос о том, можно ли с их помощью описать какое-либо явление природы, решается опытным путем. Такое описание всегда является приближенным. Более того, случайная величина, которая вполне удовлетворительно описывает какую-то физическую величину в одном классе явлений, может оказаться плохой характеристикой этой же величины при исследовании других явлений. Точно так же дорога, которую на карте страны можно считать прямой (идеальной математической прямой "без ширины"), становится полосой с изгибами на крупномасштабном плане населенного пункта.

Обычно различают три способа получения случайных величин:

- из заранее составленных таблиц случайных чисел;
- физические генераторы случайных чисел;
- с помощью формул (генераторов или датчиков) псевдослучайных чисел.

Поскольку "качество" используемых в имитационном моделировании случайных чисел проверяется с помощью специальных тестов, можно не интересоваться тем, как эти числа получены: лишь бы они удовлетворяли принятой системе тестов.

Числа, получаемые по какой-либо формуле и имитирующие значения случайной величины X, называются псевдослучайными числами. Под словом "имитирующие" подразумевается, что эти числа удовлетворяют ряду тестов так, как если бы они были значениями этой случайной величины.

Основой или «сырьем» для моделирования случайных величин с заданным законом распределения являются так называемые базовые случайные числа. Совокупность $\{R_i\}, i = 1, 2, ...$ независимых равномерно распределенных на отрезке [0, 1] случайных величин называется последовательностью базовых случайных чисел.

Мы называем эти числа псевдослучайными потому, что фактически они остаются полностью детерминированными в том смысле, что если каждое обращение к соответствующей формуле (точнее, к алгоритму) начинается с одними и теми же исходными данными (константами и начальными значениями), то на выходе получаются одинаковые последовательности чисел R.

В настоящее время почти все стандартные библиотечные программы вычисления равномерных случайных чисел основаны на конгруэнтных методах, разработанных Лемером.

Основная формула мультипликативного конгруэнтного метода Лемера имеет вид:

$$R_{i+1} = aR_i (\mathrm{mod}\ m), \tag{1.5}$$

где a и m – неотрицательные целые числа.

Согласно этому выражении, нужно взять случайное число R_i, умножить его на постоянный коэффициент a и взять модуль полученного числа m (т.е. разделить на aR_i и остаток считать как R_{i+1}). Поэтому для вычисления (или генерирования) последовательности R_i нам необходимы начальные значения R_0, множитель a и модуль m. Выбираются a, R_0 и m так, чтобы обеспечить максимальную длину (или, как говорят, период) неповторяющейся последовательности R_i и минимальную корреляцию между генерируемыми числами.

На рисунке 1.1 показан фрагмент среды MathCad, на котором проиллюстрирована математическая реализация этого метода.

Переменной A присваивается значение $a = 5^{13} = 1220703125$, переменной m – значение $m = 2^{31} + 1 = 2147483649$. Функция $\mathrm{mod}(x_1, x_2)$ вычисляет остаток от целочисленного деления первого аргумента во второй. Получаем последовательность $\{X\}$ псевдослучайных чисел, равномерно распределенных от 0 до m. Делим каждый член этой последовательности на m, получаем базовую последовательность $\{R_i\}$ – числа равномерно распределенные от 0 до 1.

Методы генерации псевдослучайных чисел с заданным законом распределения

Базовые случайные числа позволяют генерировать новые случайные последовательности, подчиняющиеся любому закону распределения.

$N := 50 \qquad n := 1..N$

$A := 5^{13} \qquad m := 2^{31} + 1$

$A = 1.220703125 \cdot 10^{9} \qquad m = 2.147483649 \cdot 10^{9} \qquad X_0 := 111111111$

$$X_n := \mathrm{mod}(A \cdot X_{n-1}, m) \qquad R_0 := \frac{X_0}{m} \qquad R_n := \frac{X_n}{m}$$

X =

	0
0	$1.11111111 \cdot 10^{8}$
1	$4.23519126 \cdot 10^{8}$
2	$1.365173986 \cdot 10^{9}$
3	$1.459022034 \cdot 10^{9}$
4	$1.442895947 \cdot 10^{9}$
5	$1.06708695 \cdot 10^{9}$
6	$6.7659776 \cdot 10^{8}$
7	$1.160823598 \cdot 10^{9}$
8	$1.586936277 \cdot 10^{9}$
9	$1.222270366 \cdot 10^{9}$

R =

	0
0	0.052
1	0.197
2	0.636
3	0.679
4	0.672
5	0.497
6	0.315
7	0.541
8	0.739
9	0.569
10	0.433
11	0.298
12	0.95
13	0.128

Рисунок 1.1 – Моделирование базовой последовательности мультипликативным конгруэнтным методом. Фрагмент среды MathCad

Существует два основных пути преобразования базовых случайных чисел $\{R_i\}$, в случайные числа $\{y_i\}$, распределенные по заданному закону распределения.

Один из них, который называется методом инверсии, состоит в реализации определенных арифметических операций над базовым числом R_i, чтобы получить y_i.

Второй метод основывается на моделировании условий соответствующей предельной теоремы теории вероятностей. Кроме указанных двух основных подходов можно также выделить эвристические способы генерирования случайных чисел.

Метод инверсии

Моделирование случайной величины, равномерной на (a, b)

Предположим, что нам необходимо составить программу для моделирования входного потока заявок распределенного по равномерному закону в интервале (a, b).

Уравнение метода инверсии (1.6) для рассматриваемого случая выглядит так:

$$\int_a^y \frac{dy}{b-a} = R, \tag{1.6}$$

где R – равномерно распределенное случайное число на (0; 1), т.е. базовое число. Это интегральное уравнение решается легко и ответ ясен:

$$\frac{y-a}{b-a} = R. \tag{1.7}$$

Отсюда мы имеем явное выражение для y:

$$y = a + R(b-a), \tag{1.8}$$

где R – как обычно, базовое случайное число.

Моделирование экспоненциальной случайной величины

Как известно, случайная величина x, распределенная по экспоненциальному закону описывается следующей плотностью распределения:

$$p(x) = \lambda e^{-\lambda x} \tag{1.9}$$

На рисунке 1.2 построены графики экспоненциальных плотностей распределения при различных параметрах λ.

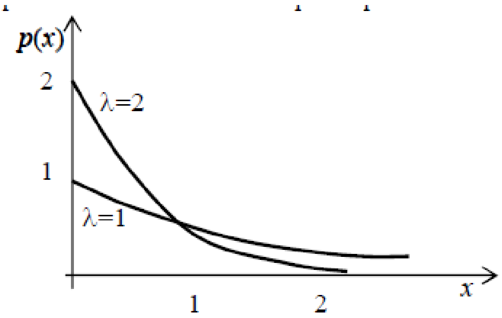


Рисунок 1.2 – Экспоненциальная плотность вероятностей $p(x) = \lambda e^{-\lambda x}$ с разными значениями параметра λ

Экспоненциальному распределению, как правило, подчиняется случайный интервал времени τ между поступлениями заявок в систему массового обслуживания. Поэтому весьма важно уметь моделировать потоки заявок разной интенсивности λ.

Напомним, что математическое $M[\tau]$ ожидание экспоненциально распределенной случайной величины τ равно:

$$M[\tau] = 1/\lambda,$$

а дисперсия: $D[\tau] = 1/\lambda^2$.

Чтобы найти алгоритм имитации экспоненциально распределенных чисел τ, применим метод инверсии:

$$\int_0^{\tau} \lambda e^{-\lambda x} = R \qquad (1.10)$$

$$1 - e^{-\lambda\tau} = R, \qquad (1.11)$$

откуда

$$\tau = -\frac{1}{\lambda}\ln(1 - R), \qquad (1.12)$$

но, поскольку случайная величина (1 - R) распределена точно так же, как R, и находится в том же интервале (0,1), то (1.12) можно заменить на более удобную формулу:

$$\tau = -\frac{1}{\lambda}\ln R, \qquad (1.13)$$

что дает искомый ответ.

Моделирование нормальной случайной величины на основе центральной предельной теоремы

Нормальное (или гауссово) распределение (рисунок 1.3) - это, несомненно, один из наиболее важных и часто используемых в имитационном моделировании видов непрерывных распределений.

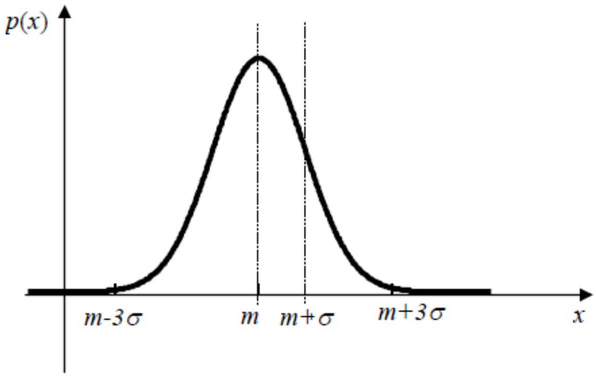


Рисунок 1.3 – Нормальная (гауссовская) плотность вероятностей

Плотность вероятности нормально распределенной случайной величины записывается так:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (1.14)$$

где m и σ - параметры нормального распределения $m = M_x$ - математическое ожидание; σ - среднеквадратическое отклонение.

Интегральная функция распределения нормальной случайной величины равна

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}} dx. \qquad (1.15)$$

Поэтому алгоритмы моделирования нормальных случайных чисел базируются на предельных теоремах теории вероятностей. Центральная предельная теорема говорит о том, что сумма n одинаково распределенных независимых случайных величин x со средним M_x и дисперсией D_x стремится к нормально распределенной величине с параметрами nM_x и nD_x при бесконечном увеличении n. Следствием теоремы является, в частности, и то, что для получения нормальной выборки, можно воспользоваться базовыми случайными числами R. Идея алгоритма состоит в следующем. Определим новую случайную величину s в виде суммы базовых чисел R_i, (i=1, 2, 3, …, n):

$$s = R_1 + R_2 + ... + R_n. \tag{1.16}$$

Тогда, согласно утверждению центральной предельной теоремы, случайная величина s является асимптотически нормальной величиной с математическим ожиданием M_s и дисперсией D_s равными соответственно:

$$M_s = n/2, \tag{1.17}$$

и

$$D_s = n/12. \tag{1.18}$$

Для практического использования формула (1.16) неудобна (поясните почему), поэтому введем вспомогательную случайную величину z равную

$$z = \frac{(s - n/2)}{\sqrt{n/12}} \tag{1.19}$$

Из (1.19) следует, что z – случайная величина, распределенная по нормальному закону с нулевым средним и единичной дисперсией. Тогда для любого нормального распределения со средним μ и дисперсией σ^2 случайное отклонение y, соответствующее указанным выше n случайным числам, получается из формулы

$$\frac{(y-\mu)}{\sigma} = z = \frac{\left(s - \frac{n}{2}\right)}{\sqrt{n/12}} \tag{1.20}$$

Следовательно,

$$y = \mu + \frac{\sigma(s + n/2)}{\sqrt{n/12}} = \mu + \frac{\sigma}{\sqrt{n/12}}\left(\sum_{i=1}^{n} R_i + n/2\right). \tag{1.21}$$

Согласно той же предельной теореме, нормальность достигается быстро даже при сравнительно небольших значениях n. В практических задач n обычно принимается равным 12. При этом последняя формула упрощается и принимает вид:

$$y = \mu + \sigma\left(\sum_{i=1}^{12} R_i + 6\right). \tag{1.22}$$

Формула (1.22) и дает алгоритм моделирования нормальных случайных чисел с требуемыми параметрами μ и σ.

Описанный метод считается малоэффективным, так как требует генерации нескольких случайных базовых чисел R для получения одного нормального выборочного значения y.

Оценка статистических характеристик случайных величин

При решении многих прикладных задач необходимые вероятностные характеристики соответствующих случайных величин неизвестны исследователю и должны определяться по экспериментальным данным. Такое статистическое описание результатов наблюдений, построение и проверка различных математических моделей, использующих понятие вероятности, составляют основное содержание математической статистики. Фундаментальными понятиями статистической теории являются понятия генеральной совокупности и выборки.

Генеральная совокупность - совокупность всех мыслимых (возможных) результатов наблюдений над случайной величиной, которые в принципе могут быть проведены при данных условиях.

Содержательный смысл этого понятия состоит в том, что предполагается существование некоторых вполне определенных свойств, неслучайных закономерностей, присущих данной совокупности. Эти свойства и должны быть определены исследователем. Фактически эти свойства являются объективным отображением вероятностных свойств изучаемого объекта, которые могут быть охарактеризованы с помощью соответствующих законов распределения вероятностей или связанных с ними числовых параметров. Как правило, считается, что указанные свойства не изменяются во времени.

Выборка - это конечный набор $x_1, x_2, ..., x_N$ значений случайной величины, полученный в результате наблюдений. Число элементов N выборки называется ее объемом или размером.

Заметим, что выборка может иметь и совпадающие значения x_i случайной величины X. Интуитивно понятно, что чем больше объем выборки, тем более точно она должна отражать статистические свойства случайной величины. Определение. Выборка называется репрезентативной (представительной), если она достаточно полно характеризует свойства генеральной совокупности.

Для обеспечения репрезентативности выборки чаще всего используют метод случайного выбора элементов. Предполагается, что при таком выборе каждая возможная выборка фиксированного объема имеет одну и ту же вероятность выбора, а последовательные наблюдения взаимно независимы.

Оцениванием в статистике называется указание приближенного значения интересующего нас параметра (или функции от некоторых параметров) на основе наблюдаемых (экспериментальных) данных, представленных в виде выборки ограниченного объема.

Оценка - это правило вычисления приближенного значения параметра (или функции от некоторых параметров) по наблюдаемым данным.

При многократном извлечении выборок одного и того же объема и последующем нахождении множества оценок одного и того же параметра получаются различные числовые значении этих оценок, изменяющиеся от одной выборки к другой случайным образом.

Иными словами, любая оценка произвольного параметра есть случайная величина. В этом состоит принципиальное отличие оценки от самого параметра.

Элементарные статистические процедуры

В случае гауссовского распределения для истинного математического ожидания m_x существует его оценка $\tilde{m}_x$, вычисляемая по выборке объема n случайной величины $X = (x_1, x_2, ..., x_n)$:

$$\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} x_i. \qquad (1.23)$$

Для истинной дисперсии D_x (характеристика рассеивания случайной величины око ее математического ожидания) ее оценка $\tilde{D}_x$ при известном математическом ожидании m_x вычисляется так:

$$\tilde{D}_x = \tilde{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_x)^2, \qquad (1.24)$$

где σ_x является среднеквадратическим отклонением.

В случае неизвестного математического ожидания дисперсию $\tilde{D}_x$ нужно вычислять по формуле:

$$\tilde{D}_x = \tilde{\sigma}_x^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right) \qquad (1.25)$$

Приведенные оценки являются несмещенными и асимптотически эффективными.

Не будем забывать о том, что оценки сами являются случайными величинами, а значит, обладают некоторым разбросом, который оценивается дисперсией. Дисперсии D {} вышеуказанных оценок соответственно таковы:

дисперсия оценки среднего:

$$D\{\tilde{m}_x\} = \sigma_x^2 / n; \qquad (1.26)$$

дисперсия оценки $\tilde{D}_x$ в случае известного математического ожидания:

$$D\{\tilde{D}_x\} = 2\sigma_x^4 / n; \qquad (1.27)$$

в случае, если не известно математическое ожидание:

$$D\{\tilde{D}_x\} = 2\sigma_x^4 /(n-1). \qquad (1.28)$$

После вычисления точечных оценок обычно переходят к построению вариационного ряда, диаграммы накопленных частот и гистограммы выборки.

Пусть имеется набор (выборка) экспериментальных данных $x_1, x_2, ..., x_n$. Вариационный ряд (или ряд распределения) $z_1, z_2, ..., z_n$ получают из исходных данных путем расположения x_m $(m = 1, 2, ..., n)$ в порядке возрастания от x_{min} до x_{max} так, чтобы $x_{min} = z_1 \le z_2 \le ... \le z_n = x_{max}$.

Диаграмма накопленных частот $P_n(x)$ является эмпирическим аналогом интегрального закона распределения P(x) и ее строят в соответствии с формулой

$$P_n(x) = \sum_{j=1}^{\mu_n(x)} \frac{1}{n} \qquad (1.29)$$

где $\mu_n(x)$ - число элементов в выборке, для которых значение $x_j < x$.

Практически это делается так. На оси абсцисс указывают значения наблюдений x_{min} (или z_1). Значение по оси ординат равно нулю левее точки x_{min}; в точке x_{min} и далее во всех других точках x_m диаграмма имеет скачок, равный 1/n. Если существует λ совпадающих значений x_m, то в этом месте на диаграмме происходит скачок, равный λ / n. Ясно, что для величин $x > x_{max}$ значение диаграммы накопленных частот равно 1. Отметим, что если $n \to \infty$, то $P_n(x) \to P(x)$.

Пример. Пусть имеется выборка объема 5:

$$x_1 = 5; x_2 = 2; x_3 = 4; x_4 = 5; x_5 = 7.$$

Вариационный ряд для данной выборки будет таким:

$$z_1 = 2; z_2 = 4; z_3 = 5; z_4 = 5; z_5 = 7.$$

Соответствующая диаграмма накопительных частот представлена на рисунке 1.4.

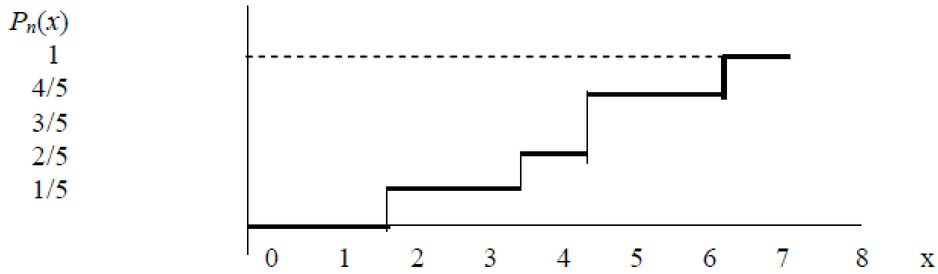


Рисунок 1.4 – Диаграмма накопленных частот

Гистограмма $f_n(x)$ является эмпирическим аналогом функции плотности распределения f(x). Последовательность построения гистограммы такова.

По оценочной формуле находят предварительное количество квантов (интервалов) K, на которое нужно разбить на ось Ox:

$$K=1+3.2\lg n;$$

найденное значение K округляется до ближайшего целого числа.

Формула для K является эмпирической, что означает примерное значение. Ее величина связана с той целью, чтобы в один квант попало хотя бы одно выборочное значение x_i. Интересно, что зависимость количества интервалов K от объема выборки n равна (таблица 1.1)

Таблица 1.1 - Зависимость количества интервалов K от объема выборки n

n	100	200	300	500	1000
K	7	8	9	10	11

Далее определяют длину каждого кванта (интервала):

$$\Delta x = (x_{max} - x_{min}) / K,$$

которую для удобства построений можно несколько округлить в ту или иную сторону.

Середину области изменения выборки (центр распределения)

$$(x_{max} + x_{min}) / 2$$

принимают за центр некоторого интервала, после чего находят границы и окончательное количество указанных интервалов так, чтобы в совокупности они перекрывали всю область от x_{min} и x_{max}.

Далее подсчитывают количество наблюдений n_m, попавшее в каждый квант: n_m равно числу членов вариационного ряда, для которых справедливо неравенство

$$x_m \le z_1 < x_m + \Delta x.$$

Здесь x_m и $x_m + \Delta x$ - границы m-го интервала. Отметим, что при использовании этой формулы значения z_1, попавшие на границу между (m-1) и m-м интервалами, относят к m-му интервалу.

Далее подсчитывают относительное количество (относительную частоту) наблюдений n_m / n, попавших в данных квант.

Наконец, строят гистограмму, представляющую собой ступенчатую кривую, значение которой на m-й интервале $(x_m, x_m + \Delta)$ (m=1, 2, …, K) постоянно и равно n_m / n, или с учетом условия $\int_{-\infty}^{\infty} p_n(z)\,dz = 1$, равно $(n_m / n) \cdot \Delta x$.

Практическая часть

1. Смоделируйте базовую последовательность объемом N=1000 мультипликативным конгруэнтным методом.
2. Напишите одну комплексную программу моделирования выборки случайных чисел, оценки математического ожидания и дисперсии для всех ниже перечисленных распределений:

a) равномерное на интервале (a, b);

b) экспоненциальное с параметром λ;

c) нормальное с параметрами μ и σ, используя метод суммирования или какой-либо один из эвристических методов.

3. Самостоятельно задав параметры распределений, промоделируйте выборки всех вышеуказанных распределений. Объем каждой выборки принять N=1000.
4. Вычислите оценки математического ожидания и дисперсии каждой из полученных в п. 2 последовательностей случайных чисел для следующих объемов выборки N_1=10, N_2=20, N_3=50, N_4=100 и N_5=1000. Сравните полученные оценки с заданными в пп. 2 параметрами. Постройте графики зависимостей оценок от объема выборки. Оцените относительные погрешности для какой-либо одной выборки.
5. Для всех выборок разных распределений, рассчитайте и постройте:

- диаграммы накопленных частот;
- гистограммы распределений.

6. Сравните гистограммы с графиками теоретических распределений. Для сравнения постройте также вышеуказанные гистограммы на одном графике с функциями распределений.

Варианты лабораторной работы. Параметры распределений

№ вар	a	b	λ	μ	σ
1	3	11	1	0	1
2	10	16	5	0	2
3	1	6	1	0	3
4	4	7	2	0	4
5	9	17	4	0	5
6	1	11	1	0	6
7	8	13	4	1	1

8	3	9	1	1	2
9	2	7	1	1	3
10	9	14	5	1	4
11	6	11	3	1	5
12	3	7	2	1	6
13	7	12	3	2	1
14	5	12	3	2	2
15	7	13	4	2	3
16	3	12	1	2	4
17	2	12	2	2	5
18	4	13	2	10	6
19	4	8	5	10	1
20	10	14	4	10	2
21	9	16	1	10	3
22	0	7	3	10	4
23	6	6	5	10	5
24	10	11	3	10	6
25	6	11	4	5	1
26	9	18	2	5	2
27	4	7	1	5	3
28	2	6	2	5	4
29	4	8	5	5	5
30	9	13	4	5	6

Пример построения гистограммы и диаграммы накопленных частот в среде MathCad 5.0 для базовой последовательности R. Оценки математического ожидания и дисперсии приведены ниже на рисунках.

Вычисляем количество квантов (интервалов), на которые разбиваем ось ох

$K := \mathrm{floor}(1 + 3.2 \cdot \log(N))$ $\quad 1 + 3.2 \cdot \log(N) = 10.6$ $\quad K = 10$ $\quad k := 1..K$

границы области построения гистограммы $\quad \max(R) = 0.999$ $\quad \min(R) = 8.945 \cdot 10^{-4}$

ширина одного интервала $\quad \Delta x := \dfrac{\max(R) - \min(R)}{K}$

массив границ интервалов $\quad x_k := k \cdot \Delta x + \min(R)$ $\quad x_0 := \min(R)$

$G := \mathrm{hist}(x, R)$

Функция построения гистограммы. Вычисляет сколько значений массива R попадает в интервалы, определенные массивом x.
Возвращает массив, размерность которого на единицу меньше размерности массива x, поскольку количество интервалов на единицу меньше количества точек, ограничивающих эти интервалы

G =

110
87
86
99
114
85
109
115
99
96

x =

$8.945 \cdot 10^{-4}$
0.101
0.2
0.3
0.4
0.5
0.6
0.699
0.799
0.899
0.999

Рисунок 1.5 – Количество квантов, на которые разбивается ось ОХ

Теоретическое значение закона распределения (плотности вероятности)

$p(x) := \text{if}(x < 0, 0, \text{if}(x > 1, 0, 1))$ $\qquad xx := -0.2, -0.19 .. 1.2$

Теоретическое значение интегральной функции распределения

$F(x) := \text{if}(x < 0, 0, \text{if}(x > 1, 1, x))$

График, на котором построена теоретическая функция распределения, совмещенная с гистограмой. Гистограмма нормирована, относительная частота попадания в интервал делится на ширину интервала, тогда высота столбика может быть соотнесена со значением теоретической функци распределения в пределах каждого интервала.

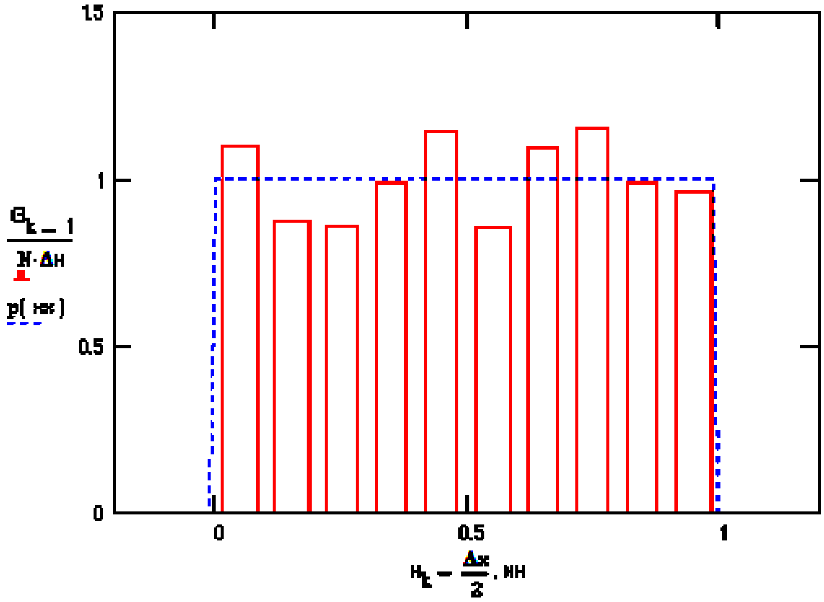


Рисунок 1.6 – Теоретическая функция распределения

$$Gn_k := \sum_{i=1}^{k} G_{i-1}$$ Вычисление массива значений для диаграммы накопленных частот

График, на котором диаграмма накопленных частот совмещена с интегральной функцией распределения

$$G = \begin{pmatrix} 110 \\ 87 \\ 86 \\ 99 \\ 114 \\ 85 \\ 109 \\ 115 \\ 99 \\ 96 \end{pmatrix} \qquad Gn = \begin{pmatrix} 0 \\ 110 \\ 197 \\ 283 \\ 382 \\ 496 \\ 581 \\ 690 \\ 805 \\ 904 \\ 1 \cdot 10^3 \end{pmatrix}$$

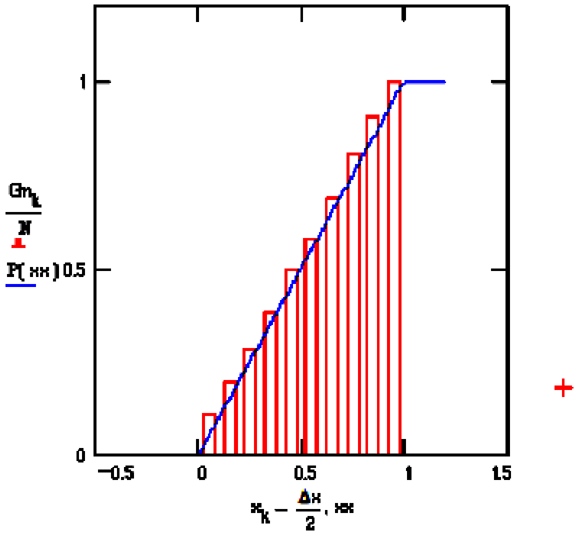


Рисунок 1.7 – Вычисление массива значений для диаграммы накопленных частот

оценка матожидания в зависимости от объема выборки

$$M(N,R) := \frac{1}{N}\cdot\sum_{i=1}^{N} R_i \qquad M(10,R) = 0.528 \qquad M(N,R) = 0.507 \qquad MR := M(N,R)$$

несмещенная оценка дисперсии в зависимости от объема выборки

$$D(N,R) := \frac{1}{N-1}\cdot\sum_{i=1}^{N}\left(R_i - MR\right)^2 \qquad m := 2..N$$

графики данных зависимостей

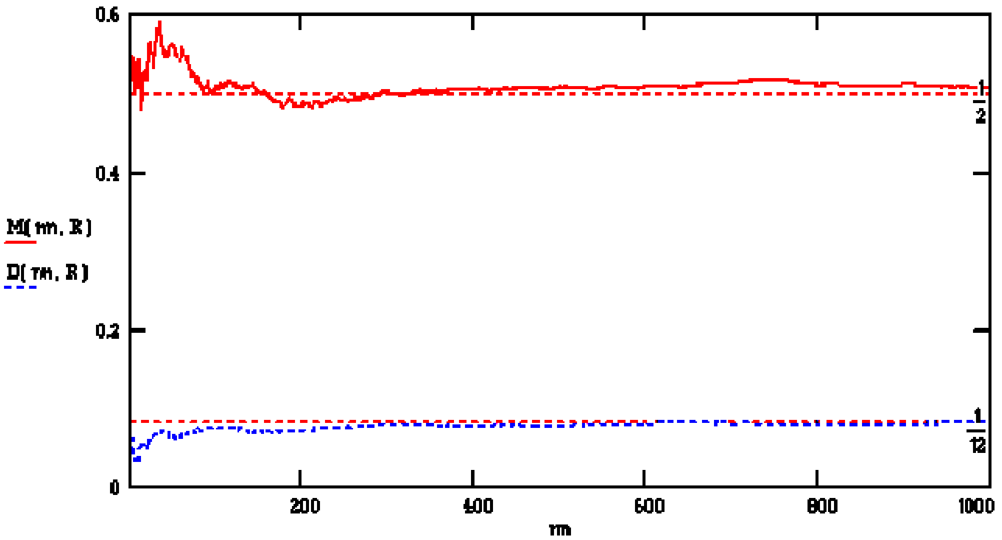


Рисунок 1.8 – Оценка математического ожидания в зависимости о объема выборки

Контрольные вопросы

1. Что такое распределение случайной дискретной величины?
2. Что такое дискретная случайная величина? Дайте развернутый ответ.
3. Каков алгоритм получения случайных величин на ЭВМ?
4. Раскройте понятие «методы генерации псевдослучайных чисел с заданным законом распределения».
5. Что такое метод инверсии?
6. Что значит «моделирование случайной величины равномерной на (a, b)»?
7. Что значит «моделирование экспоненциальной случайной величины»?
8. Что значит «моделирование нормальной случайной величины на основе центральной предельной теоремы»?

9. Что такое оценка статистических характеристик случайных величин?

10. Что такое элементарные статистические процедуры?

Самостоятельная работа №4
«Рассмотрение организма с позиций медицинской системы»

1 Краткие теоретические сведения

Говоря о биологических системах, часто применительно к ним употребляют термин “организм”. Но с позиций системного анализа организмом называют любую систему, обладающую собственными целями и способностью (ресурсами) для их достижения, т.е. целенаправленными действиями. С другой стороны, в настоящее время любой организм рассматривают как биохимическую машину с кибернетическим управлением, функционирование которой осуществляется за счет взаимодействия двух подсистем – метаболической и кибернетической.

Организм образуется из множества различных, качественно неоднородных элементов, каждый из которых играет в процессе жизнедеятельности строго определенную, зачастую незаменимую роль. В общем виде взаимодействие основных элементов организма можно представить в виде схемы (рисунок 10) [1, 9]. В верхней части схемы расположена информационно-кибернетическая часть, в нижней – метаболическая подсистема (МП).

Верхняя часть представляет кибернетическую цепочку, состоящую из трех подсистем: рецепторная подсистема – управляющая подсистема – эффекторная подсистема. Именно эта часть системы

формирует поведение организма, осуществляя восприятие, хранение, переработку и использование информации. Информация из внешней среды поступает в рецепторную подсистему (рецепторы внешней среды), а далее сигналы от рецепторов доставляются в управляющую подсистему, функции которой выполняют рефлексы, а у высших животных – ЦНС. Управляющая подсистема является лидирующей и играет наиболее сложную роль в процессе управления всей системой.

Доставляемые сведения позволяют поддерживать поведение организма в соответствии с условиями окружающей среды, но не менее важной является информация о состоянии самого организма, доставляемая из метаболической подсистемы через рецепторы внутренней среды.

Эта информация корректирует воздействия, вырабатываемые управляющей подсистемой в зависимости от потребностей организма.

Реализация поведенческих актов осуществляется через эффекторную подсистему. Как часть кибернетической системы у высших животных она включает органы движения, органы звуковой сигнализации (у человека – органы речи) и т. п. К ней же относят и все органы, оказывающие различного рода воздействия на внутреннюю сферу организма – железы внутренней секреции (поджелудочная, тимус), железы внешней секреции (например, потовые) и др.

Таблица 1 – Кибернетические функции организма

Функция	Системы, реализующие функции
Восприятие информации Обработка информации, принятие решений, формирование программ поведения Реализация принятых программ	Сенсорные системы организма (зрение, слух, осязание и т.д.); рецепоры внутренней среды (хеморецепторы, барорецептроры и т.п.) Центральная нервная система (ЦНС) Эффекторные системы организма (скелетно-мышечная, нервная, эндо- кринная, репродуктивная, органы речи или внешней сигнализации)

Кибернетическая система, представленная специальной системой органов, существует не у всех живых организмов. У низших животных из-за простоты управленческих функций также нет и постоянных элементов, которые выполняли бы сложные процессы управления. Подобные специальные элементы появляются у более сложных организмов, хотя при этом сохраняются и процессы управления без них. Усложнение кибернетической подсистемы организма идет параллельно с развитием специализированных органов, обеспечивающих все более дифференцированное снабжение метаболической подсистемы.

Нижняя часть схемы иллюстрирует метаболическую подсистему организма, которая представляет собой совокупность функционально и структурно связанных процессов преобразования химических веществ, протекающих в клетках организма, и их транспортировки с целью обеспечения организма веществом и энергией для его жизнедеятельности, роста и размножения. В качестве входных величин для нее будут выступать вещества, доставляемые из окружающей среды. Дальнейшее перемещение и переработка доставленных продуктов и энергии в МП направлены на обеспечение функционирования кибернетической цепочки (и своего собственного).

Метаболическую подсистему организма иногда называют еще метаболическим “котлом”. В ней вещества, поступающие с пищей

(субстраты), – белки, жиры, витамины, микроэлементы − преобразуются в более необходимые для организма, иногда сложные, а иногда и в более простые соединения.

Та часть процессов метаболизма, которая направлена на получение из простых веществ более сложных, называется анаболизмом. Компоненты, получаемые при анаболизме, идут на сборку элементов и ремонт клеточных структур. Процессы сборки можно относить к метаболической подсистеме, считая, что она осуществляет как синтез нужных веществ, так и формирование из них клеточных структур (мембран и других элементов), самих клеток и многоклеточных структур, необходимых для жизнедеятельности организма. Сам процесс сборки направлен не только на реставрацию отмирающих клеток или тканей. В молодом организме возникают новые органы, он растет и развивается. В организме, способном к деторождению, идет сборка половых клеток и синтезируются специальные вещества – секреты, необходимые для нормального оплодотворения. Эти процессы протекают непрерывно в течение довольно длительного времени в жизни живого организма.

Если это вещества, которые нельзя непосредственно использовать для синтеза нужных макромолекул или для сборки новых конструкций, подобные отходы разлагаются на более простые компоненты. Процессы разложения сложных веществ в метаболической подсистеме на более простые получили название катаболизма. Вещества, образующиеся при катаболизме, могут использоваться непосредственно или выступать в качестве строительных блоков, а ненужные вещества приобретают такие свойства (например, растворимость), которые облегчают их окончательное выведение из организма.

Процессы анаболизма и катаболизма контролируются кибернетической подсистемой, поэтому сам факт протекания определенных процессов в МП и их скорость определяются именно

управлением.

Функции организма как открытой метаболической системы, обменивающейся с окружением веществами и энергией таблица 2.

Таблица 2 – Метаболические функции организма

Функция	Вещества	Системы организма
1	2	3
1.Доставка из окружащей среды вещества	Субстраты: белки, липиды, углеводы, витамины, незаменимые аминокислоты Топливо: углеводы Окислитель: кислород Транспорт внутри организма	Физиологический комплекс: жедудочно-кишечный тракт, система пищеварения Те же системы Дыхательная система Сердечнососудистая система
2.Производство энергии	АТФ	Биохимическая система: митохондрии
3. Обмен веществ	Анаболизм: биополимеры Катаболизм: высокомолекулярные отходы, некротические элементы	Биохимическая система: аппарат Гольджи, эндоплазматическая сеть Клеточные мембраны

4.Выведение конечных продуктов из организма	Отходы: азот, креатинин, мочевина, вода, углекислота, билирубин Наследственный материал	Физиологический комплекс: почки, системы дыхания и кровообращения, печень и желчная система Репродуктивная система

Причем для управления поведением и состоянием организма, как объекта управления, используются различные управляющие воздействия.

К основным методам управления, используемым в живых системах, относят энергетический, вещественный и информационный.

Энергетическое управление предполагает воздействие на биологическую систему в целом или на ее подсистемы физических управляющих агентов, не изменяющих количества вещества биологического объекта. К этим агентам относятся физические поля: электрическое, магнитное, тепловое, акустическое, радиационное, электромагнитное. Воздействие может осуществляться как контактно через электроды, так и бесконтактным способом. Суть энергетического управления заключается в стимулировании функционирования отдельных подсистем организма или подавлении некоторых патологических процессов, протекающих в больном организме.

В отличие от энергетического управления, исключающего введение в организм каких-либо материальных управляющих агентов, вещественное управление использует самые различные фармакологические, гормональные, химические и другие агенты в твердом, жидком и газообразном состоянии для управления состоянием живого организма и его отдельными функциональными подсистемами. Таким образом, вещественное управление предусматривает непременное изменение количества вещества, содержащегося в организме.

Информационное управление – это управление состоянием человека с помощью воздействия специально сформированных потоков информации. Этот метод управления является наиболее эффективным, но и наименее количественно формализованным.

2 Задание

Изучить основ системного анализа применительно к живым биологическим организмам

3 Цель работы

Целью занятия является изучение основ системного анализа применительно к живым биологическим организмам.

Выполнение работы

кибернетическая подсистема организма при рассмотрении его с позиций системного анализа;

метаболическая подсистема организма при рассмотрении его с позиций системного анализа;

виды биосистем;

основные виды управления биотехническими системами.

Контрольные вопросы

1. Дайте понятие “организма” с позиции системного анализа.
2. Что такое информационное управление и как оно осуществляется?
3. Что такое энергетическое управление и как оно осуществляется?
4. Как контролируются процессы анаболизма и катаболизма?
5. Опишите метаболическую подсистему организма
6. Опишите кибернетическая подсистема организма
7. Нарисуйте и опишите обобщенную структуру живого организма

Список использованных источников

1. Гольденберг Л.М. и др. Цифровая обработка сигналов: справочник / Л.М. Гольденберг, Б.Д. Матюшкин, М.Н. Поляк. – М.: Радио и связь, 1985. – 312 с.

2. Гутников В.С. Фильтрация измерительных сигналов. – Л.: Энергоатомиздат. Ленингр. отд-ние, 1990. – 192 с.

3. Каппелини В., Константинидис А.Дж., Эмилиани П. Цифровые фильтры и их применение. – М.: Энергоатомиздат, 1983.

Самостоятельная работа №5
«Интервальное оценивание»

1.1 Понятие доверительного интервала

Пусть имеется выборка $X = (X_1, \ldots, X_n)$ из распределения $\boldsymbol{F}_\theta$ с неизвестным параметром $\theta \in \Theta \subseteq R$. Задача интервального оценивания заключается в том, чтобы найти интервал, который накрывает оцениваемый параметр с заданной наперед вероятностью.

Интервал (θ^-, θ^+) называется доверительным интервалом для параметра θ уровня доверия $1-\varepsilon$, если для любого $\theta \in \Theta$

$$P_\theta\left(\theta^- < \theta < \theta^+\right) \geq 1-\varepsilon \qquad (1.1)$$

Если в соотношении (1.1) вероятность в точности равна $1-\varepsilon$ (или стремится к $1-\varepsilon$), то интервал называют точным (или асимптотически точным) доверительным интервалом уровня доверия $1-\varepsilon$.

1.2Построение точных доверительных интервалов для параметров нормального распределения

Пусть $X = (X_1, \ldots, X_n)$ - выборка объема n из нормального распределения N_{a,σ^2}. Рассмотрим возможные задачи интервального оценивания:

1. Пусть известно σ^2, а $a \in R$- неизвестный параметр. Требуется построить точный доверительный интервал для параметра ***a*.**

$$P_{a,\sigma^2}\left(\overline{X} - \frac{\tau_{1-\varepsilon/2} \cdot \sigma}{\sqrt{n}} < a < \overline{X} + \frac{\tau_{1-\varepsilon/2} \cdot \sigma}{\sqrt{n}}\right) = 1-\varepsilon, \qquad (1.2)$$

где $\tau_{1-\varepsilon/2}$ - квантиль уровня $1-\varepsilon/2$ стандартного нормального распределения (см. Приложение Б, таблица 1).

2. Пусть известен параметр a, требуется построить доверительный интервал для σ^2.

$$P_{a,\sigma^2}\left(\frac{n\cdot s_1^2}{g_2}<\sigma^2<\frac{n\cdot s_1^2}{g_1}\right)=1-\varepsilon \quad , \qquad (1.3)$$

где s_1^2 - выборочная дисперсия, g_1 и g_2 - квантили распределения "хи- квадрат" с n степенями свободы ($\chi^2_{\alpha,n}$) уровня $a=\varepsilon/2$ и $a=1-\varepsilon/2$ соответственно (см. Приложение Б, таблица 3).

3. Доверительный интервал для σ^2 при неизвестном a:

$$P_{a,\sigma^2}\left(\frac{(n-1)\cdot s_0^2}{g_2}<\sigma^2<\frac{(n-1)\cdot s_0^2}{g_1}\right)=1-\varepsilon \quad , \qquad (1.4)$$

где s_0^2 - несмещенная выборочная дисперсия.

4. Доверительный интервал для а при неизвестном σ^2:

$$P_{a,\sigma^2}\left(\overline{X}-\frac{t_{1-\varepsilon/2,n-1}\cdot s_0}{\sqrt{n}}<a<\overline{X}+\frac{t_{1-\varepsilon/2,n-1}\cdot s_0}{\sqrt{n}}\right)=1-\varepsilon \quad ,(1.5)$$

где $t_{1-\varepsilon/2,n-1}$- квантиль распределения Стьюдента с $n-1$ степенью свободы уровня $1-\varepsilon/2$ (см. Приложение Б, таблица 2).

Приведем пример построения доверительного интервала.

Пусть имеется выборка объема N из нормального распределения, для которого известен параметр $\sigma^2=4$, а параметр а неизвестен. Требуется построить точный доверительный интервал для параметра а. На рисунке 1.1 приведен текст программы, реализующей метод построения доверительного интервала в среде Mathcad.

Ввод исходных данных

x - вектор, представляющий выборку из нормального распределения
dx - известное среднеквадратическое отклонение
q=1-e - уровень доверия

x :=

C:\..\v1.txt — вектор, представляющий выборку считывается из файла

$x^T =$

	0	1	2	3	4	5	6	7	8	9
0	0.432	-0.501	2.332	0.494	3.503	1.233	-1.313	1.969	2.207	1.188

вывод на экран элементов выборки (их можно просмотреть, щелкнув по вектору x и воспользовавшись линейкой прокрутки)

$dx := 2$ $\quad q := 0.9$ — ввод значений среднеквадратического отклонения и уровня доверия

$N := \mathrm{length}(x)$ $\quad N = 700$ — нахождение обьема выборки

$Mx := \dfrac{\sum_{i=0}^{N-1} x_i}{N}$ $\quad Mx = 1.97$ — вычисление среднего выборочного

$p := 1 - \dfrac{1-q}{2}$

$v := \mathrm{qnorm}(p,0,1)$ $\quad v = 1.645$ — нахождение квантиля стандартного нормального распределения уровня 1-e/2

$upper := Mx + \dfrac{v \cdot dx}{\sqrt{N}}$ $\quad upper = 2.094$

нахождение границ доверительного интервала

$lower := Mx - \dfrac{v \cdot dx}{\sqrt{N}}$ $\quad lower = 1.846$

$length := upper - lower$ $\quad length = 0.249$ — вычисление длины доверительного интервала

Рисунок 1.1. Построение доверительного интервала для параметра а нормального распределения при известном σ

1.3 Задание к самостоятельной работе

Даны две выборки одной случайной величины с нормальным распределением N_{a,σ^2} объема n_1 и n_2 соответственно.

Для вариантов с нечетным номером:

1. Для обеих выборок построить точный доверительный интервал уровня доверия q_0 для параметра a, считая:

а) σ неизвестным,

б) σ известным и равным σ_0**.**

2. В одной системе координат построить графики зависимости длины доверительного интервала от уровня доверия q для всех четырех случаев (объем выборки равен n_1, σ неизвестно; объем выборки равен n_1, σ известно; объем выборки равен n_2, σ неизвестно; объем выборки равен n_2, σ известно). При этом q придать минимум 50 разных значений через равные промежутки.

Проанализировать взаимное расположение полученных графиков и объяснить его.

Для вариантов с четным номером:

1. Для обеих выборок построить точный доверительный интервал

уровня доверия q_0 для параметра σ^2, считая:

а) a неизвестным,

б) a известным и равным a_0.

2. В одной системе координат построить графики зависимости длины доверительного интервала от уровня доверия q для всех четырех случаев (объем выборки равен n_1, a неизвестно; объем выборки равен n_1, a известно; объем выборки равен n_2, a неизвестно; объем выборки равен n_2, a известно). При этом q придать минимум 50 разных значений через равные промежутки.

Проанализировать взаимное расположение полученных графиков и объяснить его.

Указания: Выборки необходимо считать с двух текстовых файлов - "di-V.txt "di-V-1 .txt"(V - номер вашего варианта).

Для написания программы можно пользоваться следующими встроенными функциями: функцией length(x) для определения объема выборки, функциями для нахождения значений квантилей распределений
qnorm(p,a,a) (возвращает значение квантиля нормального распределения N, a, σ^2 уровня p), qchisq(p,d) (возвращает значение квантиля распределения "хи-квадрат" с d степенями свободы уровня p), ***qt(p,d)*** (возвращает значение квантиля распределения Стьюдента с d степенями свободы уровня p). Остальные статистические функции должны быть запрограммированы.

Варианты заданий

1. $\sigma_0 = 2; q_0 = 0,9$
2. $\sigma_0 = 0; q_0 = 0,8$
3. $\sigma_0 = 3; q_0 = 0,7$
4. $\sigma_0 = 2; q_0 = 0,5$
5. $\sigma_0 = 1; q_0 = 0,6$
6. $\sigma_0 = 3; q_0 = 0,9$
7. $\sigma_0 = 0,5; q_0 = 0,8$
8. $\sigma_0 = -1; q_0 = 0,8$
9. $\sigma_0 = 1,5; q_0 = 0,7$
10. $\sigma_0 = 0,5; q_0 = 0,8$
11. $\sigma_0 = 1; q_0 = 0,5$
12. $\sigma_0 = -5; q_0 = 0,6$
13. $\sigma_0 = 1,2; q_0 = 0,7$
14. $\sigma_0 = 4; q_0 = 0,8$
15. $\sigma_0 = 2,5; q_0 = 0,75$
16. $\sigma_0 = 10; q_0 = 0,6$
17. $\sigma_0 = 3,2; q_0 = 0,9$
18. $\sigma_0 = 0; q_0 = 0,75$
19. $\sigma_0 = 3; q_0 = 0,75$
20. $\sigma_0 = 3; q_0 = 0,5$

Самостоятельная работа №6

«Дисперсионный анализ»

Дисперсионный анализ - это статистический метод, предназначенный для выявления влияния отдельных факторов на результат эксперимента. Суть метода заключается в том, что общая вариация результирующего показателя расчленяется на части, соответствующие совместному и раздельному влиянию различных качественных факторов, и остаточную вариацию, аккумулирующую влияние неучтенных факторов. Статистическое изучение этих частей позволяет делать выводы о том, действительно ли тот или иной качественный фактор оказывает влияние на результирующий показатель.

Дисперсионный анализ основан на следующих допущениях:

1) наблюдения результирующего фактора ξ - это нормально распределенная случайная величина с центром распределения $M\xi = \phi(b_1, \ldots, b_m)$, где $b_1, \ldots, b_m$ - это m независимых управляющих качественных факторов;

2) дисперсия единичного наблюдения, обусловленная случайными ошибками, постоянна во всех опытах и не зависит от $b_1, \ldots, b_m$.

По числу факторов, влияние которых исследуется, различают однофакторный и многофакторный дисперсионный анализ.

1 Однофакторный дисперсионный анализ

Как следует из названия, данным методом исследуется влияние на результирующий признак одного качественного показателя.

Пусть в результате эксперимента получено r групп выборочных значений результирующего признака $X_{ij}\left(j = 1, \ldots, n_i, i = 1, \ldots, r\right)$, соответствующих r значениям качественного фактора; n_i - это количество наблюдений для i -го значения качественного фактора $\left(\sum_{i=1}^{r} n_i = n\right)$.

Пусть $a_i \ (i = 1, \ldots, r)$ - групповые средние, а $a = \frac{1}{r}\sum_{i=1}^{r} a_i$ - общее (генеральное) среднее.

Будем проверять гипотезу $H_0 = \{a_1 = \ldots = a_r = a\}$ о том, что качественный фактор не влияет на результирующий признак против альтернативной гипотезы $H_1 = \{H_0$ неверна$\}$.

Определим общее и групповые выборочные средние (соответственно $\overline{X}$ и $\overline{X}_i$):

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ji}, \overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ji}.$$

Как известно, выборочные групповые средние являются несмещенными и состоятельными оценками средних a_i.

Представим полную сумму квадратов отклонений результирующего признака от общего среднего в виду двух сумм квадратов отклонений:

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ji} - \overline{X}\right)^2 = \sum_{i=1}^{r} n_i\left(\overline{X}_i - \overline{X}\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ji} - \overline{X}_i\right)^2 = Q_1 + Q_2$$

Сумма Q_1 представляет собой сумму квадратов отклонений групповых средних значений от общего среднего значения ("сумма квадратов между группами"), т.е. вариацию, обусловленную качественным фактором, а сумма Q_2 является суммой квадратов отклонения каждой величины от соответствующего группового среднего значения ("сумма квадратов внутри групп"), т.е. остаточную вариацию, обусловленную случайными отклонениями от групповых средних.

Теорема 7.1 В случае справедливости гипотезы H_0 величина

$$F = \frac{Q_1/(r-1)}{Q_2/(n-r)}$$

имеет распределение Фишера с $r-1, n-r$ степенями свободы.

Отсюда, для проверки гипотезы H_0 при уровне значимости q получаем следующий критерий:

$$\rho(X) = \begin{cases} H_0, & \text{если} \quad F \le F_{1-q,r-1,n-r,} \\ H_1, & \text{в} \cdot \text{противном} \cdot \text{случае} \end{cases}$$

На практике для вычисления сумм Q_1, Q_2, Q бывает удобнее пользоваться формулами

$$Q_1 = \sum_{i=1}^{r} \frac{\left(\sum_{j=1}^{n_i} X_{ji}\right)^2}{n_i} - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ji}\right)^2}{n},$$

$$Q_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ji}^2 - \sum_{i=1}^{r} \frac{\left(\sum_{j=1}^{n_i} X_{ji}\right)^2}{n_i}, \quad Q = \sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ji}^2 - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_i} X_{ji}\right)^2}{n}$$

Приведем пример. Предположим, на экспертную оценку отправлено 15 видов товара. Каждого вида товара оправлялось по 20 образцов. Оценив каждый образец, эксперт должен был дать среднюю оценку каждому виду товара. Экспертиза проводилась двумя экспертами. Необходимо выяснить, насколько субъективной была эта экспертиза. Экспертные оценки приведены в следующей таблице:

Средние оценки экспертов по каждому виду товара	Виды товара														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

1-й эксперт	4	3,2	4,6	3,8	3,4	3,2	3,5	4,6	3,7	4,1	5	3,3	4	5	4,9
2-й эксперт	3,9	3,9	4,1	4,3	1,9	3,2	2,3	5	4,9	2,7	4,1	5	3,6	4	2,5

На рисунке 7.1 приведен текст программы в среде Mathcad, проверяющей гипотезу о том, что личность эксперта не влияет на оценку товаров. По результатам статистического анализа эта гипотеза была принята.

$n1 := 15 \qquad n2 := 15 \qquad n := n1 + n2$

Вводим количество наблюдений для каждого из значений качественного фактора и вычисляем общее число наблюдений

$x := (4\ \ 3.2\ \ 4.6\ \ 3.8\ \ 3.4\ \ 3.2\ \ 3.5\ \ 4.6\ \ 3.7\ \ 4.1\ \ 5\ \ 3.3\ \ 4\ \ 5\ \ 4.9)^T$

$y := (3.9\ \ 3.9\ \ 4.1\ \ 4.3\ \ 1.9\ \ 3.2\ \ 2.3\ \ 5\ \ 4.9\ \ 2.7\ \ 4.1\ \ 5\ \ 3.6\ \ 4\ \ 2.5)^T$

Вводим значения результирующего признака при каждом значении качественного фактора

$$Q1 := \frac{\left(\sum_{i=0}^{n1-1} x_i\right)^2}{n1} + \frac{\left(\sum_{i=0}^{n2-1} y_i\right)^2}{n2} - \frac{\left[\sum_{i=0}^{n1-1} x_i + \left(\sum_{i=0}^{n2-1} y_i\right)\right]^2}{n} \qquad Q1 = 0.8$$

Вычисляем "сумму квадратов между группами" - Q1, "сумму квадратов внутри групп" - Q2 и "общую сумму квадратов" - Q

$$Q2 := \sum_{i=0}^{n1-1} (x_i)^2 + \sum_{i=0}^{n2-1} (y_i)^2 - \left[\frac{\left(\sum_{i=0}^{n1-1} x_i\right)^2}{n1} + \frac{\left(\sum_{i=0}^{n2-1} y_i\right)^2}{n2}\right] \qquad Q2 = 19.613$$

$$Q := \sum_{i=0}^{n1-1} (x_i)^2 + \sum_{i=0}^{n2-1} (y_i)^2 - \frac{\left(\sum_{i=0}^{n1-1} x_i + \sum_{i=0}^{n2-1} y_i\right)^2}{n} \qquad Q = 20.414 \qquad Q1 + Q2 = Q = 1$$

$R1 := Q1 \qquad R2 := \dfrac{Q2}{n-3} \qquad R := \dfrac{Q}{n-1}$

Находим среднее квадратов - R1, R2, R3

$R1 = 0.8 \qquad R2 = 0.726 \qquad R = 0.704$

$F := \dfrac{R1}{R2} \qquad F = 1.102$

Вычисляем дисперсионное отношение F

$F\_кр := qF(0.95, 1, n-2) \qquad F\_кр = 4.196$

Т. к. F<F_кр, то делаем вывод о том, личность эксперта существенно не влияет на оценку товаров

Рисунок 7.1. Проверка гипотезы об отсутствии влияния одного качественного фактора на результирующий показатель

2 Двухфакторный дисперсионный анализ

В данном случае исследуется наличие или отсутствие влияния на результирующий признак двух качественных показателей.

Пусть рассматривается два фактора - A и B. Фактор A может принимать r значений $(A = \{A_1, \ldots, A_r\})$ а фактор B - s значений $(B = \{B_1, \ldots, B_s\})$.

В результате эксперимента получены выборочные значения результирующего признака $Xj_{ik}, \ j = 1, \ldots, n_{ik}, \ i = 1, \ldots, r, \ k = 1, \ldots, s$; n_{ik} - это количество наблюдений при i-м значения качественного фактора A и k-м значении качественного фактора $B\left(\sum_{i=1}^{r}\sum_{k=1}^{s} n_{ik} = n\right)$.

По указанной выборке будем проверять справедливость следующих гипотез:

H_A - о том, что качественный фактор A не влияет на результирующий признак,

H_B - о том, что фактор B не влияет на результирующий признак,

H_{AB} - о том, что взаимодействие факторов A и B не влияет на результирующий признак.

Для этого вводим общее и групповое выборочные средние:

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{n_{il}} X_{jik}, \overline{X}_{ik} = \frac{1}{n_{ik}}\sum_{j=1}^{n_{ik}} X_{jik}.$$

Вычисляем значения Q, Q_1, Q_2, Q_3, Q_4 по формулам:

$$Q = \sum_{k=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{n_{ik}}\left(X_{jik} - \overline{X}\right)^2, Q_1 = \sum_{k=1}^{s}\sum_{i=1}^{r} n_{ik}\left(\frac{1}{s}\sum_{k=1}^{s}\overline{X}_{ik} - \overline{X}\right)^2,$$

$$Q_2 = \sum_{k=1}^{s}\sum_{i=1}^{r} n_{ik}\left(\frac{1}{r}\sum_{i=1}^{r}\overline{X}_{ik} - \overline{X}\right)^2, \qquad (7.1)$$

$$Q_3 = \sum_{k=1}^{s}\sum_{i=1}^{r} n_{ik}\left(\overline{X}_{ik} - \frac{1}{s}\sum_{k=1}^{s}\overline{X}_{ik} - \frac{1}{r}\sum_{i=1}^{r}\overline{X}_{ik} + \overline{X}\right)^2,$$

$$Q_4 = \sum_{k=1}^{s}\sum_{i=1}^{r}\sum_{j=1}^{n_{ik}}\left(X_{jik} - \overline{X}_{ik}\right)^2 .$$

Проверку гипотез проводим по следующему критерию:

а) Если $\frac{Q_1/(r-1)}{Q_4(n-rs)} \ge F_{1-q,r-1,n-rs}$, то гипотеза H_A отвергается;

б) Если $\frac{Q_2/(s-1)}{Q_4(n-rs)} \ge F_{1-q,s-1,n-rs}$,то гипотеза H_B отвергается;

в) Если $\frac{Q_3/((r-1)(s-1))}{Q_4/(n-rs)} \ge F_{1-q,(r-1)(s-1),n-rs}$, то гипотеза H_{AB} отвергается (см. Приложение Б, таблица 4).

Задание к самостоятельной работе

Исследовать влияние на результирующий показатель: а) одного качественного фактора, б) двух качественных факторов. Уровень значимости для проверки гипотез взять равным $q = 0.05$.

Указание: Под буквой б) данные необходимо считывать из текстового файла. Данные в файле располагаются в виде матрицы следующим образом: в 1-м столбце данные соответствуют значениям факторов (A_1, B_1), во 2-м столбце - значениям (A_2, B_1), в 3-м - (A_1, B_2), в 4-м - (A_2, B_2).

Соответственно, чтобы можно было воспользоваться формулами (7.1), необходимо правильно считать данные из файла, как это сделано, например, в программе, текст которой приведен на рисунке 7.2.

Рисунок 7.2. Пример считывания данных из файла

Если вы считали данные подобным образом, то в дальнейшем, чтобы обратиться к элементу Xj_{ik}, пользуйтесь записью $(X_{i,k})_j$.

Варианты заданий

1. а) Исследовать влияние посещения секций и кружков во внеклассное время на успеваемость школьников. Качественный фактор - количество часов, проводимых школьниками на дополнительных занятиях. Результирующий признак - средние баллы учеников по совокупности предметов за год. Согласно значениям качественного фактора ученики были поделены на 3 группы (по 16, 12 и 10 человек соответственно).
Данные о средних баллах школьников приведены в таблице:

б) Исследовать влияние на скорость прорастания семян томатов следующих факторов: температуры воздуха (фактор А) и влажности воздуха (фактор В). Значения фактора $A : A = A_1$ - температура воздуха ниже 22^o, $A = A_2$ - температура воздуха выше 22°. Значения фактора $B : B = B_1$ - влажность воздуха ниже 90 процентов, $B = B_2$ - влажность воздуха выше 90 процентов. Время прорастания семян (в часах) при различных значениях факторов приведено в файле disp-1.txt.

Количество часов T, проводимое на дополнительных занятиях	Средний балл за год															
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}	X_{16i}
T=0	3,74	3,35	4,48	5,00	3,93	3,88	3,40	4,10	4,21	3,81	4,64	4,58	3,99	3,72	2,97	3,70
0<T≤3	4,36	4,06	3,52	4,44	3,22	4,27	4,42	3,89	3,68	3,95	4,17	4,29				
T>3	2,55	4,03	3,48	3,46	3,74	3,35	3,87	3,26	4,65	2,34						

2. а) Исследовать влияние поведения цены за барель нефти на курс акций некоторого предприятия. Результирующий признак - цена акции предприятия. Согласно значениям качественного фактора (поведения цены на нефть) данные были поделены на 3 группы (по 15, 14 и 13 значений соответственно). Данные о курсе акций предприятия приведены в следующей таблице:

Поведение цены на нефть	Курс акции														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
цена растет	106,12	97,22	99,55	100,07	99,13	98,98	98,13	102,12	99,55	105,00	100,18	101,60	96,35	99,35	100,60
цена стабильна	97,96	102,12	102,07	103,75	98,46	100,92	98,84	99,50	103,10	103,64	96,82	100,98	95,11	98,55	
цена падает	100,58	97,99	101,65	101,45	100,41	99,83	99,04	100,81	96,69	100,46	100,73	101,11	101,22		

б) Исследовать влияние на успеваемость студентов по физике следующих факторов: посещаемости лекционных занятий (фактор А) и активности при работе на практических занятиях (фактор В).Значения фактора $A : A = A_1$ - посещаемость ниже 60 процентов, $A = A_2$ - посещаемость ниже 60 процентов. Значения фактора $B : B = B_1$- низкая активность, $B = B_2$- высокая активность. Средние баллы студентов по физике за 4 семестра при различных значениях факторов приведены в файле disp-2.txt.

3. а) Исследовать влияние на урожайность огурцов уровня влажности воздуха в теплице. Результирующий признак - количество килограммов огурцов, собранных с одного куста за сезон. Согласно значениям качественного фактора данные были поделены на 3 группы (по 16, 14 и 15 значений соответственно). Данные об урожайности огурцов приведены в таблице

Влажность воздуха	Урожай с одного куста огурцов															
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}	X_{16i}
низкая	19,84	18,64	19,90	22,47	20,72	21,45	24,50	21,95	17,92	22,05	20,75	20,52	20,04	19,12	20,27	23,99
средняя	18,20	19,96	23,91	18,51	23,05	17,24	20,69	19,34	17,64	17,81	20,57	24,69	16,08	17,03		
высокая	20,05	20,74	21,82	20,39	18,33	22,59	18,53	21,82	19,74	23,85	20,90	18,71	17,68	18,22	21,25	

б) Исследовать влияние на размер выпускаемой на заводе детали следующих факторов: станка, на котором производится деталь (фактор А) и рабочего, изготавливающего деталь (фактор В). Значения фактора $A : A = A_1$ - 1-й станок, $A = A_2$ - 2-й станок. Значения фактора $B : B = B_1$ - 1-й рабочий, $B = B_2$ - 2-й рабочий. Размеры получаемых деталей (в миллиметрах) при различных значениях факторов приведены в файле disp-3.txt.

Сезон	Количество потребленной энергии														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
зима	255,95	295,30	324,45	292,69	278,12	291,27	290,55	305,73	330,64	269,81,	329,45	283,01	298,33	314,13	327,66
лето	319,14	274,65	325,24	288,73	309,48	326,89	320,48	314,63	293,57	314,86	275,53	329,28	287,35	304,11	
межсезонье	301,13	331,65	330,68	318,07	335,80	338,58	332,63	307,37	343,26	308,51	352,80	315,01	334,53	280,54	364,32

4. а) Исследовать влияние фактора сезонности на среднее за сезон количество потребляемой энергии семьей из 3-х человек. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 14 и 15 значений соответственно). Данные о количестве потребленной энергии (в кВт) приведены в таблице

б) Исследовать влияние на всхожесть семян следующих факторов: свежести семян (фактор А) и освещенности теплицы (фактор В). Значения фактора $A : A = A_1$ - свежие семена (прошлого урожая), $A = A_2$ - не свежие семена. Значения фактора $B : B = B_1$ - хорошая освещенность, $B = B_2$ - плохая освещенность. Доли взошедших семян из каждой пачки приведены в файле disp-4.txt.

4 а) Исследовать влияние уровня кислотности почвы на урожайность свеклы. Согласно значениям качественного фактора данные были поделены на 3 группы (по 13, 14 и 15 значений соответственно). Данные о среднем урожае свеклы с 1 Га земли для каждого хозяйства приведены в таблице

ивы	Средняя урожайность свеклы в 1 Га (в Цт.)												
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}
я	546,1	594,9	570,2	605,7	565,6	601,7	592,7	598,4	576,2	600,9	579,9	603,8	572,3

	0	5	0	9	6	7	7	3	4	0	3	1	9
ная	592,60	615,74	589,76	583,63	584,38	619,75	581,59	577,47	611,24	586,72	609,74	611,95	609,35
ьная	629,24	621,44	611,20	612,20	617,45	630,55	650,25	616,24	625,16	625,03	649,04	641,62	573,71

б) Исследовать влияние на заболеваемость гриппом и ОРЗ следующих факторов: количества времени, выделяемых человеком на сон (фактор А) и регулярность занятий спортом (фактор В). Значения фактора $A : A = A_1$ - сон менее 8 часов в сутки, $A = A_2$ - сон более 8 часов. Значения фактора $B : B = B_1$ - занятие спортом не реже 1 раза в неделю, $B = B_2$ - реже 1 раза в неделю. Исследования проводились в нескольких городах. Данные о том, сколько раз в среднем болеет за год житель каждого города приведены в файле disp-5.txt.

5 а) Исследовать влияние времени суток на количество вызовов скорой помощи. Анализ проводился в 15 городах с примерно одинаковой численностью населения в течение года. Данные о среднем количестве вызовов бригад скорой помощи в каждом городе приведены в таблице

Время суток	Среднее число вызовов скорой помощи за указанный период														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
8.00 - 16.00	62,98	89,72	86,03	95,21	91,80	88,17	73,01	81,72	85,98	81,28	91,88	91,29	70,46	49,99	85,75
16.00- 24.00	94,45	95,84	91,08	82,14	97,79	100,08	96,74	103,86	105,55	112,58	81,53	125,17	110,20	99,24	95,50
0.00 - 8.00	132,64	133,18	132,43	141,26	135,21	127,13	133,41	134,68	146,26	133,27	133,86	145,24	132,56	155,29	137,19

б) Исследовать влияние на количество крупных ДТП (с участием более 2-х машин или с наличием пострадавших) следующий факторов: плотности транспортного потока (фактор А) и наличие гололеда (фактор В). Значения фактора $A: A = A_1$ - плотный транспортный поток, $A = A_2$ - свободное движение. Значения фактора $B: B = B_1$ - наличие гололеда, $B = B_2$ - отсутствие гололеда. Исследования проводились в нескольких городах. Данные о среднем количестве крупных ДТП в час приведены в файле disp-6.txt.

6 а) Исследовать влияние прослушиваемой водителем музыки на скорость движения автомобиля по незагруженной транспортом трассе. Согласно значениям качественного фактора данные были поделены на 3 группы (по 10, 14 и 15 значений соответственно). Данные о среднем скорости водителей приведены в таблице

зыки	Средняя скорость автомобиля на трассе													
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X
еская	74,75	82,83	88,62	82,04	84,69	93,39	90,19	92,05	76,44	76,95				
ирная	100,69	92,44	106,97	98,18	93,56	92,77	94,83	105,58	97,55	100,90	112,62	92,94	100,02	1
к	104,81	105,32	107,65	109,92	107,38	98,16	88,85	119,11	114,72	93,44	106,44	121,97	101,80	8

б) Исследовать влияние на посещаемость человеком кинотеатров следующих факторов: возраста (фактор А) и активности человека как читателя художественной литературы (фактор В). Значения фактора $A: A = A_1$ - возраст от 18 до 30 лет, $A = A_2$ - от 31 года. Значения фактора $B: B = B_1$ - количество прочитанных за год книг 0-3, $B = B_2$ - более 3-х. Опрос проводился в нескольких кинотеатрах. Данные о том, сколько раз в год в среднем человек посещает кинотеатры, приведены в файле disp-7.txt.

7 а) Исследовать влияние уровня дохода семьи из 4-х человек на количество потребляемого за неделю хлеба. Согласно значениям качественного фактора данные были поделены на 3 группы (по 10, 15 и 11 значений соответственно). Данные о среднем потреблении хлеба (в кг) за неделю приведены в таблице

Уровень	Среднее количество потребляемого за неделю хлеба (в кг.)													
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_1	X_1	X_1	X_1	X_1	X_1

дохода семьи	i	i	i	i	i	i	i	i	i	0i	1i	2i	3i	4i	5i
низкий	2,79	3,44	2,72	2,55	2,57	3,55	2,49	2,37	3,31	2,63					
средний	3,06	3,02	2,81	2,67	3,05	2,98	2,43	3,30	3,26	3,42	3,31	3,21	2,78	2,86	3,24
высокий	3,30	3,08	2,79	2,82	2,97	3,33	3,88	2,93	3,18	3,18	3,75				

б) Исследовать влияние на объем входящего интернет-трафика следующих факторов: отношения пользователя к компьютерным играм (фактор А) и наличия в домашней сети пользователя бесплатных ресурсов (фактор В). Значения фактора $A : A = A_1$ - пользователь увлекается компьютерными играми, $A = A_2$ - не увлекается. Значения фактора $B : B = B_1$ - большой объем бесплатных ресурсов в домашней сети, $B = B_2$ - бесплатных ресурсов не имеется, либо их мало. Данные о среднем входящем интернет-трафике (в Мб.) приведены в файле disp-8.txt.

8 а) Исследовать влияние возраста покупателя на количество покупаемых в магазине глазированных сырков. Согласно значениям качественного фактора данные были поделены на 3 группы (по 12, 15 и 11 значений соответственно). Данные о среднем количестве покупаемых сырков за одно посещение магазина приведены в таблице

Возраст покупателя	Среднее количество покупаемых за 1 раз глазированных сырков														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
до 25	1,26	1,37	0,99	1,14	1,08	1,21	0,76	0,65	2,52	0,86	0,36	1,38			
25-45	3,40	3,07	2,50	2,54	2,56	3,14	2,80	3,65	2,31	3,19	3,22	3,88	2,68	3,28	3,77
старше 45	0,59	2,15	0,98	1,73	0,97	0,88	1,71	0,54	1,44	1,71	2,23				

б) Исследовать влияние на склонность к сердечно-сосудистым заболеваниям у людей старше 50 лет следующих факторов: пола пациента (фактор А) и курения (фактор В). Значения фактора $A: A = A_1$ - пациент - женщина, $A = A_2$ - пациент - мужчина. Значения фактора $B: B = B_1$ - пациент курит, $B = B_2$ - пациент не курит. Данные о среднем количестве инфарктов на 100 человек старше 50 лет приведены в файле disp-9.txt.

9 а) Исследовать влияние климата на количество потребляемых мясных продуктов. Анализ проводился в нескольких регионах. Согласно значениям качественного фактора данные были поделены на 3 группы (по 13, 15 и 9 значений соответственно). Данные о среднем количестве потребляемых в неделю мясных продуктов (в кг.) жителями каждого региона приведены в таблице

Климатические условия	Среднее количество мяса, употребляемого жителем региона за неделю (в кг.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
северные регионы	0,46	0,52	0,86	0,91	0,54	0,64	0,67	0,74	0,69	0,61	0,65	0,53	0,76		
средняя полоса	0,41	0,33	0,41	0,28	0,39	0,36	0,40	0,67	0,39	0,33	0,56	0,61	0,18	0,59	0,23
южные регионы	0,09	0,22	0,27	0,40	0,27	0,06	0,30	0,36	0,42						

б) Исследовать влияние на спрос на фотоуслуги следующих факторов: дня недели(фактор А) и времени года (фактор В). Значения фактора $A: A = A_1$ - рабочий день, $A = A_2$ - выходной. Значения фактора В : $B = B_1$ - теплое время года, $B = B_2$ - холодное время года. Данные о среднем количестве отпечатываемых за день фотографий по нескольким фотолабораториям приведены в файле disp-10.txt.

10 а) Исследовать влияние возраста на продолжительность разговоров по мобильному телефону. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 15 и 14 значений соответственно). Данные о средней продолжительности телефонных разговоров в день (в минутах) приведены в таблице

Возраст	Средняя продолжительность телефонных разговоров в день (в мин.)

абонента	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
16-25 лет	4,91	4,64	4,67	5,05	5,10	4,82	5,33	5,12	5,26	5,10	4,75	5,20	4,77	4,99	4,94
25-45 лет	3,78	4,14	4,02	4,10	4,21	3,99	4,11	4,30	3,67	3,90	4,04	3,95	4,04	4,13	4,05
45-60 лет	1,17	0,81	0,94	1,23	1,00	1,04	0,91	1,23	1,33	0,65	1,38	1,13	1,51	1,03	

б) Исследовать влияние на количество потребляемой человеком минеральной воды следующих факторов: образа жизни (фактор А) и пола (фактор В). Значения фактора $A: A = A_1$ - занимается спортом, $A = A_2$ - не занимается спортом. Значения фактора $B: B = B_1$ - мужчина, $B = B_2$ - женщина. Данные о среднем количестве выпиваемой за неделю минеральной воды (в литрах) приведены в файле disp-11.txt.

11 а) Исследовать объективность судей, оценивавших спортсменов на некотором соревновании. Перед судьями выступало 15 спортсменов, совершивших по 10 подходов. Данные о средних баллах спортсменов (по 6-ти бальной шкале) приведены в таблице

эксперт	Средние баллы спортсменов														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1-й	3,9	4,5	4,0	4,7	3,8	4,4	4,7	4,0	3,8	4,3	4,5	4,5	4,5	4,1	4,0
2-й	4,8	3,9	4,4	4,7	4,5	3,9	4,1	4,7	4,8	4,6	4,2	3,8	4,7	4,8	3,6
3-й	5,2	3,9	5,1	4,1	4,7	4,0	4,6	4,9	4,5	4,5	5,1	4,9	4,4	4,4	4,9

б) Исследовать влияние на заболеваемость ОРЗ следующих факторов: как человек провел летний отпуск (фактор А) и принимает ли он витаминные комплексы (фактор В). Значения фактора $A: A = A_1$ - совершил длительный выезд на природу/отдыхал на курорте, $A = A_2$ - провел отпуск дома. Значения фактора $B: B = B_1$ - принимает витаминные комплексы, $B = B_2$ - не принимает. Данные о среднем количестве перенесенных за год ОРЗ приведены в файле disp-12.txt.

12 а) Исследовать влияние "возраста" йогурта на содержание в нем молочнокислых бактерий. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 13 и 13 значений соответственно). Данные о среднем содержании молочнокислых бактерий в 1 г. йогурта (в 10^6 КОЕ) приведены в таблице

Возраст йогурта	Содержание молочнокислых бактерий (в 10^6 КОЕ)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1/3 от срока годности	22,04	21,99	22,08	21,92	22,10	21,92	21,98	22,21	21,99	22,11	21,94	22,09	22,02	22,05	22,00
2/3 от срока годности	16,09	15,94	15,89	16,14	16,16	15,93	16,09	15,95	15,95	16,09	15,89	16,06	15,94		
конец срока годности	9,98	10,10	9,93	10,10	9,68	9,88	10,02	10,08	10,09	10,12	10,05	9,96	10,31		

б) Исследовать влияние на срок службы стиральных машин следующих факторов: жесткости воды (фактор А) и используемое число оборотов центрифуги (фактор В). Значения фактора $A: A = A_1$ - жесткая вода, $A = A_2$ - ph нейтральная или мягкая вода. Значения фактора $B: B = B_1$ - 1000 оборотов, $B = B_2$ - 800 оборотов. Данные о средней продолжительности работы стиральных машин некоторой марки (в часах) приведены в файле disp-13.txt.

14. а) Исследовать влияние вида применяемых удобрений на содержание калия в фасоли. Согласно значениям качественного фактора данные были поделены на 3 группы (по 12, 11 и 13 значений соответственно). Данные о среднем калия в фасоли (в мг. на 100 г.) приведены в таблице

Вид удобрения	Среднее содержание калия в фасоли (в мг. на 100 г.)												
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	
1	1062,	1059,	1059,	1061,	1060,	1060,	1058,	1059,	1059,	1060,	1060,	1059,	

	1	3	7	3	3	6	3	7	1	8	8	0	
2	1062,9	1061,9	1059,2	1060,7	1060,1	1061,1	1059,7	1060,1	1058,4	1061,5	1057,9		
3	1059,7	1059,2	1059,0	1058,8	1059,8	1058,5	1061,4	1058,5	1059,3	1059,7	1061,3	1060,2	1

б) Исследовать влияние на количество детей у женщин следующих факторов: места проживания (фактор А) и наличия высшего образования (фактор В). Значения фактора $A : A = A_1$ - сельская местность, $A = A_2$ - город. Значения фактора $B : B = B_1$ - есть высшее образование, $B = B_2$ - высшего образования нет. Данные были собраны по 30 сельским пунктам и 30 городам. Данные о среднем количестве детей у женщин 40 лет приведены в файле disp-14.txt.

13 а) Исследовать влияние года обучения студента в ВУЗе на посещаемость им лекционных занятий. Согласно значениям качественного фактора данные были поделены на 3 группы - (по 15, 14 и 13 значений соответственно). Данные посещаемости лекционных занятий студентами каждой специальности (в процентах) приведены в таблице

Год обучения	Средняя посещаемость лекционных занятий (в %)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1-й	76,32	67,52	70,51	75,23	68,72	66,11	72,25	66,93	72,51	66,23	75,77	73,84	67,83	68,68	66,66
2-й – 3-й	63,06	61,53	58,18	61,09	61,24	63,21	62,58	55,92	56,45	58,92	66,67	63,11	63,02	62,15	
4-й – 5-й	51,82	49,86	48,87	51,81	54,51	58,69	48,14	52,42	53,48	53,23	57,99	54,53	55,78		

б) Исследовать влияние на курс акций некоторого предприятия следующих факторов: поведения курса акций предприятия А (фактор А) и поведения курса акций предприятия Б (фактор В). Значения фактора $A : A = A_1$ - курс растет, $A = A_2$ - курс падает. Значения фактора $B : B = B_1$ - курс растет, $B = B_2$ - курс падает. Данные о цене акции изучаемого предприятия приведены в файле disp-15.txt.

14 а) Исследовать влияние сорта яблок на содержание в них железа. Согласно значениям качественного фактора данные были поделены на 3 группы - (по 15, 13 и 15 значений соответственно).

Данные о содержание железа в яблоках (в мг. на 100 г.) приведены в таблице

Сорт яблок	Содержание железа в разных сортах яблок (в мг. на 100 г.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
сорт «А»	2,51	2,51	2,50	2,50	2,50	2,50	2,50	2,49	2,53	2,50	2,49	2,51	2,50	2,49	2,49
сорт «Б»	2,46	2,45	2,44	2,44	2,44	2,45	2,45	2,46	2,44	2,45	2,45	2,47	2,44		
сорт «С»	2,38	2,41	2,39	2,40	2,39	2,40	2,38	2,40	2,40	2,40	2,41	2,39	2,40	2,40	2,40

б) Исследовать влияние на урожайность пшеницы следующих факторов: количества осадков за сезон (фактор А) и давность содержания земли "под паром"(фактор В). Значения фактора $A: A = A_1$ - количество осадков выше нормы, $A = A_2$ - количество осадком ниже нормы. Значения фактора $B: B = B_1$ - в прошлом году земля находилась "под паром", $B = B_2$ - не находилась. Данные об урожайности пшеницы (в ц. с Га) приведены в файле disp-16.txt.

15 а) Исследовать влияние возраста покупателя на то, сколько денег он тратит на покупку мясной продукции в магазине. Согласно значениям качественного фактора данные были поделены на 3 группы (по 12, 15 и 13 значений соответственно). Данные о среднем количестве средств, потраченных покупателями магазина на мясную продукцию за 1 посещение, приведены в таблице

Возраст	Средние затраты покупателей на мясную продукцию за 1 посещение (в руб)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
18-25	159	158	139	146	166	150	147	145	154	155	149	151			

25-40	145	162	177	151	168	159	149	167	163	151	141	150	147	156	150
старше40	74	89	88	77	62	87	53	78	66	72	74	78	83		

б) Исследовать влияние на успеваемость школьников по биологии следующих факторов: участия в олимпиадах (фактор А) и прохождения летней практики на школьном приусадебном участке(фактор В). Значения фактора $A : A = A_1$- принимал участие в олимпиадах, $A = A_2$ не принимал. Значения фактора $B : B = B_1$- проходил практику на приусадебном участке, $B = B_2$- не проходил. Данные средней оценке школьников за 2 года приведены в файле disp-17.txt.

16 а) Исследовать объективность экспертов, оценивавших качество "Докторской" колбасы от разных производителей. На экспертизу было представлено по 10 образцов колбасы от 15 производителей. Данные о средних баллах, выставленных экспертами каждому производителю (по 10-ти бальной шкале), приведены в таблице

Эксперт	Средние оценки, выставленные «Докторской» колбасе от 15 пользователей														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
1-й	7,0	6,4	7,4	6,3	6,0	7,7	5,4	6,5	6,4	6,7	7,6	7,7	6,2	5,8	6,4
2-й	5,9	7,7	7,1	7,5	8,1	7,0	7,6	8,5	5,4	6,5	7,2	6,8	7,2	7,7	7,2
3-й	8,5	4,1	7,3	6,6	7,6	8,1	6,7	8,1	5,1	9,2	8,5	7,1	7,1	5,4	5,7

б) Исследовать влияние на формирование цены на товар "С" следующих факторов: изменение спроса на товар "А" (фактор А) и изменение спроса на товар "В" (фактор В). Значения фактора $A : A = A_1$- спрос растет, $A = A_2$- спрос падает. Значения фактора $B : B = B_1$- спрос растет, $B = B_2$- спрос падает. Данные о средней цене на товар "С" по городу (в руб.) приведены в файле disp-18.txt.

17 а) Исследовать влияние сорта томатов на содержание в

нем витамина С. Согласно значениям качественного фактора данные были поделены на 3 группы (по 15, 14 и 13 значений соответственно). Данные о среднем содержании витамина С в томатах (в мг. на 100 г.) приведены в таблице

Сорт томатов	Содержание витамина С в томатах (в мг. на 100 г.)														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15}
сорт «А»	34,93	34,81	35,54	34,04	35,34	34,87	34,99	35,30	34,38	34,74	35,03	34,82	35,21	34,14	35,56
сорт «В»	34,06	35,42	34,73	35,72	34,60	35,60	35,35	35,51	34,90	35,58	35,00	35,66	34,79	34,76	
сорт «С»	35,34	35,06	34,31	34,98	35,17	35,73	35,13	34,73	35,10	35,41	35,26	34,31	35,14		

б) Исследовать влияние на склонность к инфекционным заболеваниям у детей следующих факторов: пола ребенка (фактор А) и употребления витаминных комплексов (фактор В). Значения фактора $A : A = A_1$ - девочка, $A = A_2$ - мальчик. Значения фактора $B : B = B_1$ - ребенок принимает витаминные комплексы, $B = B_2$ - не принимает. Данные о среднем количестве перенесенных ребенком инфекционных заболеваний за год приведены в файле disp-19.txt.

18 а) Исследовать влияние поведения атмосферного давления на количество вызовов скорой помощи. Анализ проводился в 15 городах с примерно одинаковой численностью населения в течение года. Данные
о среднем количестве вызовов бригад скорой помощи (в день) в каждом городе приведены в таблице

атм.давление	Среднее число вызовов скорой помощи в день														
	X_{1i}	X_{2i}	X_{3i}	X_{4i}	X_{5i}	X_{6i}	X_{7i}	X_{8i}	X_{9i}	X_{10i}	X_{11i}	X_{12i}	X_{13i}	X_{14i}	X_{15i}
падает	97,2	97,9	95,5	91,1	98,9	100,0	98,4	101,9	102,8	106,3	90,8	112,6	105,1	99,6	97,8
стабильно	86,5	76,7	88,7	84,9	80,8	82,5	79,2	82,5	90,9	84,3	89,6	89,5	82,3	78,0	79,5
растет	102,2	94,6	98,5	95,8	95,7	100,5	101,6	97,6	106,4	96,1	106,3	98,0	95,6	97,7	102,7

б) Исследовать влияние на продолжительность исходящих звонков с сотового телефона следующих факторов: пола абонента(фактор А) и количество отправляемых абонентом смс в день (фактор В). Значения фактора $A : A = A_1$ - женский пол, $A = A_2$ - мужской пол. Значения фактора $B : B = B_1$ - абонент отправляет до 5 смс в день, $B = B_2$ - больше 5 смс в день. Данные о средней продолжительности одного разговора абонента по исходящей связи приведены в файле disp-20.txt.

Самостоятельная работа №7
«Проверка гипотезы о виде распределения с помощью критерия согласия Смирнова»

1 Статистическая гипотеза и статистический критерий. Критерий согласия.

Пусть в результате некоторого эксперимента получена выборка $X = (X_1, X_2, \ldots, X_n)$ из некоторого распределения F.

Статистической гипотезой H называется любое утверждение о виде неизвестного распределения, или о параметрах известного распределения наблюдаемой в эксперименте случайной величины.

Гипотеза H называется простой, если она однозначно определяет распределение выборки: $H = \{F = F_1\}$, иначе H называют сложной, например, $H = \{F \in \{F\}\}$, где $\{F\}$- некоторое семейство распределений: $\{F\} = \{F(x, \theta), \theta \in \Theta\}$.

Если выдвигаются всего две гипотезы, то одну из них принято называть основной (H_0), а другую альтернативной (конкурирующей) (H_1).

Правило, согласно которому проверяемая гипотеза H_0 принимается или отвергается, называется статистическим критерием. Дадим формальное определение критерия.

Статистическим критерием для проверки гипотез $H_1, \ldots, H_k$ называется любое измеримое отображение $\rho : R^n \to \{H_1, \ldots, H_k\}$.

Если $\rho(X) = H_i$, то мы принимаем гипотезу H_i (или считаем $\theta = \theta_i$ в параметрическом случае).

Качество критерия характеризуется набором вероятностей ошибочных решений.

Будем говорить, что произошла ошибка i-го рода, если гипотеза H_i отвергнута, когда она верна. Вероятностью ошибки i-го рода критерия р называется

$$q_i(\delta) = P_{H_i}(p(X) \neq H_i).$$

Если удастся выбрать критерий p так, что все числа $q_i(p)$ малы, то мы будем объявлять, что верна гипотеза H_k, если $\rho(X) = H_k$. При этом мы будем ошибаться примерно в доле случаев q_i, если верна гипотеза H_i.

Уровнем значимости статистического критерия называют вероятность ошибочно отвергнуть основную проверяемую гипотезу, когда она верна (вероятность ошибки первого рода).

Рассмотрим случай, когда о распределении наблюдений X имеется две гипотезы: $H_0 = \{F = F_1\}$ при альтернативе $H_1 = \{F \neq F_1\}$. В этом случае любой критерий $\rho(X): R^n \to \{H_0, H_1\}$ принимает не более двух значений, то есть область R^n делится на две части

$$R^n = S \cup (R^n \setminus S)$$

так, что

$$\rho(X) = \begin{cases} H_0, & X \in R^n \setminus S, \\ H_1, & X \in S. \end{cases}$$

Область S, в которой принимается альтернативная гипотеза, называется критической областью критерия ρ.

Обозначим q = q(ρ) - уровень значимости критерия ρ, тогда

$$q = q_1(\rho) = P_{H_0}(\rho(X) \neq H_0) = P_{H_0}(\rho(X) = H_1) = P_{H_0}(X \in S).$$

По своему смыслу критическая область должна строиться так, чтобы событие $X \in S$ было маловероятным. В конкретных задачах ρ выбирают обычно равной 0,1; 0,05; 0, 01 и так далее.

Критериями согласия называют критерии для проверки простой гипотезы H_0 при сложной альтернативе $H_1 = \{H_0$ неверна$\}$.

Критерии согласия принимают или отвергают основную гипотезу исходя из величины некоторой статистики T(X), характеризующей отклонение эмпирических данных от соответствующих (гипотезе H_0) гипотетических значений, распределение которой в случае справедливости H_0 можно было бы определить.

Предположим, такая статистика и ее распределение при гипотезе H_0 найдены. Пусть T - множество всевозможных значений статистики T. Определим для фиксированного заранее достаточно малого числа $\alpha > 0$ подмножество $\tilde{T} \subset T$, так чтобы вероятность осуществления события $T(X) \in \tilde{T}$ в случае справедливости гипотезы H_0 удовлетворяла условию

$$P\left(T(X) \in \tilde{T} \mid H_0\right) \le \alpha$$

Тогда правило проверки гипотезы H_0 можно сформулировать следующим образом: Если для данной выборки ***X*** значение статистики $T(X) \in \tilde{T}$, то в предположении справедливости гипотезы ***H_0*** произошло маловероятное событие и гипотеза должна быть отвергнута как противоречащая статистическим данным. В противном случае нет основания отказываться от принятия гипотезы ***H_0*. *На языке критериев это можно записать так:***

$$\rho(X) = \begin{cases} H_0, & T(X) \in T \setminus \tilde{T} \\ H_1, & T(X) \in \tilde{T} \end{cases}$$

2.1 Критерий Смирнова для проверки гипотезы о виде распределения

Имеется выборка $X = (X_1, X_2, \ldots, X_n)$ из некоторого распределения F. Проверяется простая гипотеза $H_0 = \{F = F_1\}$ против сложной альтернативы $H_1 = \{F \ne F_1\}$. В том случае, когда распределение F_1 имеет непрерывную функцию распределения F_1, для проверки гипотезы можно воспользоваться критерием Смирнова.

Пусть $F_n^*(x)$ - эмпирическая функция распределения. Рассмотрим статистику

$$T_n = \sqrt{n} \sup_x \left(F_n^*(x)\right) = \sqrt{n} \max_{X_i} \left(F_n^*(X_i) - F_1(X_i)\right).$$

Теорема 4.1 ***В случае справедливости гипотезы*** H_0 при $n \to +\infty$ ***статистика*** T_n ***имеет распределение Смирнова:***

$$\lim_{n\to\infty} P(T_n < x) = S(x),$$

здесь $S(x) \equiv 1 - e^{-2x^2}$ - функция Смирнова.

Зададим некоторый уровень значимости q и найдем пороговое значение статистики C_q из условия $S(C_q) = 1 - q$. Построим критерий Смирнова для проверки гипотезы о виде распределения:

$$\rho(X) = \begin{cases} H_0, & T_n(X) \le C_q, \\ H_1, & T_n(X) > C_q. \end{cases}$$

Приведем пример применения этого критерия. Пусть в результате эксперимента получена выборка объема N=200 из распределения некоторой случайной величины. Проверим гипотезу о том, что эта случайная величина имеет стандартное нормальное распределение. Текст программы, реализующий проверку данной гипотезы с помощью критерия Смирнова, приведен на рисунке 2.1.

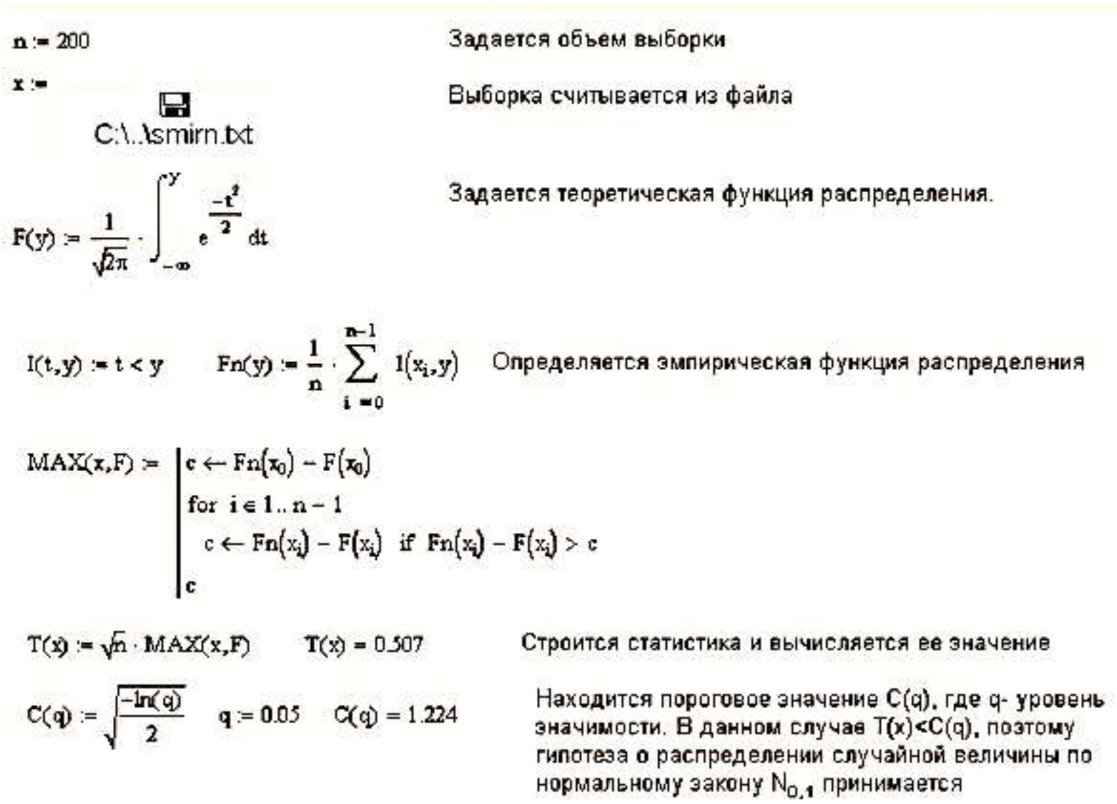


Рисунок 2.1. Проверка гипотезы о распределении случайной величины по нормальному закону $N_{0,1}$ с помощью критерия Смирнова.

2.2 Задание к лабораторной работе

В файле smirn-V.txt (V - номер вашего варианта) задана выборка из некоторого распределения. Задав некоторый уровень значимости q, с помощью критерия Смирнова проверить следующие гипотезы:

Варианты заданий

1. а) $H_0 = \{F = N_{1,2}\}$ против $H_1 = \{F \neq N_{1;2}\}$;
   б) $H_0 = \{F = E_3\}$ против $H_1 = \{F \neq E_3\}$;
   в) $H_0 = \{F = F_{3,5}\}$ против $H_1 = \{F \neq F_{3,5}\}$.
2. а) $H_0 = \{F = U_{2,4}\}$ против $H_1 = \{F \neq U_{2,4}\}$;
   б) $H_0 = \{F = K_{0,1}\}$ против $H_1 = \{F \neq K_{0,1}\}$;
   в) $H_0 = \{F = \Gamma_{2,2}\}$ против $H_1 = \{F \neq \Gamma_{2,2}\}$.
3. а) $H_0 = \{F = \beta_{3,2}\}$ против $H_1 = \{F \neq \beta_{3,2}\}$;
   б) $H_0 = \{F = K_{1,1}\}$ против $H_1 = \{F \neq K_{1,1}\}$;
   в) $H_0 = \{F = E_4\}$ против $H_1 = \{F \neq E_4\}$.
4. а) $H_0 = \{F = \beta_{3,4}\}$ против $H_1 = \{F \neq \beta_{3,4}\}$;
   б) $H_0 = \{F = F_{5,5}\}$ против $H_1 = \{F \neq F_{5,5}\}$;
   в) $H_0 = \{F = U_{-1,1}\}$ против $H_1 = \{F \neq U_{-1,1}\}$.
5. а) $H_0 = \{F = K_{2,4}\}$ против $H_1 = \{F \neq K_{2,4}\}$;
   б) $H_0 = \{F = N_{2,4}\}$ против $H_1 = \{F \neq N_{2,4}\}$;
   в) $H_0 = \{F = U_{0,1}\}$ против $H_1 = \{F \neq U_{0,1}\}$.
6. а) $H_0 = \{F = \beta_{5,4}\}$ против $H_1 = \{F \neq \beta_{5,4}\}$;
   б) $H_0 = \{F = F_{2,4}\}$ против $H_1 = \{F \neq F_{2,4}\}$;
   в) $H_0 = \{F = E_3\}$ против $H_2 = \{F \neq E_3\}$.
7. а) $H_0 = \{F = U_{3,6}\}$ против $H_1 = \{F \neq U_{3,6}\}$;
   б) $H_0 = \{F = K_{1,1}\}$ против $H_1 = \{F \neq K_{1,1}\}$;
   в) $H_0 = \{F = \Gamma_{2,3}\}$ против $H_1 = \{F \neq \Gamma_{2,3}\}$.

8. а) $H_0 = \{F = N_{-1,3}\}$ против $H_1 = \{F \neq N_{-1,3}\}$;

б) $H_0 = \{F = E_{1.5}\}$ против $H_1 = \{F \neq E_{1.5}\}$;

в) $H_0 = \{F = \beta_{3,5}\}$ против $H_1 = \{F \neq \beta_{3,5}\}$.

9. а) $H_0 = \{F = F_{2,3}\}$ против $H_1 = \{F \neq F_{2,3}\}$;

б) $H_0 = \{F = K_{0,2}\}$ против $H_1 = \{F \neq K_{0,2}\}$;

в) $H_0 = \{F = \Gamma_{2,2}\}$ против $H_1 = \{F \neq \Gamma_{2,2}\}$.

10. а) $H_0 = \{F = \beta_{3,4}\}$ против $H_1 = \{F \neq \beta_{3,4}\}$;

б) $H_0 = \{F = E_5\}$ против $H_1 = \{F \neq E_5\}$;

в) $H_0 = \{F = U_{0,5}\}$ против $H_1 = \{F \neq U_{0,5}\}$.

11. а) $H_0 = \{F = \Gamma_{4,4}\}$ против $H_1 = \{F \neq \Gamma_{4,4}\}$;

б) $H_0 = \{F = F_{2,4}\}$ против $H_1 = \{F \neq F_{2,4}\}$;

в) $H_0 = \{F = E_3\}$ против $H_1 = \{F \neq E_3\}$.

12. а) $H_0 = \{F = N_{5,1}\}$ против $H_1 = \{F \neq N_{5,1}\}$;

б) $H_0 = \{F = E_{2.5}\}$ против $H_1 = \{F \neq E_{2.5}\}$;

в) $H_0 = \{F = U_{3,5}\}$ против $H_1 = \{F \neq U_{3,5}\}$.

13. а) $H_0 = \{F = \beta_{3,2}\}$ против $H_1 = \{F \neq \beta 3{,}2\}$;

б) $H_0 = \{F = K_{0,1}\}$ против $H_1 = \{F \neq K_{0,1}\}$;

в) $H_0 = \{F = E_2\}$ против $H_1 = \{F \neq E_2\}$.

14. а) $H_0 = \{F = U_{0,2}\}$ против $H_1 = \{F \neq U_{0,2}\}$;

б) $H_0 = \{F = \Gamma_{3,3}\}$ против $H_1 = \{F \neq \Gamma_{3,3}\}$;

в) $H_0 = \{F = K_{2,1}\}$ против $H_1 = \{F \neq K_{2,1}\}$.

15. а) $H_0 = \{F = F_{5,3}\}$ против $H_1 = \{F \neq F_{5,3}\}$;

б) $H_0 = \{F = E_2\}$ против $H_1 = \{F \neq E_2\}$;

в) $H_0 = \{F = \Gamma_{2,4}\}$ против $H_1 = \{F \neq \Gamma_{2,4}\}$.

16. а) $H_0 = \{F = N_{0,4}\}$ против $H_1 = \{F \neq N_{0,4}\}$;

б) $H_0 = \{F = E_{3.5}\}$ против $H_1 = \{F \neq E_{3.5}\}$;

в) $H_0 = \{F = F_{4,5}\}$ против $H_1 = \{F \neq F_{4,5}\}$.

17. а) $H_0 = \{F = \Gamma_{4,2}\}$ против $H_1 = \{F \neq \Gamma_{4,2}\}$;

б) $H_0 = \{F = E_5\}$ против $H_1 = \{F \neq E_5\}$;

в) $H_0 = \{F = U_{-2,2}\}$ против $H_1 = \{F \neq U_{-2,2}\}$.

18. а) $H_0 = \{F = \beta_{3,6}\}$ против $H_1 = \{F \neq \beta_{3,6}\}$;

б) $H_0 = \{F = F_{2,2}\}$ против $H_1 = \{F \neq F_{2,2}\}$;

в) $H_0 = \{F = E_{4.5}\}$ против $H_1 = \{F \neq E_{4.5}\}$.

19. а) $H_0 = \{F = F_{3,4}\}$ против $H_1 = \{F \neq F_{3,4}\}$;

б) $H_0 = \{F = N_{5,5}\}$ против $H_1 = \{F \neq N_{5,5}\}$;

в) $H_0 = \{F = U_{-1,0}\}$ против $H_1 = \{F \neq U_{-1,0}\}$.

20. а) $H_0 = \{F = K_{2,3}\}$ против $H_1 = \{F \neq K_{2,3}\}$;

б) $H_0 = \{F = E_{1.5}\}$ против $H_1 = \{F \neq E_{1.5}\}$;

в) $H_0 = \{F = U_{0,6}\}$ против $H_1 = \{F \neq U_{0,6}\}$.

Самостоятельная работа №8
«Вычисление числовых характеристик выборки»

Цель: изучить возможности Mathcad по вычислению числовых характеристик выборки

Задание: решить представленную задачу.

Краткие теоретические сведения

Моменты

В теории вероятностей и математической статистике, помимо математического ожидания и дисперсии, используются и другие числовые характеристики случайных величин. В первую очередь это начальные и центральные моменты.

Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени случайной величины X, т.е. $\alpha_k = M(x^k)$.

Центральным моментом k-го порядка случайной величины ξ называется величина μ_k, определяемая формулой

$$\mu_k = M[(x - M)^k].$$

Заметим, что математическое ожидание случайной величины – начальный момент первого порядка $\alpha_1 = M$, а дисперсия – центральный момент второго порядка:

$$\mu_2 = M[(x - M)^2] = D(x).$$

Существуют формулы, позволяющие выразить центральные моменты случайной величины через ее начальные моменты. Одна из таких формул приведена выше:

$$D = M(x - M)^2 = \mu_2 - \alpha_1^2.$$

В дальнейшем будет использована формула

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3.$$

Асимметрия

В теории вероятностей и математической статистике в качестве меры асимметрии распределения служит коэффициент асимметрии, который определяется формулой:

$$\beta = \frac{\mu_3}{\sigma^3}$$

где μ_3 - центральный момент третьего порядка;

$\sigma = \sqrt{D} = \sqrt{\mu_2}$ - среднеквадратичное отклонение

Коэффициент асимметрии – безразмерная величина, а по его знаку можно судить о характере асимметрии. Для симметричной СВ $\mu_3 = 0$.

Указание. Для того чтобы определить точность коэффициента асимметрии, выделенное выражение для него щелкните в строке Floating Point в меню Symbolics и укажите в окне диалога число десятичных знаков в выводе.

Эксцесс

Нормальное распределение наиболее часто используется в теории вероятностей и математической статистке, и поэтому график плотности вероятностей нормального распределения стал своего рода эталоном, с которым сравнивают другие распределения. Одним из параметров, определяющих отличие сравниваемого распределения от нормального, является эксцесс.

Эксцесс γ случайной величины ξ определяется равенством

$$\gamma = \frac{\mu_4}{(D)^2} - 3.$$

По известным нам свойствам математического ожидания и определению центрального момента получим формулу для μ_4

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4.$$

У нормального распределения, естественно, $\gamma = 0$. Если $\gamma > 0$, то это означает, что график плотности вероятностей p(x) сильнее «заострен», чем у нормального распределения, если же $\gamma < 0$, то «заостренность» графика p(x) меньше, чем у нормального распределения.

Порядок выполнения работы

1. Изучить теоретический материал.

2. Вычислить для выборки заданной случайным образом выборочные моменты 3 и 4-го порядков выборочный эксцесс E, коэффициент асимметрии.

3. Оформить отчет.

Пример выполнения задания приведен на рисунке 1.

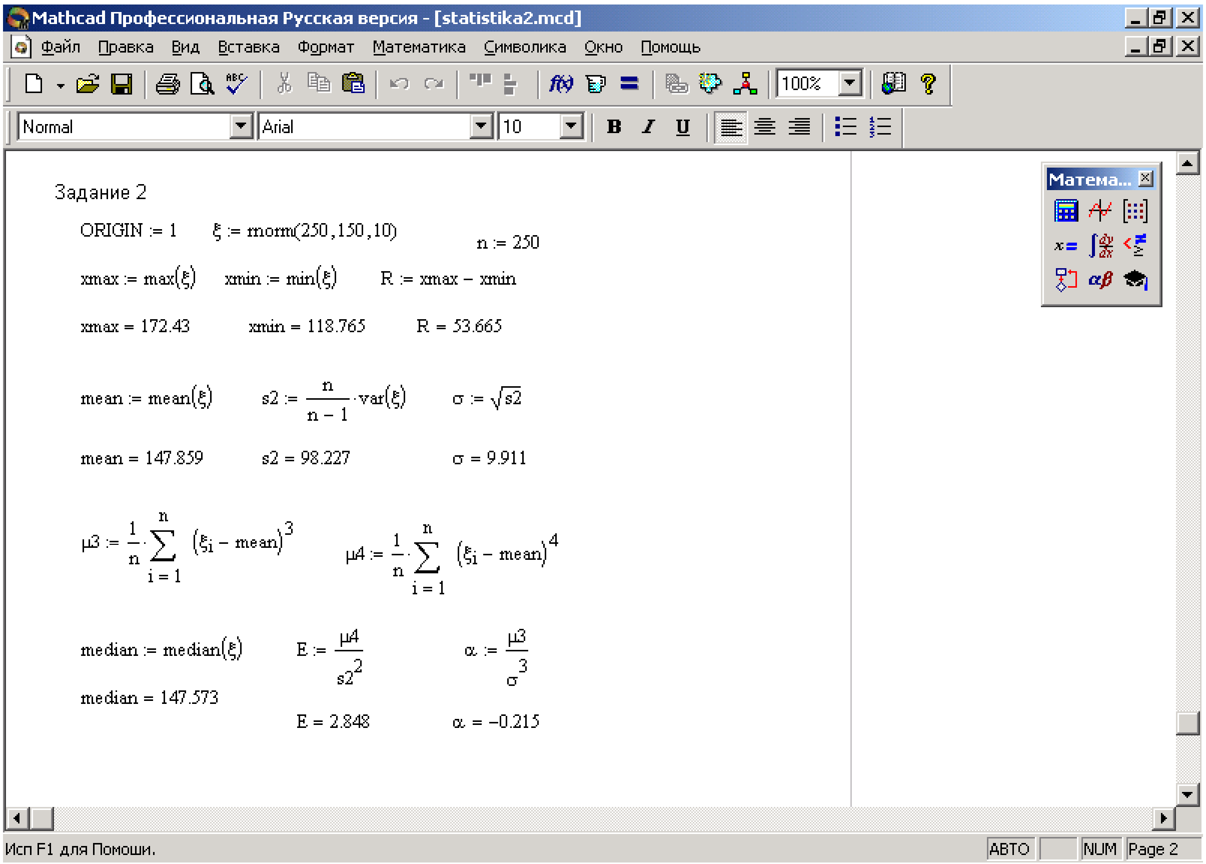


Рисунок 1 – Пример выполнения задания в MathCad

Контрольные вопросы

1. Раскройте понятие «моменты».
2. Что такое центральный момент?
3. Что такое начальный момент?
4. Что такое асимметрия?
5. Что такое коэффициент асимметрии?
6. Что такое эксцесс?

1. Что такое математическое ожидание?

Самостоятельная работа №9
«Применения MATHCAD для решения задач теории вероятности»

Цель работы:

1. Закрепить знания об основных законах распределения случайной величины.
2. Научить применять MathCad для построения функций распределения и плотности распределения случайной величины.
3. Закрепить навыки расчета основных числовых характеристик случайных величин.
4. Научиться применять MathCad для расчета основных числовых характеристик случайных величин.

Функции и инструменты MathCad

Прежде чем приступать к решению задач теории вероятностей в MathCad, познакомимся с инструментами, которые предоставляет пакет для их решения.

Напомним, что дискретная случайная величина x, принимающая значения $x_1 < x_2 < ... < x_i < ...$ с вероятностями $p_1, p_2, ..., p_i, ...$, может быть задана распределением – таблицей вида (X_i - значение случайной величины, p – вероятность появления именно этого значения).

X	x_1	x_2	…	x_i	…	x_n
p	p_1	p_2	…	p_i	…	p_n

Для дальнейшей работы ряд распределения необходимо сделать вариационным.

Вариационный ряд – это дискретный ряд распределения, у которого значения X_i располагаются в порядке возрастания.

Такие таблицы в среде MathCad записывается в виде матрицы размерности $2 \times n$.

Указания. Распределение случайной величины в MathCad записывается в виде матрицы A

$$\begin{pmatrix} A_{1.1} & \vline & A_{1.2} & \ldots & \ldots & A_{1.n} \\ A_{2.1} & \vline & \ldots & \ldots & \ldots & A_{2.n} \end{pmatrix},$$

где $A_{1.i}$ - значения случайной величины (x_i);

$A_{2.i}$ - соответствующие вероятности (p_i);

i=1, 2 , 3 , 4, 5.

Значения X в матрицу следует записывать в порядке возрастания.

По умолчанию первый элемент матрицы имеет индекс 0 (i=0). Используйте системную переменную ORIGIN для задания нижней границы индексации в матрице: ORIGIN:=1.

Функцией распределения случайной величины X называется функция F(x), значение которой в точке x равно вероятности того, что случайная величина X будет меньше этого значения x, то есть

$$F(x) = P\{X < x\}.$$

Функция F(x) для дискретной случайной величины вычисляется по формуле

$$F(x) = \sum_{x_i < x} p_i.$$

где суммирование ведется по всем значениям i, для которых $x_i < x$.

Функция распределения случайной величины, имеющей приведенное выше распределение, имеет вид:

$$F(x) = \begin{cases} 0, \quad x < x_x \\ p_1, \quad x_x < x < x_2 \\ p_1 + p_2, x_2 < x < x_3 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_1 + p_2 + \ldots + p_{n-1}, \quad x_{n11} < x < x_n \\ 1, \quad x_n < x \end{cases}$$

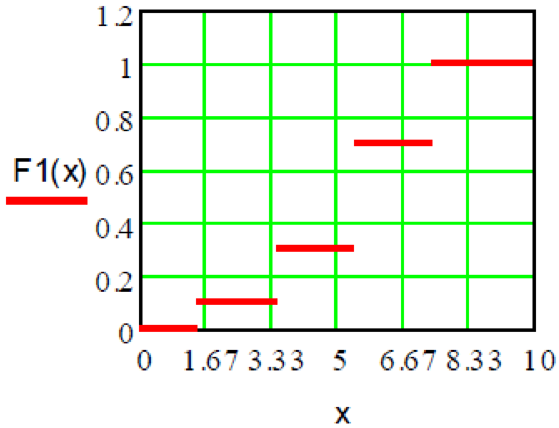
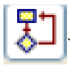


Рисунок 5.1 – Функция распределения

Функцию распределения, заданную разными выражениями на разных интервалах изменения аргументов, можно определить следующим образом:

1. Введите имя функции переменной x и знак присваивания (знак присваивания «:=»).
2. Используйте панель программных элементов (Programming) (кнопка ). С помощью кнопки Add Line добавьте необходимое число строк для задания функции распределения (рисунок 5.1).
3. Введите в помеченной позиции нуль, щелкните по кнопке if и введите неравенство, определяющие первый интервал изменения аргумента (символ ∞ можно ввести щелчком по соответствующей кнопке в панели  (Calculus)).
4. Затем перейдите во вторую строку определения функции, введите $A_{2,1}$ - имя переменной, содержащей значение p_1, или число 0.2 – значение.
5. Введите неравенство, определяющие второй интервал изменения аргумента (знак можно ввести щелчком по соответствующей кнопке в панели отношений (Boolean)); выделите, нажимая клавишу <SPACE>, вторую строку определения функции, щелкните по кнопке Add Line и введите, действуя, как описано выше определение функции на следующем интервале.

В результате у Вас должно получиться следующее:

$$F(x) := \begin{vmatrix} 0 \ \text{ if } \ -\infty < x < A_{1,1} \\ A_{2,1} \ \text{ if } \ A_{1,1} \le x < A_{1,2} \\ A_{2,1} + A_{2,2} \ \text{ if } \ \ldots \\ \ldots \ \ldots \ \ldots \\ \ldots \ \ldots \ \ldots \\ 1 \ \text{ if } \ A_{1,5} \le x < \infty \end{vmatrix}$$

Рисунок 5.2 – Функция распределения случайной величины

На приведенном рисунке 5.2 функция распределения определена с использованием имен переменных.

Замечание. Следует помнить, что MathCad не совсем корректно строит графики ступенчатых функций, соединяя отрезками прямых значения функции в точке скачка. Более точный график функции распределения представляет собой отрезки, параллельные оси абсцисс, с «выколотым» правым концом.

Случайные величины. Функции распределения

Каждая случайная величина полностью определяется своей функцией распределения.

Функция распределения любой случайной величины обладает следующими свойствами:

- F(x) определена на всей числовой прямой R;
- F(x) не убывает, т.е. если $x_1 \le x_2$, то $F(x_1) \le F(x_2)$;
- $F(-\infty) = 0$;
- $F(+\infty) = 1$;
- P(a<x<b)=F(b)-F(a).

Важно помнить, что функция распределения является «паспортом» случайной величины: она содержит всю информацию об этой случайной величине, и поэтому изучение случайной величины заключается в исследовании ее функции распределения, которую часто называют просто распределением.

Для проведения вычислений со случайными величинами (непрерывными и дискретными) в MathCad есть богатая библиотека встроенных функций наиболее распространенных стандартных распределений. Каждое распределение представлено в библиотеке тремя функциями – плотностью вероятностей, функцией распределения и функцией, обратной к функции распределения. Имена всех встроенных функций, определяющих плотности вероятностей, начинаются с буквы d, определяющих функции распределения – с буквы p.

Например, для работы с нормальным распределением предназначены функции dnorm(x, h, s), p(norm(x, h, s) и qnorm(x, h, s).

Наиболее распространенные распределения дискретных случайных величин

Познакомимся с дискретными случайными величинами, которые чаще всего используются при решении практических задач. Эти случайные величины имеют биномиальные, геометрические и пуассоновские распределения.

Биномиальное распределение (схема Бернулли)

Пусть проводится серия из n независимых испытаний, каждое из которых заканчивается либо «успехом», либо «неуспехом». Пусть в каждом испытании (опыте) вероятность успеха p, а вероятность неудачи – q=1-p. С таким испытанием можно связать случайную величину x, равную числу успехов в серии из n испытаний. Эта величина принимает целые значения от 0 до n.

Ее распределение называется биномиальным и определяется формулой Бернулли

$$p_k = P(\xi = k) = C_n^k p^k q^{n-k}$$

где 0<p1, q=1-p, k=0,1, …, n, $C_n^k = \frac{n!}{k!(n-k)!}$

В MathCad для вычисления плотности вероятности и функции распределения случайной величины, имеющей биномиальное распределение, предназначены функции dbinom(k, n,p) и pbinom(k, n, p), значения которых – соответственно p(k) и F(k).

Геометрическое распределение

Со схемой испытаний Бернулли можно связать еще одну случайную величину: h – число испытаний до первого успеха. Эта величина принимает бесконечное множество значений от 0 до $+\infty$, и ее распределение определяется формулой

$$p_k = P(\eta = k) = p^k q$$

где 0<p<1, q=1-p, k=0, 1, …, n.

В MathCad для вычисления плотности вероятности и функции распределения случайной величины, имеющей геометрическое распределение, предназначены функции dgeom(k, p) и pgeom(k, p), значения которых – соответственно p(k) и F(k).

Пуассоновское распределение

Пуассоновское распределение имеет случайная величина m, принимающая значения k=0, 1, 2, … с вероятностями

$$p_k = P(\mu = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, ...,$$

где $\lambda > 0$ - параметр пуассоновского распределения.

В MathCad для вычисления вероятности и функции распределения случайной величины, имеющей пуассоновское распределение, предназначены функции dpois(k, λ) и ppois(k, λ), значения которых – соответственно p(k) и F(k).

Наиболее распространенные частные распределения непрерывных случайных величин

Равномерное распределение

Непрерывная случайная величина ξ, принимающая значение на отрезке [a, b], распределена равномерно на [a, b], если плотность распределения p(x) и функция распределения случайной величины ξ имеют соответственно вид

$$p(x)=\begin{cases}0, & x \notin [a,b] \\ \dfrac{1}{b-a}, & x \in [a,b]\end{cases} \quad F(x)=\begin{cases}0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1, & x > b\end{cases}$$

В MathCad значения в точке x плотности распределения и функции распределения случайной величины имеющее равномерное распределение на отрезке [a, b], вычисляются встроенными функциями соответственно dunif(x, a, b) и punif(x, a, b).

Экспоненциальное (показательное) распределение

Непрерывная случайная величина ξ имеет показательное распределение с параметром $\lambda > 0$, если плотность распределения имеет вид

$$p(x)=\begin{cases}0, & x < 0 \\ \lambda e^{-\lambda x}, & x \ge 0\end{cases} \quad F(x)=\begin{cases}0, & x \le 0 \\ 1-\lambda e^{-\lambda x}, & x > 0\end{cases}$$

В MathCad значения в точке x плотности распределения и функции распределения случайной величины, имеющей экспоненциальное распределение с параметром λ, вычисляются встроенными функциями соответственно dexp(x, l) и pexp(x, l).

Нормальное распределение

Это распределение играет исключительно важную роль в теории вероятностей и математической статистике. Случайная величина ξ нормально распределена с параметрами m и σ, $(\sigma > 0)$, если ее плотность распределения имеет вид

$$p_{\xi}(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

В MathCad значения в точке x плотности распределения и функции распределения нормальной случайной величины с параметрами a, σ вычисляются встроенными функциями соответственно dnorm(x, a,s) и pnorm(x, a,s).

Распределение Стьюдента

Пусть случайная величина ξ имеет стандартное нормальное распределение, а случайная величина $\chi_n^2 - \chi^2$ распределение с n степенями свободы. Если ξ и χ_n^2 независимы, то про случайную величину $\tau_n = \dfrac{\xi}{\sqrt{\chi_n^2 / n}}$ говорят, что она имеет распределение Стьюдента с числом степеней свободы n. Доказано, что плотность вероятности этой величины вычисляется по формуле

$$p_\tau n(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}, x \in R$$

При больших n распределение Стьюдента практически не отличается от N (0, 1).

В MathCad значения в точке x плотности распределения и функции Стьюдента с n степенями свободы вычисляются встроенными функциями соответственно dt(x, n) и pt(x, n).

Числовые характеристики случайных величин

Каждая случайная величина полностью определяется своей функцией распределения. В то же время при решении практических задач достаточно знать несколько числовых параметров, которые позволяют представить основные особенности случайной величины в сжатой форме.

К таким величинам относятся, в первую очередь, математическое ожидание и дисперсия.

Математическое ожидание случайной величины

Математическое ожидание – число, вокруг которого сосредоточены значения случайной величины.

Если ξ - дискретная случайная величина с распределением,

ξ	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

то ее математическим ожиданием – оно обозначается M – называется величина

$$M[x] = \sum_{i=1}^{n} p_i x_i.$$

Математическое ожидание непрерывной случайной величины с плотностью вероятностей p(x) вычисляется по формуле

$$M[x] = \int_{-\infty}^{\infty} xp(x)dx.$$

Дисперсия случайной величины

Дисперсия случайной величины характеризует меру разброса значений случайной величины около ее математического ожидания. Если случайная величина ξ имеет математическое ожидание M, то дисперсией случайной величины ξ называется величина $D = M[(X - M)^2]$. Легко показать, что $D = M[x^2] - (M[x])^2$. Эта универсальная формула одинаково хорошо применима как для дискретных случайных величин, так и для непрерывных. Величина M^2 вычисляется по формулам:

$$M[x^2] = \sum_{i=1}^{n} p_i x_i^2 \quad M[x^2] = \int_{-\infty}^{\infty} x^2 p(x)dx,$$

для дискретных и непрерывных случайных величин соответственно.

Еще одним параметром для определения меры разброса значений случайной величины является среднеквадратическое отклонение σ, связанное с дисперсией соотношением $\sigma = \sqrt{D}$.

Моменты

В теории вероятностей и математической статистике, помимо математического ожидания и дисперсии, используются и другие числовые характеристики случайных величин. В первую очередь это начальные и центральные моменты.

Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени случайной величины X, т.е. $\alpha_k = M(x^k)$.

Центральным моментом k-го порядка случайной величины ξ называется величина μ_k, определяемая формулой

$$\mu_k = M[(x - M)^k].$$

Заметим, что математическое ожидание случайной величины – начальный момент первого порядка $\alpha_1 = M$, а дисперсия – центральный момент второго порядка:

$$\mu_2 = M[(x - M)^2] = D(x).$$

Существуют формулы, позволяющие выразить центральные моменты случайной величины через ее начальные моменты. Одна из таких формул приведена выше:

$$D = M(x - M)^2 = \mu_2 - \alpha_1^2.$$

В дальнейшем будет использована формула

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3.$$

Асимметрия

В теории вероятностей и математической статистике в качестве меры асимметрии распределения служит коэффициент асимметрии, который определяется формулой:

$$\beta = \frac{\mu_3}{\sigma^3}$$

где μ_3 - центральный момент третьего порядка;

$$\sigma = \sqrt{D} = \sqrt{\mu_2} \text{ - среднеквадратичное отклонение}$$

Коэффициент асимметрии – безразмерная величина, а по его знаку можно судить о характере асимметрии. Для симметричной СВ $\mu_3 = 0$.

Указание. Для того чтобы определить точность коэффициента асимметрии, выделенное выражение для него щелкните в строке Floating Point в меню Symbolics и укажите в окне диалога число десятичных знаков в выводе.

Эксцесс

Нормальное распределение наиболее часто используется в теории вероятностей и математической статистке, и поэтому график плотности вероятностей нормального распределения стал своего рода эталоном, с которым сравнивают другие распределения. Одним из параметров, определяющих отличие сравниваемого распределения от нормального, является эксцесс.

Эксцесс γ случайной величины ξ определяется равенством

$$\gamma = \frac{\mu_4}{(D)^2} - 3.$$

По известным нам свойствам математического ожидания и определению центрального момента получим формулу для μ_4

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4.$$

У нормального распределения, естественно, $\gamma = 0$. Если $\gamma > 0$, то это означает, что график плотности вероятностей p(x) сильнее «заострен», чем у нормального распределения, если же $\gamma < 0$, то «заостренность» графика p(x) меньше, чем у нормального распределения.

Порядок выполнения работы

Задание 1.

Постройте с помощью MathCAD график функции распределения для случайной величины:

X	1	0	7	4	-2
P	0.1	0.5	0.1	0.1	0.2

Порядок выполнения задания

1. Задайте случайную величину A в виде матрицы.
2. Определите функцию распределения случайной величины F(x).
3. Определите функцию распределения G(x) той же случайной величины с использованием конкретных значений переменных.
4. Постройте графики функций распределений F(x) и G(x). Отредактируйте и сравните графики.

В отчете представить два графика F(x) и G(x), сделать выводы о их сходстве и объясните, почему так произошло.

Задание 2.

Порядок выполнения задания

1. Задайте случайные величины, имеющие биномиальное распределение. Параметры распределения задайте самостоятельно. Постройте графики распределения и функции распределения случайной величины. (Подберите удобные параметры отображения графиков). Удобно строить оба графика в одной системе координат.

2. Проверьте для них $\sum_{k=0}^{n} p_k = 1$.

3. Вычислите вероятность попадания значений случайной величины в выбранный интервал.

4. Найдите значение k, для которого величина $P(\xi = k)$ максимальна (медиану). Исследуйте (понаблюдайте) зависимость этой вероятности от параметров распределения.

Указание. Для того, чтобы определить по графику распределения наиболее вероятное значение случайной величины, щелкните в меню Format (Формат) в пункте Graph (График) по строке Trace (Следование), установите перекрестье маркера на точке максимума распределения и выведите в рабочий документ вероятность значения, указанного в окне X-Value (Величина X).

5. Измените значения параметров распределения и повторите вычисления. Сравните полученные результаты. Выводы привести в отчете.

Повторить п. 1-5 для других распределений (геометрическое и пуассоновское).

В отчете представьте по одному варианту для каждого распределения: параметры распределения, графики вероятности и функции распределения, значение наиболее вероятного значения СВ и значение вероятности попадания СВ в указанный диапазон. В отчет вставлять фрагменты из MathCAD.

Задание 3.

Порядок выполнения задания

1. Введите параметры равномерного распределения. (Можно использовать любые).
2. Определите плотность вероятности и функцию распределения случайной величины.
3. Постройте графики.
4. Поэкспериментируйте с параметрами различных распределений.
5. Сделайте выводы о зависимости выходных данных от входных параметров.

Повторить п. 1-5 для других распределений (экспоненциального, нормального распределения и распределения Стьюдента).

В отчете представьте по одному варианту для каждого распределения: параметры распределения, графики вероятности и функции распределения и выводы о зависимости выходных данных от входных параметров. В отчет вставлять фрагменты из MathCad.

Задание 4.

Порядок выполнения задания

1. Вычислите математическое ожидание и дисперсии для случайных величин, имеющих дискретные распределения (биномиальные, геометрические и пуассоновские распределения). В качестве исходных параметров возьмите:

n=10 – число испытаний;

p=0.1*h – вероятность наступления события;

$\lambda = 4 + k$ - параметр пуассоновского распределения;

N=1000 – число испытаний стремится к ∞;

k – номер варианта студента по журналу;

2. Сверите полученные значения со справочными данными.

В отчете привести все расчетные формулы и результаты сравнения для каждого распределения в виде фрагментов из MathCad.

Задание 5.

Порядок выполнения задания

1. Вычислите математические ожидания и дисперсии для случайных величин, имеющих непрерывные распределения (равномерное, экспоненциальное, нормальное, распределение Стьюдента). В качестве исходных параметров возьмите (при желании можно использовать любые другие) следующие параметры:

a=1*h, b=3*h, n=5, λ=0.1*h, σ=0.05*h – необходимые параметры распределений;

h – порядковый номер студента по журналу.

2. Сверите полученные значения со справочными данными.

В отчете привести все расчетные формулы и результаты для каждого распределения в виде фрагментов из MathCad.

Задание 6 (Дополнительное).

Вычислите коэффициент асимметрии случайной величины X с равномерным распределением (или любого другого).

Порядок выполнения задания

1. Определите значения параметров распределения случайной величины.
2. Вычислите коэффициент асимметрии.
3. Сверьте полученное значение со справочными данными для данного распределения.
4. Постройте график плотности вероятности.
5. Сделайте выводы, опираясь на значение коэффициента асимметрии и форму распределения.

Задание 7. (Дополнительное)

Вычислите эксцесс случайной величины ξ с равномерным распределением (или любым другим).

Порядок выполнения задания

1. Определите значения параметров распределения случайной величины.
2. Вычислите коэффициент асимметрии.
3. Сверьте полученное значение со справочными данными для данного распределения.
4. Постройте график плотности вероятности.
5. Сделайте выводы, опираясь на значение коэффициента асимметрии и форму распределения.

Справочный материал приведен ниже.

Законы распределения случайных величин и их характеристики

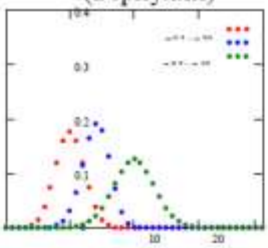
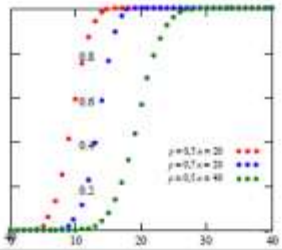
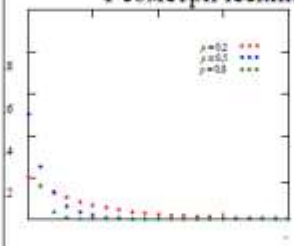
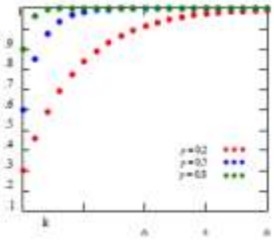
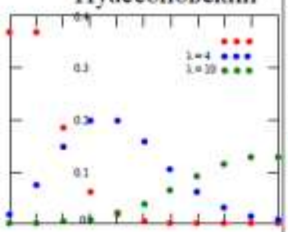
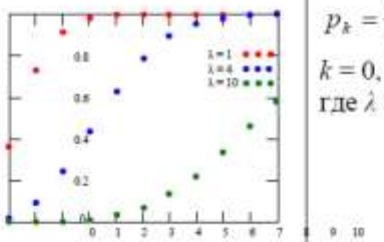
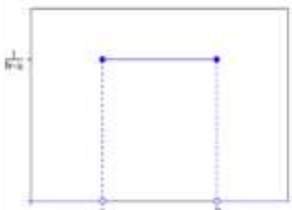
Вид закона распределения		Аналитическое выражение	Основн. число-вые хар-ки
Плотность распределения	Функция распределения		
Дискретные случайные величины			
Биноминальный (Бернулли) 		$p_k = C_n^k p^k q^{n-k}$, где $0 < p < 1,\ q = 1-p,$ $k = 0,\ 1,\ \dots,\ n,$ $C_n^k = \dfrac{n!}{k!(n-k)!}$	$M(X) = np$, $D(X) = npq$, $A = \dfrac{q-p}{\sqrt{npq}}$, $E = \dfrac{1-6pq}{npq}$
Геометрический 		$p_k = p^k q$ где $0 < p < 1,\ q = 1-p,$ $k = 0,\ 1,\ \dots,\ n$	$M(X) = \dfrac{1-p}{p}$, $D(X) = \dfrac{q}{p^2}$, $A = \dfrac{2-p}{\sqrt{1-p}}$, $E = 6 + \dfrac{p^2}{1-p}$.
Пуассоновский 		$p_k = \dfrac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots,$ где $\lambda > 0$.	$M(X) = \lambda$, $D(X) = \lambda$, $A = \lambda^{-1/2}$, $E = \lambda^{-1}$
Непрерывные случайные величины			
 Равномерный		$p(x) = \begin{cases} 0, & x \notin [a,b] \\ \dfrac{1}{b-a}, & x \in [a,b] \end{cases}$ $F(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases}$	$M(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$ $A = 0$, $E = -\dfrac{6}{5}$.

Рисунок 5.2 – Законы распределения случайных величин

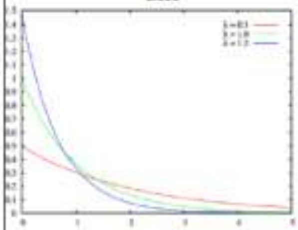
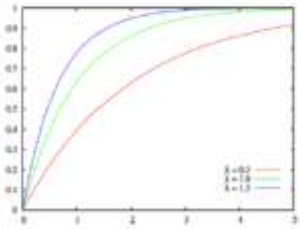
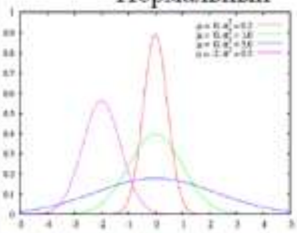
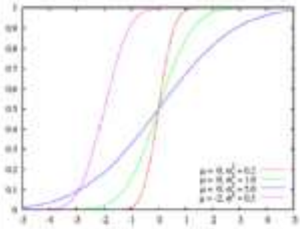
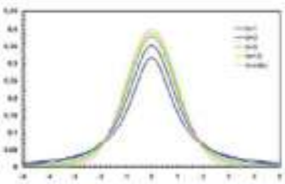
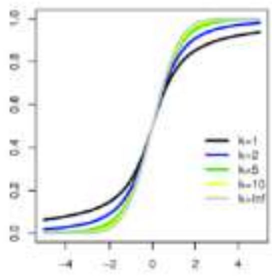
Экспоненциальн ый 		$p(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\ge 0\end{cases}$ $F(x)=\begin{cases}0, & x\le 0\\ 1-\lambda e^{-\lambda x}, & x>0\end{cases}$ $\lambda > 0.$	$M(X)=\lambda^{-1}$, $D(X)=\lambda^{-2}$, $A=2$, $E=6$.
Нормальный 		$p(x)=$ $=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)$ $F(x)=\Phi(x)$, где $\Phi(x)$ – функция Лапласа: $\Phi(x)=$ $=\dfrac{1}{\sqrt{2\pi\sigma^2}}\int\limits_{-\infty}^{x}\exp\left(-\dfrac{(z)^2}{2\sigma^2}\right)dz$ где $z=x-m$	$M(X)=m$, $D(X)=\sigma^2$, $A=0$, $E=0$.
Стьюдента 		$p(x)=$ $=\dfrac{1}{\sqrt{n\pi}}\dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$ $x\in R$ где $\Gamma\left(\dfrac{n}{2}\right)$ - гамма-функция Эйлера	$M(X)=0$, если n>1 $D(X)=\dfrac{n}{n-2}$, если n>2, $A=0$, если n>3, $E=\dfrac{6}{n-4}$, если n>4.

Рисунок 5.3 – Законы распределения случайных величин