

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Невинномысский технологический институт (филиал)

ЛАБОРАТОРНЫЙ ПРАКТИКУМ
по дисциплине
«ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Методические указания к выполнению лабораторных
работ для студентов направления подготовки
15.03.04 Автоматизация технологических процессов и производств
Направленность (профиль) Информационно-управляющие системы
Часть 1

Невинномысск 2021

Методические указания предназначены для студентов направления подготовки 15.03.04 Автоматизация технологических процессов и производств и других технических специальностей. Они содержат основы теории, порядок проведения лабораторных работ. Работы подобраны и расположены в соответствии с методикой изучения дисциплины «Обработка экспериментальных данных». Объем и последовательность выполнения работ определяются преподавателем в зависимости от количества часов, предусмотренных учебным планом дисциплины, как для очной, так и для заочной форм обучения.

Составитель:

ст. преподаватель Д.В. Самойленко

Ответственный редактор:

канд. техн. наук А.А. Евдокимов

Лабораторная работа 1

Основные приемы работы с MathCAD

Цель и содержание: изучить основные приемы работы с системой MathCAD, использование MathCAD для решения различных задач (экономических, инженерных и т.д.).

Теоретическое обоснование

Система MathCAD относится к так называемым инструментальным средствам, которые предназначены для решения задач в различных областях человеческой деятельности (экономических, инженерных, физических и других).

Окно системы MathCAD называется рабочий лист. Оно содержит области для ввода формул и текстовых блоков. Основным блоком рабочего листа является область для ввода формул. Каждая из созданных на рабочем листе формул выполняется последовательно, то есть в системе MathCAD расчет может производиться либо слева направо, либо сверху вниз.

На рабочем листе MathCAD размещение информации определяется курсором. Если курсор является крестообразным, то в этом месте может быть создан следующий объект. При вводе формулы курсор является угловым, а при вводе текста – текстовый курсор в виде вертикальной черты.

В текстовых блоках на рабочем листе MathCAD осуществляется ввод комментариев, поясняющих решение задач в тех или иных областях человеческой деятельности.

Для создания текстового блока на рабочем листе MathCAD следует выполнить следующие действия:

- 1) активизировать главное меню MathCAD,
- 2) перейти в опцию Insert, в которой затем выбрать команду Text Region (Вставка → Текстовый блок).

Текстовый блок можно создать и автоматическом режиме, то есть блок создается после нажатия на клавишу Space (Пробел). Если текст введенный в данный блок не может быть воспринят как формула, то он сразу же становится текстовым.

Формула является основным элементом рабочего листа MathCAD. При открытии документа MathCAD каждый новый элемент уже является формулой. Для ввода формулы следует расположить крестообразный курсор в необходимое место рабочего листа MathCAD, а затем вводить буквы, цифры и знаки соответствующих операций.

Элементы формул в MathCAD можно вводить несколькими способами: либо с помощью клавиатуры, либо с помощью специальных панелей инструментов. Панели инструментов могут быть вызваны следующим образом: View (вид) → Toolbars (панель инструментов) или активизировать панель инструментов Math (математика). Для ввода элементов формул используют панели:

- Calculator (счет) – предназначена для ввода чисел, знаков основных арифметических операций и наиболее употребляемых функций;
- Evaluation (вычисление) – служит для ввода операторов вычисления и знаков логических операций;
- Graph (график) – используется для построения графиков функций;
- Matrix (матрица) – для ввода векторов, матриц и выполнения матричных операций;
- Calculus (исчисление) – для выполнения операций математического анализа;
- Greek (греческий алфавит) – для ввода букв греческого алфавита;
- Symbolic (аналитические выражения) – для выполнения расчетов в режиме символьных вычислений.

Для внесения изменений в формулу или текстовый блок используется панель инструментов Formatting (форматирование). Для того чтобы изменить стиль формулы или создать новый стиль, необходимо выполнить команду Format → Equation (формат → выражение). При оформлении текстовых блоков используется команда Format → Style (формат → стиль).

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б)..

Методика и порядок выполнения работы

1. Загрузите систему MathCAD, используя главное меню Windows: Пуск → Программы → MathSoft Apps → MathCAD.

2. Перейдите в режим автоматических вычислений: Math → Automatic Calculation (математика → автоматические вычисления).

3. Вызовите панель инструментов Calculator (счет), выполнив щелчок по кнопке Calculator Toolbars (панель инструментов «счет»), или воспользуйтесь командой: View → Toolbars → Calculator (вид → панели инструментов → счет).

4. Установите в MathCAD английскую раскладку клавиатуры.

5. Введите исходные данные. Для ввода исходных данных в MathCAD используется операция присвоения. Для того чтобы задать операцию присвоения в пакете MathCAD, можно воспользоваться клавиатурой или панелью инструментов Calculator (счет). При использовании клавиатуры для задания операции присваивания следует набрать символ «:=». Если используется панель Calculator (счет), то следует выполнить щелчок по кнопке Assign Value (присвоить значения).

6. После того как выполнено присвоение исходных данных, необходимо начать ввод формул, при помощи которых может быть найдено значение какого-либо параметра. Каждую их формул в MathCAD следует вводить слева направо. При вводе формул следует учитывать: 1) знак «*» – между переменными обязателен, так как в противном случае MathCAD воспримет несколько переменных как единое целое; 2) если же в формуле присутствует арифметический корень, то используются кнопки панели инструментов Calculator (счет) или комбинация клавиш CTRL + «/»; 3) если же выражение, задающее формулу, является дробным, то следует дважды нажать клавишу Space (Пробел), а затем клавишу «/».

7. Для вывода на экран результата расчета следует набрать имя переменной, а затем нажать клавишу «=».

Задание. Используя режим автоматических вычислений, найдите в системе MathCAD решение задач.

Пример выполнения задания

Используя систему MathCAD, найдите решение следующей задачи. Вычислите высоту конусообразной воронки объема V , если длина ее образующей равна L .

$$V := 6 \quad L := 15$$

Ввод исходных данных

$$R := \sqrt[3]{\frac{V \cdot \sqrt{3}}{\pi}} \quad R = 1.49$$

Задание формулы для вычисления значения радиуса основания конуса R и вывод полученного значения

$$H := \sqrt{L^2 - R^2} \quad H = 14.926$$

Ввод формулы для вычисления искомой величины H и печать найденного значения

Варианты заданий

Вариант для выполнения задания, приведенного в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Задача 1. В Сберегательный банк обратился гражданин с просьбой выдать ему кредит на приобретение бытовой, телевизионной и аудио – техники. Рассчитать, на сколько увеличится сумма кредита, если наращение процентов описывается простой ставкой. Для расчета использовать формулы

$$I = P \cdot n \cdot i, \quad S = P + I = P + P \cdot n \cdot i = P(1 + n \cdot i), \quad n = \frac{t}{K},$$

где I – проценты за весь срок ссуды; P – первоначальная сумма долга; S – наращенная сумма в конце срока; i – ставка наращения (десятичная дробь); n – срок ссуды (обычно в годах); t – число дней ссуды; K – число дней в году.

Задача 2. В Сберегательный банк обратился гражданин с просьбой выдать ему кредит на приобретение бытовой, телевизионной и аудио – техники. Погашение долга с процентами производится частями на протяжении всего срока кредита. Рассчитать наращенную сумму долга S и величину разового погасительного платежа R , используя формулы:

$$S = P \cdot (1 + n \cdot i), \quad R = \frac{S}{m \cdot n},$$

где S – наращенная сумма долга; P – первоначальная сумма ссуды; n – срок кредита; m – число платежей в году; R – величина разового погасительного платежа.

Задача 3. В Сберегательный банк обратились представители организации с просьбой выдать им кредит для развития новых инновационных технологий. После беседы с управляющим банка с представителем фирмы было принято решение о выдаче кредита под сложные проценты. Рассчитать их величину и значение наращенной суммы в конце n - го года при условии, что проценты начисляются один раз в году:

$$S = P \cdot (1 + i)^n, \quad I = S - P = P \cdot (1 + i)^n - P = P \cdot [(1 + i)^n - 1],$$

где P – первоначальный размер долга; i – ставка наращения по сложным процентам; n – число лет наращения.

Задача 4. Рассчитать величину экономии от капиталовложений в автомобильный парк (X) по формуле:

$$X = \frac{A \cdot R \cdot B}{365 \cdot P},$$

где A – экономия грузооборота, т. км; R – стоимость автопарка, тыс. руб.; $B = 1.5$ – коэффициент перевода т. км нетто в т. км брутто; P – производительность.

Задача 5. Рассчитать величину заработной платы по формуле:

$$OS = KO \cdot HS,$$

где OS – основная заработная плата, тыс. руб, KO – количество отработанных часов, час, HS – часовая тарифная ставка, руб.

Задача 6. Рассчитать величину амортизационных отчислений по формуле:

$$A = \frac{PS + KR - LC}{T},$$

где PS – первоначальная стоимость, тыс. руб, KR – затраты на капитальный ремонт, тыс. руб, LS – ликвидная стоимость, руб, T – срок эксплуатации, год, $PS = 30$ тыс. руб, $KR = 1$ тыс. руб, $LS = 185$ руб, $T = 5$ лет.

Задача 7. Рассчитать величину платежа по кредитам Y , используя формулу:

$$Y = \frac{X \cdot R^t \cdot (R - 1)}{R^t - 1},$$

где $R = 1 + I/100$, X – размер кредита, тыс. руб, R – процент по кредиту, %, t – срок погашения кредита, год.

Задача 8. Рассчитать величину фондоотдачи на промышленном предприятии, используя формулу:

$$F = \frac{C \cdot P}{F_1},$$

где F – фондоотдача, C – численность рабочих, чел, P – производительность труда, F_1 – фондовооруженность.

Задача 9. Рассчитать заработную плату в бюджетной организации, используя следующие формулы:

$$C = S - P - R, \quad P = 0.01 \cdot S, \quad R = 0.13(S - M - 300 \cdot N),$$

где C – сумма к выдаче, руб; S – всего начислено, руб; P – профвзнос, руб; R – подоходный налог, руб; N – число иждивенцев, чел; M – минимальная заработная плата, руб.

Задача 10. Рассчитать дневную оплату труда тракториста – машиниста по формуле:

$$S = \frac{M \cdot V \cdot 3.6}{25.2 \cdot 100} \cdot P,$$

где M – минимальная заработная плата, руб; V – средний тарифный коэффициент; P – процент выполнения норм выработки. В зависимости от группы работ

применяются следующие тарифные коэффициенты для I – й группы $V = 3.12$, для II – й – $V = 3.53$, для III – й – $V = 3.99$.

Задача 11. Вычислить величину реализованной цены товара с учетом уровня инфляции по формуле:

$$P = S \cdot (1 + i/100)^t,$$

где P – реализованная цена товара, руб; S – реальная стоимость товара, учитывающая издержки производства и норму прибыли на момент выпуска товара в продажу, руб; i – уровень инфляции, %; t – длительность производственного цикла, мес.

Задача 12. Используя известное значение номинальной ставки и условие, что проценты по вкладу начисляются ежедневно, по формуле рассчитайте значение эффективной ставки:

$$i_E = \left(1 + \frac{i_{\text{nom}}}{N}\right)^N - 1,$$

где i_E – эффективная процентная ставка; i_{nom} – номинальная процентная ставка; N – количество периодов начисления процентов.

Задача 13. Рассчитайте значение будущей стоимости (BS), используя заданные значения приведенной стоимости (PS) и периодической процентной ставки (i_p):

$$BS = PS \cdot (1 + i_p)^N,$$

где i_p – периодическая процентная ставка; N – количество периодов начисления процентов; BS – будущая стоимость; PS – приведенная стоимость.

Задача 14. Какая сумма получится через 20 лет после вложения 1000 \$, если годовая процентная ставка составляет 5%?

Задача 15. Какая сумма будет получена через 20 лет, если на бирже на сумму 1000 \$ куплены акции, дающие в среднем 12,3 % прибыли в год?

Задача 16. Какая сумма останется от 1000 \$ через 20 лет, если эти деньги хранятся в кубышке, а действующая годовая ставка инфляции равна 2,7 %?

Примечание. В качестве расчетной формулы для решения задач №№ 14 – 16 используйте формулу из задачи № 13.

Задача 17. По заданному значению эффективной годовой ставки (i_E), используя формулу $i_p = \left[\sqrt[N]{i_E + 1} \right] - 1$, рассчитайте значения периодической процентной ставки i_p .

Задача 18. Какую сумму из годового бюджета необходимо зарезервировать, чтобы купить оборудование стоимостью 30000 \$, если оно будет получено только через 15 месяцев? Предполагается, что на зарезервированные деньги приходится 7% годовой прибыли.

Задача 19. Какую сумму необходимо ежемесячно выплачивать по ссуде на покупку автомобиля, если срок ссуды – 4 года, номинальная процентная ставка составляет 9 % с ежемесячным начислением процентов, а полная стоимость ссуды – 15000 \$?

Задача 20. Сколько денег должно быть в бюджете компании в начале года, чтобы она имела возможность ежемесячно выплачивать по 600 \$ за оборудование (выплаты производятся в конце месяца), если бюджетные деньги обеспечивают компании прибыль по эффективной годовой ставке 5%?

Примечание. К данной лабораторной работе студенту следует решить две задачи, первая из которых определяется номером персонального компьютера, а вторая – к номеру компьютера необходимо прибавить их количество в аудитории.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Для чего используется система MathCAD?
2. Как называется рабочая область в MathCAD, и на какие области она подразделяется?
3. Каким образом создается комментарий в MathCAD?
4. Последовательность действий при создании формулы.
5. Перечислите и охарактеризуйте основные панели инструментов системы MathCAD для ввода элементов формул.

6. Использование системы MathCAD для решения простейших задач.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 2

Использование MathCAD для обработки экспериментальных данных, заданных в табличном виде

Цель и содержание: изучить основные методы обработки статической информации, заданной в табличном виде, рассмотреть реализацию этих методов в пакете MathCAD.

Теоретическое обоснование

Обработка статистической информации, заданной в табличном виде, основывается на двух концепциях: интерполяция и аппроксимация.

Основная задача *интерполяции* заключается в следующем: найти значения заданной функции в точках интервала, в которых она не задана. Чаще всего при интерполяции считается, что функция задана таблично. Совместно с процессом интерполяции функции рассматривают более широкое понятие *экстраполяции*. Экстраполяция функции заключается в восстановлении функции в точках за пределами заданного интервала.

При решении задач интерполяции и экстраполяции выполняется построение интерполяционным многочленом $L(x)$, который приближенно заменяет исходную функцию $f(x)$ и проходит через точки, в которых известно значение этой функции. Эти точки называют *узлами интерполяции*. Используя найденный многочлен, можно вычислить значение исходной функции вне узлов интерполяции.

При реализации численных методов рассматривают интерполяцию – многочленами или сплайнами.

Основной задачей *аппроксимации* является построение приближенной функции, которая наиболее близко проходит около данных точек или заданной непрерывной функции. Аппроксимацией функции также называют получение аналитической зависимости, описывающей данную функцию.

Одним из основных методов аппроксимации является метод наименьших квадратов, который заключается в следующем: найти такие значения параметров, входящих в аналитическую зависимость, для которых имеет место равенство: $\sum (y_{набл} - y_{табл})^2 \rightarrow \min$.

При аппроксимации функции методом наименьших квадратов необходимо решить нормальную систему линейных алгебраических уравнений:

$$Ax = b,$$

где $A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \dots & \dots \\ t_n & 1 \end{pmatrix}$, $b = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$, $x = \begin{pmatrix} c \\ d \end{pmatrix}$

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

1. Запустите систему MathCAD и проверьте, установлен ли режим автоматического выполнения вычислений

2. Введите полученные в ходе наблюдений экспериментальные данные. В пакете MathCAD экспериментальные (эмпирические) данные, заданные таблично, представляются в виде матрицы, состоящей из двух столбцов, количество строк в такой матрице определяется количеством опытов. В MathCAD матрица может быть создана одним из следующих способов: 1) при помощи панели инструментов Matrix; 2) используя главное меню системы: Insert → Matrix. Обозначим полученные экспериментальные данные следующим образом: X – некоторый показатель; Y – экспериментальное значение, соответствующее значению показателя X.

3. Определите количество точек в наборах исходных данных при помощи функции rows.

4. Используя функцию `mean`, вычислите среднее значение для переменной Y .

5. Рассчитайте величину медианы для экспериментально полученных данных, соответствующих переменной Y , используя функцию `median`.

6. При помощи функции `stdev` вычислите значения среднеквадратичного отклонения и дисперсии экспериментально полученных данных.

7. Постройте функцию, которая бы являлась наилучшим приближением для полученных опытным путем данных. Эту функцию будем называть аппроксимирующей. Построение аппроксимирующей функции в простейшем случае (при условии, что между переменным существует линейная зависимость) может быть выполнена двумя способами: 1) при помощи функций `slope` – для определения коэффициента наклона прямой, `intercept` – свободного члена; 2) путем решения нормальной системы уравнений для метода наименьших квадратов, используя функцию `lsolve`.

8. Перейдите в нижнюю часть рабочего листа MathCAD и задайте аналитическое выражение для аппроксимирующей кривой.

9. Для построения графиков функций (на основании экспериментальных данных и аппроксимирующей) на первом этапе следует задать оси. Это можно сделать следующим образом: 1) используя панель инструментов Graph (график), на которой следует выбрать `X – Y Plot` (декартовы координаты); 2) при помощи клавиатуры `Shift + @`; 3) `Insert → Graph → X – Y Plot`. В результате чего на экране будет выведено окно, в котором в дальнейшем будут выведены необходимые графики.

10. На следующем этапе построения графика функции необходимо задать обозначения осей в соответствии с математическими обозначениями (ось абсцисс и ординат). Параметрами, наносимыми на ось ординат, являются значения экспериментальных данных и аппроксимирующей функции, причем ввод этих значений осуществляется через запятую. Для построения графика функций достаточно нажать клавишу `Enter`.

11. Для улучшения общего вида построенных графиков необходимо выполнить их форматирование. Форматирование полученных графиков выполня-

ется при помощи диалогового окна **Formatting Currently Selected X – Y Plot** (форматирование графика в декартовых координатах). Рассмотрим процесс форматирования графика функции на примере экспериментальных данных. В диалоговом окне **Formatting Currently Selected X – Y Plot** перейдите на вкладку **Traces** (кривые), на которой в качестве графика выберите **trace 1**. Активизируйте столбец **Legend Label** (подпись) и введите название соответствующего графика. Далее в столбце **Symbol** (маркер) из раскрывающегося списка необходимо выбрать способ обозначения отдельных точек. Для графика, построенного по экспериментальным данным, в столбце **Type** (вид линии) установите, что он состоит из отдельных точек. Аналогичным образом можно изменить вид и аппроксимирующей функции.

11. Установка флажка в поле **Hide Arguments** (скрыть параметры) диалогового окна **Formatting Currently Selected X – Y Plot** позволяет не отображать названия осей.

12. Для того чтобы в окне графика функций отображались подписи кривых, следует установить флажок в поле **Hide Legend** (скрыть подписи).

13. Перейдите в поле **Title** (заголовок) на вкладке **Labels** (надписи) и задайте в нем название графика. Для отображения названия следует пометить поле **Show Title** (показать заголовок)

Задание. По результатам исследования некоторых экономических показателей были получены экспериментальные данные, которые можно представить в виде таблицы. Рассчитайте основные статистические показатели для экспериментальных данных. При помощи математического пакета **MathCAD** выполните аппроксимацию функции. Используя метод наименьших квадратов, вычислите значения коэффициентов при условии, что между эмпирическими данными существует линейная зависимость. По полученному аналитическому выражению найдите значения параметра **y** в точке $x = N + 0.55$, **N** – номер варианта. Постройте графики исходной и аппроксимирующей функции.

Пример выполнения задания

По результатам исследований некоторого экономического показателя были получены следующие статистические данные (таблица 2.1).

Таблица 2.1

X	2.5	1.3	3	5.8	5	17	11	21	15.5	5.5
Y	3.8	12.5	7	4	25	30	22	20	27	45

При помощи математического пакета MathCAD выполните аппроксимацию функции. Используя метод наименьших квадратов, вычислите значения коэффициентов при условии, что между эмпирическими данными существует линейная зависимость. Постройте графики исходной и аппроксимирующей функций.

$$\text{data} := \begin{pmatrix} 2.5 & 3.8 \\ 1.3 & 12.5 \\ 3 & 7 \\ 5.8 & 4 \\ 5 & 25 \\ 17 & 30 \\ 11 & 22 \\ 21 & 20 \\ 17.5 & 27 \\ 5.5 & 55 \end{pmatrix}$$

Ввод исходных данных.

$$X := \text{data}^{\langle 0 \rangle} \quad Y := \text{data}^{\langle 1 \rangle}$$

$$n := \text{rows}(\text{data}) \quad n = 10$$

Подсчет количества строк экспериментальных данных.

$$\text{mean}(Y) = 20.63$$

$$\text{mean}(X) = 8.96$$

$$\text{median}(X) = 5.65$$

$$\text{median}(Y) = 21$$

$$\text{stdev}(X) = 6.784$$

$$\text{stdev}(Y) = 14.638$$

Расчет значений статистических функций для экспериментальных значений X и Y.

$$\text{stdev}(Y)^2 = 214.272$$

$$\text{stdev}(X)^2 = 46.026$$

$$A^{(0)} := X \quad A^{(1)} := \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \end{pmatrix} := \text{lsolve}(\text{data}^T \cdot \text{data}, \text{data}^T \cdot Y)$$

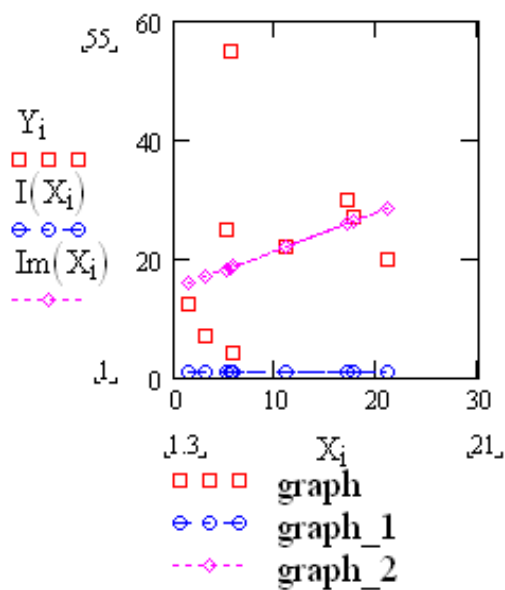
$$I(x) := b + a \cdot x$$

$$a := \text{slope}(X, Y) \quad a = 0.638$$

$$b := \text{intercept}(X, Y) \quad b = 14.916$$

$$\text{Im}(x) := \text{slope}(X, Y) \cdot x + \text{intercept}(X, Y)$$

$$i := 1..10$$



Задание нормальной системы уравнений для определения коэффициентов a и b .

Решение нормальной системы уравнений с помощью функции `lsolve`.

Расчет коэффициентов a и b при помощи функций `slope` и `intercept`.

Построение графика функции по исходным данным, аппроксимирующих кривых

Варианты заданий

Вариант для выполнения задания, приведенного в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Исходные данные для выполнения задания к лабораторной работе получить у преподавателя.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. В чем заключаются концепции интерполяции и экстраполяции функции?
2. Перечислите основные методы интерполяции функции.
3. Перечислите и охарактеризуйте основные методы интерполяции функции при помощи многочленов.
4. Охарактеризуйте концепцию аппроксимации функции.
5. Каким образом организуется аппроксимация функции в среде MathCAD?

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 3

Использование сплайн - интерполяции при обработке и анализе экспериментальных данных, заданных в табличном виде

Цель и содержание: изучить метод приближения функции сплайнами, рассмотреть использование пакета MathCAD для реализации сплайн - интерполяции.

Теоретическое обоснование

Сплайн-интерполяция функции одной переменной относится к многоинтервальной интерполяции. При выполнении сплайновой интерполяции значения полученного многочлена будут совпадать со значениями исходной функции в узлах интерполяции, но и обеспечивает непрерывность заданного числа первых производных на границе частных интервалов.

Сплайн представляет собой функцию, которая на межузловом интервале совпадает с некоторым полиномом. Этот полином является единственным для каждого из интервалов.

Кубический сплайн – это совокупность полиномов третьей степени, каждый из которых для i – го участка можно представить в виде:

$$y = a_i \cdot (x - x_i)^3 + b_i \cdot (x - x_i)^2 + c_i \cdot (x - x_i) + d_i, \quad (1)$$

где $x_{i+1} - x_i = const = h$.

Кубическим сплайном, задаваемым локально равенством (1), называется интерполирующая функция, значения которой можно рассчитать при помощи формулы:

$$i = \text{int} \frac{x - a}{h}, \quad (2)$$

$$y(x) = \frac{(x_{i+1} - x)^2(2(x - x_i) + h)}{h^3} y_i + \frac{(x_{i+1} - x)(2(x_{i+1} - x) + h)}{h^3} y_{i+1} + \frac{(x_{i+1} - x)^2(x - x_i)}{h^3} m_i - \frac{(x - x_i)^2(x - x_{i+1})}{h^3} m_{i+1}, \quad (3)$$

где m_i, m_{i+1} – первые производные функции $y(x)$. Их значения могут быть вычислены двумя способами.

1 способ – производные m_i, m_{i+1} вычисляются при помощи формул численного дифференцирования по трем точкам:

$$m_i = \frac{y_{i+1} - y_{i-1}}{2h} \quad (i = \overline{1, n}), \quad m_0 = \frac{4y_1 - y_2 - 3y_0}{2h} \quad (i = 0), \quad m_n = \frac{3y_n + y_{n-2} - 4y_{n-1}}{2h} \quad (i = n). \quad (4)$$

2 способ – значения производных m_i вычисляются аналитически, а затем представляются в виде массива данных.

Для построения сплайн - интерполяции в системе MathCAD используются следующие функции:

1) $\text{cspline}(VX, VY)$ – выполняет построение вектора вторых производных VS при приближении исходной функции в узлах интерполяции к кубическому полиному;

2) $\text{pspline}(VX, VY)$ – вычисляет компоненты вектора вторых производных VS при приближении функции в узлах интерполяции к параболе;

3) $\text{lspline}(VX, VY)$ – позволяет вычислить значения компонент вектора вторых производных VS при приближении в узловых точках интервалов интерполяции к линейной функции;

4) $\text{interp}(VS, VX, VY, x)$ – вычисляет значения функции $y(x)$ для заданных векторов VS, VX, VY и заданных значений переменной x .

В математическом пакете MathCAD сплайн-интерполяция производится в два этапа. На первом этапе при помощи одной из функций cspline , pspline и lspline выполняется построение вектора вторых производных (VS) для функции заданной векторами VX и VY . На втором же этапе для каждой из точек, используя функцию interp , вычисляются значения $y(x)$.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б)..

Методика и порядок выполнения работы

1. Проверьте, установлен ли в системе MathCAD режим автоматических вычислений.

2. Введите исходные данные. Исходные данные задайте в виде двух матриц – столбцов: матрица X – эмпирически полученные значения переменной x ; а матрица Y – переменной y .

3. При помощи одной из встроенных функций (cspline , pspline или lspline) постройте вектор вторых производных, соответствующих функции $y(x)$.

4. Используя функцию interp , вычислите значение исходной функции вне узлов интерполяции.

5. Построить графики исходной функции $y(x)$ и соответствующего сплайнового приближения.

Задание. Выполните сплайн-интерполяцию по эмпирическим данным, представленным в виде таблицы, используя среду MathCAD. Исходная функция $y(x)$ приближается к линейной функции, параболе, кубическому полиному.

Пример выполнения задания

Экспериментальные данные, полученные в результате наблюдений, представлены в таблице 3.1.

Таблица 3.1

X	1	2	3	4	4.5	5
Y	0	3	4	1	2	5

Выполните приближение сплайнами при условии, что в узлах интерполяции график, построенный по экспериментальным данным, приближается к линейной функции, параболе, кубическому полиному.

$$data := \begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 3 & 4 \\ 4 & 1 \\ 4.5 & 2 \\ 5 & 5 \end{pmatrix} \quad X := data^{(0)} \quad Y := data^{(1)}$$

Ввод исходных данных – формирование векторов X и Y.

$$S := cspline(X, Y) \quad fits(x) := int\ erp(S, x)$$

Построение вектора вторых производных при приближении в узлах интерполяции к кубическому полиному; вычисление значений функции в заданных точках.

$$fits(2.5) = 4.148 \quad fits(4.5) = 2$$

$$S1 := pspline(X, Y) \quad fits1(x) := int\ erp(S1, x)$$

Вычисление компонент вектора вторых производных при приближении функции в узлах интерполяции к параболе; расчет значений функции $y(x)$

$$fits1(2.5) = 4.088 \quad fits1(4.5) = 2$$

$$S2 := lpline(X, Y) \quad fits2(x) := int\ erp(S2, x)$$

Вычисление значе-

$$fits2(2.5) = 4.092 \quad fits2(4.5) = 2$$

ний компонент вектора вторых производных функции при ее приближении в узлах интерполяции к линейной функции и значений $y(x)$

$$\begin{aligned} i &:= 0..length(X) - 1 \\ scale &:= 100 \quad j := 0..scale \\ x_j &:= \min(X) + j \cdot \frac{\max(X) - \min(X)}{scale} \end{aligned}$$

Подготовительный этап к графической интерпретации выполненных выше операций приближения сплайнами и построение соответствующих графиков

Варианты заданий

Вариант для выполнения задания, приведенного в лабораторной работе, определяется номером персонального компьютера, за которым работает студент. В качестве варианта задания к лабораторной работе следует выбрать соответствующий вариант из лабораторной работы № 2.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. В чем заключается приближение функции сплайнами?
2. Что такое сплайн? Каким образом можно задать кубический сплайн?
3. Основные функции для выполнения сплайн - интерполяции в пакете MathCAD.
4. Алгоритм сплайн - интерполяции в пакете MathCAD.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 4

Использование инструментальных средств для формирования последовательности случайных чисел

Цель и содержание: познакомить студентов с основными видами распределений случайных чисел, основными инструментами, используемыми при формировании последовательностей случайных чисел, рассмотреть реализацию методов формирования этих последовательностей при помощи различных инструментальных средств (MathCAD, Excel).

Теоретическое обоснование

В табличном процессоре Excel для формирования последовательности случайных чисел используется надстройка (пакет) «Анализ данных». Его подключение осуществляется следующим образом.

1. Войдите в главное меню Excel.
2. Активизируйте в нем опцию «сервис».
3. В появившемся подменю выберите команду «надстройки» и нажмите клавишу Enter.
4. После чего на экран будет выведено диалоговое окно «надстройки»; в нем выберите «Пакет анализа» и нажмите кнопку ОК.

Если перечисленная выше последовательность действий выполнена верно, то в меню соответствующем опции **Сервис** добавится команда **Анализ данных**.

Надстройка «Анализ данных» содержит основные инструменты и функции для выполнения анализа данных, в том числе: научных, экономических и т.д.

Инструмент Пакета анализа «Генерация случайных чисел»

С помощью инструмента «генерация случайных чисел» происходит заполнение интервала независимыми случайными числами.

Если в диалоговом окне «Генерация случайных чисел» указан параметр «число переменных», то можно получить многомерную выборку. Для получения многомерной выборки необходимо ввести число столбцов в выходной таб-

лице. Если не будет указано количество столбцов, то Excel заполнит все столбцы.

Параметр «число случайных чисел» определяет число точек данных, которое будет сгенерировано для каждой из переменных.

В инструменте «Генерация случайных чисел» пакета анализа предусмотрен выбор закона распределения. В состав инструмента «Генерация случайных чисел» входят следующие виды распределения:

- нормальное распределение;
- равномерное распределение;
- дискретное распределение;
- распределение Пуассона;
- биномиальное распределение.

Выбор закона распределения в диалоговом окне «Генерация случайных чисел» осуществляется из ниспадающего списка «распределение». Каждый из законов распределения характеризуется набором параметров.

Для формирования последовательности случайных чисел в пакете MathCAD используются встроенные функции, объединенные в раздел Random Numbers (случайные числа):

- $\text{rbeta}(k, s_1, s_2)$ – моделирует β - распределение;
- $\text{rbinom}(k, n, p)$ – формирует последовательность случайных чисел, подчиняющихся биномиальному закону распределения;
- $\text{rcauchy}(k, \ell, s)$ – распределение Коши;
- $\text{rchisq}(k, d)$ – χ^2 - распределение;
- $\text{rexp}(k, r)$ – моделирование последовательности случайных чисел, подчиняющихся экспоненциальному закону распределения;
- $\text{rF}(k, m, n)$ – распределение Фишера (F – Фишера);
- $\text{rgamma}(k, s)$ – моделирование γ - распределения;
- $\text{rgeom}(k, p)$ – формирование последовательности случайных чисел, распределенных по геометрическому закону;
- $\text{rlnorm}(k, \mu, \sigma)$ – логнормальное распределение;

- $\text{rlogis}(k, \ell, s)$ – логистическое распределение;
- $\text{rbinom}(k, n, p)$ – формирует последовательность случайных чисел, соответствующую отрицательному биномиальному распределению;
- $\text{rnorm}(k, \mu, \sigma)$ – моделирует последовательность случайных чисел, подчиняющихся нормальному закону распределения;
- $\text{rpois}(k, \lambda)$ – распределение Пуассона;
- $\text{rt}(k, t)$ – распределение Стьюдента;
- $\text{runif}(k, a, b)$ – формирование последовательности случайных чисел, подчиняющихся равномерному закону распределения;
- $\text{rweibull}(k, s)$ – распределение Вейбулла.

Все перечисленные выше функции для построения последовательности случайных чисел можно вводить также с клавиатуры.

В лабораторной работе рассмотрим два наиболее часто используемых вида распределения: нормальное и равномерное.

Равномерное распределение характеризуется верхней и нижней границей. Вероятность попадания переменной в отрезок фиксированной длины зависит только от длины отрезка и не зависит от его расположения на интервале. Обычно для этого распределения на практике берут интервал $[0, 1]$.

Нормальное распределение характеризуется средним значением и стандартным отклонением. На практике обычно принимают среднее значение равное 0 , а стандартное отклонение равное 1 .

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б)..

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Активизируйте команду «Анализ данных» в пункте меню «сервис».

2. В диалоговом окне «Анализ данных» выберите инструмент «Генерация случайных чисел» и нажмите кнопку ОК. После этого на экран будет выведено диалоговое окно «Генерация случайных чисел».

3. В нем в поле ввода «число переменных» введите – 1, а в поле ввода «число случайных чисел» – количество случайных чисел (согласно варианту).

4. При помощи мыши перейдите в поле ввода «распределение». Для задания типа распределения нажмите на кнопку прокрутки и выберите «нормальное», после чего в окне диалога появится область для ввода его параметров.

5. В области «параметры» в поле ввода «среднее» введите – 0, а в поле «стандартное отклонение» – 1.

6. Перейдите с помощью мыши в область «параметры вывода».

7. Затем перейдите на поле «выходной интервал». Активизируйте ячейку, начиная с которой будет выведена сформированная последовательность случайных чисел. При этом имя этой ячейки автоматически перенесено в поле «выходной интервал».

8. Для завершения формирования последовательности случайных чисел нажмите кнопку ОК.

Задание 1. Используя пакет анализа, сформируйте последовательность случайных чисел, распределенных по нормальному закону распределения.

Методика и порядок выполнения задания 2

1. Войдите в главное меню Excel, активизируйте опцию **Сервис**.

2. В ней выберите команду **Анализ данных**.

3. В диалоговом окне «анализ данных» выберите инструмент «генерация случайных чисел» и нажмите кнопку ОК. После этого на экран будет выведено диалоговое окно «генерация случайных чисел».

4. В нем в поле ввода «число переменных» введите – 1, а в поле ввода «число случайных чисел» – количество случайных чисел (согласно варианту).

5. При помощи мыши перейдите в поле ввода «распределение». Для задания типа распределения нажмите на кнопку прокрутки и выберите «равномерное», после чего в окне диалога появится область для ввода его параметров.

6. После выбора закона распределения введите основные его характеристики. Для равномерного распределения – это отрезок, которому будет принадлежать вся последовательность.

7. Затем укажите выходной интервал, определяющий место расположения данной последовательности на экране.

8. Для завершения формирования последовательности случайных чисел нажмите кнопку ОК.

Задание 2. Сформируйте последовательность случайных чисел, распределенных по равномерному закону распределения.

Методика и порядок выполнения задания 3

1. Запустите систему MathCAD и проверьте, установлен ли в ней режим автоматических вычислений.

2. Присвойте произвольным переменным функции: $\text{rnorm}(k, \mu, \sigma)$ и $\text{runif}(k, a, b)$. Для нормального закона распределения величину μ , считать равной нулю, а $\sigma = 1$, при формировании последовательности, подчиняющейся равномерному закону распределения, выбрать интервал $[0, 1]$.

3. Выведите полученные последовательности случайных чисел на экран.

Задание №3. Используя пакет MathCAD, сформировать последовательности случайных чисел, подчиняющихся нормальному и равномерному закону распределения.

Пример выполнения задания 3

Используя пакет MathCAD, сформировать последовательности случайных чисел, подчиняющихся нормальному и равномерному закону распределения, если эта последовательность состоит из 10 чисел.

$x := \text{rnorm}(10, 0, 1)$

Формирование последовательности случайных чисел X , подчиняющихся нормальному закону распределения

	0
0	-0.439
1	-0.679
2	-0.473
3	-0.951
4	-1.686
5	0.044
6	-0.121
7	0.556
8	2.192
9	0.809

$y := \text{runif}(10, 0, 1)$

	0
0	$8.817 \cdot 10^{-3}$
1	0.276
2	0.588
3	0.838
4	0.485
5	0.744
6	0.458
7	0.744
8	0.599
9	0.735

Моделирование последовательности случайных чисел Y , подчиняющихся равномерному закону распределения

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент. Варианты заданий приведены в таблице 4.1.

Таблица 4.1

№	N	№	N	№	N	№	N
1	15	6	40	11	65	16	90
2	20	7	45	12	70	17	95
3	25	8	50	13	75	18	100
4	30	9	55	14	80	19	105
5	35	10	60	15	85	20	110

В таблице 4.1 № – номер варианта, а N – количество чисел в моделируемой последовательности.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Каким образом можно подключить настройку «Анализ данных»?
2. Перечислите основные виды распределений, которые можно смоделировать в пакете «Анализ данных».
3. Алгоритм формирования последовательности случайных чисел, удовлетворяющей нормальному закону распределения, в пакете «Анализ данных».
4. Последовательность действий в пакете «Анализ данных» при формировании последовательности случайных чисел, соответствующей равномерному закону распределения.
5. В каком из разделов функций системы MathCAD сгруппированы основные функции для формирования последовательности случайных чисел?
6. Перечислите основные функции для формирования последовательности случайных чисел.
7. Последовательность действий при формировании последовательности случайных чисел в системе MathCAD.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 5

Применение инструментальных средств для решения основных задач математической статистики

Цель и содержание: познакомить студентов с основными задачами математической статистики, рассмотреть реализацию этих задач в MathCAD и Excel.

Теоретическое обоснование

Главной задачей математической статистики является анализ и извлечение максимума информации из полученных опытным путем (или эмпирических) данных.

Основными понятиями математической статистики являются **генеральная совокупность** и **выборка**.

Генеральная совокупность представляет собой вероятностное пространство с определенной на нем случайной величиной ξ .

Функцию распределения ξ $F_{\xi}(x)$ называют *теоретической функцией распределения*.

Если над случайной величиной ξ провести ряд экспериментов, то в результате будет получен ряд выборочных значений. Например, x_i , где $i = \overline{1, n}$. Тогда элементы x_i будем называть выборочными, а саму последовательность этих значений *выборкой*.

Первоначальная обработка статистической информации заключается в отыскании максимального и минимального элементов в выборке и построении вариационного ряда.

Для более полной обработки статистической информации необходимо соблюдение следующих этапов: группировка и ее графическое представление.

Группировка некоторой выборки объема n выполняется следующим образом:

- промежуток $[x_{\min}, x_{\max}]$ разбивается на заданное количество интервалов группировки одинаковой длины;
- вычисляется число выборочных значений n_j , попавших в j – й интервал группировки.

Следует отметить, что каждый из интервалов группировки будут описываться своей левой и правой границами, то есть $\Delta_j = (a_j, b_j)$, и числом выборочных значений попавших в него.

Основными формами графического представления выборки являются гистограмма и полигон частот.

Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1; n_1)$, $(x_2; n_2)$, ..., $(x_k; n_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i , а на оси ординат – соответствующие им частоты n_i . Точки $(x_i; n_i)$ соединяют отрезками прямых и получают полигон частот.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки $(x_1; W_1)$, $(x_2; W_2)$, ..., $(x_k; W_k)$. Для построения полигона относительных частот на оси абсцисс откладывают варианты x_i , а на оси ординат – соответствующие им относительные частоты W_i . Точки $(x_i; W_i)$ соединяют отрезками прямых, получают полигон относительных частот.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высоты равны отношению n_i/h (плотность частоты).

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной h , а высоты равны отношению W_i/h (плотность относительной частоты).

Для построения гистограммы частот и относительных частот в пакете MathCAD используется функция $hist(\Delta, A)$.

При первичной обработке статистической информации следует соблюдать следующие правила:

- 1) перед этапом группировки следует упорядочить выборку по возрастанию, то есть построить вариационный ряд;
- 2) выбрать число интервалов группировки случайной величины, причем следует учитывать тот факт, что наилучшие результаты будут получены, если их число находится в пределах от 10 до 20;
- 3) для наибольшей точности проводимого анализа следует использовать интервалы одинаковой длины;
- 4) анализ статистической информации должен охватывать всю область;
- 5) недопустимо использование полуоткрытых промежутков и интервалы группировки не должны накладываться или перекрывать друг друга.

Использование инструмента «Гистограмма» в пакете анализа

Для графического представления результатов статистической выборки в табличном процессоре Excel используется надстройка «Анализ данных».

Инструмент «Гистограмма» вычисляет частоты появления данных. Числовой промежуток между наименьшим и наибольшим значениями делится на равные интервалы, образуя интервал карманов.

Если в области окна «Гистограмма» установлен флажок в поле «парето», то на экран будет выведена отсортированная гистограмма. В этой гистограмме значения расположены по возрастанию значений данных.

При установке флажка в поле «интегральный процент» на гистограмме будет выведен график интегрального процента.

При установке флажка «вывод графика» на экран будет выведено графическое представление гистограммы.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Выполните построение последовательности случайных чисел, распределенных по заданному закону распределения. Для данного задания по нормальному и равномерному.

2. Активизируйте пункт меню «Сервис»; в нем выберите команду «Анализ данных».

3. В диалоговом окне «Анализ данных» выберите инструмент «гистограмма».

4. В диалоговом окне «гистограмма» вначале заполните область «входные данные». В поле «входной интервал» введите интервал, в котором расположена последовательность случайных чисел.

5. В поле «интервал карманов» введите карман. Интервал карманов формируется в отдельном столбце. Например, в качестве кармана можно рассматривать последовательность чисел: -3, -2, -1, 0, 1, 2, 3.

6. Затем с помощью мыши перейдите в область окна «параметры вывода».

7. В поле «выходной интервал» укажите ячейку, с которой начнется вывод результатов.

8. Перейдите с помощью мыши в поле «вывод графика» и нажмите кнопку ОК. Построение гистограммы завершено.

Задание 1. Используя инструмент «Гистограмма» надстройки «Анализ данных», постройте гистограмму частот для последовательностей случайных чисел, распределенных по нормальному и равномерному закону.

Методика и порядок выполнения заданий 2 и 3

1. Проверьте, установлен ли в системе MathCAD режим автоматических вычислений.

2. Задайте исходные статистические данные, представив их в виде матрицы.

3. Найдите наибольший и наименьший элементы в выборках, соответствующих признакам X и Y, и отсортируйте выборки.

4. Вычислите размах выборки.

5. Введите количество интервалов разбиения, и рассчитайте длину каждого из интервалов.

6. Вычислите значения длин интервалов.

7. Используя функцию $\text{hist}(\Delta, A)$, постройте гистограмму частот.

Задание 2. Используя пакет MathCAD, постройте гистограмму частот для последовательностей случайных чисел, распределенных по нормальному и равномерному законам распределения.

Пример выполнения задания 2

Смоделируйте последовательность случайных чисел, если количество элементов в ней равно 10, и постройте для этого распределения гистограмму частот.

$$\xi := \text{morm}(10,0,1)\xi =$$

	0
0	-0.439
1	-0.679
2	-0.473
3	-0.951
4	-1.686
5	0.044
6	-0.121
7	0.556
8	2.192
9	0.809

$$\begin{aligned} X_{\min} &:= \min(\xi) & X_{\max} &:= \max(\xi) \\ X_{\max} &= 2.192 & X_{\min} &= -1.686 \\ R &:= X_{\max} - X_{\min} \\ R &= 3.877 \end{aligned}$$

$$\xi := \text{sort}(\xi) \quad \xi =$$

	0
0	-1.686
1	-0.951
2	-0.679
3	-0.473
4	-0.439
5	-0.121
6	0.044
7	0.556
8	0.809
9	2.192

$$\begin{aligned} m &:= 5 & \Delta\xi &:= \frac{R}{m} \\ \Delta\xi &= 0.775 & j &:= 0..m \\ k &:= 0..m-1 \end{aligned}$$

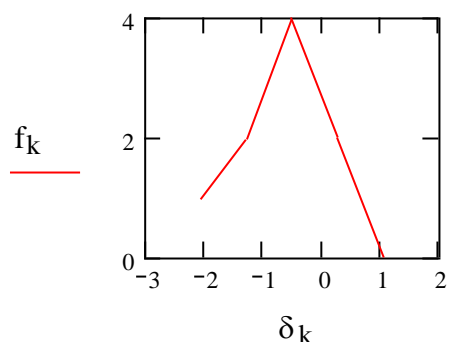
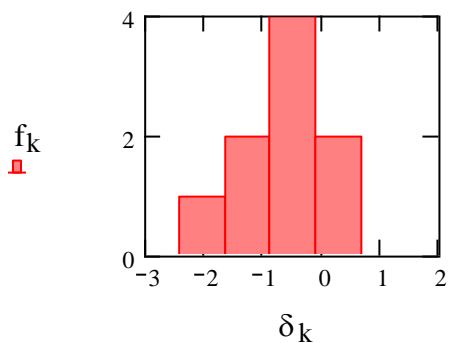
$$\delta_j := X_{\min} + \frac{\Delta\xi}{2} \cdot (2 \cdot j - 1)$$

$$f := \text{hist}(\delta, \xi) \quad f = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 2 \\ 0 \end{pmatrix}$$

Формирование последовательности случайных чисел, распределенных по нормальному закону распределения

Поиск наибольшего и наименьшего элементов в выборке, вычисление ее размаха, то есть разности между минимальным и максимальными элементами выборки
Сортировка полученной выборки

Построение гистограммы частот



Задание 3. С помощью пакета MathCAD выполните построение гистограммы для статистической информации, представленной в таблице 1.

Методика и порядок выполнения работы

1. Перейдите в системе MathCAD в режим автоматического выполнения вычислений.
2. Смоделируйте последовательность случайных чисел, распределенных по нормальному или равномерному законам распределения.
3. Сформируйте на основании данной выборки вариационный ряд.
4. Определите максимальный, минимальный элементы и размах полученной выборки.
5. Задайте число интервалов группировки, и вычислите длину каждого из них.
6. Вычислите середину каждого из интервалов группировки, а затем сформируйте вектор – столбец, содержащий полученные значения.
7. При помощи функции hist рассчитайте значения частот для каждого из интервалов группировки, а потом представьте найденные значения в виде матрицы – столбца.

8. Сформируйте вектор – столбец, значениями в котором являются накопленные частоты.

9. Постройте гистограмму, полигон частот, полигон накопленных и полигон относительных накопленных частот.

Задание 4. Смоделируйте последовательность случайных чисел, распределенных по нормальному закону распределения. Вычислите максимальный и минимальный элементы, размах данной выборки. Выполните группировку для значений $m = 10, m = 20, m = 50$. Постройте для каждой из группировок гистограмму, полигон частот и полигон накопленных частот.

Пример выполнения задания 4

Смоделируйте последовательность, состоящую из 200 случайных чисел, распределенных по нормальному закону распределения. Вычислите максимальный и минимальный элементы, размах данной выборки. Выполните группировку для значений $m = 10, m = 20, m = 50$. Постройте для каждой из группировок гистограмму, полигон частот и полигон накопленных частот.

origin := 1

$\xi := \text{morm}(200, 100, 15)$

$\xi =$

	0
0	93.415
1	89.809
2	92.901
3	85.728
4	74.715
5	100.653
6	98.19
7	108.346
8	132.877
9	112.131
10	114.777
11	112.933
12	113.733
13	110.095
14	84.335
15	101.036

Формирование последовательности случайных чисел, распределенных по нормальному закону

$x_{\max} := \max(\xi)$

$x_{\max} = 145.722$ $x_{\min} := \min(\xi)$

$x_{\min} = 56.487$ $R := x_{\max} - x_{\min}$

$R = 89.235$

Расчет основных статистических функций

	0
0	56.487
1	57.394
2	59.111
3	65.588
4	67.643
5	69.532
6	71.251
7	72.675
8	74.593
9	74.715
10	75.778
11	76.362
12	78.567
13	78.832
14	79.329
15	80.12

$$\xi := \text{sort}(\xi)$$

$$\xi =$$

Построение вариационного ряда для элементов полученной выборки

$$n := 200$$

$$m := 10$$

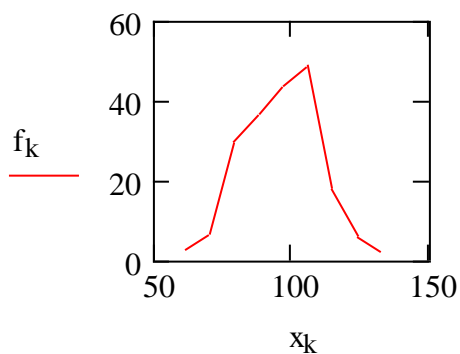
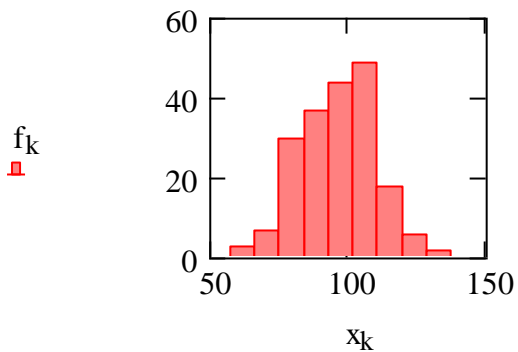
$$\Delta := \frac{R}{m}$$

Построение гистограммы частот для 10 интервалов группировки.

$$\Delta = 8.924 \quad j := 1..m$$

$$k := 1..m-1 \quad x_j := x_{\min} + \frac{\Delta}{2} \cdot (2 \cdot j - 1)$$

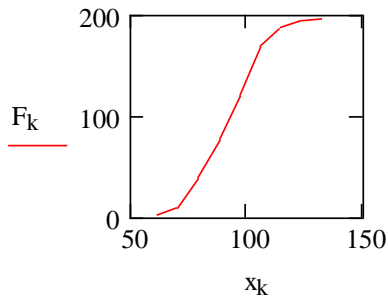
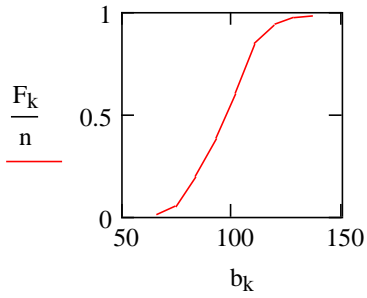
$$f := \text{hist}(x, \xi)$$



$$a_j := x_{\min} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta$$

$$\Delta = 8.924 \quad F_k := \sum_{j=1}^k f_j$$

Построение полигона частот для $m = 10$



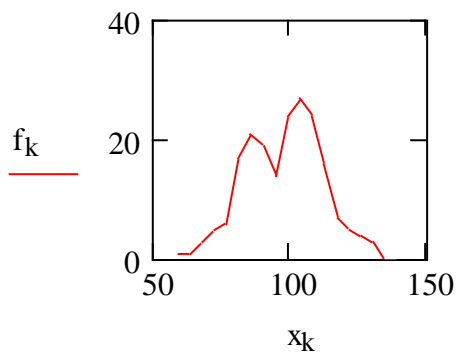
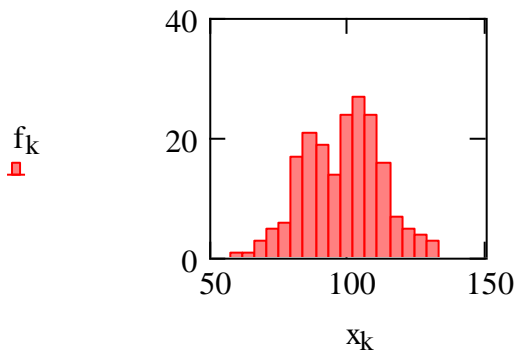
$$m := 20 \quad \Delta := \frac{R}{m} \quad \Delta = 4.462$$

$$j := 1..m \quad k := 1..m-1$$

$$x_j := x_{\min} + \frac{\Delta}{2} \cdot (2 \cdot j - 1)$$

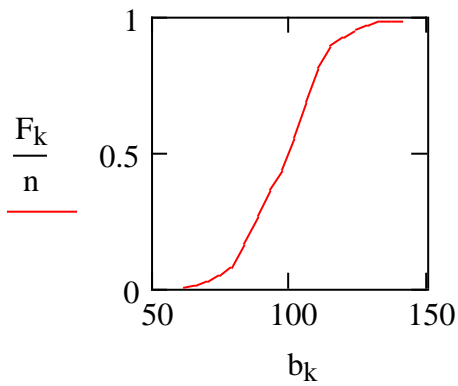
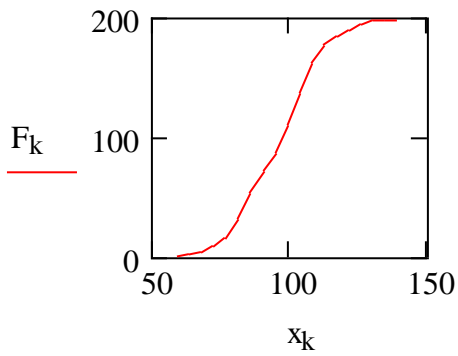
$$f := \text{hist}(x, \xi)$$

построение гистограммы частот и полигона частот и относительных частот для 20 отрезков группировки данных



$$a_j := x_{\min} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta$$

$$F_k := \sum_{j=1}^k f_j$$



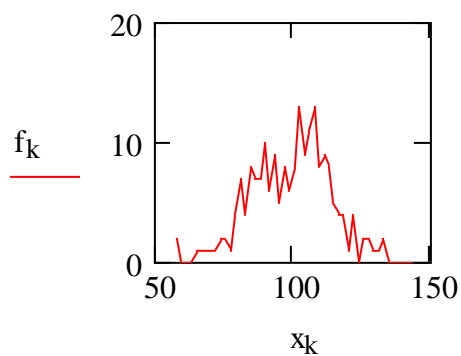
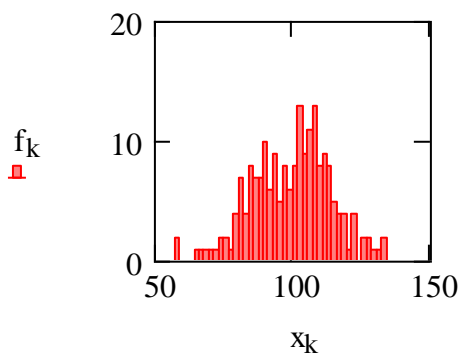
$$m := 50 \quad \Delta := \frac{R}{m} \quad \Delta = 1.785$$

$$j := 1..m \quad k := 1..m-1$$

$$x_j := \text{xmin} + \frac{\Delta}{2} \cdot (2 \cdot j - 1)$$

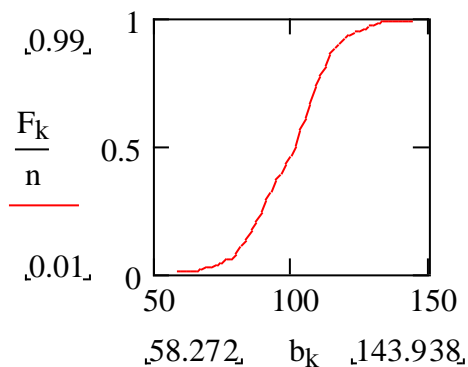
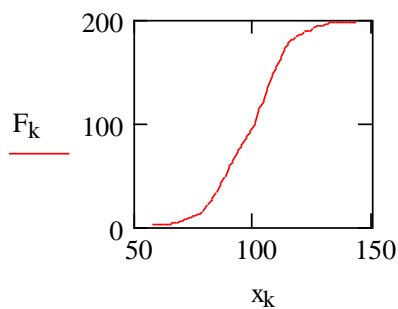
$$f := \text{hist}(x, \xi)$$

Построение гистограммы, полигона частот и полигона относительных частот для 50 интервалов группировки выборки



$$a_j := \text{xmin} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta$$

$$F_k := \sum_{j=1}^k f_j$$



Варианты заданий

Вариант для выполнения заданий, приведенный ниже, определяется номером персонального компьютера студента.

Задания № 1 и №2 выполняются по исходным лабораторной работы № 4.

В качестве варианта задания № 3 следует выбрать соответствующий вариант из лабораторной работы №2.

При выполнении задания № 4 следует сначала сформировать последовательность, состоящую из 250 случайных чисел, распределенных по нормальному закону распределения: $\mu = 150$, $\sigma = 10$. Все расчеты произвести для 100 элементов этой последовательности. Выбор чисел начинается с порядкового номера числа, указанного в таблице.

Таблица 5.2

№	n	№	n	№	n	№	n	№	n
1	10	5	50	9	90	13	95	17	135
2	20	6	60	10	170	14	105	18	145
3	30	7	70	11	75	15	115	19	155
4	40	8	80	12	85	16	125	20	165

В таблице 5.2 № – номер варианта, n – порядковый номер числа, с которого начинается отсчет.

Примечание. Для построения гистограммы частот, полигона частот и полигона относительных частот представить в виде отдельной матрицы, состоящей из ста строк и одного столбца. Эту матрицу следует вводить вручную, а затем уже над ее элементами выполнить последовательность действий, описывающую процесс построения гистограммы частот, полигона частот и полигона относительных частот.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Сформулируйте основную задачу математической статистики.
2. Дайте определения основных понятий математической статистики.
3. Алгоритм выполнения группировки выборочных данных.
4. Основные формы графического представления выборки. Определения.
5. Перечислите основные правила первичной обработки статистической информации.
6. Использование табличного процессора Excel для графического представления результатов выборки.
7. Алгоритм построения гистограммы и полигона частот в пакете MathCAD.
8. Последовательность действий при построении гистограммы в табличном процессоре Excel.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 6

Применение инструментальных средств для расчета числовых характеристик выборки

Цель и содержание: изучить основные функции для расчета числовых характеристик выборки, показать возможность расчета этих характеристик при помощи различных информационных технологий (пакет MathCAD, табличный процессор Excel).

Теоретическое обоснование

Каждая из выборок характеризуется следующими показателями: положения, разброса, асимметрии.

Показатели положения выборки описываются характеристиками.

Средним значением выборки будем называть величину, значение которой можно найти по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i . \quad (1)$$

Величина, являющаяся решением уравнения $F_n(x) = p$, называется **выборочной квантилью уровня p**.

Если $p = 0.5$, то выборочную квантиль называют **выборочной медианой**. Выборочная медиана разбивает выборку пополам. Для расчета ее значений используют формулу:

$$m_e = \begin{cases} x_{k+1} & k = 2 \cdot n + 1 \\ \frac{x_k + x_{k+1}}{2} & k = 2 \cdot n \end{cases} . \quad (2)$$

Для вычисления показателей положения в пакете MathCAD используются функции: mean(A), median(A), max(A), min(A), где A – сформированная выборка.

Аналогами этих функций в табличном процессоре Excel являются функции: СРЗНАЧ(диапазон _ ячеек), МЕДИАНА (диапазон _ ячеек), МИН(диапазон _ ячеек), МАХ(диапазон _ ячеек).

К **показателям разброса** сформированной выборки относятся выборочная дисперсия (дисперсия выборки), стандартное отклонение, размах выборки и выборочный эксцесс.

Выборочной дисперсией будем называть величину, значения которой может быть найдено по формуле:

$$D_s = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2, \quad (3)$$

где x_i – значение элемента выборки, \bar{x} – среднее значение.

Чаще всего при решении задач математической статистики используют следующую формулу для расчета значений выборочной дисперсии:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2. \quad (4)$$

Стандартное отклонение – это величина, значение которой равно

$$\sigma = \sqrt{D_s}. \quad (5)$$

Размахом выборки называется величина, равная разности значений наибольшего и наименьшего элементов выборки:

$$R = x_{\max} - x_{\min}. \quad (6)$$

Для расчета значения **выборочного эксцесса** следует вначале вычислить величину выборочного центрального момента четвертого порядка, используя формулу $\mu_4 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^4$, а только потом определить значение выборочного эксцесса по формуле:

$$E = \frac{\mu_4}{s^2} - 3. \quad (7)$$

В пакете MathCAD для расчета выборочной дисперсии используется функция var, в Excel для расчета этих характеристик применяют функции: ДИСП(диапазон _ ячеек), СТАНДОТКЛ(диапазон _ ячеек).

К показателям асимметрии при исследовании выборки относится **коэффициент асимметрии**, значение которого можно найти по формуле:

$$a = \frac{\mu_3}{\sigma^3}, \quad (8)$$

где μ_3 – выборочный центральный момент третьего порядка $\mu_3 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^3$.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Загрузите табличный процессор Excel.
2. Сформируйте последовательности случайных чисел, состоящие из заранее определенного числа элементов и подчиняющиеся одному из известных законов распределения.
3. Используя встроенные функции Excel, рассчитайте значения соответствующих характеристик выборки.

Задание 1. Используя основные функции табличного процессора Excel, найдите значения основных числовых характеристик выборки, при условии, что выборка является последовательностью случайных чисел, подчиняющейся:

- а) нормальному закону распределения;
- б) равномерному закону распределения.

Методика и порядок выполнения заданий 2

1. Проверьте, установлен ли в MathCAD режим автоматического выполнения вычислений.
2. Сформируйте последовательность случайных чисел, подчиняющуюся нормальному и равномерному законам распределения.
3. Отсортируйте полученную выборку.
4. Найдите для полученной выборки максимальный, минимальный элемент, размах полученной выборки.
5. Используя встроенные статистические функции пакета MathCAD и расчетные формулы из раздела «Теоретическое обоснование» вычислите значения основных числовых характеристик выборки: выборочное среднее, медиану,

выборочную дисперсию, стандартное отклонение, выборочные моменты 3 и 4 – го порядков, выборочный эксцесс и коэффициент асимметрии

Задание 2. Используя пакет MathCAD, вычислите значения основных числовых характеристик выборки, при условии, что выборка является последовательностью случайных чисел, подчиняющихся:

- а) нормальному закону распределения;
- б) равномерному закону распределения.

Пример выполнения задания 2

Смоделируйте последовательность, состоящую из 200 случайных чисел, распределенных по нормальному закону распределения. Эта последовательность является некоторой выборкой. Для данной выборки рассчитайте значения основных числовых характеристик.

ORIGIN := 1

$\xi := \text{rnorm}(200, 100, 10)$

$\xi =$

	1
1	95.61
2	93.206
3	95.267
4	90.485
5	83.143
6	100.435
7	98.794
8	105.564
9	121.918
10	108.087
11	109.851
12	108.622
13	109.156
14	106.73
15	89.557
16	100.691

Моделирование последовательности случайных чисел, подчиняющихся нормальному закону распределения

$$\xi := \text{sort}(\xi)$$

$$\xi =$$

	1
1	70.991
2	71.596
3	72.741
4	77.059
5	78.429
6	79.688
7	80.834
8	81.784
9	83.062
10	83.143
11	83.852
12	84.241
13	85.711
14	85.888
15	86.219
16	86.747

$$X_{\max} := \max(\xi) \quad X_{\max} = 130.482$$

$$X_{\min} := \min(\xi) \quad X_{\min} = 70.991$$

$$\text{mean} := \text{mean}(\xi) \quad \text{mean} = 99.829$$

$$\text{median} := \text{median}(\xi) \quad \text{median} = 100.892$$

$$\text{var} := \text{var}(\xi) \quad \text{var} = 98.787$$

$$R := X_{\max} - X_{\min} \quad R = 59.49$$

$$n := 200$$

$$s_2 := \frac{n}{n-1} \cdot \text{var} \quad s_2 = 99.283$$

$$o := \sqrt{s_2} \quad o = 9.964$$

$$\mu_3 := \frac{1}{n} \cdot \sum_{i=1}^n (\xi_i - \text{mean})^3$$

$$\mu_3 = -150.009$$

$$\mu_4 := \frac{1}{n} \cdot \sum_{i=1}^n (\xi_i - \text{mean})^4$$

$$\mu_4 = 3.183 \times 10^4$$

$$E := \frac{\mu_4}{(s_2)^2} - 3 \quad E = 0.229$$

$$\alpha := \frac{\mu_3}{o^3} \quad \alpha = -0.152$$

Построение
вариацион-
ного ряда по
сформиро-
ванной вы-
борке

Расчет зна-
чений чис-
ловых ха-
рактеристик,
описываю-
щих показа-
тели разбро-
са

Вычисление
значений
числовых
характери-
стик, описы-
вающих по-
казатели по-
ложения

Расчет ко-
эффициента
асимметрии

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Выборками для заданий 1 и 2 считать последовательности случайных чисел, сформированные в лабораторных работах № 4 и № 6.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Какими показателями характеризуется статистическая выборка?
2. Какие параметры описывают показатели положения?
3. Перечислите и дайте определение основных числовых характеристик, описывающих показатели разброса.
4. Показатель асимметрии.
5. Алгоритм расчета основных числовых характеристик выборки в табличном процессоре Excel.
6. Последовательность действий при решении задачи отыскания значений числовых характеристик выборки в пакете MathCAD.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 7

Использование инструмента «Описательная статистика» для обработки экспериментальных данных

Цель и содержание: изучить инструмент «описательная статистика» и рассмотреть его применение для обработки экономической информации.

Теоретическое обоснование

Исходные данные для анализа могут быть представлены на рабочем листе в виде списка значений. Для идентификации массива значений используются названия столбцов – метки, а также создаются именованные блоки.

После выполнения команды «Анализ данных» из пункта меню «сервис» вызывается диалоговое окно «Анализ данных», в котором необходимо выбрать режим «описательная статистика». В одноименном диалоговом окне задаются установки для анализа данных.

Входной интервал – блок ячеек, содержащий значения исследуемого экономического показателя. Можно указать один или несколько смежных столбцов/строк ячеек, содержащих числовые данные для анализа, которые находятся в произвольном месте (этот же, другой лист этой или любой открытой рабочей книги).

Группирование – определяет ориентацию блока исходных данных на рабочем листе (по строкам или по столбцам). Если для параметра «группирование» выбран переключатель «по столбцам/по строкам», то MS Excel запишет двухколоночную таблицу статистических данных для каждого столбца/строки данных во входном интервале.

Метки – наличие имен в блоке ячеек.

Вывод результатов описательной статистики осуществляется по месту указания (блок ячеек на листе с исходными данными, другой лист или новая рабочая книга).

Вывод описательной статистики осуществляется на тот же рабочий лист, тогда указывается начальная ячейка выходного диапазона; на новый лист, для которого можно указать имя, или в новую рабочую книгу. Содержание вывода определяется установками.

Флажок «итоговая статистика» предназначен для формирования данных в выходной таблице. «Итоговая статистика» обеспечивает полный вывод показателей описательной статистики.

Для включения в выходную таблицу строки уровень надежности установите флажок в поле «уровень надежности». В поле ввода «уровень надежности»

вводится требуемое значение. В нем указывается процент надежности данных для вычисления доверительного интервала.

Установите флажок k - й наименьший, если хотите включить в выходную таблицу строку для k - го наименьшего значения для каждого интервала данных. В строку для поля ввода k - й наименьший введите число k_1 . k -й наименьший указывает на порядковый номер наименьшего после минимального значения.

Установите флажок k - й наибольший, если хотите включить в выходную таблицу строку для k - го наибольшего значения для каждого интервала данных. В строку для поля ввода k - й наибольший введите число k_2 . k - й наибольший указывает на порядковый номер наибольшего после максимального значения.

Числа, которые будут получены, продемонстрируют близость основных параметров выборок к заданным законам распределения.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б)..

Методика и порядок выполнения работы

1. Загрузите табличный процессор Excel.
2. При помощи инструмента анализа «Генерация случайных чисел» сформируйте две выборки. Количество случайных чисел определяется вариантом индивидуального задания. В столбце А разместите последовательность случайных чисел, подчиняющихся нормальному закону распределения, а в столбце В – равномерному закону.
3. Присвойте этим столбцам имена. Вызовите окно диалога инструмента «Описательная статистика».
4. Начинаем вводить параметры. Введите входной интервал. Входной интервал представляет собой совокупность ячеек, содержащих случайные числа.
5. Для параметра «группирование» выберите опцию «по столбцам».

6. Установите флажок «метки» в первой строке.
7. Установите флажок «уровень надежности» и в поле ввода «уровень надежности» введите значение 95%.
8. В поле ввода k - й наибольший введите число 1.
9. В поле ввода k - й наименьший введите число 1.
10. В поле ввода «выходной интервал» введите адрес ячейки, начиная с которой будут выведены результаты обработки статистической информации при помощи инструмента «Описательная статистика».
11. Установите флажок «итоговая статистика».
12. Параметры окна диалога «описательная статистика» установлены.
13. Нажмите кнопку ОК. На экране появится таблица описательной статистики для двух выборок.

Задание. Примените инструмент «описательная статистика» к двум последовательностям случайных чисел, распределенным в соответствии с нормальным и равномерным законами распределения.

Варианты заданий

Вариант для выполнения задания, приведенного в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

В качестве варианта задания к лабораторной работе следует выбрать соответствующий вариант из лабораторной работы № 4.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Каким образом можно произвести загрузку инструмента «Описательная статистика»?
2. Перечислите и охарактеризуйте основные параметры «Описательная статистика».
3. Порядок использования инструмента «Описательная статистика».

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 8

Использование MathCAD для оценки функции распределения выборки

Цель и содержание: изучить методы оценки эмпирической функции распределения (функции распределения выборки), рассмотреть решение этой задачи в пакете MathCAD.

Теоретическое обоснование

Пусть X – некоторое статистическое распределение количественного признака для исследуемого экономического показателя.

Эмпирической функцией распределения (или функцией распределения выборки) будем называть такую функцию $F^*(x)$, которая для каждого значения признака x определяет относительную частоту события $X < x$:

$$F^*(x) = \frac{n_x}{N}. \quad (1)$$

Если распределение количественного признака X представляет собой генеральную совокупность, то функцию $F(x)$ называют **теоретической функцией распределения**.

Пусть $F^*(x)$ – эмпирическая функция распределения, а $F(x)$ – теоретическая функция распределения. Тогда в качестве отклонения значений теоретической функции от эмпирической будем рассматривать величину:

$$D_n(x) = \sup_x |F^*(x) - F(x)|. \quad (2)$$

Функцию $D_n(x)$ называют **статистикой Колмогорова**. Эта функция также является случайной величиной, и ее распределение не зависит от неизвестной функции распределения $F(x)$, в том случае, если последняя непрерывна. Тогда имеет место **теорема Колмогорова**: если теоретическая функция распределения случайной величины $F(x)$ непрерывна, а $F^*(x)$ – выборочная функция распределения (или функция распределения выборки), тогда при $n \rightarrow \infty$ имеет место равенство:

$$P\left(\sup_x |F^*(x) - F(x)| < \frac{z}{\sqrt{n}}\right) \rightarrow K(z) = \begin{cases} 0, & z \leq 0 \\ \sum_{k=-\infty}^{\infty} (-1)^k \cdot e^{-2k^2 \cdot z^2}, & z > 0 \end{cases}. \quad (3)$$

Функция $K(z)$ является функциональным рядом, значения членов которого можно найти путем табулирования.

Если ε – число, задающее точность вычисления значений функции $K(z)$, то количество членов N в частичной сумме можно найти по формуле:

$$N = \text{int}\left(\frac{1}{z} \cdot \sqrt{\frac{1}{2} \cdot \ln\left(\frac{1}{\varepsilon}\right)}\right) + 1, \quad (4)$$

где функция int задает целую часть числа. В пакете MathCAD для отыскания целой части числа используют функцию floor .

Пусть $p_1 = \alpha$ – вероятность события, которое считается достоверным при вероятности $p_2 = 1 - \alpha$. Решив уравнение $1 - K(z) = \alpha$, найдем значение z_α . Тогда имеет место неравенство

$$F(x) - \frac{z_\alpha}{\sqrt{n}} < F^*(x) < F(x) + \frac{z_\alpha}{\sqrt{n}}. \quad (5)$$

Неравенство (5) имеет место для всех $x \in \mathbb{R}$, вероятность которых близка к величине $1 - \alpha$.

Следовательно, в окрестности функции распределения выборки построен «коридор», в котором расположена теоретическая функция распределения $F(x)$. При возрастании n «ширина» этого коридора стремится к нулю, то есть при $n \rightarrow \infty$ $\ell \rightarrow 0$.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Проверьте, установлен ли в пакете MathCAD режим автоматических вычислений.

2. Сформируйте выборку, предполагая, что она представляет собой последовательность случайных чисел, подчиняющихся нормальному или равномерному закону распределения.

3. Задайте статистику Колмогорова и постройте функцию $K(z)$.

4. Определите значение параметра α .

5. Рассчитайте значение «коридора» для теоретической функции распределения и постройте его график.

Задание 1. Сформируйте выборку, которая представляет собой последовательность случайных чисел, при условии что она подчиняется: а) нормальному, б) равномерному закону распределения

Пример выполнения задания 1

Сформируйте выборку, которая представляет собой последовательность случайных чисел, подчиняющихся нормальному закону распределения. Количество элементов в выборке – 200, $\mu = 100$, $\sigma = 10$. Для полученной выборки постройте 95% - й «коридор» для функции ее распределения

ORIGIN := 1

$\xi := \text{rnorm}(200, 100, 10)$ $\xi =$

	1
1	95.61
2	93.206
3	95.267
4	90.485
5	83.143
6	100.435
7	98.794
8	105.564
9	121.918
10	108.087
11	109.851
12	108.622
13	109.156
14	106.73
15	89.557
16	100.691

Моделирование выборки, как последовательности случайных чисел, распределенных по нормальному закону распределения

$$\xi := \text{sort}(\xi)$$

 $\xi =$

	1
1	70.991
2	71.596
3	72.741
4	77.059
5	78.429
6	79.688
7	80.834
8	81.784
9	83.062
10	83.143
11	83.852
12	84.241
13	85.711
14	85.888
15	86.219
16	86.747

Формирование вариационного ряда

$$X_{\max} := \max(\xi) \quad X_{\max} = 130.482$$

$$X_{\min} := \min(\xi) \quad X_{\min} = 70.991$$

$$R := X_{\max} - X_{\min} \quad R = 59.49$$

$$n := 200 \quad m := 200 \quad \Delta := \frac{R}{m} \quad \Delta = 0.297$$

$$j := 1..m \quad k := 1..m - 1$$

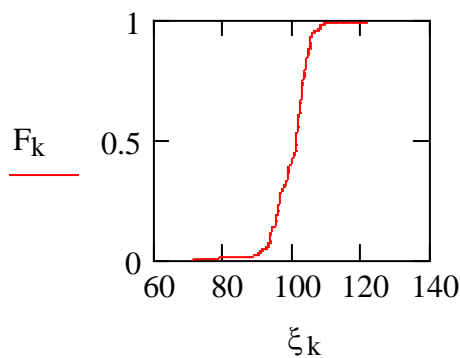
$$x_j := X_{\min} + \frac{\Delta}{2} \cdot (2 \cdot j - 1) \quad f := \text{hist}(x, \xi)$$

$$a_j := X_{\min} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta$$

$$F_k := \sum_{j=1}^k \frac{f_j}{n}$$

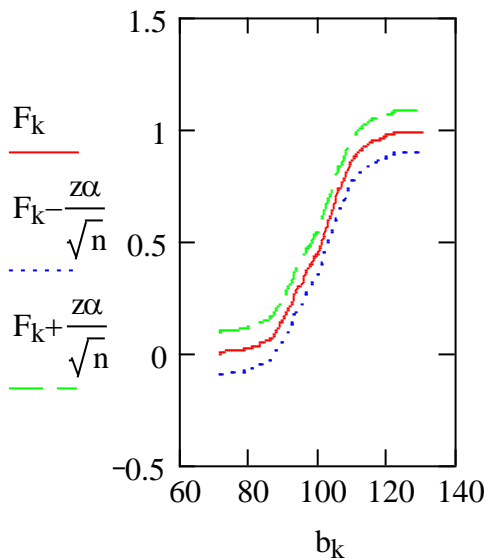
Выполнение первичной обработки статистической информации

Построение гистограммы частот и полигона частот



$$z\alpha := 1.358$$

Формирование и построение «коридора» для теоретической функции распределения.



Методика и порядок выполнения заданий 2

1. Перейдите в пакете MathCAD в режим автоматических вычислений.
2. Перейдите в главное меню Math и активизируйте команду (режим) Optimization.
3. Определите количество элементов, соответствующее числу случайных чисел в последовательности.
4. Для выбранного распределения постройте графики плотности распределения и функции распределения.
5. Для полученного распределения вычислите значения: математического ожидания, дисперсии, среднеквадратичного отклонения, медианы, моментов третьего и четвертого порядка, асимметрии и эксцесса.
6. Сформируйте выборку объема N значений случайной величины, подчиняющейся заданному распределению.
7. Определите как функции переменной N выборочные значения среднего, среднеквадратичного отклонения, центральных моментов третьего и четвертого порядков, асимметрии и эксцесса и вычислите их значения.
8. Для данного распределения постройте гистограмму, полигон частот и график накопленных относительных частот.
9. Для полученного распределения 95 % - й «коридор» для теоретической функции распределения.

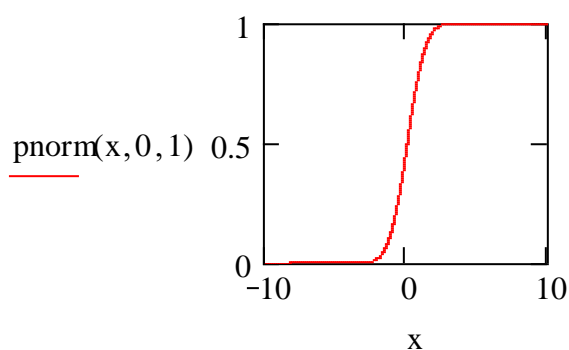
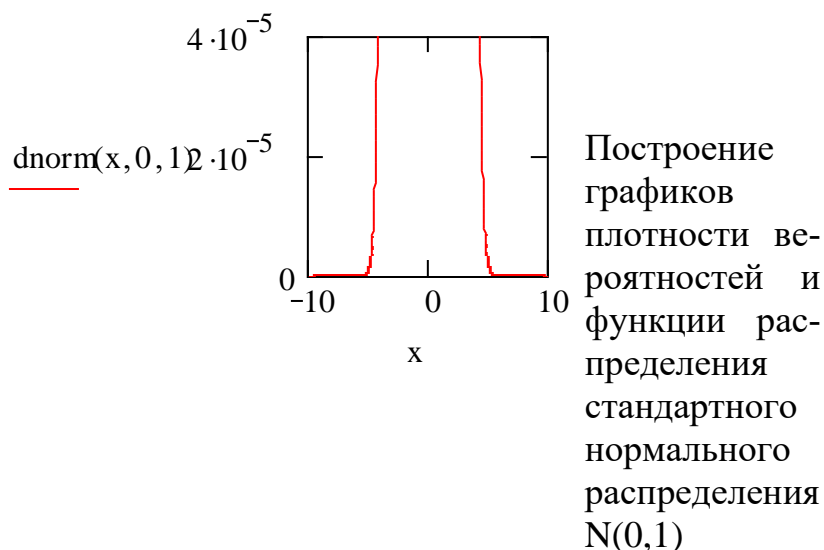
10. Сравните полученные теоретические и выборочные значения параметров.

11. Вычисления повторите для $n = 50$, $n = 100$, $n = 150$.

Задание 2. Смоделируйте выборку значений случайной величины с заданным непрерывным распределением (равномерным или нормальным) и выполните её полный предварительный анализ для указанных значений объема выборки, числа интервалов группировки и доверительной вероятности. Постройте графики плотности вероятностей и функции распределения и сравните их с полученными графиками соответствующих выборочных функций.

Пример выполнения задания 2

Смоделируйте выборку значений случайной величины с нормальным распределением: $n = 150$, $\mu = 0$, $\sigma = 1$. Выполните её полный предварительный анализ для указанных значений объема выборки, числа интервалов группировки и доверительной вероятности. Постройте графики плотности вероятностей, функции распределения и сравните их с полученными графиками соответствующих выборочных функций.



$$M\zeta := 0 \quad D\zeta := 1 \quad \sigma\zeta := 1$$

$$\mu\zeta_2 := D\zeta + (M\zeta)^2 \quad \mu\zeta_2 = 1$$

$$\mu\zeta_3 := \int_{-\infty}^{\infty} x^3 \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{-x^2}{2}\right) dx$$

$$\mu\zeta_3 = 0$$

$$\mu\zeta_4 := \int_{-\infty}^{\infty} x^4 \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{-x^2}{2}\right) dx$$

$$\mu\zeta_4 = 3$$

$$a := \frac{\mu\zeta_3}{\sqrt{(\mu\zeta_2)^3}} \quad a = 0$$

$$E := \frac{\mu\zeta_4}{\left[(\mu\zeta_2)^2\right]} - 3$$

$$E = -3.109 \times 10^{-8}$$

$$\text{median}\zeta := \text{qnorm}(0.5, 0, 1)$$

$$\text{median}\zeta = 0$$

$$\text{ORIGIN} := 1$$

$$\zeta := \text{rnorm}(150, 0, 1)$$

	1
1	-0.439
2	-0.679
3	-0.473
4	-0.951
5	-1.686
6	0.044
7	-0.121
8	0.556
9	2.192
10	0.809
11	0.985
12	0.862
13	0.916
14	0.673
15	-1.044
16	0.069

$\zeta =$

$$n := 150$$

$$\zeta := \text{sort}(\zeta)$$

Определение и вычисление для $N(0,1)$ значений математического ожидания, дисперсии, среднеквадратичного отклонения, моментов второго, третьего, четвертого порядка, асимметрии, эксцесса, медианы

Моделирование последовательности случайных чисел из стандартного нормального распределения

	1
1	-2.901
2	-2.84
3	-2.157
4	-2.031
5	-1.917
6	-1.822
7	-1.694
8	-1.686
9	-1.615
10	-1.429
11	-1.411
12	-1.378
13	-1.325
14	-1.313
15	-1.28
16	-1.252

$\zeta =$

Построение
вариационно-
го ряда для
смоделиро-
ванной вы-
борки

$$X_{\max} := \max(\zeta) \quad X_{\max} = 3.048$$

$$X_{\min} := \min(\zeta) \quad X_{\min} = -2.901$$

$$R := X_{\max} - X_{\min} \quad R = 5.949$$

$$\text{mean} := \text{mean}(\zeta) \quad \text{mean} = -0.109$$

$$\text{var} := \text{var}(\zeta) \quad \text{var} = 0.956$$

$$s_2 := \frac{n}{n-1} \cdot \text{var} \quad s_2 = 0.962$$

$$\sigma := \sqrt{s_2} \quad \sigma = 0.981$$

$$\text{median} := \text{median}(\zeta)$$

$$\text{median} = -7.214 \times 10^{-3}$$

$$\mu_3 := \frac{1}{n} \cdot \sum_{i=1}^n (\zeta_i - \text{mean})^3$$

$$\mu_3 = -9.342 \times 10^{-3}$$

$$\mu_4 := \frac{1}{n} \cdot \sum_{i=1}^n (\zeta_i - \text{mean})^4$$

$$\mu_4 = 3.021 \quad E_1 := \frac{\mu_4}{(s_2)^2}$$

$$E_1 = 3.264$$

$$a_1 := \frac{\mu_3}{\sigma^3} \quad a_1 = -9.9 \times 10^{-3}$$

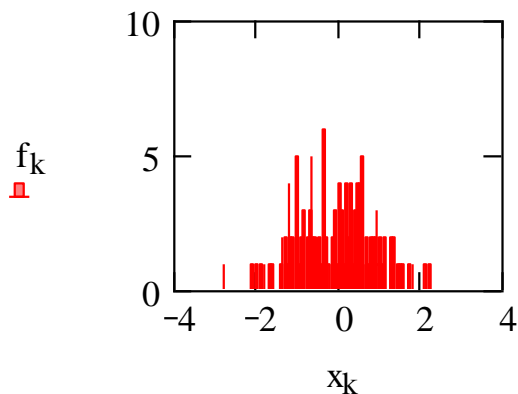
$$m := 150 \quad \Delta := \frac{R}{m} \quad \Delta = 0.04$$

$$j := 1..m \quad k := 1..m-1$$

Первичный
анализ статисти-
ческих
данных и
расчет выбо-
рочных харак-
теристик для
смоделиро-
ванной после-
довательности

$$x_j := X_{\min} + \frac{\Delta}{2} \cdot (2 \cdot j - 1)$$

$$f := \text{hist}(x, \zeta)$$



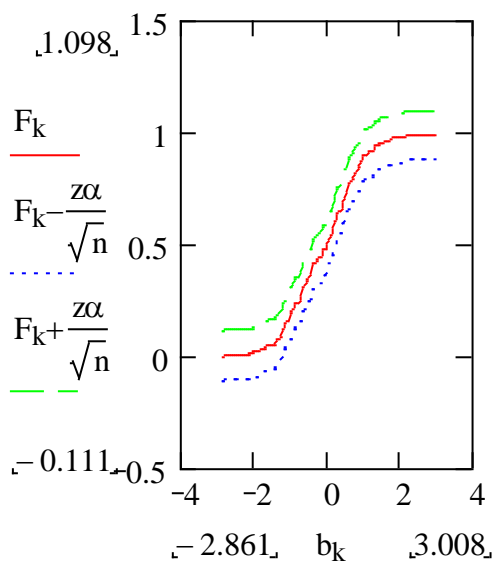
Построение
гистограммы
частот

$$a_j := X_{\min} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta$$

$$F_k := \sum_{j=1}^k \frac{f_j}{n}$$

Построение
эмпирической
функции рас-
пределения

$$z\alpha := 1.358$$



Построение
эмпирической
функции рас-
пределения и
95 % - ого ко-
ридора для
нее

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Выборками для задания 1 считать последовательности случайных чисел, сформированные в лабораторных работах № 4 и № 7. В качестве выборки для

выполнения задания 2 считать последовательность случайных чисел, сформированную в лабораторной работе № 7.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Дайте определения эмпирической и теоретической функций распределения.
2. Чем отличаются эмпирическая и теоретическая функции распределения?
3. Сформулируйте определение статистики Колмогорова и теорему Колмогорова.
4. Порядок построения коридора для теоретической функции распределения.
5. Алгоритм оценки функции распределения в пакете MathCAD.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Список рекомендуемой литературы

Основная литература

1. Карманов, Ф. И. Статистические методы обработки экспериментальных данных. Лабораторный практикум с использованием MathCad: Учебное пособие / Ф. И. Карманов, В. А. Острейковский. – М.: Абрис, 2012. – 208 с.

2. Охорзин, В. А. Прикладная математика в системе MathCAD: учебное пособие / В. А. Охорзин – 3-е изд., стер. – СПб: изд-во «Лань», 2009. – 352 с.

Дополнительная литература

1. Горелова, Г. В. Теория вероятностей и математическая статистика в примерах и задачах с применением Excel : учеб. пособие для студентов вузов / Г. В. Горелова, И. А. Кацко . - 4-е изд. - Ростов н/Д : Феникс, 2006. - 480 с.

2. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – М.: Физматлит, 2006. – 816 с.

3. Зеньковский, В. А. Применение Excel в экономических и инженерных расчетах. Серия «Про Пк» / В. А. Зеньковский. – М.: Солон-Пресс, 2009. – 192 с.

4. Дьяконов, В. П. Mathcad 11/12/13 в математике : справочник / В. П. Дьяконов. - М. : Горячая линия-Телеком, 2007. - 958 с.

5. Каганов, В. И. Компьютерные вычисления в средах EXCEL и MathCAD / В. И. Каганов. - М. : Горячая линия-Телеком, 2003. - 328 с.

Интернет-ресурсы

- 1) <http://www.exponenta.ru> – образовательный математический сайт;
- 2) <http://www.intuit.ru> – Национальный открытый университет «ИНТУ-ИТ»;
- 3) <http://www.window.edu.ru> – Единое окно доступа к образовательным ресурсам.

Приложение А

Аппаратура и материалы

Для выполнения лабораторной работы необходим персональный компьютер со следующими характеристиками: процессор Intel с тактовой частотой 2000 МГц и выше; оперативная память – не менее 1024 Мбайт; свободное дисковое пространство – не менее 1,2 Гбайт; устройство для чтения компакт-дисков; монитор типа Super VGA (число цветов – 256) с диагональю не менее 17 ". **Программное обеспечение** – операционная система WINDOWS 2000 / XP, пакет автоматизации математических вычислений MathCAD, MS Excel.

ложение Б

Указания по технике безопасности.

1. Самостоятельно не производить ремонт персонального компьютера, установку и удаление программного обеспечения.
2. В случае неисправности персонального компьютера сообщить об этом обслуживающему персоналу лаборатории (оператору, администратору).

3. Соблюдать правила техники безопасности при работе с электрооборудованием.
4. Не касаться электрических розеток металлическими предметами.
5. Рабочее место пользователя персонального компьютера должно содержаться в чистоте.
6. Не разрешается возле персонального компьютера принимать пищу, напитки.

Приложение В

Содержание отчета и его форма

Отчет по лабораторной работе должен состоять из:

- 1) названия лабораторной работы;
- 2) ответов на контрольные вопросы;
- 3) формулировки индивидуальных заданий и порядка их выполнения.

Отчет о выполнении лабораторной работы в письменном виде сдается преподавателю.

Приложение Г

Защита работы

Перед выполнением лабораторной работы каждый студент получает индивидуальное задание. Защита лабораторной работы происходит только после его выполнения (индивидуального задания). При защите лабораторной работы студент отвечает на контрольные вопросы, приведенные в конце, и поясняет выполненное индивидуальное задание. Ход защиты лабораторной работы контролируется преподавателем.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Невинномысский технологический институт (филиал)

ЛАБОРАТОРНЫЙ ПРАКТИКУМ
по дисциплине
«ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Методические указания к выполнению лабораторных
работ для студентов направления подготовки
15.03.04 Автоматизация технологических процессов и производств
Направленность (профиль) Информационно-управляющие системы
Часть 2

Невинномысск 2020

Методические указания предназначены для студентов направления 15.03.04 Автоматизация технологических процессов и производств и других технических специальностей. Они содержат основы теории, порядок проведения лабораторных работ. Работы подобраны и расположены в соответствии с методикой изучения дисциплины «Обработка экспериментальных данных». Объем и последовательность выполнения работ определяются преподавателем в зависимости от количества часов, предусмотренных учебным планом дисциплины, как для очной, так и для заочной форм обучения.

Составитель:

ст. преподаватель Д.В. Самойленко

Ответственный редактор:

канд. техн. наук А.А. Евдокимов

Лабораторная работа 9

Применение пакета MathCAD для построения доверительных интервалов основных числовых характеристик выборки, параметров распределения и вероятности

Цель и содержание: изучить методы построения доверительных интервалов для основных числовых характеристик выборки, параметров распределения, вероятности; рассмотреть способы построения доверительных интервалов в пакете MathCAD.

Теоретическое обоснование

При использовании точечных оценок в процессе обработки статистической информации получается лишь приближенное значение неизвестного параметра.

Предположим, что некоторое событие A появилось с вероятностью α , то есть $p(A) = \alpha$, и эта величина достаточно мала. Введем в рассмотрение выборку x_1, x_2, \dots, x_n , элементы которой включаются в генеральную совокупность случайной величины ξ с функцией распределения $F_\xi(x, \theta)$, где θ – оцениваемый параметр.

Пусть найдено две функции $\underline{\theta}(x_1, x_2, \dots, x_n)$ и $\bar{\theta}(x_1, x_2, \dots, x_n)$, удовлетворяющие условиям: 1) $\underline{\theta}(x_1, x_2, \dots, x_n) < \bar{\theta}(x_1, x_2, \dots, x_n)$ для всех элементов выборки, 2) $P(\underline{\theta} < \theta < \bar{\theta}) = 1 - \alpha$.

Тогда интервал $(\underline{\theta}, \bar{\theta})$ будем называть доверительным интервалом для параметра θ , соответствующим доверительной вероятности $P(A) = 1 - \alpha$

Рассмотрим методы построения доверительных интервалов для основных числовых характеристик выборки, параметров распределения и вероятностей.

Доверительные интервалы для математического ожидания

Пусть ξ – случайная величина, представляющая собой последовательность случайных чисел, подчиняющихся нормальному закону распределения. Будем считать, что математическое ожидание этой случайной величины μ

$(M_{\xi} = a)$ не известно, а ее дисперсия D_{ξ} найдена. Тогда задача построения доверительного интервала для M_{ξ} сводится к отысканию величины a .

В качестве точечной оценки неизвестного параметра a будем рассматривать величину $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$. Так как выборочная средняя \bar{x} является случайной величиной, то величина ζ , определяемая равенством $\zeta = \bar{x} - a$, также является случайной. Эти величины должны обладать следующими свойствами: 1) случайная величина \bar{x} распределена нормально и ее математическое ожидание равно a , то есть $M_{\bar{x}} = a$; 2) случайная величина ζ также подчиняется нормальному закону распределения и ее математическое ожидание равно 0, то есть $M_{\bar{x}-a} = 0$; 3) $D_{\bar{x}-a} = D_{\xi}/n$, то есть дисперсия случайной величины ζ равна средней дисперсии генеральной совокупности ξ ; 4) случайная величина $\frac{\bar{x} - a}{\sqrt{D_{\xi}/n}}$ подчиняется нормальному закону распределения с математическим ожиданием, равным 0, и стандартным отклонением – 1.

Введем в рассмотрение функцию $\Phi(x)$, определяемую равенством:

$$\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^x e^{-z^2/2} \cdot dz. \quad (1)$$

С помощью равенства (1) задается функция распределения случайной величины, удовлетворяющей нормальному закону распределения.

Пусть доверительная вероятность появления некоторого события равна α , то есть $p = \alpha$. Значение параметра x_{α} , соответствующего этой вероятности, можно найти, решив уравнение $\Phi(x_{\alpha}) = 1 - 0.5\alpha$.

Если последовательность случайных чисел ξ подчиняется нормальному закону распределения с параметрами $\mu = 0$ и $\sigma = 1$, то вероятность попадания ее значений в интервал $(-x_{\alpha}, x_{\alpha})$ будет равна $1 - \alpha$; это же условие выполняется и

для случайной величины $\frac{\bar{x} - a}{\sqrt{D_{\xi}/n}}$. Таким образом, для вероятности $1 - \alpha$ имеет

место неравенство

$$\bar{x} - x_{\alpha} \cdot \sqrt{\frac{D_{\xi}}{n}} \leq a \leq \bar{x} + x_{\alpha} \cdot \sqrt{\frac{D_{\xi}}{n}}. \quad (2)$$

Из неравенства (2) следует, что при $p = 1 - \alpha$ все значения неизвестного параметра a принадлежат интервалу $\left(\bar{x} - x_{\alpha} \cdot \sqrt{\frac{D_{\xi}}{n}}, \bar{x} + x_{\alpha} \cdot \sqrt{\frac{D_{\xi}}{n}} \right)$, тогда

$$\underline{\theta} = \bar{x} - x_{\alpha} \cdot \sqrt{\frac{D_{\xi}}{n}}, \text{ а } \bar{\theta} = \bar{x} + x_{\alpha} \cdot \sqrt{\frac{D_{\xi}}{n}}.$$

Предположим теперь, что неизвестным параметром является дисперсия случайной величины. Следовательно, задача построения доверительного интервала сводится к отысканию концов интервала, в который будут попадать все значения величины D_{ξ} .

В данном случае для вычисления значения дисперсии может быть использовано равенство (7) из лабораторной работы № 10. Пусть случайная величина $\frac{\bar{x} - a}{\sqrt{D_{\xi}/n}}$ подчиняется закону распределения Стьюдента с $(n-1)$ – й степенью свободы и $p = \alpha$ – доверительная вероятность.

Для построения доверительного интервала вначале следует найти значение параметра $t_{\alpha, n-1}$, который является решением уравнения $F_{n-1}(t_{\alpha, n-1}) = 1 - 0.5\alpha$, где $F_{n-1}(t_{\alpha, n-1})$ – функция распределения Стьюдента с $(n-1)$ – й степенью свободы.

Повторив цикл рассуждений, получим доверительный интервал $\left(\bar{x} - t_{\alpha, n-1} \cdot \sqrt{\frac{S^2}{n}}, \bar{x} + t_{\alpha, n-1} \cdot \sqrt{\frac{S^2}{n}} \right)$. Таким образом, полученный интервал целиком покрывает все значения неизвестного параметра с доверительной вероятностью $1 - \alpha$, то есть имеет место двойное неравенство

$$\bar{x} - t_{\alpha, n-1} \cdot \sqrt{\frac{S^2}{n}} \leq a \leq \bar{x} + t_{\alpha, n-1} \cdot \sqrt{\frac{S^2}{n}}. \quad (3)$$

Следовательно, $\underline{\theta} = \bar{x} - t_{\alpha, n-1} \cdot \sqrt{\frac{S^2}{n}}$, а $\bar{\theta} = \bar{x} + t_{\alpha, n-1} \cdot \sqrt{\frac{S^2}{n}}$

Доверительный интервал для дисперсии

Пусть задана случайная величина ξ , которая подчиняется нормальному закону распределения. Предположим, что известно математическое ожидание этой случайной величины и оно равно a , то есть $M_\xi = a$. Необходимо построить доверительный интервал для неизвестного параметра $D_\xi = \sigma^2$.

Будем считать, что оценка неизвестной дисперсии σ^2 может быть рассчитана по формуле: $\bar{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - a)^2$. Пусть задано отношение

$\frac{n \cdot \bar{\sigma}^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - a}{\sigma} \right)^2$. Так как величины x_i ($i = \overline{1, n}$) являются элементами вы-

борки, извлеченными из генеральной совокупности случайной величины ξ , то и величина $\frac{x_i - a}{\sigma}$ также является случайной и подчиняется нормальному закону

распределения. Тогда случайная величина $\frac{n \cdot \bar{\sigma}^2}{\sigma^2}$ будет удовлетворять χ^2 - распределению с n степенями свободы.

Учитывая, что доверительная вероятность равна α , определим величины $\chi_{l,\alpha}^2$ и $\chi_{r,\alpha}^2$ таким образом, чтобы они удовлетворяли условиям:

$$P\left(\frac{n \cdot \bar{\sigma}^2}{\sigma^2} \leq \chi_{l,\alpha}^2\right) = \frac{\alpha}{2} \quad \text{и} \quad P\left(\frac{n \cdot \bar{\sigma}^2}{\sigma^2} \geq \chi_{r,\alpha}^2\right) = \frac{\alpha}{2}. \quad (4)$$

Из равенств (4) следует, что

$$P\left(\chi_{l,\alpha}^2 \leq \frac{n \cdot \bar{\sigma}^2}{\sigma^2} \leq \chi_{r,\alpha}^2\right) = 1 - \alpha. \quad (5)$$

Тогда на основании равенства (5) для вероятности $p = 1 - \alpha$ имеет место неравенство для оценки неизвестного параметра σ^2 :

$$P\left(\chi_{l,\alpha}^2 \leq \frac{n \cdot \bar{\sigma}^2}{\sigma^2} \leq \chi_{r,\alpha}^2\right) = 1 - \alpha. \quad (6)$$

Следовательно, $\left(\frac{n \cdot \bar{\sigma}^2}{\chi_{r,\alpha}^2}, \frac{n \cdot \bar{\sigma}^2}{\chi_{l,\alpha}^2}\right)$ – доверительный интервал для неизвестной дисперсии σ^2 .

Предположим теперь, что для случайной величины ξ , представляющей собой последовательность случайных чисел, распределенных в соответствии с нормальным законом, известна ее дисперсия D_ξ , а неизвестным является ее математическое ожидание M_ξ . Задача поиска доверительного интервала для D_ξ заключается в поиске интервала, которому бы принадлежали все значения σ^2 , если не известно M_ξ .

Для оценки величины σ^2 в этом случае будем использовать формулу точечной оценки (8) из лабораторной работы № 10.

$$\text{Пусть } \frac{(n-1) \cdot s^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^2, \quad \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i. \text{ По аналогии с предыдущем}$$

случаем введем в рассмотрение случайную величину $\frac{(n-1) \cdot s^2}{\sigma^2}$, которая подчиняется χ^2 - распределению с $(n-1)$ - й степенью свободы. Используя это распределение, вычислим значения величин $\chi_{l,\alpha}^2$ и $\chi_{r,\alpha}^2$, удовлетворяющих условиям:

$$P\left(\frac{(n-1) \cdot s^2}{\sigma^2} \leq \chi_{l,\alpha}^2\right) = \frac{\alpha}{2}$$

и

$$P\left(\frac{(n-1) \cdot s^2}{\sigma^2} \geq \chi_{r,\alpha}^2\right) = \frac{\alpha}{2}. \quad (7)$$

Тогда для доверительной вероятности $p = 1 - \alpha$ имеет место равенство:

$$P\left(\chi_{l,\alpha}^2 \leq \frac{(n-1) \cdot s^2}{\sigma^2} \leq \chi_{r,\alpha}^2\right) = 1 - \alpha. \quad (8)$$

Из равенства (8) следует оценка неизвестного параметра σ^2 :

$$\frac{(n-1) \cdot s^2}{\chi_{r,\alpha}^2} \leq \sigma^2 \leq \frac{(n-1) \cdot s^2}{\chi_{l,\alpha}^2}. \quad (9)$$

Следовательно, интервал $\left(\frac{(n-1) \cdot s^2}{\chi_{r,\alpha}^2}, \frac{(n-1) \cdot s^2}{\chi_{l,\alpha}^2} \right)$ – доверительный интервал неизвестной дисперсии σ^2 .

Доверительный интервал для оценки параметров пуассоновского распределения

Пусть задана случайная величина ξ , подчиняющаяся распределению Пуассона с заданным параметром λ и x_1, x_2, \dots, x_n – произвольная выборка из этой последовательности. Функция распределения вероятности этой выборки и оценка параметра λ приведены в лабораторной работе № 10. Задача построения доверительного интервала для заданного параметра λ заключается в отыскании такого интервала, в который попадают все его значения.

Если α – доверительная вероятность, а x_α – решение уравнения $\Phi(x_\alpha) = 1 - 0.5\alpha$, где $\Phi(x)$ – функция Лапласа, определяемая равенством (1), то для любого значения параметра λ имеет место равенство:

$$P \left(\left| \frac{2 \cdot \sqrt{\lambda} - 2 \cdot \sqrt{\lambda}}{\sqrt{1/n}} \right| < x_\alpha \right) \approx 1 - \alpha. \quad (10)$$

Тогда доверительный интервал параметра λ можно задать в виде $\left(\left(\sqrt{\lambda} - \frac{x_\alpha}{2\sqrt{n}} \right)^2, \left(\sqrt{\lambda} + \frac{x_\alpha}{2\sqrt{n}} \right)^2 \right)$. В этот интервал попадают все значения оцениваемого параметра λ с вероятностью $p = 1 - \alpha$.

Доверительный интервал вероятности

Формула точечной оценки вероятности появления некоторого события получена в лабораторной работе 10.

Как и при построении доверительных интервалов для основных характеристик выборки (математического ожидания, дисперсии) введем доверительную вероятность $p = 1 - \alpha$. Если параметр x_α есть решение уравнения $\Phi(x_\alpha) = 1 - 0.5\alpha$, где $\Phi(x)$ – функция, определяемая равенством (1), то с вероятностью, равной $1 - \alpha$, все значения оцениваемого параметра p будут принадлежать интервалу:

$$\left(\sin^2 \left(\arcsin(\sqrt{p}) - \frac{x_\alpha}{\sqrt{2/n}} \right), \sin^2 \left(\arcsin(\sqrt{p}) + \frac{x_\alpha}{\sqrt{2/n}} \right) \right). \quad (11)$$

Если $n \rightarrow \infty$, то концы доверительного интервала определяются с большей точностью.

Доверительный интервал для коэффициента корреляции

Пусть задана непрерывная или дискретная двумерная случайная величина (ξ, η) и элементы $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ образуют выборку из этой случайной величины. Точечной оценкой для коэффициента корреляции k , указывающего взаимосвязь между значениями случайных величин ξ и η , является выборочный коэффициент корреляции \bar{k} .

Алгоритм расчета выборочного коэффициента корреляции заключается в следующем:

1) вычислите значения выборочных средних для переменных x и y :

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i;$$

2) рассчитайте значения среднеквадратичных отклонений для переменных x и y и среднего значения произведения этих величин:

$$\bar{\sigma}_x^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2, \quad \bar{\sigma}_y^2 = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2, \quad \bar{m} = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y}) \cdot (x_i - \bar{x});$$

3) используя формулу $\bar{k} = \frac{\bar{m}}{\sqrt{\bar{\sigma}_x^2 \cdot \bar{\sigma}_y^2}}$, найдите значение выборочного ко-

эффициента корреляции.

Для оценки коэффициента корреляции можно также использовать формулу:

$$\bar{k} = \frac{\sum_{i=1}^n x_i \cdot y_i - \frac{1}{n} \cdot \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^n x_i\right)^2\right) \cdot \left(\sum_{i=1}^n y_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^n y_i\right)^2\right)}} \quad (12)$$

По аналогии с предыдущими случаями, пусть доверительная вероятность равна α , x_α – решение уравнения $\Phi(x_\alpha) = 1 - 0.5\alpha$.

Следовательно, доверительный интервал, в который попадают все значения $k_{\xi, \eta}$ с вероятностью $p = 1 - \alpha$, можно задать в виде:

$$\left(\theta \left(\operatorname{arth}(\bar{k}) - \frac{x_\alpha}{\sqrt{n-3}} \right), \theta \left(\operatorname{arth}(\bar{k}) + \frac{x_\alpha}{\sqrt{n-3}} \right) \right), \quad (!3)$$

где $\operatorname{th}(x) = \frac{\operatorname{sh}(x)}{\operatorname{ch}(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ – гиперболический тангенс, $\operatorname{arth}(x)$ – функция, обратная гиперболическому тангенсу.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Перейдите в режим автоматического выполнения вычислений.
2. Задайте выборочные значения случайной величины.
3. Используя формулы для получения точечных оценок математического ожидания и дисперсии, найдите их значения для M_ξ и D_ξ .
4. Постройте 95 % - й доверительный интервал математического ожидания, если не известна дисперсия случайной величины.
5. Найдите границы доверительного интервала математического ожидания, если известно значение дисперсии случайной величины.

6. Определите границы 90% - го доверительного интервала для дисперсии случайной величины. Рассмотреть два случая: а) известное математическое ожидание; б) неизвестное математическое ожидание.

Задание 1. Сгенерируйте последовательность случайных чисел $N = 200$, подчиняющуюся нормальному закону распределения с параметрами: $\mu = 100$, $\sigma = 10$. Из полученной последовательности сформируйте выборку объемом: $n = 20$. Постройте доверительные интервалы для M_{ξ} и D_{ξ} .

Методика и порядок выполнения задания 2

1. Проверьте, установлен ли в пакете MathCAD режим автоматического выполнения вычислений.

2. Смоделируйте последовательность случайных чисел, состоящую из 350 элементов, подчиняющихся закону распределения с заданным значением параметра λ .

3. Из полученной последовательности сгенерируйте несколько выборок различного объема: $N = 50, 100, 150, \dots, 300$.

4. Задайте значение доверительной вероятности α и вычислите квантиль уровня $1 - 0.5 \cdot \alpha$ для стандартного нормального распределения.

5. Используя формулу, приведенную в лабораторной работе № 10, найдите численное значение точечной оценки параметра λ .

6. Найдите границы доверительного интервала для параметра λ , если известно значение доверительной вероятности α .

7. Вычислите величину $\Delta\lambda$, равную разности значений правого и левого концов доверительного интервала: $\Delta\lambda = \lambda_{right} - \lambda_{left}$. Представьте эту величину как функцию объема выборки N и постройте соответствующие графики.

Задание 2. Сгенерируйте последовательность случайных чисел, состоящую из 350 элементов и подчиняющуюся распределению Пуассона с параметром $\lambda = 0.1 \cdot N$, где N – номер варианта. Сформируйте из этой последовательности выборки различного объема. Для полученных выборок постройте доверительный интервал параметра λ .

Методика и порядок выполнения задания 3

1. Установите в пакете MathCAD режим автоматического выполнения вычислений.

2. Смоделируйте последовательность случайных чисел из $N = 300$ элементов, подчиняющихся биномиальному закону распределения, и из этой последовательности сформируйте выборку.

3. Введите значение доверительной вероятности и вычислите соответствующее ему значение квантиля $1 - 0.5 \cdot \alpha$ для стандартного нормального распределения.

4. Используя формулу для оценки вероятности, приведенную в лабораторной работе 10, найдите величину оценки параметра p .

5. Постройте доверительный интервал для параметра p по заданному значению доверительной вероятности α .

Задание 3. Смоделируйте последовательность случайных чисел, состоящую из трехсот элементов и подчиняющихся биномиальному закону распределения с неизвестным параметром p . Постройте доверительный интервал для вероятности события по заданным значениям числа испытаний n и числа m появления событий в серии из N_1 испытаний.

Методика и порядок выполнения задания 4

1. Перейдите в режим автоматического выполнения вычислений.

2. Введите исходные данные

3. Используя формулы, приведенные в разделе «Теоретическое обоснование», найдите значения \bar{x} , \bar{y} , \bar{m} , σ_x^2 , σ_y^2 .

4. Задайте значение доверительной вероятности α и вычислите соответствующее ему значение квантиля x_α , решив уравнение: $\Phi(x) = 1 - 0.5 \cdot \alpha$.

5. Определите значение точечной оценки для коэффициента корреляции.

6. Вычислите правый и левый концы доверительного интервала для коэффициента корреляции k для заданного значения доверительной вероятности α .

7. Вычислите значение точечной оценки коэффициента корреляции k с помощью формулы (13).

8. Постройте доверительный интервал для этой точечной оценки.

Задание № 4. Постройте доверительный интервал для коэффициента корреляции по заданной выборке $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

В качестве исходных данных для выполнения задания 1 использовать соответствующие значения из лабораторной работы 2.

Значения параметров n и m , необходимые для выполнения задания 3, следует выбрать из таблицы 9.1.

Таблица 9.1

№	n	m	№	n	m	№	n	m	№	n	m
1	50	35	6	70	27	11	85	34	16	55	33
2	60	35	7	40	22	12	70	36	17	65	28
3	70	35	8	45	24	13	75	38	18	85	30
4	80	25	9	50	26	14	80	35	19	90	38
5	90	25	10	65	32	15	90	34	20	80	40

Значения элементов выборки $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ для выполнения задания 4 получить у преподавателя.

Задание 4 также следует выполнить для исходных данных, приведенных в лабораторной работе 2.

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Сформулируйте определение доверительного интервала.
2. Построение доверительного интервала математического случайной величины, если известно значение дисперсии.

3. Алгоритм построения доверительного интервала математического ожидания, если ее дисперсия неизвестна.
4. Доверительный интервал дисперсии случайной величины. Методика построения этого интервала, если известно математическое ожидание.
5. Последовательность действий при построении доверительного интервала дисперсии случайной величины при неизвестном математическом ожидании.
6. Алгоритм построения доверительного интервала для оценки параметров пуассоновского распределения.
7. Доверительный интервал вероятности. Алгоритм построения.
8. Алгоритм вычисления выборочного коэффициента корреляции.
9. Построение доверительного интервала коэффициента корреляции.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 10

Проверка статистических гипотез с помощью инструментальных средств

Цель и содержание: изучить классификацию статистических гипотез, сформулировать основные задачи проверки гипотез числовых характеристик выборки, рассмотреть решение этих задач в пакете MathCAD.

Теоретическое обоснование

Пусть задана случайная величина ξ , подчиняющаяся нормальному закону распределения. Если θ – один из параметров этого распределения и $\theta_0 = \theta$ – его точное значение, то величину θ можно считать его точечной оценкой.

При обработке и анализе статистической информации актуальными являются вопросы: 1) какова величина отклонения точечной оценки θ от точного значения параметра θ_0 ; 2) какова степень уверенности, что точное значение θ

не будет совпадать с θ_0 , то есть определить степень уверенности при выполнении условия $\theta_0 \neq \theta$.

Пусть $P(\theta - \theta_0) = \alpha$ – вероятность события, что отклонение $\theta - \theta_0$ больше некоторого заданного числа, причем значение этой вероятности ищется из условия $\theta = \theta_0$. Если $\alpha \rightarrow 0$, то событие $\theta - \theta_0$ является маловероятным. В этом случае говорят, что отличие точечной оценки от точного значения параметра является значимым и гипотеза о равенстве параметров θ и θ_0 отвергается. Если же $\alpha \rightarrow 1$, тогда отклонение θ от θ_0 обусловлено естественными факторами (погрешностью измерений и т. д.), а, следовательно, гипотеза о равенстве $\theta = \theta_0$ принимается.

Пусть θ – выборочное значение некоторого неизвестного или заданного оцениваемого параметра, а p_θ – функция распределения вероятности θ , при условии, что вероятность была найдена при $\theta = \theta_0$.

Нулевой гипотезой называется первоначально выдвинутая гипотеза H_0 . Конкурирующей (альтернативной) H_1 называется гипотеза, которая противоречит нулевой.

Критической областью будем называть все значения критерия, при которых отвергается нулевая гипотеза, тогда *область принятия гипотезы* – совокупность значений критерия, при которых H_0 принимается.

Критерий – случайная величина, служащая для проверки гипотезы H_0 .

Критическими точками θ_{left} и θ_{right} будем называть точки, отделяющие критическую область от области принятия гипотезы. Критические точки θ_{left} и θ_{right} должны удовлетворять условиям: $P(\theta < \theta_{\text{left}}) = \frac{\alpha}{2}$ и $P(\theta > \theta_{\text{right}}) = \frac{\alpha}{2}$.

Уровнем значимости при проверке гипотез называют вероятность α , которая использовалась для вычисления концов интервала $(\theta_{\text{left}}, \theta_{\text{right}})$. При проверке статистических гипотез выделяют *одностороннюю* или *двустороннюю* критические области.

Односторонней критической областью называют правостороннюю или левостороннюю критические области. *Левосторонняя критическая область* определяется неравенством $\theta < \theta_0$, а правосторонняя соответственно неравенством $\theta > \theta_0$. Двусторонней будем называть критическую область, удовлетворяющую условиям: $\theta < \theta_0$ и $\theta > \theta_0$.

Принцип проверки статистических гипотез заключается в следующем: если наблюдаемое значение критерия принадлежит критической области, то гипотезу отвергают; если же наблюдаемое значение принадлежит области принятия, то гипотезу принимают.

Ошибка первого рода при проверке статистических гипотез возникает, если отклоняется верная гипотеза. *Ошибка второго рода* связана с принятием ложной гипотезы.

Проверка гипотезы о числовом значении математического ожидания при известной дисперсии

Пусть ξ – случайная величина, подчиняющаяся нормальному закону распределения, M_ξ – неизвестное математическое ожидание этой случайной величины, $M_\xi = a$, D_ξ – известная дисперсия, $D_\xi = \sigma^2$. Введем нулевую гипотезу H_0 – неизвестный параметр a равен заданному числу a_0 , то есть $H_0 : a = a_0$. В этом случае альтернативная (конкурирующая) гипотеза может быть сформулирована одним из способов: 1) $H_1 : a \neq a_0$, 2) $H_1 : a > a_0$, 3) $H_1 : a < a_0$.

Проверим первую из альтернативных гипотез. Пусть $H_0 : a = a_0$ и $H_1 : a \neq a_0$, α – уровень значимости.

Если x_1, x_2, \dots, x_n является выборкой из случайной величины ξ , то значение критерия можно найти по формуле:

$$\varphi = \frac{\bar{x} - a}{\sqrt{\sigma^2/n}}, \quad (1)$$

где $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$.

Построим для критерия φ критическую область. Построение критической области заключается в определении ее граничных точек $x_{l,\alpha}$ и $x_{r,\alpha}$, которые должны удовлетворять условиям:

$$P(\varphi < x_{l,\alpha}) = \frac{\alpha}{2}, \quad P(\varphi > x_{r,\alpha}) = \frac{\alpha}{2}. \quad (2)$$

Значения $x_{l,\alpha}$ и $x_{r,\alpha}$, являются симметричными относительно нуля, причем правая граница $x_{r,\alpha}$ является решением уравнения $\Phi(x_{r,\alpha}) = 1 - 0.5 \cdot \alpha$, а левая — $x_{l,\alpha} = -x_{r,\alpha}$.

После того как найдены границы критической области, необходимо вычислить по этой выборке значение критерия φ и проверить, попадает ли его значение в критическую область.

Если $\varphi < x_{l,\alpha}$ или $\varphi > x_{r,\alpha}$, то нулевая гипотеза H_0 отвергается, а верной признается альтернативная гипотеза H_1 . Если же $\varphi \in (x_{l,\alpha}, x_{r,\alpha})$, то верной считается гипотеза H_0 .

Пусть нулевая гипотеза $H_0 : a = a_0$ остается той же самой, а в качестве альтернативной рассмотрим гипотезу $H_1 : a > a_0$.

Как и для гипотезы $H_1 : a \neq a_0$, зададим уровень значимости — α . В данном случае критерий для проверки альтернативной гипотезы можно вычислить по формуле (1). В этом случае критическая область, в которую попадают все значения критерия φ , при которых гипотеза H_0 отвергается, является правосторонней. Следовательно, критическая область удовлетворяет условию $P(\varphi > x_{r,\alpha}) = \alpha$, где $x_{r,\alpha}$ является корнем уравнения $\Phi(x_{r,\alpha}) = 1 - \alpha$.

Как только найдена граница критической области, следует проверить, принадлежит ли найденное значение критерия φ этой области. При $\varphi > x_{r,\alpha}$ гипотеза H_0 отвергается, при $\varphi < x_{r,\alpha}$ — H_0 принимается.

Для сравнения нулевой гипотезы $H_0 : a = a_0$ с конкурирующей гипотезой $H_1 : a < a_0$ введем заранее известное значение уровня значимости и вычислим значение критерия по формуле (1).

В этом случае критическая область является односторонней, в частности, левосторонней, при этом критическая точка будет удовлетворять условию $P(\varphi < x_{1,\alpha}) = \alpha$, и значение $x_{1,\alpha}$ можно вычислить, используя формулу $x_{1,\alpha} = -x_{r,\alpha}$, где $x_{r,\alpha}$ – корень уравнения $\Phi(x_{r,\alpha}) = 1 - \alpha$.

После отыскания границы критической области необходимо сравнить вычисленное значение критерия φ с найденными значениями границ и проверить выполнение условия попадания его в критическую область. Если значение критерия φ меньше $x_{1,\alpha}$, то есть $\varphi < x_{1,\alpha}$, то сформулированная нулевая гипотеза H_0 отвергается, а верной признается (принимается) альтернативная гипотеза H_1 , в противном случае $\varphi > x_{1,\alpha}$ гипотеза H_0 принимается (не отвергается).

Проверка гипотезы о числовом значении математического ожидания при неизвестной дисперсии

Пусть $M_\xi = a$ – неизвестное математическое ожидание некоторой случайной величины, $D_\xi = \sigma^2$ – неизвестная дисперсия. x_1, x_2, \dots, x_n – выборка значений случайной величины ξ . Для получения точечных оценок основных характеристик этой выборки будем использовать формулы (6) и (8) из лабораторной работы 10.

Вновь первоначально выдвинем нулевую гипотезу H_0 – неизвестный параметр a равен заданному числу a_0 , то есть $H_0 : a = a_0$, и три альтернативные (конкурирующие) гипотезы: 1) $H_1 : a \neq a_0$, 2) $H_1 : a > a_0$, 3) $H_1 : a < a_0$.

В этом случае для проверки альтернативных гипотез используется критерий, значение которого можно определить по формуле:

$$\varphi = \frac{\bar{x} - a}{\sqrt{s^2/n}}. \quad (3)$$

Если нулевая гипотеза H_0 является верной, то этот критерий является случайной величиной, подчиняющейся распределению Стьюдента с $(n - 1)$ – й степенью свободы.

Проверим каждую из альтернативных гипотез при заданном уровне значимости α .

1. Пусть $H_0 : a = a_0$ и $H_1 : a \neq a_0$. В этом случае критическая область является двусторонней, и численные значения ее границ могут быть найдены, учитывая условия (2). Так как распределение Стьюдента является симметричным, то вначале вычисляется значение $x_{r,\alpha}$, а затем $x_{l,\alpha}$. Значение $x_{r,\alpha}$ является корнем уравнения $F_{n-1}(x_{r,\alpha}) = 1 - 0.5 \cdot \alpha$, где $F_{n-1}(x)$ – функция распределения Стьюдента с $(n - 1)$ – й степенью свободы.

После того как найдены границы критической области, а, следовательно, определена и область, которой не принадлежит найденное значение критерия, необходимо проверить условие его принадлежности этой области.

Если критерий φ удовлетворяет одному из условий $\varphi < x_{l,\alpha}$ или $\varphi > x_{r,\alpha}$, то отвергается начальная гипотеза H_0 , а принимается альтернативная гипотеза $H_1 : a \neq a_0$. Если же $\varphi \in (x_{l,\alpha}, x_{r,\alpha})$, то принимается гипотеза H_0 .

Если в качестве альтернативной рассматривать гипотезу $H_1 : a > a_0$, то критическая область для значений критерия φ , при которых отвергается нулевая гипотеза H_0 , является правосторонней, то есть представляет собой полузакрытый интервал $(x_{r,\alpha}, \infty)$. Значение границы критической области, то есть величины $x_{r,\alpha}$, в этом случае можно определить из условия $P(\varphi > x_{r,\alpha}) = \alpha$, где $x_{r,\alpha}$ – корень уравнения $F_{n-1}(x_{r,\alpha}) = 1 - \alpha$ при заданном уровне значимости α .

Далее для проверки альтернативной гипотезы следует сравнить найденное по формуле (4) значение критерия с полученной границей. Если значение критерия φ удовлетворяет условию $\varphi > x_{r,\alpha}$, то нулевая гипотеза отвергается, а принимается альтернативная. При выполнении неравенства $\varphi < x_{r,\alpha}$ принимается гипотеза H_0 .

При проверке альтернативной гипотезы $H_1 : a < a_0$ критическая область для критерия φ является левосторонней, то есть описывается интервалом $(-\infty, x_{l,\alpha})$. Ее граница в этом случае удовлетворяет условию $P(\varphi < x_{l,\alpha}) = \alpha$, а

численное значение можно вычислить, используя формулу $x_{l,\alpha} = -x_{r,\alpha}$, где $x_{r,\alpha}$ – решение уравнения $\Phi(x_{r,\alpha}) = 1 - \alpha$, параметр α – уровень значимости.

Если полученное значение критерия φ попадает в критическую область, то есть удовлетворяет неравенству $\varphi < x_{1,\alpha}$, то нулевая гипотеза отвергается, а признается верной альтернативная гипотеза. При выполнении условия $\varphi > x_{1,\alpha}$ гипотеза H_0 не отвергается.

Проверка гипотезы о числовом значении дисперсии

Пусть задана величина ξ , представляющая собой последовательность случайных чисел, подчиняющихся нормальному закону распределения и x_1, x_2, \dots, x_n – произвольная выборка значений из этой случайной величины.

Сформулируем нулевую гипотезу H_0 – значение дисперсии D для нормального распределения равно заданному числу D_0 , то есть $H_0 : D = D_0$. Альтернативная (конкурирующая) гипотеза о числовом значении дисперсии может быть сформулирована одним из способов: 1) $H_1 : D \neq D_0$, 2) $H_1 : D > D_0$, 3) $H_1 : D < D_0$.

Для проверки гипотезы о числовом значении дисперсии введем параметр α – уровень значимости.

В качестве критерия для сравнения нулевой и альтернативной гипотез будем рассматривать величину φ , значение которой можно вычислить по формуле:

$$\varphi = (n - 1) \cdot \frac{s^2}{D_0}, \quad (4)$$

где величина s^2 – точечная оценка дисперсии.

При сравнении гипотез о числовом значении дисперсии критерий φ будет удовлетворять χ^2 - распределению с $(n - 1)$ - й степенью свободы.

1. Проверим гипотезу о числовом значении дисперсии, если $H_1 : D \neq D_0$. В этом случае критическая область является двусторонней и ее границы будут удовлетворять условиям (2). Пусть α – заданный уровень значимости. Построение критической области заключается в отыскании $x_{l,\alpha}$ и $x_{r,\alpha}$: $x_{l,\alpha}$ – решение

уравнения $F_{n-1}(x_{l,\alpha}) = 0.5 \cdot \alpha$, а $x_{r,\alpha}$ – корень уравнения $F_{n-1}(x_{r,\alpha}) = 1 - 0.5 \cdot \alpha$, где $F_{n-1}(x)$ – функция, подчиняющаяся χ^2 - распределению с $(n - 1)$ – й степенью свободы.

После того как найдены границы критической области и вычислено значение критерия φ , следует проверить, попадает ли оно в критическую область.

Если $\varphi < x_{l,\alpha}$ или $\varphi > x_{r,\alpha}$, то нулевая гипотеза H_0 отвергается, а верной признается альтернативная гипотеза H_1 . Если же $\varphi \in (x_{l,\alpha}, x_{r,\alpha})$, то верной считается гипотеза H_0 .

2. Если альтернативной гипотезе H_0 является гипотеза $H_1: D > D_0$, то критическая область для критерия φ , при которых гипотеза H_0 отвергается, является правосторонней, то есть $\varphi \in (x_{r,\alpha}, +\infty)$. Следовательно, критическая точка удовлетворяет условию $P(\varphi > x_{r,\alpha}) = \alpha$, причем $x_{r,\alpha}$ является решением уравнения $F_{n-1}(x_{r,\alpha}) = 1 - \alpha$, если α – заданный уровень значимости.

Используя формулу (4), следует вычислить значение критерия φ и проверить условие его принадлежности к критической области. При выполнении условия $\varphi > x_{r,\alpha}$ гипотеза H_0 отвергается, а H_1 принимается, при выполнении же обратного неравенства верной считается гипотеза H_0 , а отвергается H_1 .

В том случае, если в качестве альтернативной гипотезы рассматривается $H_1: D < D_0$, то критическая область для параметра φ представляет собой интервал $(0, x_{l,\alpha})$, то есть она является левосторонней критической областью. Граница этой области $x_{l,\alpha}$ может быть найдена из уравнения $F_{n-1}(x_{r,\alpha}) = \alpha$, $x_{l,\alpha} = -x_{r,\alpha}$ и должна удовлетворять условию $P(\varphi < x_{l,\alpha}) = \alpha$, где α – заданный уровень значимости.

Если же после расчета значения критерия φ выясняется, что оно принадлежит критической области, то есть $\varphi < x_{l,\alpha}$, то гипотеза $H_0: D = D_0$ отвергается, а принимается гипотеза $H_1: D < D_0$ при выполнении условия $\varphi > x_{l,\alpha}$

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Установите режим автоматического выполнения вычислений.
2. Сгенерируйте последовательность случайных чисел, распределенных в соответствии с нормальным законом. Число чисел – 200, математическое ожидание (μ) – 100, стандартное отклонение (σ) – 10.
3. Из построенной последовательности смоделируйте выборку, состоящую из 150 элементов.
4. Используя формулу (6) из лабораторной работы № 10, найдите значение точечной оценки математического ожидания.
5. Определите нулевую гипотезу H_0 о значении математического ожидания – $H_0: a = a_0$ и уровень значимости α .
6. С помощью формулы (6) из раздела «Теоретическое обоснование» вычислите значение критерия для проверки нулевой гипотезы.
7. Сформулируйте альтернативную гипотезу H_1 : 1) $a \neq a_0$, 2) $a > a_0$, 3) $a < a_0$.
8. Найдите границы критической области для каждой из формулировок альтернативной гипотезы H_1 .
9. Выполните сравнение полученного критерия с границами критической области.
10. Сформулируйте выводы.

Задание 1. Сгенерируйте последовательность случайных чисел, подчиняющихся нормальному закону распределения: количество чисел – 200, математическое ожидание (μ) – 100, стандартное отклонение (σ) – 10. Из элементов этой последовательности выберите выборку объемом 150. Задайте нулевую гипотезу о величине математического ожидания. Сформулируйте альтернативную гипотезу.

зу H_1 : 1) $a \neq a_0$, 2) $a > a_0$, 3) $a < a_0$. Проверьте эту гипотезу при заданном уровне значимости α при условии, что дисперсия известна.

Методика и порядок выполнения задания 2

1. Установите в пакете MathCAD режим автоматического выполнения вычислений.

2. Смоделируйте последовательность случайных чисел ξ и выборку по исходным данным задания 1.

3. С помощью формул (6) и (8) из лабораторной работы № 10 вычислите значения точечных оценок для математического ожидания и дисперсии.

4. Задайте нулевую гипотезу H_0 о значении математического ожидания – $H_0: a = a_0$ и уровень значимости α .

5. При помощи формулы (3) вычислите значение критерия для проверки нулевой гипотезы.

6. Определите конкурирующую гипотезу H_1 : 1) $a \neq a_0$, 2) $a > a_0$, 3) $a < a_0$.

7. Рассчитайте значения границ критической области для каждого из видов задания (определения) альтернативной гипотезы.

8. Сравните значение полученного критерия с полученными значениями границ.

9. Сформулируйте выводы о принятии или отвержении гипотез H_0 и H_1 .

Задание 2. Используя сформированную в задании 1 последовательность случайных чисел и выборку, сформулируйте нулевую гипотезу о величине математического ожидания. Задайте альтернативную гипотезу H_1 : 1) $a \neq a_0$, 2) $a > a_0$, 3) $a < a_0$. Проверьте эту гипотезу при заданном уровне значимости α , при условии что дисперсия не известна.

Методика и порядок выполнения задания 3

1. Перейдите в пакете MathCAD в режим автоматического выполнения вычислений.

2. При помощи формул (6) и (8) из лабораторной работы № 10 вычислите значения точечных оценок математического ожидания и дисперсии.

3. Введите заданное значение уровня значимости.

4. Определите нулевую гипотезу о значении дисперсии $H_0: D = D_0$.
5. Для сформулированной нулевой гипотезы вычислите значение критерия.
6. Задайте альтернативную гипотезу H_1 : 1) $D \neq D_0$, 2) $D > D_0$, 3) $D < D_0$.
7. Вычислите границы критической области для каждой из альтернативных гипотез.
8. Проведите сравнение полученного значения критерия с границами критической области для альтернативных гипотез и сделайте соответствующие выводы.

Задание 3. Для выборки, сформированной в задании № 1, сформируйте нулевую гипотезу $H_0: D = D_0$. Задайте три альтернативные гипотезы. Проверьте эти гипотезы для заданного уровня значимости α .

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Уровень значимости для вариантов 2, 4, 6, 8, 10, 12, 14, 16, 18 и 20 считать равным 0,95, а для вариантов 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 – $\alpha = 0,9$

Параметры μ (математическое ожидание) и σ (стандартное отклонение) для выполнения заданий 1 – 3 следует выбрать из таблицы 10.1

Таблица 10.1

№	μ	σ	№	μ	σ	№	μ	σ	№	μ	σ
1	0.1	1.5	6	3.2	5	11	4.5	3.1	16	4.2	6
2	0.2	2	7	1.5	0.1	12	5	3.2	17	3	1.2
3	1.1	2.5	8	2	0.2	13	1.2	3	18	3.5	2.1
4	2.2	4	9	2.5	1.1	14	2.1	3.5	19	5.5	4.1
5	3.1	4.5	10	4	2.2	15	4.1	5.5	20	6	4.2

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Основные определения теории статистических гипотез.
2. Проверка гипотез о числовом значении математического ожидания при известной дисперсии.
3. Алгоритм решения этой задачи в пакете MathCAD.
4. Методика проверки гипотезы о числовом значении математического ожидания, если не известно значение дисперсии.
5. Последовательность действий при проверке гипотез о числовом значении математического ожидания при неизвестной дисперсии в MathCAD.
6. Проверка гипотезы о числовом значении дисперсии.
7. Алгоритм проверки гипотезы о числовом значении дисперсии в MathCAD.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 11

Проверка статистических гипотез о равенстве математических ожиданий при помощи пакета MathCAD

Цель и содержание: сформулировать основные задачи проверки гипотез о равенстве значений математических ожиданий выборок, показать реализацию сформулированных задач в пакете MathCAD.

Теоретическое обоснование

Нулевой гипотезой называется первоначально выдвинутая гипотеза H_0 . Конкурирующей (альтернативной) H_1 называется гипотеза, которая противоречит нулевой.

Принцип проверки статистических гипотез заключается в следующем: если наблюдаемое значение критерия принадлежит критической области, то гипотезу отвергают; если же наблюдаемое значение принадлежит области принятия, то гипотезу принимают.

Проверка гипотезы о равенстве математических ожиданий при известных дисперсиях

Пусть ξ и η величины, представляющие собой последовательность случайных чисел, подчиняющихся нормальному закону распределения с параметрами a_ξ и σ_ξ , a_η и σ_η соответственно, последовательности x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_n – выборки из случайных величин ξ и η , причем математическое ожидание для каждой из выборок обоих случайных величин $M_\xi = a_\xi$, $M_\eta = a_\eta$ не известно, а параметры $\sigma_\xi^2 = D_\xi$ и $\sigma_\eta^2 = D_\eta$ считаются известными.

Учитывая тот факт, что случайные величины ξ и η являются независимыми, проверим нулевую гипотезу $H_0 : a_\xi = a_\eta$. Альтернативной гипотезой в данном случае может быть одна из гипотез: 1) $H_1 : a_\xi \neq a_\eta$, 2) $H_1 : a_\xi > a_\eta$, 3) $H_1 : a_\xi < a_\eta$.

Для получения аналитического выражения, при помощи которого можно вычислить значения критерия φ , применяемого в процессе проверки гипотезы H_0 , введем в рассмотрение случайные величины \bar{x} , \bar{y} и $\bar{x} - \bar{y}$.

При этом величины \bar{x} и \bar{y} будут подчиняться нормальному закону распределения с параметрами $M_{\bar{x}} = a_\xi$, $D_{\bar{x}} = \frac{\sigma_\xi^2}{N}$ и $M_{\bar{y}} = a_\eta$, $D_{\bar{y}} = \frac{\sigma_\eta^2}{M}$ соответственно, тогда $\bar{x} - \bar{y}$ также подчиняется нормальному закону распределения с параметрами $M_{\bar{x}-\bar{y}} = a_\xi - a_\eta$, $D_{\bar{x}-\bar{y}} = \frac{\sigma_\xi^2}{N} + \frac{\sigma_\eta^2}{M}$, где N и M – число элементов в выборке.

Пусть нулевая гипотеза $H_0 : a_\xi = a_\eta$ принимается, то есть считается верной. Тогда величина φ , определяемая равенством

$$\varphi = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_\xi^2}{N} + \frac{\sigma_\eta^2}{M}}}, \quad (1)$$

также будет подчиняться нормальному закону распределения, причем со стандартными значениями параметров $\mu = 0$ и $\sigma = 1$.

Проверим нулевую гипотезу $H_0 : a_\xi = a_\eta$, если задана альтернативная гипотеза $H_1 : a_\xi \neq a_\eta$.

Пусть α – заданный уровень значимости. Для выдвинутой альтернативной гипотезы критическая область будет двусторонней, и ее границы будут удовлетворять условиям, зада

ваемым при помощи равенств (2) из лабораторной работы № 10, то есть ее левая граница – $x_{l,\alpha} = -x_{r,\alpha}$, а правая граница $x_{r,\alpha}$ – корень уравнения $\Phi(x_{r,\alpha}) = 1 - 0.5 \cdot \alpha$ и $\Phi(x)$ – функция распределения для нормального распределения с параметрами $\mu = 0$ и $\sigma = 1$.

После отыскания границ критической области и вычисления значения критерия φ по выборкам выясняется, принадлежит ли полученное число критической области. Если значение φ удовлетворяет одному из условий $\varphi < x_{l,\alpha}$ или $\varphi > x_{r,\alpha}$, то гипотеза H_0 будет отвергнута, а гипотеза $H_1: a_\xi \neq a_\eta$ принимается. В том случае, если $\varphi \in (x_{l,\alpha}, x_{r,\alpha})$, то принимается гипотеза H_0 .

Проверка гипотезы о равенстве математических ожиданий при равных неизвестных дисперсиях

Пусть ξ и η величины, представляющие собой последовательность случайных чисел, распределенных в соответствии с нормальным законом распределения с параметрами μ_ξ и σ_ξ , μ_η и σ_η , последовательности x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_n – выборки из случайных величин ξ и η , причем математическое ожидание для каждой из выборок обеих случайных величин $\mu_\xi = M_\xi = a_\xi$, $\mu_\eta = M_\eta = a_\eta$ не известно, значения дисперсий для каждой из выборок $\sigma^2 = D_\xi$ и $\sigma^2 = D_\eta$ считается не известным и равными между собой $D_\xi = D_\eta$.

Для определения критерия проверки нулевой гипотезы H_0 – о равенстве значений математических ожиданий этих случайных величин ($H_0: a_\xi = a_\eta$) – вначале найдем значения величин \bar{x} , \bar{y} , s_ξ^2 и s_η^2 . Формулы для вычисления значений этих величин приведены в лабораторной работе № 10. Величины, являющиеся средними значениями по выборкам, \bar{x} и \bar{y} подчиняются нормальному закону распределения с параметрами: математическое ожидание – a_ξ и a_η ,

стандартное отклонение $-\frac{\sigma}{\sqrt{N}}$ и $\frac{\sigma}{\sqrt{M}}$, тогда как величины s_{ξ}^2 и s_{η}^2 будут подчиняться χ^2 - распределению с $(N-1)$ - й и $(M-1)$ - й степенями свободы соответственно.

Так как рассматриваемые случайные величины ξ и η являются независимыми, то статистическая характеристика, определяемая равенством $\frac{(N-1) \cdot s_{\xi}^2}{\sigma^2} + \frac{(M-1) \cdot s_{\eta}^2}{\sigma^2}$, будет подчиняться χ^2 - распределению с $(N+M-2)$

степенями свободы. Величина $\bar{x} - \bar{y}$ распределена в соответствии с нормальным законом с параметрами $M_{\bar{x}-\bar{y}} = a_{\xi} - a_{\eta}$, $D_{\bar{x}-\bar{y}} = \frac{\sigma^2}{N} + \frac{\sigma^2}{M}$, где N и M - число

элементов в выборке. Тогда и величина $\frac{(\bar{x} - \bar{y}) - (a_{\xi} - a_{\eta})}{\sqrt{\frac{\sigma}{N} + \frac{\sigma}{M}}}$ представляет собой

последовательность случайных величин, распределенных нормально с параметрами $\mu=0$ и $\sigma=1$, а отношение, задаваемое равенством

$$\frac{(\bar{x} - \bar{y}) - (a_{\xi} - a_{\eta})}{\sqrt{\frac{\sigma}{N} + \frac{\sigma}{M}} \cdot \sqrt{\left(\frac{(N-1) \cdot s_{\xi}^2}{\sigma^2} + \frac{(M-1) \cdot s_{\mu}^2}{\sigma^2}\right)} \cdot \frac{1}{N+M-2}},$$

подчиняется распределению

Стьюдента с $(N+M-2)$ степенями свободы.

Если нулевая гипотеза $H_0 : a_{\xi} = a_{\eta}$ верна, то параметр φ , определяемый с помощью равенства

$$\varphi = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{N} + \frac{1}{M}\right) \cdot \frac{(N-1) \cdot s_{\xi}^2 + (M-1) \cdot s_{\mu}^2}{N+M-2}}}, \quad (2)$$

подчиняется распределению Стьюдента с $(N+M-2)$ степенями свободы.

Предположим, что параметр φ является критерием для проверки гипотезы H_0 . Конкурирующую гипотезу можно сформулировать одним из способов: 1) $H_1 : a_{\xi} \neq a_{\eta}$, 2) $H_1 : a_{\xi} > a_{\eta}$, 3) $H_1 : a_{\xi} < a_{\eta}$.

Пусть альтернативной к гипотезе $H_0: a_\xi = a_\eta$ является гипотеза $H_1: a_\xi \neq a_\eta$ и величина α – известный по условию уровень значимости. Тогда вначале для проверки H_0 построим критическую область. В этом случае она является двусторонней. Границами этой области являются точки $x_{1,\alpha}$ и $x_{r,\alpha}$, значения которых можно найти, решив уравнения: $F_{N+M-2}(x_{1,\alpha}) = 0.5\alpha$ и $F_{N+M-2}(x_{r,\alpha}) = 1 - 0.5\alpha$, где $F_{N+M-2}(x)$ – теоретическая функция распределения Стьюдента с $(N + M - 2)$ степенями свободы.

После определения числовых значений границ критической области и критерия φ следует проверить условие его принадлежности этой области. Если значение критерия удовлетворяет одному из условий, определяемых с помощью неравенств $\varphi < x_{1,\alpha}$ или $\varphi > x_{r,\alpha}$, то нулевую гипотезу отвергают, а признают верной гипотезу H_1 . При $\varphi \in (x_{1,\alpha}, x_{r,\alpha})$ принимается нулевая гипотеза $H_0: a_\xi = a_\eta$.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Проверьте, установлен ли в пакете MathCAD режим автоматического выполнения вычислений.
2. Сгенерируйте последовательность случайных чисел и выборку с параметрами, приведенными в задании 1 лабораторной работы 13.
3. Используя формулу (6) из раздела «Теоретическое обоснование» лабораторной работы 10, найдите значение точечной оценки математического ожидания.
4. Рассчитайте по формуле (1) значение критерия проверки гипотезы о равенстве математических ожиданий при известной дисперсии.

5. Введите уровень значимости.

6. Сформулируйте нулевую гипотезу H_0 о равенстве значений математического ожидания при условии, что известны значения дисперсий выборок – $H_0: a_\xi = a_\eta$.

7. Определите альтернативную гипотезу H_1 : 1) $H_1: a_\xi \neq a_\eta$, 2) $H_1: a_\xi > a_\eta$, 3) $H_1: a_\xi < a_\eta$.

8. Вычислите границы критической области для каждого из способов задания альтернативной гипотезы.

9. Проверьте условия принадлежности критерия φ критической области для каждой из формулировок альтернативной гипотезы.

10. Сформулируйте соответствующие выводы.

Задание № 1. Из последовательности случайных чисел, сформированной в задании № 1 лабораторной работы 13, смоделируйте две выборки объемами 100 и 120 элементов соответственно. Эта выборка, как и вся последовательность, будет подчиняться нормальному закону распределения с заданными параметрами. Сформулируйте нулевую гипотезу H_0 о равенстве математических ожиданий и задайте альтернативную гипотезу H_1 : 1) $H_1: a_\xi \neq a_\eta$, 2) $H_1: a_\xi > a_\eta$, 3) $H_1: a_\xi < a_\eta$. Проверьте для заданного уровня значимости α сформулированную альтернативную гипотезу.

Методика и порядок выполнения задания 2

1. Установите в пакете MathCAD режим автоматического выполнения вычислений.

2. Смоделируйте выборки объемом 100 и 120 элементов из генеральной совокупности случайной величины ξ .

3. Для построенных выборок вычислите значение точечной оценки математического ожидания.

4. Используя формулу (2) из раздела «Теоретическое обоснование», найдите значение критерия для проверки гипотезы о равенстве математических ожиданий.

5. Определите нулевую гипотезу H_0 – о равенстве математических ожиданий, то есть $H_0: a_\xi = a_\eta$.

6. Сформулируйте альтернативные гипотезы: 1) $H_1: a_\xi \neq a_\eta$, 2) $H_1: a_\xi > a_\eta$, 3) $H_1: a_\xi < a_\eta$.

7. Найдите границы критической области для каждой из выдвинутых альтернативных гипотез.

8. Сравните численное значение найденного значения критерия φ с полученными значениями границ и сформулируйте выводы о принятии или опровержении гипотез H_0 или H_1 .

Задание № 2 Используя выборки, сформированные в задании № 1, сформулируйте нулевую гипотезу о равенстве математических ожиданий. Задайте альтернативную гипотезу одним из способов: 1) $H_1: a_\xi \neq a_\eta$, 2) $H_1: a_\xi > a_\eta$, 3) $H_1: a_\xi < a_\eta$. Проверьте для заданного уровня α сформулированную альтернативную гипотезу, если не известны значения дисперсий этих выборок.

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Уровнем значимости для проверки гипотез о равенстве математических ожиданий считать значения: $\alpha = 95\%$ – для вариантов $N = 1, 3, 5, 7, 9, 11, 13, 15, 17, 19$; $\alpha = 90\%$ – для вариантов $N = 2, 4, 6, 8, 10, 12, 14, 16, 18, 20$.

Значения параметров a (математическое ожидание) и σ (стандартное отклонение) для формирования последовательностей случайных чисел выбрать из таблицы 10.1

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Дайте определение нулевой и альтернативной гипотезы.
2. Сформулируйте основной принцип проверки статических гипотез.

3. Проверка гипотезы о равенстве математических ожиданий при известном значении дисперсий.
4. Реализация алгоритма проверки гипотезы о равенстве математических ожиданий при известном значении дисперсий.
5. Проверка статистических гипотез о равенстве математических ожиданий при неизвестном равных дисперсий.
6. Последовательность действий при проверке статистической гипотезы о равенстве математических ожиданий при неизвестных значениях дисперсий выборок, исходя из предположения, что дисперсии выборок могут принимать равные значения.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 12

Применение инструментальных средств для построения уравнений регрессии

Цель и содержание: рассмотреть основные принципы регрессионного анализа как одного из методов обработки статистической информации, изучить методику построения уравнения регрессии с помощью инструментальных средств (пакет MathCAD, табличный процессор Excel).

Теоретическое обоснование

Пусть в результате исследования некоторого экономического показателя была получена совокупность значений (x_i, y_i) . Требуется исследовать зависимость $y(x)$, в которой x – независимая переменная, а y – зависимая переменная.

При построении уравнения регрессии будем считать, что переменная x будет определяться точно, а при измерении y могут возникать погрешности. Следовательно, полученные результаты эксперимента, можно рассматривать в виде функции $\eta(x)$, которая является произвольной случайной величиной. В этом случае **регрессией** будем называть зави-

симось $y(x)$ условного математического ожидания случайной величины $\eta(x)$ от независимой переменной x , то есть $y(x) \equiv M(\eta/x)$.

Тогда задача построения уравнения регрессии сводится к восстановлению значений функции $y(x)$ по результатам экспериментов (x_i, y_i) , где $i = \overline{1, n}$.

Предположим, что зависимость $y(x)$ может быть аппроксимирована функцией $f(x, a_0, a_1, \dots, a_n)$. С учетом того, что при измерении y могут возникнуть незначительные погрешности, то есть

$$y_i = f(x_i, a_0, a_1, \dots, a_k) + \xi_i, \quad (1)$$

где a_0, a_1, \dots, a_k – коэффициенты регрессии, ξ_i – последовательность случайных величин, которые характеризуют погрешность измерения величины y .

Так как ξ_i представляет собой совокупность случайных величин, то считают, что она распределена нормально с параметрами $M_{\xi_i} = 0$ и $D_{\xi_i} = \sigma^2$, причем считается, что каждой из последовательностей ξ_i значения дисперсии совпадают.

При построении уравнения регрессии коэффициенты должны выбираться таким образом, чтобы отклонение значений функции f от результатов экспериментов было минимальным. Отклонение в этом случае можно описать равенством:

$$\Phi(a_0, a_1, \dots, a_k) = \sum_{i=1}^k (f(x_i, a_0, a_1, \dots, a_k) - y_i)^2. \quad (2)$$

Таким образом, задача отыскания коэффициентов a_0, a_1, \dots, a_k может быть решена методом наименьших квадратов.

Рассмотрим процесс построения уравнения регрессии для случая $f(x, a_0, a_1) = a_0 + a_1 \cdot x$. Используя метод наименьших квадратов, вычислим значения коэффициентов a_0 и a_1 . Для этого составим функцию $\Phi(a_0, a_1)$:

$\Phi(a_0, a_1) = \sum_{i=1}^k (a_0 + a_1 \cdot x - y_i)^2$, а затем составим нормальную систему уравнений:

$$\begin{cases} \frac{\partial \Phi}{\partial a_0} = 0 \\ \frac{\partial \Phi}{\partial a_1} = 0 \end{cases} \quad (3)$$

Решив систему уравнений (3), получим:

$$a_0 = \frac{\sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i \cdot y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, \quad a_1 = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}. \quad (4)$$

Значения коэффициентов a_0 и a_1 в пакете MathCAD можно вычислить с помощью функций $\text{intercept}(X, Y)$ и $\text{slope}(X, Y)$. Принцип вычисления значений этих коэффициентов a_0 и a_1 с помощью этих функций описан в лабораторной работе № 2.

Оценки коэффициентов a_0 и a_1 можно получить также с помощью метода максимального правдоподобия. Так как при исследовании экономических показателей дисперсия $D_{\xi_i} = \sigma^2$ чаще всего неизвестна, то используя метод максимального правдоподобия, можно найти величину $\bar{\sigma}^2$:

$$\bar{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{a}_0 - \bar{a}_1 \cdot x_i)^2. \quad (5)$$

Поскольку $\bar{\sigma}^2$, \bar{a}_0 и \bar{a}_1 являются случайными величинами, то в качестве точечных оценок величин \bar{a}_0 и \bar{a}_1 можно использовать равенства:

$$M_{\bar{a}_0} = a_0, \quad M_{\bar{a}_1} = a_1, \quad D_{\bar{a}_0} = \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \cdot \bar{\sigma}^2,$$

$$D_{\bar{a}_1} = \frac{\bar{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (6)$$

Так как оценка дисперсии $\bar{\sigma}^2$ является смещенной, то при построении уравнения регрессии удобнее рассматривать несмещенную оценку:

$$s^2 = \frac{n}{n-2} \cdot \bar{\sigma}^2 = \frac{1}{n-2} \cdot \left(\sum_{i=1}^n y_i^2 - \bar{a}_0 \cdot \sum_{i=1}^n y_i - \bar{a}_1 \cdot \sum_{i=1}^n x_i \cdot y_i \right). \quad (7)$$

Поэтому можно сделать вывод о том, что величина $\frac{(n-2) \cdot s^2}{\sigma^2}$ будет подчиняться χ^2 - распределению с $(n-2)$ степенями свободы.

На основании свойств случайных величин и их точечных оценок \bar{a}_0 , \bar{a}_1 и s^2 можно найти границы доверительных интервалов для параметров a_0 , a_1 и σ^2 .

Рассмотрим процесс построения доверительного интервала для a_0 . Предположим, что дисперсия σ^2 является известной, тогда случайная величина, определяемая равенством $\frac{\bar{a}_0 - a_0}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$, подчиняется нормальному закону

распределения.

Пусть α – доверительная вероятность, x_α – корень уравнения $\Phi(x_\alpha) = 1 - 0,5 \cdot \alpha$, где $\Phi(x)$ - функция, определяемая равенством (1) раздела «Теоретическое обоснование» лабораторной работы № 9. Тогда доверительный интервал для параметра a_0 имеет вид

$$\bar{a}_0 - x_\alpha \cdot \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} < a < \bar{a}_0 + x_\alpha \cdot \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (8)$$

для вероятности $p = 1 - \alpha$.

В том же случае, если значение $D_{\xi_i} = \sigma^2$ неизвестно, тогда в качестве критерия можно рассматривать величину $\frac{\bar{a}_0 - a_0}{s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$, где оценка s^2 ,

определяемая с помощью равенства (7), подчиняется распределению Стьюдента с $(n-2)$ степенями свободы.

Тогда, если известно значение доверительной вероятности α , можно вычислить корень уравнения $t_\alpha - F_{n-2}(t_\alpha) = 1 - 0,5 \cdot \alpha$, F_{n-2} – функция распределения Стьюдента. Следовательно, доверительный интервал имеет вид

$$\bar{a}_0 - t_\alpha \cdot \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} < a < \bar{a}_0 + t_\alpha \cdot \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (9)$$

Доверительный интервал для параметра \bar{a}_1 строится аналогично. Если значение дисперсии выборки σ^2 известно, то доверительный интервал, в который попадает значение параметра a_1 с вероятностью $p = 1 - \alpha$, имеет вид:

$$\left(\bar{a}_1 - \frac{x_\alpha \cdot \sigma}{\sum_{i=1}^n (x_i - \bar{x})^2}, \bar{a}_1 + \frac{x_\alpha \cdot \sigma}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \quad (10)$$

Если же значение дисперсии неизвестно, то доверительный интервал, который покрывает все значения параметра a_1 с вероятностью $p = 1 - \alpha$, принимает вид:

$$\left(\bar{a}_1 - \frac{t_\alpha \cdot s}{\sum_{i=1}^n (x_i - \bar{x})^2}, \bar{a}_1 + \frac{t_\alpha \cdot s}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \quad (11)$$

Найдем границы доверительного интервала для неизвестной дисперсии σ^2 . Если α – доверительная вероятность, причем $\alpha \rightarrow 0$, то решив уравнения $F_{n-2}(\chi_{l,\alpha}) = 0,5 \cdot \alpha$, $F_{n-2}(\chi_{r,\alpha}) = 1 - 0,5 \cdot \alpha$, получим численные значения границ доверительного интервала $\chi_{l,\alpha}$ и $\chi_{r,\alpha}$. Следовательно, все значения случайной

величины $\frac{(n-2) \cdot s^2}{\sigma^2}$ принадлежат интервалу $(\chi_{l,\alpha}, \chi_{r,\alpha})$ с доверительной вероятностью $p = 1 - \alpha$. Таким образом, доверительный интервал для неизвестного

$$\text{параметра } \sigma^2 - \left(\frac{(n-2) \cdot s^2}{\chi_{r,\alpha}}, \frac{(n-2) \cdot s^2}{\chi_{l,\alpha}} \right).$$

Предположим, что регрессия, описывающая зависимость между показателями x и y , задана функцией: $f(x, a_0, a_1) = \bar{a}_0 + \bar{a}_1 \cdot x$. Выберем из совокупности показателей x некоторое значение x_0 и, подставив его в выражение для функции $f(x, a_0, a_1)$, найдем $\bar{y}_0 = \bar{a}_0 + \bar{a}_1 \cdot x_0$.

Так как величина \bar{y}_0 является случайной, то ее значение будет изменяться в зависимости от выбранных значений. Следовательно, ее математическое

ожидание будет равно истинному значению функции $f(x)$ в точке x_0 , то есть $y_0 = a_0 + a_1 \cdot x_0$. Построим доверительный интервал для y_0 . Введем в рассмотрение статистику $\frac{(\bar{a}_0 + a_1 \cdot x_0) - (a_0 + a_1 \cdot x_0)}{s \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$, которая будет подчиняться распреде-

$$s \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

лению Стьюдента с $(n - 2)$ степенями свободы. Следовательно, доверительный интервал, в который попадает точное значение $y_0 = a_0 + a_1 \cdot x_0$ с вероятностью $p = 1 - \alpha$, имеет вид

$$\left(\bar{a}_0 + \bar{a}_1 \cdot x - t_\alpha \cdot s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \bar{a}_0 + \bar{a}_1 \cdot x + t_\alpha \cdot s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right). \quad (12)$$

Найденные границы доверительных интервалов для значения x_0 образуют доверительную полосу или доверительный коридор. Однако следует заметить, что данный коридор является доверительной областью не для всей линии регрессии, а определяют концы доверительных интервалов для значения функции регрессии y в каждой из точек совокупности x . Используя коридор регрессии, нельзя одновременно построить доверительные интервалы для x_0 и x_1 .

Доверительная область, которой принадлежит вся линия регрессии целиком, определяется при помощи равенств, соответствующих верхней и нижней границам:

$$y = \bar{a}_0 + \bar{a}_1 \cdot x - 2 \cdot f_\alpha \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (13)$$

$$y = \bar{a}_0 + \bar{a}_1 \cdot x + 2 \cdot f_\alpha \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (14)$$

где f_α – решение уравнения $F_{2,n-2}(f_\alpha) = 1 - \alpha$, α – доверительная вероятность, $F_{2,n-2}(x)$ – функция распределения Фишера с 2 и $(n - 2)$ степенями свободы.

Для построения уравнения регрессии в табличном процессоре Excel используют следующие способы::

- 1) с помощью встроенных функций Excel;
- 2) используя трендовые модели;
- 3) при помощи инструмента «регрессия» из пакета «Анализ данных».

При построении уравнения регрессии, представляемого в виде линейной зависимости $y = a_0 + a_1 \cdot x$, с помощью встроенных функций табличного процессора Excel следует вычислить значения углового коэффициента и точку пересечения с осью ординат. В табличном процессоре Excel для решения этой задачи применяются функции НАКЛОН(диапазон_ячеек_1, диапазон_ячеек_2), ОТРЕЗОК(диапазон_ячеек_1, диапазон_ячеек_2). Эти функции являются аналогами функций MathCAD – slope(X, Y) и intercept(X, Y). С помощью функции КВПИРСОН(диапазон_ячеек_1, диапазон_ячеек_2) можно определить квадрат коэффициента корреляции $-R^2$. Эта величина показывает взаимосвязь между параметрами регрессии.

Значение некоторых экономических показателей, имеющих случайный характер, можно рассматривать в виде временных рядов. Временным рядом будем называть эмпирически полученную последовательность данных, наблюдаемых в определенные моменты времени t_i , где t_i — порядковый номер значения эмпирического ряда на временной оси. Каждый такой ряд характеризуется некоторой тенденцией развития процессов во времени, называемой трендом.

Линии тренда широко используются для решения задач прогнозирования с помощью методов регрессионного анализа. Подбор функции тренда $f(t)$ осуществляется методом наименьших квадратов, при котором минимизируется сумма квадратов отклонений между эмпирическими значениями и соответствующими значениями функции. Для оценки точности модели используют коэффициент детерминации, построенный на основе оценок дисперсии эмпирических данных и значений трендовой модели. Трендовая модель адекватна эмпирическому процессу и отражает тенденции его развития при значениях коэффициента детерминации, близких к 1.

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Перейдите в пакете MathCAD в режим автоматического выполнения вычислений.
2. Задайте в виде матрицы полученные в результате экспериментов данные. Эта матрица должна состоять из двух строк и числа столбцов, соответствующих количеству проведенных испытаний.
3. Используя формулы (6) и (8) из лабораторной работы 10, найдите значения точечной оценки математического ожидания и несмещенную оценку неизвестной дисперсии.
4. При помощи функции `intercept` и `slope` вычислите значения коэффициентов регрессии.
5. Постройте график уравнения регрессии и график экспериментальных данных.
6. Вычислите значения критериев для оценки коэффициентов регрессии a_0, a_1 и оценки дисперсии σ^2 .
7. Найдите границы доверительных интервалов для a_0, a_1 и дисперсии σ^2 .
8. Определите границы доверительного коридора для уравнения регрессии в заданной точке.
9. Представьте графически уравнение (линию) регрессии и границы коридора для нее в известной точке.
10. Постройте доверительную область для всего уравнения в целом.
11. Задайте графически полученное уравнение регрессии и его доверительную область.

Задание № 1. По результатам исследования экономического показателя была получена некоторая выборка. Для этой выборки рассчитайте значения коэффициентов регрессии. По полученным значениям определите для заданной доверительной вероятности α границы доверительных интервалов: коэффици-

ентов регрессии и неизвестной дисперсии. Постройте коридор и доверительную область для полученного уравнения регрессии. Представьте графически выборку эмпирических значений, уравнение регрессии, коридор и доверительную область регрессии, причем все зависимости должны быть приведены на одном графике.

Методика и порядок выполнения задания 2

1. Используя главное меню Windows, запустите табличный процессор Excel.

2. Сформируйте в рабочей книге документ, содержащий исходные данные. В столбец А занесите значения, соответствующие переменной X, а в столбец В – переменной Y.

3. Используя функции НАКЛОН и ОТРЕЗОК, вычислите значения коэффициента наклона уравнения регрессии к оси абсцисс и свободного члена этого уравнения.

4. С помощью функции КВПИРСОН определите величину квадрата коэффициента корреляции.

5. На основании коэффициентов, определенных с помощью функций НАКЛОН и ОТРЕЗОК, запишите уравнение регрессии.

Задание № 2. Используя встроенные функции табличного процессора Excel, постройте линейное уравнение регрессии.

Методика и порядок выполнения задания 3

1. Запустите табличный процессор Excel.

2. В столбцы А и В введите эмпирические данные необходимые для построения уравнения регрессии, то есть значения переменных X и Y.

3. Постройте диаграмму, на которой изобразите зависимость эмпирически полученных значений Y от переменной X. При построении диаграммы выберите тип «график».

4. Перейдите в область диаграммы и выполните щелчок правой кнопкой манипулятора «мышь» на построенном графике. В появившемся контекстном меню выберите команду «Добавить линию тренда ...» и нажмите кнопку ОК.

5. После чего на экран будет выведено диалоговое окно «Линия тренда». В нем перейдите на вкладку «Тип», в которой выберите необходимый вид линии тренда и укажите для какого параметра следует выполнить построение. Завершите процесс построения линии тренда нажатием на ОК.

6. Для получения окончательного вида линии тренда, задающей уравнение регрессии, выполните двойной щелчок любой из кнопок манипулятора «мышь» по этой линии, в результате чего на экран будет выведено диалоговое окно «Формат линии тренда».

7. В этом окне переместитесь на вкладку «Параметры» и установите флажки в поля «показывать уравнение на диаграмме» и «поместить на диаграмму величину достоверности аппроксимации R^2 ».

Задание № 3. При помощи линий тренда выполните построение уравнений регрессии для заданных экспериментальных данных. Вид линии тренда для уравнения регрессии: 1) линейная, 2) полиномиальная.

Методика и порядок выполнения задания 4

1. В табличном процессоре Excel активизируйте опцию главного меню «Сервис».

2. В появившемся подменю выберите команду «Анализ данных», после чего на экран будет выведено диалоговое окно «Анализ данных», в котором перейдите в поле «Инструменты анализа» и выберите инструмент «Регрессия».

3. После нажатия на ОК появляется диалоговое окно «Регрессия», которое состоит из четырех областей.

4. Процесс построения уравнения регрессии состоит в последовательном заполнении каждой из областей. В область «входные данные» следует ввести значения переменных Y и X , а также установить уровень значимости при построении уравнения регрессии, по умолчанию 95%.

5. В области «параметры вывода» следует определить диапазон ячеек, в который будут выведены результаты регрессионного анализа. В качестве этого параметра можно рассматривать свободную ячейку на текущем рабочем листе, новый рабочий лист или новую рабочую книгу.

6. После заполнения области «Параметры вывода» необходимо перейти в подокно «остатки», в котором следует установить флажки в поля «график подбора» и «график остатков».

7. Для завершения процесса построения уравнения регрессии следует нажать ОК.

Задание № 4. Используя пакет «Анализ данных», проведите регрессионный анализ и постройте уравнение регрессии для заданных экспериментальных данных.

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Исходные данные для выполнения заданий к лабораторной работе студент должен получить у преподавателя.

Уровень значимости при построении уравнения регрессии выбрать следующим образом: для вариантов с нечетными номерами выбрать величину $\alpha = 95\%$, а для четных вариантов – 90%

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Дайте определение регрессии.
2. Алгоритм отыскания коэффициентов уравнения регрессии.
3. Оценка параметров коэффициентов регрессии и неизвестной дисперсии.
4. Построение доверительных интервалов для коэффициентов регрессии и неизвестной дисперсии.
5. Алгоритм отыскания границ доверительного интервала и «коридора» для регрессии.
6. Последовательность действий при построении уравнений регрессии и оценки его параметров в пакете MathCAD.

7. Способы построения уравнения регрессии в Excel.
8. Последовательность действий при построении уравнения регрессии в Excel с помощью встроенных функций.
9. Охарактеризуйте понятие *тренд*.
10. Алгоритм построения уравнения регрессии, используя линию тренда.
11. Охарактеризуйте процесс построения уравнения регрессии в пакете «Анализ данных».

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 13

Применение инструментальных средств для обработки экспериментальных данных методом однофакторного дисперсионного анализа

Цель и содержание: изучить основные принципы дисперсионного анализа для обработки экспериментальных данных, рассмотреть возможность обработки экспериментальных данных методом однофакторного дисперсионного анализа с помощью

различных инструментальных средств (Excel, MathCAD).

Теоретическое обоснование

Под *дисперсионным анализом* понимается такая совокупность статистических методов, при помощи которых можно обработать эмпирически полученные данные (экономические, инженерные и т. д.).

Основной целью дисперсионного анализа является не только установление взаимосвязи между параметрами, а также и сравнение эффектов воздействия на полученном экспериментальном материале.

Методы дисперсионного анализа применяются для проверки гипотез о наличии взаимодействий между результативным признаком и признаками –

факторами (исследуемыми факторами), а также для установления силы влияния факторов друг на друга и их дальнейшего взаимодействия.

Таким образом, **дисперсионный анализ** – статистический анализ результатов, зависящих от качественных факторов.

В теории статистики выделяют однофакторный и двухфакторный дисперсионный анализ.

При рассмотрении однофакторного дисперсионного анализа будем предполагать, что результаты наблюдений для различного числа уровней представляют собой выборки из нормально распределенных генеральных совокупностей.

При проведении однофакторного дисперсионного анализа каждая из генеральных совокупностей имеет свои средние значения и дисперсии, которые предполагаются равными и не зависят от уровней. Тогда задачей однофакторного анализа является проверка гипотезы о равенстве средних для рассматриваемых совокупностей.

Рассмотрим выполнение однофакторного дисперсионного анализа на следующем примере: исследовать влияние рекламных кампаний на спрос.

В данном случае проводимая рекламная кампания будет считаться фактором. Обозначим T – каждую из рекламных кампаний, тогда T_i ($i = \overline{1, m}$), m – число проводимых кампаний.

Пусть x_{ij} – спрос на товар в результате проведения i – й рекламой кампании в j – м году ($j = \overline{1, n_i}$, n_i – число лет, в течении которых проводилась рекламная кампания). Представим все данные в виде таблицы (таблица 13.1).

Таблица 13.1

	1	2	3	...	n_i
T_1	x_{11}	x_{12}	x_{13}	...	x_{1n_1}
T_2	x_{21}	x_{22}	x_{23}	...	x_{2n_2}
...
T_m	x_{m1}	x_{m2}	x_{m3}	...	x_{mn_m}

Для обработки статистической информации методом однофакторного дисперсионного анализа построим математическую модель, которая включает-

ся в следующем. Каждое из значений x_{ij} , представленное в таблице 16.1, может быть описано равенством: $x_{ij} = a_j + \varepsilon_{ij}$, где a_j – величина, определяющая спрос на товар после проведения рекламной кампании, ε_{ij} – последовательность независимых случайных величин, которые описывают суммарный вклад совокупности случайных факторов.

Далее задача применения однофакторного дисперсионного анализа сводится к уточнению вопроса: влияет ли выбор рекламной кампании на спрос или нет? С математической точки зрения сформулированная выше задача означает: на основании эмпирических данных проверить нулевую гипотезу H_0 о том, что все проводимые рекламные кампании T_i одинаково эффективны, то есть $H_0: a_1 = a_2 = \dots = a_n$. Дисперсионный анализ в этом случае основывается на сравнении двух точечных оценок неизвестной дисперсии $D = \sigma^2$.

Предположим, что одной из найденных оценок не будет зависеть от того верна ли гипотеза H_0 , а для другой оценки это же предположение оказывается существенным. Если же полученные точечные оценки принимают близкие по величине значения, то нулевая гипотеза H_0 принимается. В противном случае, гипотеза H_0 отвергается.

Выведем формулы для получения этих оценок. Для каждой из строк, содержащей значения фактора, найдем значения

\bar{x}_i . Затем, используя найденные величины, получим формулу для вычисления численного значения оценки S_1 :

$$S_1^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2. \quad (1)$$

Для получения оценки S следует вначале найти значение по совокупности x_{ij}

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n x_{ij}, \quad (2)$$

где $n = n_1 + n_2 + \dots + n_m$, а затем на основании полученного значения можно рассчитать величину оценки S :

$$S^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2. \quad (3)$$

Если гипотеза H_0 принимается, то точечные оценки S_1^2 и S^2 будут независимыми, а случайная величина $\frac{S^2}{\sigma^2}$ подчиняется χ^2 -распределению с $(n - 1)$ степенью свободы.

Проведем сравнение значений величин точечных оценок S_1^2 и S^2 . Если гипотеза H_0 принимается, то величина F_H , значение которой можно вычислить по формуле:

$$F_H = \frac{S^2 \cdot (n - m)}{S_1^2 \cdot (n - 1)} \quad (4)$$

будет подчиняться распределению Фишера с $(n - 1)$ и $(n - m)$ степенями свободы.

Пусть α – уровень значимости, причем $\alpha \rightarrow 0$. Найдем величину x_α , являющуюся корнем уравнения:

$$F_{n-1, n-m}(x_\alpha) = 1 - \alpha. \quad (5)$$

Для реализации методов дисперсионного анализа, как одного из способов обработки статистической информации, следует сравнить найденные значения x_α и F_H . Если выполняется неравенство $F_H > x_\alpha$, то нулевая гипотеза H_0 отвергается, в противном же случае гипотеза H_0 считается верной.

Оценка степени влияния качественных факторов на результативный признак осуществляется при помощи статистической величины выборочный коэффициент детерминации r^2 , значение которого можно найти по формуле:

$$r^2 = \frac{S^2 - S_1^2}{S^2}, \quad (6)$$

где S^2 – оценка полной дисперсии, рассчитываемая по формуле (3).

Если нулевая гипотеза H_0 принимается, то есть $a_1 = a_2 = \dots = a_n = a$, то оценкой математического ожидания выборки служит величина \bar{x} , тогда оценкой дисперсии $\sigma^2 - \frac{S_1^2}{n - m}$. В противном случае оценкой математического ожи-

дания всей совокупности a_i является величина x_i , а дисперсий σ^2 для всех уровней испытания – $\frac{S_1}{n - m}$, а эффект влияния i – го на результативный признак можно найти по формуле: $\bar{x}_i - \bar{x}$, где \bar{x}_i – выборочное среднее значение i – го фактора, \bar{x} – выборочное среднее по всей совокупности.

В пакете MathCAD для реализации методов однофакторного дисперсионного анализа используются стандартные (основные) панели инструментов.

Однофакторный дисперсионный анализ при обработке статистической информации может быть выполнен и при помощи табличного процессора Excel. Однофакторный дисперсионный анализ в Excel осуществляется при помощи пакета «Анализ данных», в котором выбирается инструмент «Однофакторный дисперсионный анализ».

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Запустите систему MathCAD и проверьте, установлен ли в ней режим автоматического выполнения вычислений.

2. Введите исходные данные для проведения однофакторного дисперсионного анализа. Эти данные представьте в виде матрицы с элементами x_{ij} , где i – уровень фактора, j – номер наблюдения.

3. Определите значения количества факторов.

4. Рассчитайте объем выборки для каждого из уровня факторов и общий объем выборки.

5. Используя формулу (6) из раздела «Теоретическое обоснование» лабораторной работы № 10 и формулу (3) из этого же раздела данной работы, вычислите значения выборочных средних для каждой из групп факторов и по всей совокупности в целом.

6. При помощи формул $S_1^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$ и $S_2^2 = \frac{1}{n} \sum_{i=1}^m n_i \cdot (\bar{x}_i - \bar{x})^2$

определите численные значения этих оценок.

7. С помощью формулы (4) из раздела «Теоретическое обоснование» вычислите значение критерия Фишера F_H .

8. Задайте уровень значимости α .

9. Для заданного уровня значимости α осуществите проверку нулевой гипотезы о независимости полученного конечного значения от уровня фактора.

10. Если нулевая гипотеза H_0 независимости от фактора отклоняется, то вычислите значение коэффициента детерминации r^2 .

11. Рассчитайте значения оценок параметров распределения величины для каждого из факторов.

12. Изобразите на одном графике плотности вероятностей соответствующих распределений.

Задание № 1. Уровень продаж в зависимости от проведения рекламных кампаний представлен в виде таблицы. Установите влияние типа используемой рекламы на объем продаж товара для заданного уровня значимости α и меру влияния каждого типа рекламы. Определите параметры и постройте плотности вероятностей распределения числа продаж для каждого типа рекламы.

Методика и порядок выполнения задания 2

1. Загрузите табличный процессор Excel.
2. Введите исходные данные для проведения однофакторного дисперсионного анализа.
3. Активизируйте главное меню Excel, в нем выберите опцию **Сервис**, а затем команду Анализ данных и нажмите клавишу Enter.
4. В диалоговом окне «Анализ данных» выберите инструмент «Однофакторный дисперсионный анализ» и нажмите кнопку ОК.
5. Для его выполнения заполните одноименное окно следующим образом:
1) в поле «входной интервал» введите диапазон ячеек, задающих исходные данные для проведения дисперсионного анализа; 2) укажите способ группировки данных по «строкам» или по «столбцам»; 3) введите заданный уровень

значимости; 4) в поле «выходной интервал» укажите ячейку, в которую будут выведены результаты дисперсионного анализа.

Задание № 2. Используя пакет «Анализ данных», проведите дисперсионный анализ для установления зависимости объема продаж от проведения рекламных кампаний.

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает студент.

Индивидуальное задание соответствующее своему варианту студенту необходимо получить у преподавателя.

Уровень значимости для выполнения однофакторной дисперсионного анализа выбрать следующим образом: для вариантов с нечетными номерами выбрать величину $\alpha = 95\%$, а для четных вариантов 90%

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. Что такое дисперсионный анализ?
2. Охарактеризуйте основные методы дисперсионного анализа.
3. Сформулируйте основную цель дисперсионного анализа
4. Однофакторный дисперсионный анализ. Общее представление.
5. Алгоритм выполнения дисперсионного анализа в пакет MathCAD.

6. Алгоритм реализации методов однофакторного дисперсионного анализа в табличном процессоре Excel.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Лабораторная работа 14

Использование инструментальных средств для обработки экспериментальных данных методом двухфакторного дисперсионного анализа

Цель и содержание: изучить основные принципы двухфакторного дисперсионного анализа как одного из методов обработки экспериментальных данных, показать реализацию двухфакторного дисперсионного анализа с помощью различных инструментальных средств.

Теоретическое обоснование

Предположим, что случайная величина ζ зависит от двух факторов. Пусть A_i , $i=1,2,\dots,m_A$ и B_j , $j=1,2,\dots,m_B$ – уровни факторов для некоторых экономических показателей. Результаты измерений случайной величины ζ приведены в таблице (таблица 14.1), причем для проведения двухфакторного дисперсионного анализа будем считать, что для каждого сочетания уровней факторов приведен результат только одного наблюдения. В этом случае общее число наблюдений можно рассчитать по формуле:

$$n = m_A \cdot m_B. \quad (1)$$

Таблица 14.1

	1	2	3	...	m_B
1	x_{11}	x_{12}	x_{13}	...	x_{1m_B}
2	x_{21}	x_{22}	x_{23}	...	x_{2m_B}
...
m_A	$x_{m_A 1}$	$x_{m_A 2}$	$x_{m_A 3}$...	$x_{m_A m_B}$

Введем обозначения: a_i – значение математического ожидания случайной величины ζ для уровня A_i ($i = \overline{1, m_A}$), b_j – математическое ожидание случайной величины ζ для уровня B_j ($j = \overline{1, m_B}$). Если при изменении значений фактора A величины математического ожидания на всех его уровнях одинаковы, то говорят, что величина ζ не зависит от фактора A , тогда как в противном случае

она зависит от фактора А. Аналогичное предложение можно сформулировать и для фактора В.

Будем считать, что значения математических ожидания случайной величины ζ для факторов А и В a_i и b_j неизвестны.

Тогда задача применения двухфакторного дисперсионного анализа для исследования зависимости между факторами А и В сводится к проверке двух гипотез: $H_A : a_1 = a_2 = \dots = a_{m_A}$; $H_B : b_1 = b_2 = \dots = b_{m_B}$. Для решения задачи проверки гипотез H_A и H_B предположим, что выполнены условия:

- 1) наблюдения при различных сочетаниях уровней факторов независимы;
- 2) при всех сочетаниях уровней факторов случайная величина ζ нормально распределена с одной и той же дисперсией σ^2 .

Изменчивость наблюдаемых факторов при переходе от одного значения случайной величины к другому обуславливается как изменениями на уровнях факторов, так и случайными неконтролируемыми факторами. При этом изменчивость, вызванную случайными неконтролируемыми факторами, называют *остаточной*.

Для проведения двухфакторного дисперсионного анализа рассчитаем значения общей средней, средней изменчивости, вызываемой факторами А и В, общей характеристики изменчивости случайной величины, обусловленной случайными факторами, общей изменчивости случайной величины ζ :

- 1) общая средняя:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m_A} \sum_{j=1}^{m_B} x_{ij}, \quad (2)$$

- 2) средняя изменчивость, вызванная фактором А:

$$\sigma_A^2 = \frac{S_A^2}{n} = \frac{m_B}{n} \sum_{i=1}^{m_A} (\bar{x}_i - \bar{x})^2, \quad (3)$$

где $\bar{x}_i = \frac{m_B}{n} \sum_{j=1}^{m_B} x_{ij}$.

3) средняя изменчивость, вызванная фактором В:

$$\sigma_B^2 = \frac{S_B^2}{n} = \frac{m_A}{n} \sum_{j=1}^{m_B} (\bar{x}_{\cdot j} - \bar{x})^2, \quad (4)$$

где $\bar{x}_{\cdot j} = \frac{m_A}{n} \sum_{i=1}^{m_A} x_{ij}$

4) общая характеристика изменчивости, обусловленная случайными факторами:

$$\sigma^2 = \frac{S_0^2}{n} = \frac{1}{n} \sum_{i=1}^{m_A} \sum_{j=1}^{m_B} (\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2, \quad (5)$$

5) общая изменчивость случайной величины ζ :

$$\sigma^2 = \frac{S^2}{n} = \frac{1}{n} \sum_{i=1}^{m_A} \sum_{j=1}^{m_B} (\bar{x}_{ij} - \bar{x})^2. \quad (6)$$

Проверка гипотезы H_A основывается на сравнении величин σ_A^2 и σ_0^2 , а $H_B - \sigma_B^2$ и σ_0^2 . Если гипотеза H_A оказывается верной, то величина, определяемая равенством

$$F_A = \frac{\sigma_A^2}{\sigma_0^2}, \quad (7)$$

подчиняется распределению Фишера со степенями свободы $k = m_A - 1$ и $\ell = (m_A - 1) \cdot (m_B - 1)$.

Пусть α – заданный уровень значимости, тогда найдем для него правостороннюю критическую точку x_α , которая является решением уравнения $F_{k,\ell}(x_\alpha) = 1 - \alpha$. Если значение параметра F_A , полученное в результате измерений, удовлетворяет неравенству $F_A < x_\alpha$, то гипотеза H_A принимается. В противном случае эта гипотеза отвергается и можно сделать вывод о том, что изменение фактора А влияет на изменение случайной величины ζ .

Мерой этого влияния является **коэффициент детерминации**, значение которого можно вычислить по формуле:

$$r_A = \frac{\sigma_A^2}{\sigma^2}. \quad (8)$$

Этот коэффициент показывает, какая доля общей изменчивости случайной величины ζ обусловлена изменением фактора А.

Аналогичным образом осуществляется проверка гипотезы H_B .

Используя механизм двухфакторного дисперсионного анализа, можно получить более четкое представление о природе случайной величины ζ . В этом случае её модель на i – м уровне фактора А и j – м уровне фактора В можно представить в виде:

$$\zeta_{ij} = a + \alpha_i + \beta_j + \varepsilon_{ij}, \quad (9)$$

где $i = \overline{1, m_A}$, $j = \overline{1, m_B}$, a – генеральное среднее случайной величины ζ , α_i – величина, описывающая эффект влияния фактора А на величину ζ на i – м уровне фактора А, β_j – параметр, описывающий эффект влияния фактора В на случайную величину ζ на j – м уровне фактора В, ε_{ij} – случайные величины, описывающие эффект влияния случайных факторов, причем все значения ε_{ij} подчиняются нормальному закону распределения с $\mu = 0$ и неизвестным значением σ .

Если гипотезы H_A и H_B признаются верными, тогда в рассматриваемой модели $\alpha_1 = \alpha_2 = \dots = \alpha_{m_A} = 0$ и $\beta_1 = \beta_2 = \dots = \beta_{m_B} = 0$.

Величина \bar{x} является оценкой параметра a , а $\frac{S_0^2}{(m_A - 1) \cdot (m_B - 1)}$ – несмещенная оценка σ^2 . В том случае, если гипотезы H_A и H_B отвергаются, то есть не признаются верными, то говорят о выполнении условий:

- 1) оценка параметра a равна \bar{x} ,
- 2) оценка параметра $\alpha_i = \bar{x}_{.i} - \bar{x}$;
- 3) оценка $\beta_j = \bar{x}_{.j} - \bar{x}$;

4) несмещенная оценка параметра $\sigma^2 = \frac{S_0^2}{(m_A - 1) \cdot (m_B - 1)}$.

Для проведения двухфакторного дисперсионного анализа в табличном процессоре Excel используется надстройка «Анализ данных», в которой выбирается инструмент «Двухфакторный дисперсионный анализ с повторениями» или «Двухфакторный дисперсионный анализ без повторений».

Аппаратура и материалы, необходимые для выполнения лабораторной работы приведены в приложении А.

Указания по технике безопасности. Техника безопасности при выполнении лабораторной работы совпадает с общепринятой для пользователей персональных компьютеров (приложение Б).

Методика и порядок выполнения работы

Методика и порядок выполнения задания 1

1. Запустите MathCAD, и проверьте установлен ли в нем режим автоматического выполнения вычислений.

2. Задайте матрицу наблюдений и число уровней для каждого фактора.

3. Используя формулы (1) и (2) из раздела «Теоретическое обоснование», рассчитайте значения объема выборки и генеральной средней.

4. С помощью формул $\bar{x}_i = \frac{m_B}{n} \sum_{i=1}^{m_A} x_{i,j}$ и $\bar{x}_j = \frac{m_A}{n} \sum_{j=1}^{m_B} x_{i,j}$ найдите значения групповых средних.

5. По формулам (4) – (6) раздела «Теоретическое обоснование» вычислите значение оценок для дисперсии.

6. Введите значение уровня значимости α .

7. Для заданного уровня значимости найдите величину F_H , используя встроенную функцию $qF(1 - \alpha, k, \ell)$, то есть найдите решение уравнения $F(x) = 1 - \alpha$ для распределения Фишера F с соответствующим числом степеней свободы.

8. С помощью формулы (7) из раздела «Теоретическое обоснование» вычислите значение критерия для фактора А.

9. Сравните значения критерия для фактора А и величину F_H , а затем сформулируйте вывод о справедливости гипотезы H_A .

10. Если гипотеза H_A отвергается (отклоняется), то найдите, используя формулу (8) из раздела «Теоретическое обоснование», коэффициент детерминации для первого фактора (фактора А).

11. Повторите цикл расчетов для фактора В.

12. Найдите значения оценок параметров распределения исследуемой случайной величины.

Задание № 1. Используя пакет MathCAD, проведите двухфакторный дисперсионный анализ данных, приведенных в таблице.

Методика и порядок выполнения задания 2

1. Запустите табличный процессор Excel.

2. Введите исходные данные для проведения дисперсионного анализа.

3. Загрузите пакет «Анализ данных», выполнив следующую последовательность действий: **F10** → **Сервис** → **Анализ данных**.

4. В диалоговом окне «Анализ данных» выберите инструмент «Двухфакторный дисперсионный анализ без повторений».

5. В появившемся диалоговом окне в поле «**входной интервал**» укажите диапазон ячеек, для которого следует провести двухфакторный дисперсионный анализ, в поле «**выходной интервал**» - ячейку (или группу ячеек), в которую (или которые) следует поместить результаты дисперсионного анализа.. Отметим, что при проведении двухфакторного дисперсионного анализа в табличном процессоре Excel. Уровень значимости по умолчанию устанавливается равным 95%.

6. Для завершения нажмите кнопку ОК.

Задание № 2. При помощи надстройки «Анализ данных» проведите двухфакторный дисперсионный анализ данных, приведенных в таблице.

Варианты заданий

Вариант для выполнения заданий, приведенных в лабораторной работе, определяется номером персонального компьютера, за которым работает сту-

дент. Индивидуальное задание соответствующее своему варианту студенту необходимо получить у преподавателя.

Уровень значимости для выполнения двухфакторного дисперсионного анализа выбрать следующим образом: для вариантов с нечетными номерами выбрать величину $\alpha = 95\%$, а для четных вариантов 90% .

Содержание отчета и его форма

Требования к содержанию отчета по лабораторной работе и его форме приведены в приложении В.

Контрольные вопросы и защита работы

1. В чем заключается двухфакторный дисперсионный анализ?
2. Основная задача обработки статистической информации методом двухфакторного дисперсионного анализа.
3. Последовательность действий при обработке статистической информации методами двухфакторного дисперсионного анализа в MathCAD.
4. Алгоритм выполнения двухфакторного дисперсионного анализа в Excel.

Защита работы

Механизм защиты лабораторной работы приведен в приложении Г.

Список рекомендуемой литературы

Основная учебная литература

1. Королев, В. Ю. Теория вероятностей и математическая статистика : учебник / В. Ю. Королев ; МГУ им. М. В. Ломоносова, фак. вычислит. матем. и кибернетики . - М. : Проспект, 2006. - 160 с.

2. Горелова, Г. В. Теория вероятностей и математическая статистика в примерах и задачах с применением Excel : учеб. пособие для студентов вузов / Г. В. Горелова, И. А. Кацко. - 4-е изд. - Ростов н/Д : Феникс, 2006. - 480 с.

3. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – М.: Физматлит, 2006. – 816 с.

Дополнительная литература

1. Дьяконов, В. П. Mathcad 11/12/13 в математике : справочник / В. П. Дьяконов. - М. : Горячая линия-Телеком, 2007. - 958 с.

2. Каганов, В. И. Компьютерные вычисления в средах EXCEL и MathCAD / В. И. Каганов . - М. : Горячая линия-Телеком, 2003. - 328 с.

3. Семакин, И. Г. Информационные системы и модели : учеб. пособие / И. Г. Семакин, Е. К. Хеннер . - 2-е изд. - М. : БИНОМ. Лаборатория знаний, 2007. - 303 с.

4. Плис А. И., Сливина Н. А. MathCAD. Математический практикум для инженеров и экономистов: Учеб. пособие. – 2 – е изд., переработ. и доп. – М.: Финансы и статистика, 2003. – 656 с.

Приложение А

Аппаратура и материалы

Для выполнения лабораторной работы необходим персональный компьютер со следующими характеристиками: процессор Intel с тактовой частотой 2000 МГц и выше; оперативная память – не менее 1024 Мбайт; свободное дисковое пространство – не менее 1,2 Гбайт; устройство для чтения компакт-дисков; монитор типа Super VGA (число цветов – 256) с диагональю не менее 17 ". **Программное обеспечение** – операционная система WINDOWS 2000 / XP, пакет автоматизации математических вычислений MathCAD, MS Excel.

Приложение Б

Указания по технике безопасности.

1. Самостоятельно не производить ремонт персонального компьютера, установку и удаление программного обеспечения.
2. В случае неисправности персонального компьютера сообщить об этом обслуживающему персоналу лаборатории (оператору, администратору).
3. Соблюдать правила техники безопасности при работе с электрооборудованием.
4. Не касаться электрических розеток металлическими предметами.

5. Рабочее место пользователя персонального компьютера должно содержаться в чистоте.

6. Не разрешается возле персонального компьютера принимать пищу, напитки.

Приложение В

Содержание отчета и его форма

Отчет по лабораторной работе должен состоять из:

- 1) названия лабораторной работы;
- 2) ответов на контрольные вопросы;
- 3) формулировки индивидуальных заданий и порядка их выполнения.

Отчет о выполнении лабораторной работы в письменном виде сдается преподавателю.

Приложение Г

Защита работы

Перед выполнением лабораторной работы каждый студент получает индивидуальное задание. Защита лабораторной работы происходит только после его выполнения (индивидуального задания). При защите лабораторной работы студент отвечает на контрольные вопросы, приведенные в конце, и поясняет выполненное индивидуальное задание. Ход защиты лабораторной работы контролируется преподавателем.