**С.Н. Мухина**

**Методы математической**

**обработки информации.**

**Реализация в среде MathCAD**

**Учебное пособие**

**Калининград**

**Издательство БГАРФ**

**2019**

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО РЫБОЛОВСТВУ

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

«Калининградский государственный технический университет»

Балтийская государственная академия рыбопромыслового флота

С.Н. Мухина

**Методы математической**

**обработки информации.**

**Реализация в среде MathCAD**

Учебное пособие

для студентов, обучающихся

по специальности 10.05.03

«Информационная безопасность

автоматизированных систем»

Калининград

Издательство БГАРФ

2019

УДК 519.22

**Мухина С.Н.** Методы математической обработки информации. Реализация в среде Mathcad: учебное пособие / *С.Н. Мухина*. – Калининград: Изд-во БГАРФ, 2019. – с.

Учебное пособие включает следующие разделы: математическую статистику, корреляционный и регрессионный анализ, теорию массового обслуживания. Структура пособия ориентирована на использование пакета компьютерной математики. Задачи снабжены подробными решениями и демонстрационными примерами.

Пособие предназначено для организации компьютерных лабораторных работ и самостоятельной работы студентов, обучающихся по специальности «Информационная безопасность автоматизированных систем».

Печатается по решению редакционно-издательского совета Балтийской государственной академии рыбопромыслового флота.

**Рецензенты:**

**Ганиева Р.А.,**кандидат педагогических наук,

зав. кафедрой математики филиала ВУНЦ ВМФ

«Военно-морская академия» в г. Калининграде;

**Руденко А.И*.***, кандидат физико-математических наук,

доцент секции прикладной математики

кафедры высшей математики КГТУ БГАРФ

© «БГАРФ» ФГБОУВПО «КГТУ» (БГАРФ), 2019

**ОГЛАВЛЕНИЕ**

**Введение**

**Глава 1. Математические основы выборочного метода**

1.1. Выборка и ее представление

1.2. Числовые характеристики вариационного ряда

*Компьютерная лабораторная работа № 1*

*«Вариационные ряды, их числовые характеристики и графическое изображение»*

**Глава 2. Точечное и интервальное оценивание параметров**

**распределения**

2.1. Методы получения точечных оценок

2.2 Интервальные оценки

*Компьютерная лабораторная работа № 2*

*«Точечное и интервальное оценивание параметров распределения»*

**Глава 3. Статистические гипотезы**

3.1. Статистическая гипотеза и общая схема ее проверки

3.2. Построение теоретического закона распределения по опытным данным. Проверка гипотез о законе распределения

*Компьютерная лабораторная работа № 3*

*«Проверка гипотезы о виде закона распределения СВ по критерию Пирсона»*

3.3. Гипотезы о значениях и сравнении числовых характеристик и однородности выборок

3.3.1. Проверка гипотез о числовых значениях параметров

3.3.2. Проверка гипотез о равенстве числовых характеристик двух совокупностей

3.3.3. Проверка гипотез об однородности выборок

*Компьютерная лабораторная работа № 4*

*«Статистические гипотезы»*

**Глава 4. Основы регрессионного и корреляционного анализа**

4.1.Функциональная, статистическая и корреляционная зависимость

4.2. Линейная регрессия

4.3. Нелинейная однофакторная регрессия

*Компьютерная лабораторная работа № 5*

*«Корреляционно-регрессионный анализ»*

**Глава 5. Системы массового обслуживания (СМО)**

5.1. Структура и классификация систем массового обслуживания

5.2. Марковский случайный процесс в СМО. Уравнения

Колмогорова

5.3. Расчет показателей эффективности СМО с отказами

5.3.1. Одноканальные СМО с отказами

5.3.2. Многоканальные СМО с отказами

5.4. Расчет показателей эффективности СМО с ожиданием

(с очередью)

5.4.1. Одноканальные СМО с неограниченной очередью

5.4.2. Многоканальные СМО с неограниченной очередью

5.4.3. Системы массового обслуживания с ограниченной очередью

(одноканальные и многоканальные)

*Компьютерная лабораторная работа № 6*

*«Расчет показателей эффективности систем массового*

*обслуживания»*

**Список использованной литературы**

**Приложение 1.** Типы числовых характеристик

**Приложение 2.** Библиотека стандартных распределений

**Приложение 3.** Библиотека типовых функций регрессии Mathcad

*Статистика - это математическая теория того,*

*как узнать нечто о мире через опыт*

W. Thompson

*Первая задача математической статистики состоит*

*в указании способов сбора и группировки (если данных очень много) статистических сведений.*

*Вторая задача математической статистики состоит*

*в разработке методов анализа статистических данных*

*в зависимости от целей исследования.    
. . . Итак, задача математической статистики состоит*

*в создании методов сбора и обработки статистических данных*

*для получения научных и практических выводов.*

В. Е. Гмурман

*Статистика как наука является*

*одним из разделов прикладной математики,*

*и ее можно рассматривать как математику,*

*применяемую при разработке результатов массового наблюдения. . .*

*Статистику можно рассматривать как:*

*1) учение о совокупностях,*

*2) учение о вариации и*

*3) учение о методах приведения данных в компактной форме*

Р. А. Фишер

**Введение**

Термин «статистика» происходит от латинского слова «статус» (*status*) - состояние. Как научная дисциплина статистика оформилась в 18 веке. Термин статистика связывался с системой описания фактов, характеризующих состояние государства.

«В настоящее время статистика включает в себя и большее, и в то же время более определенное содержание. А именно, можно сказать, что статистика состоит из следующих трех разделов:

- сбор статистических сведений, т. е. сведений, характеризующих отдельные единицы каких-либо массовых совокупностей;

- статистическое исследование полученных данных, заключающееся в выяснении тех закономерностей, которые могут быть установлены на основе данных массового наблюдения;

3) разработка приемов статистического наблюдения и анализа статистических данных. Последний раздел, собственно, и составляет содержание математической статистики» (Б. В. Гнеденко).

«В наше время принято считать, что математическая статистика есть наука, изучающая теорию принятия решений в условиях неопределенности. Это определение математической статистики выкристаллизовывалось в результате многих лет ее развития. Достоинство этого определения состоит в том, что оно в сжатой и ясной форме излагает научное существо статистики» (Г. Чернов, Л. Мозес. Элементарная теория статистических решений).

В данном пособии представлено решение основных задач математической статистики с помощью компьютерной математики.

Пособие состоит из пяти глав. Первая глава посвящена задачам описательной статистики, вторая глава посвящены задачам оценивания, третья глава - проверке статистических гипотез, в четвертой рассматриваются методы исследования зависимостей в экспериментальных данных. В пятой главе рассматриваются основы теории массового обслуживания.

В начале каждой главы излагаются теоретические сведения, при необходимости описываются особенности программной реализации того или иного метода. Далее излагаются примеры решения подобных задач в пакете Mathcad, сопровождаемые подробными комментариями. В конце каждой главы предлагаются варианты заданий для выполнения компьютерной работы.

Разработанный компьютерный практикум позволит приобрести студентам практические навыки решения вероятностно-статистических задач в среде Mathcad**.**

Предполагается, что студенты знают основы теории вероятностей и имеют начальное представление о работе в математических пакетах. В данном пособии принципы программной реализации изложены на примере пакета Mathcad, однако все предлагаемые лабораторные работы могут быть выполнены и в других математических пакетах (Maple, Mathlab и др.).

**Глава 1. Математические основы выборочного метода**

*Статистика в общем смысле должна изучать явления*

*с целью наблюдения массовых характеристик*

*путем группировки и подсчета,*

*если при этом удается выявить то,*

*что поддается дальнейшему изучению.*

J. Н. van Zanten. Lehrbuch der statistische methode

**1.1. Выборка и ее представление**

Совокупность возможных значений случайной величины *Х (*СВ *Х)*, которые принципиально могут быть получены в результате экспериментов, называют **генеральной совокупностью**.

В математической статистике свойства СВ *Х* изучаются на основе некоторого ограниченного множества данных, полученных в результате экспериментов. Это ограниченное множество называют **выборка (выборочная совокупность)**.

Объем выборки – число элементов выборки. Различают малые и большие выборки. Примерной границей служит

Важным принципом формирования выборки является обеспечение случайного отбора в одинаковыхусловиях**.**

**Вариационный ряд** – совокупность *всех* элементов выборки, записанных в неубывающем порядке. В вариационном ряду могут присутствовать одинаковые элементы.

**Статистический ряд** – последовательность различных элементов, записанных в возрастающем порядке и соответствующие им частоты или относительные частоты. Удобно представлять в виде таблицы.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | … |  |  |
|  |  | … |  |  |
|  |  | … |  |  |

Решение задач в условиях большого объема выборки может потребовать «сжатия» полученных данных и перехода к интервальному, а затем к группированному статистическому ряду.

**Группированный статистический ряд –** совокупность середин промежутков и соответствующих им частот.

Группированный статистический ряд также удобно представлять в виде таблицы. Группированный статистический ряд преимущественно используется при анализе вида распределения СВ *Х* по данным наблюдений. Искажая информацию, ведет к появлению дополнительных погрешностей, однако уровень таких погрешностей не является значительным.

Важным вопросом является определение ширины промежутков (интервалов) *h* и их числа *m*.

***Рекомендации по выбору взаимосвязанных значений h и m***

1. ; .

. – формула Стерджесса, или

, тогда

***Замечание.*** *В формулах квадратные скобки**- целая часть числа.*

Графическим изображением статистического ряда служит **полигон частот** или относительных частот – аналог многоугольника распределения дискретной случайной величины. Интервальный вариационный ряд графически изображают с помощью **гистограммы.**

Гистограмма (от др.-греч. ἱστός - столб и γράμμα - черта, буква, написание) - способ графического представления табличных данных.

*Термин «гистограмма» был введен знаменитым статистиком Карлом Пирсоном для обозначения «общей формы графического представления».*

Гистограмма – это ступенчатая фигура из прямоугольников, основание которых равно *h*, а высота каждого прямоугольника

Имеет место следующая теорема о приближении гистограммы к функции плотности распределения.

**Теорема.** *Если плотность распределения элементов выборки является непрерывной функцией, то при неограниченном увеличении числа интервалов группировки имеет место сходимость по вероятности гистограммы к плотности распределения.*

Гистограммы обычно используются для предварительного анализа законов распределения исследуемой СВ *Х*, выдвижения гипотезы о виде распределения. Площадь всей гистограммы, построенной с использованием равна единице, что позволяет непосредственно сопоставлять такую гистограмму с графиками функций плотности.

*Диаграммы могут облегчить предварительное исследование*

*большинства исходных данных.*

*Они ничего не доказывают, но наглядно показывают*

*отличительные черты.*

*Они поэтому не заменяют таких критических исследований,*

*которым могут быть подвергнуты исходные данные,*

*но они ценны тем, что подсказывают такие исследования и*

*поясняют выводы, основанные на них.*

Роналд Фишер (1925)

**Историческая справка**

|  |  |
| --- | --- |
| ***Август Рональд Эйлмер Фишер***  *(17.2. 1890, Ист-Финчли, Лондон, Великобритания, - 29.7. 1962, Аделаида, Австралия) - английский статистик, биолог-эволюционист и генетик. Его называли*«величайшим последователем Дарвина». | ***Карл Пирсон***  *(27.3.1857, Лондон, Великобритания - 27.4.1936, там же) - английский математик, биолог, философ-позитивист. Профессор прикладной математики и механики (с 1884), а затем евгеники (с 1911) Лондонского университета.* |
| Ð Ð¾Ð½Ð°Ð»ÑÐ´ Ð¤Ð¸ÑÐµÑ | *http://upload.wikimedia.org/wikipedia/commons/2/21/Karl_Pearson_2.jpg* |

В среде Mathcad для обработки выборочных данных и построения гистограммы используются функции, перечисленные в таблицах 1.1, 1.2.

Таблица 1.1

**Формирование массива данных**

|  |  |
| --- | --- |
| ***Параметры матрицы*** | |
| число строк массива М | rows(M) |
| число столбцов массива М | cols(M) |
| максимальный элемент массива М | max(M) |
| минимальный элемент массива М | min(M) |
| выноска столбца матрицы М (находится в  панели «матрицы») |  |
| ***Образование новых матриц*** | |
| объединение матриц А и В «бок в бок» (матрицы должны иметь одинаковое число строк) | augment(A,B) |
| объединение матриц А и В «сверху вниз» (матрицы должны иметь одинаковое число столбцов) | stack(A,B) |
| ***Сортировка массива*** | |
| сортировка элементов массива по возрастанию | sort(M) |
| перестановка элементов в обратном порядке | reverse(M) |

Таблица 1.2

**Группированный статистический ряд**

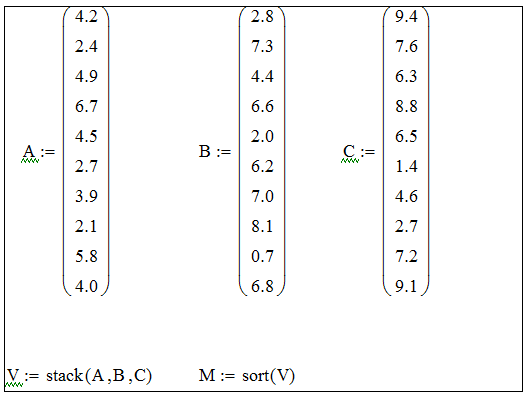
|  |  |
| --- | --- |
| ***Встроенная функция***  ***Mathcad*** | ***Описание*** |
| hist(x,M) | Вектор частот попадания данных в заданные интервалы  x – матрица значений случайной величины;  М – массив данных после  ранжирования |
| histogram(m,M) | Матрица гистограммы, где  m – число интервалов случайной величины;  М – массив данных после  ранжирования |

**Пример 1.1.** Имеются следующие данные о размерах основных фондов (в млн руб.) 30 предприятий:

4,2; 2,4; 4,9; 6,7; 4,5; 2,7; 3,9; 2,1; 5,8; 4,0; 2,8; 7,3; 4,4; 6,6; 2,0; 6,2; 7,0; 8,1; 0,7; 6,8; 9,4; 7,6; 6,3; 8,8; 6,5; 1,4; 4,6; 2,0; 7,2; 9,1.

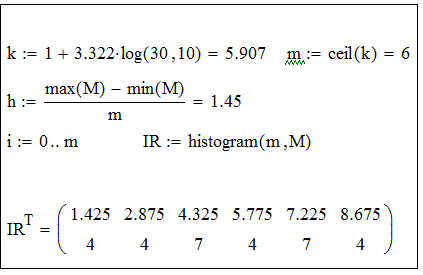
**Решение**

1. Формируем массив данных (рис. 1.1). Для этого вводим данные с помощью, например, трех матриц A, B, C. Используя встроенную функцию **stack(A,B,С)**, создаем матрицу *V*, содержащую все статистические данные. Выполняем сортировку элементов матрицы *V* по возрастанию с помощью функции **sort(*V*).**

****

*Рис. 1.1. Формирование массива данных*

2. Составляем группированный статистический ряд *IR* (рис. 1.2), рассчитав число интервалов *т* по формуле Стерджеса, округлив до ближайшего целого числа. Находим шаг *h.* Нумеруем концы интервалов, используем встроенную функцию **histogram(m,M).**

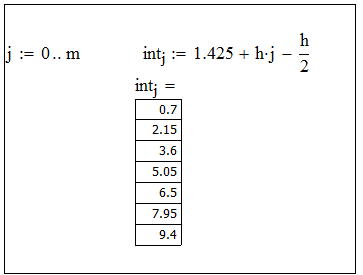
****

*Рис. 1.2. Построение группированного*

*статистического ряда*

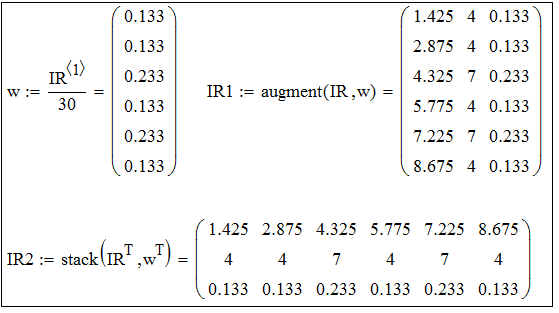
Получен ряд, разбитый на 6 интервалов. Элементы первой строки – середины интервалов; элементы второй строки – частоты попадания (количество вариант, попавших в данный интервал).

3. Зададим границы интервалов. Значения вариант – концов интервалов имеют вид **intj** (рис. 1.3).



*Рис. 1.3. Значения вариант – концов интервалов*

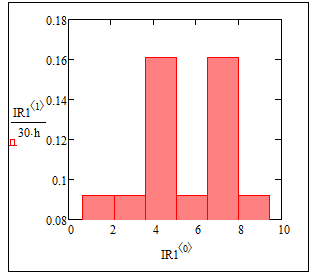
4. Вычислим – относительные частоты попадания вариант в интервалы (рис. 1.4).



*Рис. 1.4. Группированный статистический ряд частот*

*и относительных частот*

5. Определяя для каждого интервала величины приведенных относительных частот , строим гистограмму (площадь гистограммы равна 1), распределив по горизонтали интервалы, по вертикали приведенные частоты. Затем форматируем график (тип линии – solidbar – панель заливок) (рис. 1.5).



*Рис. 1.5. Построение гистограммы с помощью*

*приведенных относительных частот*

***Замечание 1.*** *Первая и третья строки (рисунок 1.4) есть аналог закона распределения дискретной случайной величины в теории вероятностей.*

***Замечание 2.*** *Обратите внимание! Нумерация строк и столбцов матрицы начинается с нуля*.

**1.2. Числовые характеристики вариационного ряда**

В теории вероятностей рассматривались числовые характеристики случайных величин.

**Числовыми характеристиками** случайной величины называют **неслучайные** численные параметры, позволяющие в предельно сжатой, компактной форме представить основные особенности закона распределения. Типы числовых характеристик и формулы для их вычисления приведены в Приложении 1.

Применительно к выборочным распределениям в математической статистике введены аналогичные характеристики (статистические), значения которых вычисляется по результатам преобразований выборочных данных. В Mathcad существуют встроенные функции, которые реализуют соответствующие вычисления выборочных числовых характеристик (таблица 1.3).

Таблица 1.3

**Встроенные функции**

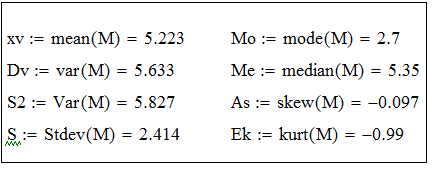
**для вычисления числовых характеристик**

|  |  |  |
| --- | --- | --- |
| ***Характеристика*** | ***Формула*** | ***Встроенная функция***  ***Mathcad*** |
| 1 | 2 | 3 |
| средняя арифметическая (выборочная средняя) |  | mean(M) |
| выборочная мода | значение признака, которому соответствует наибольшая частота (существует не всегда) | mode(M) |
| выборочная  медиана | значение признака, приходящееся на середину ранжированного ряда наблюдений и зависит от четности объема выборки | median(M) |
| размах вариации |  |  |
| выборочная  дисперсия |  | var(M) |
| исправленная  выборочная  дисперсия  (для |  | Var(M) |
| исправленное среднее квадратическое отклонение |  | Stdev(M) |
| выборочная  асимметрия |  | skew(M) |
| выборочный  эксцесс |  | kurt(M) |

**Пример 1.2.** Найтичисловые характеристики массива(среднее значение, моду, медиану, исправленную дисперсию, исправленное среднее квадратическое отклонение, коэффициенты асимметрии и эксцесса) по данным примера 1.1.

**Решение**

Используем соответствующие встроенные функции (рис. 1.6).



*Рис. 1.6. Вычисление*

*статистических характеристик*

**Компьютерная лабораторная работа № 1**

***Вариационные ряды, их числовые характеристики***

***и графическое изображение***

**Задание 1.** Провести статистическую обработку выборочных данных: сформировать массив, составить вариационный ряд, группированный статистический ряд, построить гистограмму с помощью приведенных относительных частот.

**Задание 2.** Вычислить с помощью встроенных функций числовые характеристики выборки.

Статистические данные по вариантам представлены ниже.

**1.** 176; 184; 169; 176; 182; 162; 181; 167; 176; 167; 166; 176; 162; 172; 175; 182; 166; 171; 176; 174; 174; 176; 178; 179; 178; 175; 169; 176; 172; 168; 186; 166; 163; 175; 164; 171; 178; 170; 168; 168; 180; 167; 165; 177; 169; 176; 177; 176; 174; 173.

**2.** 333,2; 332,5; 328,3; 325,5; 332,9; 331,5; 320,7; 337,9; 337,2; 340,4, 338,1; 336,2; 327,5, 329,1; 336,8; 337,9; 333,6; 327,9; 328,4; 331,4; 338,6; 349,6; 330,7; 335,6; 335,6; 348,9; 344,8; 307,1; 336,6; 325,6; 325,3; 330,9; 323; 333,5; 321,3; 349; 302,7; 338,7; 351,8; 340,8; 341,8; 320,3; 314,2; 334,6; 344,7; 341,7; 322,4; 335,1; 335; 322.

**3.** 117; 118; 121; 119; 128; 120; 121; 142; 128; 152; 105; 126; 130; 146; 133; 103; 107; 156; 135; 125; 114; 112; 104; 111; 140; 137; 135; 111; 120; 146; 125; 121; 119; 130; 131; 124; 126; 130; 115; 123; 106; 128; 147; 113; 135; 123; 110; 134; 129; 114.

**4.** 61,1; 50,5, 61,1; 44,4, 59,1; 55,2; 60,4; 93,9; 58,6; 51,1; 79,7; 86,4; 32,1; 84,5; 37,8; 33,4; 50,2; 56,6; 72,2; 56,1; 30,1; 59,6; 67; 74,8; 90,7; 48,4; 62,6; 37,4; 10,7; 88,7; 51,4; 25,7; 28,9; 64,3; 69,8; 42,6; 91,2; 71,3; 84,6; 69,2; 35,9; 78,9; 37,8; 59,5; 53,8; 41,7; 27,9; 66,6; 72,2; 55,9.

**5.** 27,39; 19,59; 17,04; 4,06; 12,38; 28,65; 18,62; 17,75; 28,97; 21,84; 20,26; 15,53; 23,3; 14,63; 12,24; 25,9; 7,13; 16,38; 15,34; 17,72; 25,14; 25,33; 13,51; 10,31; 15,03; 11,27; 25,63; 20,63; 24,13; 28,55; 19,7; 24,43; 32,19; 6,82; 16,19; 21,75; 18,06; 21,65; 25,39; 21,85; 16,54; 16,72; 30,89; 12,13; 17,02; 10,22; 16,81; 3,64; 29,77; 11,72.

**6.** 418; 455; 521; 517; 476; 473; 506; 398; 509; 464; 431; 529; 430; 436; 588; 507; 549; 511; 534; 477; 561; 385; 511; 486; 524; 543; 490; 545; 425; 590; 561; 503; 504; 438; 450; 494; 485; 543; 423; 527; 489; 500; 524; 450; 479; 502; 486; 517; 434; 545.

**7.** 100,86; 99,05; 98,61; 100,09; 99,43; 97,44; 101,52; 99,22; 97,59; 97,93; 99,96; 100,06; 100,53; 100,98; 100,48; 98,63; 96,77; 102,82; 101,94; 97,67; 100,29; 103,39; 99,36; 95,94; 97,18; 97,35; 98,97; 100,12; 98,81; 99,34; 101,08; 100,44; 100,81; 97,69; 97,79; 97,79; 99,18; 97,37; 94,32; 102,55; 101,58; 100,29; 97,99; 98,15; 98,24; 100,56; 99,20; 102,60; 97,24; 100,76.

**8.** 50,88;53,52; 48,71; 51,10; 53,09; 46,36; 52,63; 47,91; 50,92; 47,89; 47,52; 50,85; 46,17; 49,75; 50,83; 52,93; 47,54; 49,23; 50,95; 50,37; 50,39; 51,04; 51,65; 51,86; 51,79; 50,52; 48,61; 50,93; 49,70; 48,27; 54,23; 47,53; 46,61; 50,77; 47,14; 49,18; 51,75; 48,86; 48,18; 48,27; 52,19; 47,95; 47,50; 51,25; 48,53; 51,08; 51,33; 51,04; 50,29; 49,90.

**9.** 27,14; 22,62; 21,52; 40,24; 23,57; 18,60; 28,80; 23,04; 18,98; 19,82; 24,90; 25,16; 26,33; 27,46; 26,19; 21,58; 16,93; 32,06; 29,86; 19,22; 25,72; 33,49; 23,40; 14,84; 17,94; 18,37; 22,42; 25,31; 22,02; 23,35; 27,70; 26,09; 27,03; 19,22; 19,47; 19,48; 22,96; 18,43; 10,80; 31,38; 28,96; 25,72; 19,97; 20,37; 20,61; 26,39; 22,99; 31,51; 18,11; 26,90.

**10.** 10,63; 10,50; 9,66; 9,09; 10,59; 10,30; 8,13; 11,59; 11,43; 12,09; 11,62; 11,23; 9,51; 9,83; 11,37; 11,58; 10,72; 9,58; 9,69; 10,27; 11,73; 13,92; 10,14; 11,13; 11,11; 13,78; 12,96; 5,41; 11.33; 9,12; 9,05; 10,18; 8,60; 10,69; 8,25; 13,80; 4,55; 11,74; 14,36; 12,17; 12,35; 8,06; 6,85; 10,93; 12,95; 12,33; 8,47; 11,02.

**11.** 176; 184; 169; 176; 182; 162; 181; 167; 176; 167; 166; 176; 162; 172; 175; 182; 166; 171; 176; 174; 174; 176; 178; 179; 178; 175; 169; 176; 172; 168; 186; 166; 163; 175; 164; 171; 178; 170; 168; 168; 180; 167; 165; 177; 169; 176; 177; 176; 174; 173.

**12.** 333,2; 332,5; 328,3; 325,5; 332,9; 331,5; 320,7; 337,9; 337,2; 340,4, 338,1; 336,2; 327,5, 329,1; 336,8; 337,9; 333,6; 327,9; 328,4; 331,4; 338,6; 349,6; 330,7; 335,6; 335,6; 348,9; 344,8; 307,1; 336,6; 325,6; 325,3; 330,9; 323; 333,5; 321,3; 349; 302,7; 338,7; 351,8; 340,8; 341,8; 320,3; 314,2; 334,6; 344,7; 341,7; 322,4; 335,1; 335; 322.

**13.** 117; 118; 121; 119; 128; 120; 121; 142; 128; 152; 105; 126; 130; 146; 133; 103; 107; 156; 135; 125; 114; 112; 104; 111; 140; 137; 135; 111; 120; 146; 125; 121; 119; 130; 131; 124; 126; 130; 115; 123; 106; 128; 147; 113; 135; 123; 110; 134; 129; 114.

**14.** 61,1; 50,5, 61,1; 44,4, 59,1; 55,2; 60,4; 93,9; 58,6; 51,1; 79,7; 86,4; 32,1; 84,5; 37,8; 33,4; 50,2; 56,6; 72,2; 56,1; 30,1; 59,6; 67; 74,8; 90,7; 48,4; 62,6; 37,4; 10,7; 88,7; 51,4; 25,7; 28,9; 64,3; 69,8; 42,6; 91,2; 71,3; 84,6; 69,2; 35,9; 78,9; 37,8; 59,5; 53,8; 41,7; 27,9; 66,6; 72,2; 55,9.

**15.** 27,39; 19,59; 17,04; 4,06; 12,38; 28,65; 18,62; 17,75; 28,97; 21,84; 20,26; 15,53; 23,3; 14,63; 12,24; 25,9; 7,13; 16,38; 15,34; 17,72; 25,14; 25,33; 13,51; 10,31; 15,03; 11,27; 25,63; 20,63; 24,13; 28,55; 19,7; 24,43; 32,19; 6,82; 16,19; 21,75; 18,06; 21,65; 25,39; 21,85; 16,54; 16,72; 30,89; 12,13; 17,02; 10,22; 16,81; 3,64; 29,77; 11,72.

**Задание 3.** Сгенерировать выборку из 50 элементов, имеющих указанное в варианте распределение. На одном графике построить гистограмму и теоретическую функцию плотности распределения (использовать встроенные функции), сравнить полученные графики и оценить, действительно ли гистограмма является приближением функции плотности. Основные характеристики распределений и соответствующие им встроенные функции приведены в Приложении 2.

|  |  |
| --- | --- |
| **Вариант** | **Распределение** |
| 1 | выборка из равномерного распределения |
| 2 | выборка из показательного распределения |
| 3 | выборка из распределения Фишера |
| 4 | выборка из нормального распределения |
| 5 | выборка из распределения Стьюдента |
| 6 | выборка из распределения |
| 7 | выборка из логнормального распределения с параметрами |
| 8 | выборка из гамма-распределения |
| 9 | выборка из -распределения |
| 10 | выборка из распределения Фишера |
| 11 | выборка из -распределения |
| 12 | выборка из распределения Стьюдента |
| 13 | выборка из распределения |
| 14 | выборка из логнормального распределения с параметрами |
| 15 | выборка из гамма-распределения |

**Глава 2. Точечное и интервальное оценивание**

**параметров распределения**

*Узловым вопросом математической статистики*

*является вопрос: как далеко могут отклоняться величины,*

*вычисленные по выборке, от соответствующих идеальных значений?* Б. Л. Ван дер Варден (1960)

При заданном виде закона распределения СВ *Х* неизвестные параметры этого распределения можно оценить. Оценки параметров генеральной совокупности, полученные на основе выборки, называются **статистическими**.

***Виды статистических оценок***

1. Точечные – оценка характеризуется одним числом.

2. Интервальные – оценка характеризуется двумя числами - концами интервала.

**2.1. Методы получения точечных оценок**

**Метод моментов**

Метод моментов предложен К. Пирсоном в 1894 г.

**Суть метода**: сопоставление выборочных начальных и центральных моментов с соответствующими теоретическими моментами и . Число уравнений должно соответствовать числу неизвестных параметров:

;

,

Оценки, получаемые с помощью метода моментов, обладают следующими свойствами: являются состоятельными; в случае нормального распределения – эффективными; в общем случае могут быть смещенными.

Получение точечных оценок методом моментов при заданном законе распределения приведено в таблице 2.1.

Таблица 2.1

**Получение точечных оценок методом моментов**

|  |  |  |
| --- | --- | --- |
| ***Закон*** | ***Параметры*** | ***Уравнения*** |
| Нормальный | 2 параметра – |  |
| Показательный | 1 параметр – λ |  |
| Равномерный | 2 параметра – |  |
| Гамма-распределение | 2 параметра – |  |
| Биномиальный | 1 параметр – *р* |  |
| Пуассона | 1 параметр – λ |  |

**Пример 2.1.** На предприятии изготавливается определенный вид продукции. Ежемесячный объем выпуска этой продукции является случайной величиной, для характеристики которой принят показательный закон распределения

В течение шести месяцев проводился замер объемов выпуска, продукции, получены следующие данные:

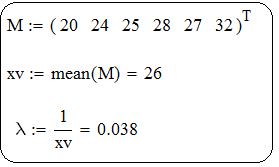
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| месяц | 1 | 2 | 3 | 4 | 5 | 6 |
| объем выпуска | 20 | 24 | 25 | 28 | 27 | 32 |

Найти оценку параметра

**Решение**

Так как закон распределения содержит лишь один параметр то для его оценки требуется составить одно уравнение.

На рисунке 2.1 приведены результаты оценки параметра в среде Mathcad.



*Рис. 2.1. Оценка параметра*

*показательного распределения*

*методом моментов*

**Метод максимального правдоподобия**

Этот универсальный метод был предложен английским математиком Р. Фишером в 1921 г.

**Суть метода**: в качестве «наиболее правдоподобной» оценки неизвестного параметра берут значение, максимизирующее вероятность получить при *п* опытах данную выборку

Пусть

***Функцией правдоподобия*** называется функция вида

***Логарифмической функцией правдоподобия*** называется функция вида

***Оценкой максимального правдоподобия*** неизвестного параметра называется такое значение , при котором достигает максимума. Часто удобнее максимизировать логарифмическую функцию правдоподобия, так как и достигают максимума в одних и тех же точках.

Достаточным условием существования локального максимума в служит неравенство вида:

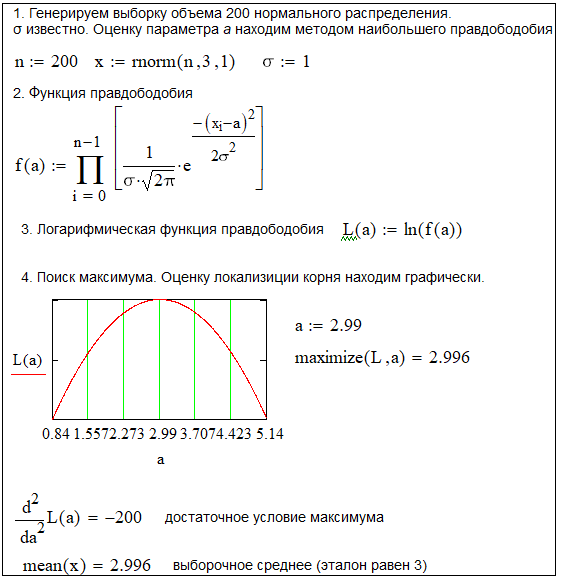
В среде Mathcad поиск безусловных и условных экстремумов производится с применением функции **maximize**.

**Пример 2.2.** Пусть имеется выборкаобъема 200 из нормального распределения с известным средним квадратическим отклонением. Требуется методом максимального правдоподобия оценить параметр *а.*

**Решение**

Структура максимизации в среде Mathcad функции правдоподобия *Ln(L)* представлена на рисунке 2.2.

Вторая производная от логарифмической производной в критической точке строго отрицательна, откуда следует, что полученная оценка является оценкой максимального правдоподобия.

**

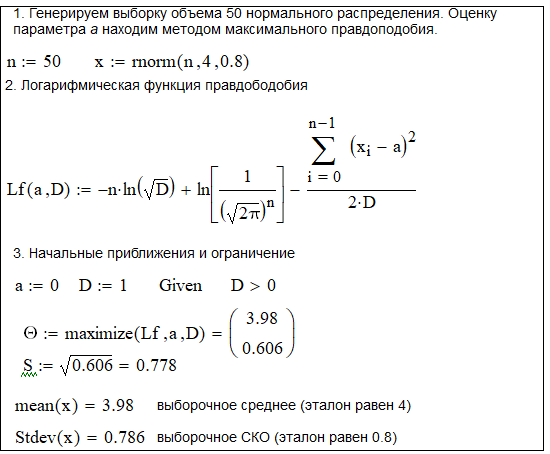
*Рис. 2.2. Максимизация функции правдоподобия*

*в среде Mathcad*

**Пример 2.3.** Найти оценкимаксимального правдоподобия параметров нормального распределения, если нормально распределенная случайная величина *Х* в результате испытаний принимает значения

**Решение**

В данном примере решается задача на условный экстремум. Условие используется как ограничение. В качестве *Ln(L)* рассматривается функция *Lf (a, D)* – функция двух переменных (рис. 2.3).



*Рис. 2.3. Метод максимального правдоподобия для СВХ,*

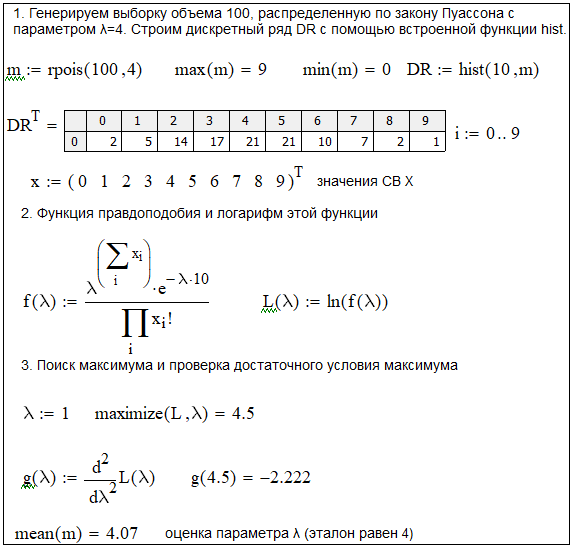
*распределенной по нормальному закону*

**Пример 2.4.** По данной выборке с элементами найти параметр распределения Пуассона.

**Решение**

В этом примере необходимо определить значения, которые принимает случайная величина *Х*. Для этого строится дискретный вариационный ряд.

Дальнейшее решение примера 2.4 приведено на рисунке 2.4



*Рис. 2.4. Метод максимального правдоподобия*

*для СВХ, распределенной по закону Пуассона*

**2.2. Интервальные оценки**

Пусть точечная оценка, использующая данные выборки, получена.

**Точность** такой оценки можно охарактеризовать неравенством

,

причем чем меньше значение , тем выше точность полученной оценки.

**Надежность** (степень доверия) оценки характеризуется величиной , с которой выполняется неравенство

Перейдя к двустороннему неравенству:

.

Вероятность - доверительная вероятность (надежность). В практике решения задач интервального оценивания наибольшее распространение получили значения , равные 0,9; 0,95; 0,99.

Часто вместо доверительной вероятности задается ее дополнение до единицы или **уровень значимости** Уровень значимости определяет вероятность нахождения неизвестного параметра за границами доверительного интервала.

***Алгоритм нахождения границ доверительного интервала***

1. Получить точечную оценку искомого параметра.

2. Задать статистику, содержащую параметр и его точечную оценку таким образом, чтобы функция распределения статистики была известна (точно или приближенно).

3. Задать уровень значимости . Принять вероятности за пределами доверительного интервала слева и справа равные .

4. Определить квантили и распределения случайной величины .

5. Записать неравенство , выразить из него искомый параметр .

Многолетней практикой решения задач математической статистики выработаны рекомендации по заданию статистик в каждой конкретной задаче (например, стандартное нормальное распределение, хи-квадрат, Фишера, Стьюдента).

Использование средств компьютерной математики позволяет в некоторых случаях упростить алгоритм.Средствами компьютерной математики можно определять квантили для распределений с произвольными параметрами, поэтому применение статистик в виде стандартного нормального распределения становится необязательным.

Интервальные оценки неизвестных параметров основных законов распределения приведены в таблице 2.2.

Таблица 2.2

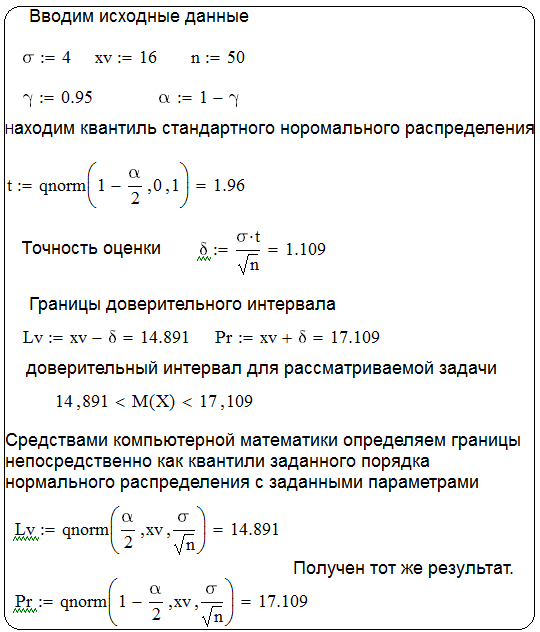
**Получение интервальных оценок**

|  |  |
| --- | --- |
| ***Интервальная оценка математического ожидания***  ***нормального распределения с известной дисперсией*** | |
| С помощью нормированной  статистики | Применение средств  компьютерной математики |
| Статистика t имеет нормальное  распределение  Определяются квантили  стандартного нормального распределения и , связанные в данном случае соотношением  .  Квантили можно определить по специальной таблице или с помощью встроенной функции Mathcad **qnorm(,0,1)**.  Доверительный интервал | Статистика z имеет  нормальное распределение  Определяются квантили и нормального распределения.  Квантили можно определить с помощью встроенной функции Mathcad  **qnorm(,,)**,  **qnorm(,,)**, которые и являются границами интервала. |
| ***Интервальная оценка математического ожидания для***  ***случайной величины с произвольным законом распределения с***  ***известной дисперсией*** | |
| Построение интервальной оценки может осуществляться двумя способами, как и в предыдущем случае по тем же алгоритмам при условии большой выборки | |
| ***Интервальная оценка дисперсии***  ***(среднего квадратического отклонения)***  ***нормального распределения*** | |
| Статистика распределена по закону хи-квадрат с (*п*-1) степенями свободы  Определяются квантили  , .  Квантили можно определить по специальной таблице или с помощью встроенной функции Mathcad  **qchisq, qchisq**  Доверительный интервал | |
| ***Интервальная оценка дисперсии (среднего квадратического***  ***отклонения) произвольно распределенной случайной величины*** | |
| Статистика *m* имеет нормальное распределение Определяются квантили стандартного нормального распределения и , связанные в данном случае соотношением  .  Квантили можно определить по специальной таблице или с помощью встроенной функции Mathcad  m **= qnorm(,0,1)**.  Вычисляем параметр , где - выборочные центральные моменты.  Доверительный интервал  Если , то | |
| ***Интервальная оценка математического ожидания нормального распределения при неизвестной дисперсии и произвольного распределения (при большом объеме выборки*)** | |
| Статистика распределена по закону Стьюдента с (*п*-1) степенями свободы. Определяются квантили t-распределения  и ,  связанные соотношением .  Квантили можно определить по специальной таблице или с помощью встроенной функции Mathcad **qt(,n-1)**.  Доверительный интервал | |

**Пример 2.5.** Найти доверительный интервал с надежностью 0,95 для оценки математического ожидания нормально распределенной СВ *Х*, если известны ее среднее квадратическое отклонение, выборочная средняя и объем выборки *n=*16.

**Решение**

На рисунке 2.5 приведены результаты решения (традиционный подход) и средствами компьютерной математики.

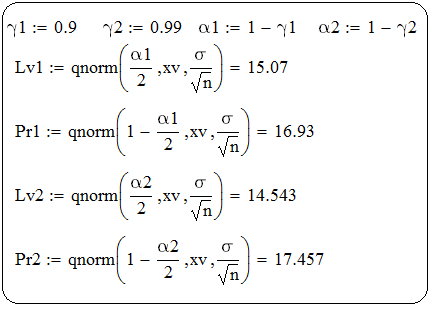


*Рис. 2.5. Решение задачи*

*интервального оценивания*

**Пример 2.6.** Рассмотрим пример 2.5 при (рис. 2.6).

**Решение**



*Рис. 2.6. Вычисление границ интервалов*

Результаты сведем в таблицу.

|  |  |  |  |
| --- | --- | --- | --- |
| Границы | Значение доверительной вероятности | | |
|  |  |  |
| левая | 15,07 | 14,891 | 14,543 |
| правая | 16,93 | 17,109 | 17,457 |

Анализируя данные, отмечаем тенденцию к расширению границ доверительных интервалов при увеличении значения .

**Задание для аудиторного решения**

Обследование оплаты труда 50 рабочих данного завода дало следующие результаты (в условных единицах):

214, 204, 212, 201, 190, 222, 226, 216, 228, 240,

224, 220, 260, 204, 240, 190, 218, 232, 254, 224,

204, 221, 256, 260, 228, 232, 204, 182, 230, 214,

242, 222, 260, 198, 216, 198, 232, 242, 216, 226,

208, 221, 202, 204, 222, 196, 222, 238, 224, 223.

1. Провести статистическую обработку выборочных данных: сформировать массив, составить вариационный ряд, группированный статистический ряд, построить гистограмму оценки функции плотности. Вычислить с помощью встроенных функций числовые характеристики выборки.

2. Провести визуальный сравнительный анализ неизвестного распределения данной СВ *Х* с нормальным распределением. Для этого на одном графике построить гистограмму оценки функции плотности изучаемой СВ *Х* и график функции плотности нормального закона. В качестве параметров нормального закона взять их наилучшие оценки по выборке.

***Замечание.*** *Близость средней арифметической величины, медианы и моды указывает на вероятное соответствие изучаемого распределения нормальному закону.*

3. Найти доверительные интервалы для оценки математического ожидания и дисперсии с надежностью 0,99.

**Компьютерная лабораторная работа № 5**

***Точечное и интервальное оценивание***

***параметров распределения***

**Задание 1.** При заданном виде закона распределения СВ *Х* оценить неизвестные параметры этого распределения. Точечные оценки параметров найти **методом моментов (обязательно)** и *методом наибольшего правдоподобия (дополнительно).*

1. При условии показательного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *хi* | 3 | 4 | 10 | 12 | 15 |
| *ni* | 3 | 3 | 6 | 4 | 4 |

Найти оценку параметра λ.

2. При условии равномерного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *хi* | 2 | 3 | 4 | 5 | 6 |
| *ni* | 4 | 6 | 5 | 5 | 8 |

Найти оценку параметров *a* и *b*.

3. При условии показательного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *хi* | 3 | 5 | 6 | 8 | 10 |
| *ni* | 2 | 3 | 5 | 10 | 10 |

Найти оценку параметра λ.

4. При условии равномерного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 |
| *ni* | 4 | 6 | 5 | 6 | 4 | 5 | 5 | 6 | 7 | 5 |

Найти оценку параметров *a* и *b*.

5. При условии равномерного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 3 | 5 | 7 | 9 | 11 | 13 | 15 |
| *ni* | 6 | 9 | 6 | 5 | 7 | 6 | 8 |

Найти оценку параметра

6. СВ *Х* распределена по биномиальному закону. Статистическое распределение выборки представлено в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *ni* | 2 | 3 | 10 | 22 | 26 | 20 | 12 | 5 |

Найти оценку параметра *р*.

7. СВ *Х* распределена по закону Пуассона. Статистическое распределение выборки представлено в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *ni* | 112 | 87 | 42 | 21 | 9 | 3 | 1 | 1 |

Найти оценку параметра λ.

8. СВ *Х* распределена по биномиальному закону. Статистическое распределение выборки представлено в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *ni* | 1 | 3 | 9 | 17 | 20 | 18 | 12 | 5 |

Найти оценку параметра *р*.

9. При условии показательного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| *ni* | 26 | 19 | 16 | 15 | 8 | 6 | 4 |

Найти оценку параметра λ

10. При условии равномерного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| *ni* | 2 | 2 | 3 | 4 | 2 | 3 | 3 | 4 | 2 | 1 |

Найти оценку параметров *a* и *b*.

11. СВ *Х* распределена по биномиальному закону. Статистическое распределение выборки представлено в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *ni* | 1 | 3 | 5 | 12 | 10 | 8 | 4 | 5 |

Найти оценку параметра *р*.

12. СВ *Х* распределена по закону Пуассона. Статистическое распределение выборки представлено в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *ni* | 45 | 30 | 12 | 8 | 7 | 3 | 1 | 1 |

Найти оценку параметра λ.

13. СВ *Х* распределена по биномиальному закону. Статистическое распределение выборки представлено в таблице:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *ni* | 1 | 3 | 8 | 15 | 20 | 16 | 5 | 5 |

Найти оценку параметра *р*.

14. При условии показательного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| *ni* | 26 | 19 | 16 | 15 | 8 | 6 | 4 |

Найти оценку параметра λ

15. При условии равномерного распределения СВ *Х* произведена выборка

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *хi* | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 |
| *ni* | 2 | 2 | 3 | 4 | 2 | 3 | 3 | 4 | 2 | 1 |

Найти оценку параметров *a* и *b*.

**Задание 2.** Провести визуальный сравнительный анализ неизвестного распределения данной СВ *Х* с нормальным распределением. Для этого на одном графике построить гистограмму изучаемой СВ *Х* и график функции плотности нормального закона. В качестве параметров нормального закона взять их наилучшие оценки по выборке.

С помощью вычисленных числовых характеристик определить, является ли неизвестное распределение близким к нормальному.

*Указание. Если выборочное распределение близко к нормальному (или является таковым), то:*

*- в интервалы должны попадать соответственно приблизительно 68%, 95% и 100% выборочных значений;*

*- в выборке величина коэффициента вариации должна быть не более 33%;*

*- оценки эксцесса и коэффициента асимметрии должны быть близки к нулю;*

*- .*

**Задание 3.** Найти доверительные интервалы для оценки математического ожидания и дисперсии. Надежность выбрать самостоятельно.

**Задание 4.** Используя образец решения примера 2.6 проведите анализ влияния на поведение границ доверительного интервала

- объема выборки;

- значения доверительной вероятности.

Статистические данные по вариантам для заданий 2-4 представлены ниже.

**1.** 176; 184; 169; 176; 182; 162; 181; 167; 176; 167; 166; 176; 162; 172; 175; 182; 166; 171; 176; 174; 174; 176; 178; 179; 178; 175; 169; 176; 172; 168; 186; 166; 163; 175; 164; 171; 178; 170; 168; 168; 180; 167; 165; 177; 169; 176; 177; 176; 174; 173.

**2.** 333,2; 332,5; 328,3; 325,5; 332,9; 331,5; 320,7; 337,9; 337,2; 340,4, 338,1; 336,2; 327,5, 329,1; 336,8; 337,9; 333,6; 327,9; 328,4; 331,4; 338,6; 349,6; 330,7; 335,6; 335,6; 348,9; 344,8; 307,1; 336,6; 325,6; 325,3; 330,9; 323; 333,5; 321,3; 349; 302,7; 338,7; 351,8; 340,8; 341,8; 320,3; 314,2; 334,6; 344,7; 341,7; 322,4; 335,1; 335; 322.

**3.** 117; 118; 121; 119; 128; 120; 121; 142; 128; 152; 105; 126; 130; 146; 133; 103; 107; 156; 135; 125; 114; 112; 104; 111; 140; 137; 135; 111; 120; 146; 125; 121; 119; 130; 131; 124; 126; 130; 115; 123; 106; 128; 147; 113; 135; 123; 110; 134; 129; 114.

**4.** 61,1; 50,5, 61,1; 44,4, 59,1; 55,2; 60,4; 93,9; 58,6; 51,1; 79,7; 86,4; 32,1; 84,5; 37,8; 33,4; 50,2; 56,6; 72,2; 56,1; 30,1; 59,6; 67; 74,8; 90,7; 48,4; 62,6; 37,4; 10,7; 88,7; 51,4; 25,7; 28,9; 64,3; 69,8; 42,6; 91,2; 71,3; 84,6; 69,2; 35,9; 78,9; 37,8; 59,5; 53,8; 41,7; 27,9; 66,6; 72,2; 55,9.

**5.** 27,39; 19,59; 17,04; 4,06; 12,38; 28,65; 18,62; 17,75; 28,97; 21,84; 20,26; 15,53; 23,3; 14,63; 12,24; 25,9; 7,13; 16,38; 15,34; 17,72; 25,14; 25,33; 13,51; 10,31; 15,03; 11,27; 25,63; 20,63; 24,13; 28,55; 19,7; 24,43; 32,19; 6,82; 16,19; 21,75; 18,06; 21,65; 25,39; 21,85; 16,54; 16,72; 30,89; 12,13; 17,02; 10,22; 16,81; 3,64; 29,77; 11,72.

**6.** 418; 455; 521; 517; 476; 473; 506; 398; 509; 464; 431; 529; 430; 436; 588; 507; 549; 511; 534; 477; 561; 385; 511; 486; 524; 543; 490; 545; 425; 590; 561; 503; 504; 438; 450; 494; 485; 543; 423; 527; 489; 500; 524; 450; 479; 502; 486; 517; 434; 545.

**7.** 100,86; 99,05; 98,61; 100,09; 99,43; 97,44; 101,52; 99,22; 97,59; 97,93; 99,96; 100,06; 100,53; 100,98; 100,48; 98,63; 96,77; 102,82; 101,94; 97,67; 100,29; 103,39; 99,36; 95,94; 97,18; 97,35; 98,97; 100,12; 98,81; 99,34; 101,08; 100,44; 100,81; 97,69; 97,79; 97,79; 99,18; 97,37; 94,32; 102,55; 101,58; 100,29; 97,99; 98,15; 98,24; 100,56; 99,20; 102,60; 97,24; 100.76.

**8.** 50,88;53,52; 48,71; 51,10; 53,09; 46,36; 52,63; 47,91; 50,92; 47,89; 47,52; 50,85; 46,17; 49,75; 50,83; 52,93; 47,54; 49,23; 50,95; 50,37; 50,39; 51,04; 51,65; 51,86; 51,79; 50,52; 48,61; 50,93; 49,70; 48,27; 54,23; 47,53; 46,61; 50,77; 47,14; 49,18; 51,75; 48,86; 48,18; 48,27; 52,19; 47,95; 47,50; 51,25; 48,53; 51,08; 51,33; 51,04; 50,29; 49,90.

**9.** 27,14; 22,62; 21,52; 40,24; 23,57; 18,60; 28,80; 23,04; 18,98; 19,82; 24,90; 25,16; 26,33; 27,46; 26,19; 21,58; 16,93; 32,06; 29,86; 19,22; 25,72; 33,49; 23,40; 14,84; 17,94; 18,37; 22,42; 25,31; 22,02; 23,35; 27,70; 26,09; 27,03; 19,22; 19,47; 19,48; 22,96; 18,43; 10,80; 31,38; 28,96; 25,72; 19,97; 20,37; 20,61; 26,39; 22,99; 31,51; 18,11; 26,90.

**10.** 10,63; 10,50; 9,66; 9,09; 10,59; 10,30; 8,13; 11,59; 11,43; 12,09; 11,62; 11,23; 9,51; 9,83; 11,37; 11,58; 10,72; 9,58; 9,69; 10,27; 11,73; 13,92; 10,14; 11,13; 11,11; 13,78; 12,96; 5,41; 11.33; 9,12; 9,05; 10,18; 8,60; 10,69; 8,25; 13,80; 4,55; 11,74; 14,36; 12,17; 12,35; 8,06; 6,85; 10,93; 12,95; 12,33; 8,47; 11,02.

**11.** 2,6; 2,3; 2,5; 2; 1,3; 3; 2,9; 3,6; 5,2; 3,8; 4; 3,9; 3,9; 3,7; 2; 3,1; 2,2; 3,7; 2,8; 2,4; 2,3; 2,5; 3,6; 2,8; 3,1; 4,3; 2,3; 3; 4,1; 3,9; 0,1; 0,8; 3,2; 2,4; 1,8; 3,1; 3,8; 3,3; 3; 2,2; 2,6; 2,3; 2,9; 4; 3,2; 3,5; 3,7; 3; 1,8; 3,1; 3,8; 3,3; 3; 2,2; 2,6; 2,3; 2,9; 4; 3,2; 3,5; 3,7; 3; 3,3; 3,2.

**12.** 12,2; 8; 7,5; 11,2; 8,5; 11,1; 11,3; 10,3; 9,9; 10,9; 13,5; 8,7; 11,1; 13,1; 6,4; 12,6; 7,9; 10,9; 7,9; 7,5; 10,8; 6,2; 9,8; 10,8; 12,9; 7,5; 9,2; 11; 10,4; 10,4; 11; 11,6; 11,9; 11,8; 10,5; 8,6; 10,9; 9,7; 8,3; 14,2; 7,5; 6,6; 10,8; 7,1; 9,2; 11,7; 8,9; 8,2; 8,3.

**13.** 5,3; 5,2; 4,8; 4,5; 5,3; 5,2; 4,1; 5,8; 5,7; 6; 5,8; 5,6; 4,8; 5,7; 5,8; 5,4; 4,8; 4,8; 5,1; 5,9; 7; 5,1; 5,6; 5,6; 6,9; 6,5; 2,7; 5,7; 4,6; 4,5; 5,1; 4,3; 5,3; 4,1; 6,9; 2,3; 5,9; 7,2; 6,1; 6,2; 4; 3,4; 5,5; 6,5; 6,2; 4,2; 5,5; 5,5; 4,2.

**14.** 21,2; 21,7; 22,9; 22,2; 26,5; 22,8; 23,4; 32,7; 26,2; 37,5; 15,6; 25,5; 27,5; 34,8; 28,7; 14,7; 16,8; 39,4; 29,4; 25; 19,9; 18,9; 15,2; 18,7; 31,7; 30,7; 29,6; 18,6; 22,5; 34,6; 25,1; 23; 22,3; 27,4; 28; 24,5; 25,5; 27,5; 20,4; 24,2; 16,2; 26,2; 35,1; 19,6; 29,7; 24,2; 18,3; 29,3; 26,7; 19,7.

**15.** 3,1; 5,1; 4,3; 6,2; 5; 5,8; 8,9; 8,6; 6,3; 3,4; 8,3; 1,6; 6,5; 4,3; 5,4; 5,8; 6,5; 7,7; 9,3; 5,8; 5,8; 4,8; 5,9; 6,4; 6,7; 5; 6,4; 6,3; 4,7; 6,8; 6,5; 3,9; 7,5; 7,8; 1,8; 4,6; 5,2; 8,6; 9,1; 5,4; 6,4; 6,7; 7,4; 6,9; 6,1; 6,5; 5,3; 7,6; 5,5;6.

**Глава 3. Статистические гипотезы**

*Во время экспериментирования*

*начальная гипотеза никогда не доказывается*

*и не устанавливается,*

*но возможно не одобряется.*

*Можно сказать, что каждый опыт*

*существует лишь для того,*

*чтобы дать фактам возможность*

*проверить начальную гипотезу.*

Рональд Фишер (1935)

**3.1. Статистическая гипотеза и общая схема ее проверки**

***Статистической гипотезой*** называется любое предположение о виде или параметрах неизвестного закона распределения.

Проверяемую гипотезу называют **нулевой** . Противоположную ей гипотезу называют **альтернативной, конкурирующей**.

*Нулевая и альтернативная гипотезы – это две возможности выбора, осуществляемого в задачах проверки статистических гипотез.*

***Схема проверки нулевой гипотезы***

1. Рассматривая выборочные данные и учитывая конкретные условия задачи, выдвигают нулевую гипотезу и альтернативную.

2. Так как решение о справедливости нулевой гипотезы принимается на основе выборочных данных, могут возникнуть ошибки двух родов.

Таблица 3.1

**Типы ошибок**

**при проверке статистических гипотез**

|  |  |  |
| --- | --- | --- |
| ***Гипотеза*** | ***Принимается*** | ***Отвергается*** |
| верна | правильное решение | ошибка 1-го рода;  вероятность ошибки 1-го рода равна уровню значимости α |
| неверна | ошибка 2-го рода;  вероятность ошибки 2-го рода равна β | правильное решение |

***Замечание.*** *Возможно* ***одновременное*** *уменьшение α и β лишь при увеличении объема выборки.*

Пользуясь терминологией статистического контроля качества продукции:

вероятность α – риск поставщика (по результатам выборочного контроля вся партия, удовлетворяющая стандарту, забракована);

вероятность *β* – риск потребителя (принятие по анализу выборки партии, не удовлетворяющей стандарту).

Применяя юридическую терминологию:

α – вероятность вынесение судом обвинительного приговора невиновному;

*β* – вероятность вынесение судом оправдательного приговора виновному.

Зарождения понятия об ошибках 1-го и 2-го родов можно проследить в следующих записях.

*«Любой скорее оправдает преступника, чем признает виновным невинного».* Аристотель (951b 0).

*«Много лучше освободить десять виновных, чем приговорить к смерти одного невинного».* (Воинский устав России, 1716).

3. Задается уровень значимости α, значение которого обычно находится в интервале (0,001; 0,1). Задание таких значений связано с необходимостью использовать таблицы функций распределений. Применение средств компьютерной математики позволяет находить характеристики распределений при произвольных значениях α.

4. Используя выборочные данные, вводится статистический критерий – некоторая функция *К*. Эти функции подчинены известному закону распределения (хи-квадрат распределение, *t*-распределение, нормальное и др.).

5. Определяется тип критической области и критические точки.

Таблица 3.2

**Типы критической области**

|  |  |
| --- | --- |
|  | **Критическая область** |
|  | **-** левосторонняя |
|  | **-** правосторонняя |
|  | **-** двусторонняя |

6. По выборочным данным вычисляется числовое значение критерия . Применяется правило:

- ***,*** то гипотеза отвергается и принимается альтернативная;

- ***,*** то гипотеза принимается.

*Принятие гипотезы следует расценивать не как раз и навсегда установленный, абсолютно верный содержащийся в ней факт, а лишь как достаточно правдоподобное, не противоречащее опыту утверждение.*

По своему прикладному содержанию статистические гипотезы можно подразделить на несколько основных типов:

- о законе распределения;

- о числовых значениях параметров;

- о равенстве числовых характеристик генеральных совокупностей;

- об однородности выборок (т.е. принадлежности их одной и той же генеральной совокупности).

**3.2. Построение теоретического закона распределения**

**по опытным данным. Проверка гипотез о законе**

**распределения**

При решении практических задач модель закона распределения в общем случае заранее неизвестна, возникает необходимость выбора модели закона распределения, согласующейся с результатами выборочных наблюдений. Существует много подходов к решению этой задачи. Наибольшее распространение получила схема, разработанная Пирсоном (1900 г.), дополненная Фишером (1924 г.).

***Схема проверки гипотезы по критерию Пирсона***

1**.** По выборке объема *п* строят статистический ряд (дискретный или интервальный) и его графическое изображение – полигон или гистограмму.

2. По данным предыдущих исследований или по предварительным данным выдвигают гипотезу о модели закона распределения СВ *Х* с функцией распределения *F(x)*.

3. Параметры выбранного закона неизвестны, поэтому их заменяют наилучшими оценками по выборке.

4. Используя гипотетическую функцию распределения *F(x)*, определяют теоретические значения вероятностей или

5. Рассчитывают теоретические частоты ,

6**.** Рассчитывают наблюдаемое значение критерия Пирсона

где - эмпирические частоты,

7. Задаваясь уровнем значимости α, находят критическую область (она всегда правосторонняя) Значениеопределяют по специальным таблицам или с помощью встроенной функции MathCAD – **qchisq**(, где *m* - количество промежутков, *r* - число параметров выбранного закона, оцениваемых по выборке.

8. Если *w*, то гипотеза отклоняется; если  *w* **-** проверяемая гипотеза согласуется с выборочными данными, то она принимается.

***Замечание 1.*** *Если в каком-нибудь интервале число теоретических частот <5, необходимо объединить соседние интервалы.*

***Замечание 2.*** *На практике если решение об отклонении нулевой гипотезы принято при близких значениях и****,*** *перед переходом к проверке другой гипотезы целесообразно повторить проверку с увеличением объема выборки.*

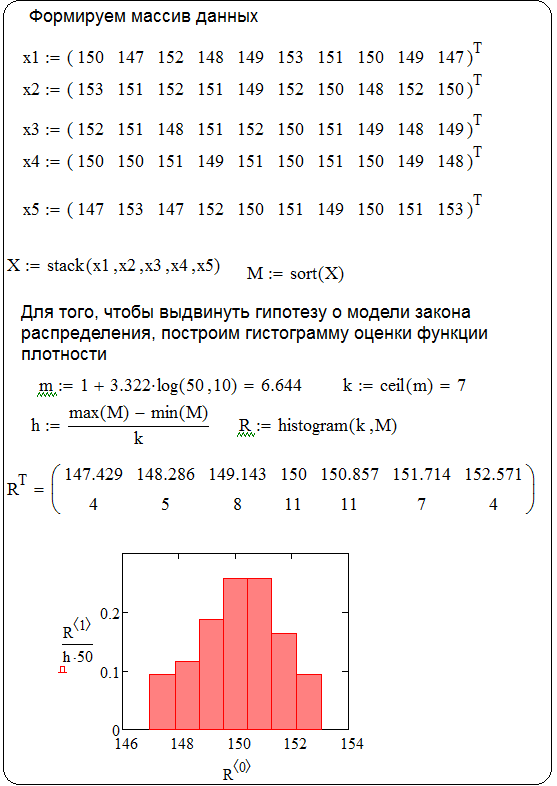
**Пример 3.1.** Результаты взвешивания 50 случайно отобранных пачек чая приведены ниже (в граммах). Оценить закон распределения СВ *Х* – массы пачки чая – для уровня значимости 0,05.

**Решение**

Формирование массива данных и построение гистограммы отображено на рисунке 3.1.

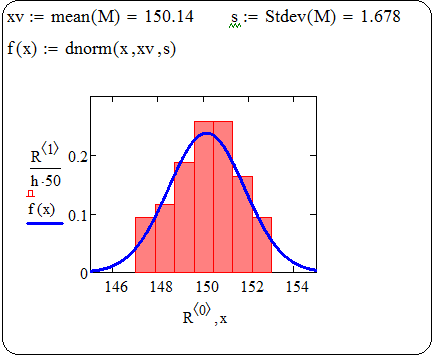
По виду гистограммы и по условию задачи в качестве модели закона распределения выберем нормальный закон, число параметров которого равно 2: *а* – математическое ожидание, σ – среднее квадратическое отклонение. По выборочным данным получим оценки параметров нормального закона распределения:

На одном графике строим гистограмму оценки функции плотности и график плотности распределения нормального закона с найденными параметрами (рисунок 3.2).



*Рис. 3.1. Построение гистограммы*

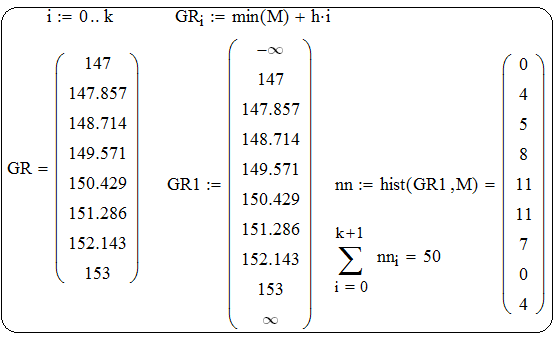
*оценки функции плотности*



*Рис. 3.2. Визуальный сравнительный анализ*

*теоретического закона с эмпирическим*

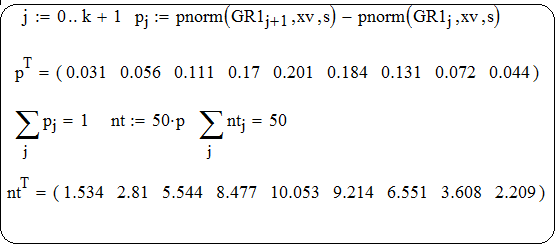
Определим границы интервалов GR и наблюдаемые частоты nn (рис.3.3).

**

*Рис. 3.3. Предварительный анализ*

*наблюдаемых частот*

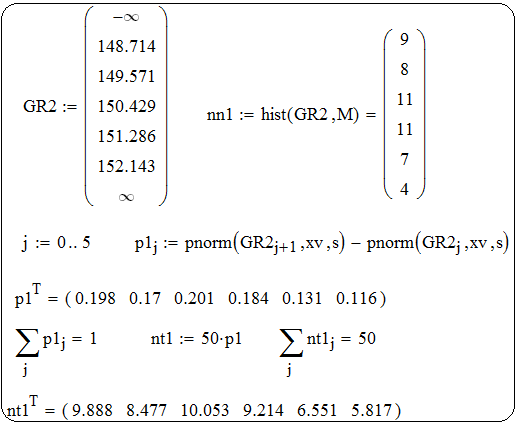
Определим вероятности *p* попаданийв отдельные промежутки и теоретические частоты nt (рис. 3.4).



*Рис. 3.4. Предварительный анализ*

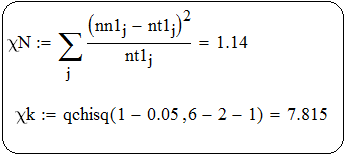
*теоретических частот*

Анализ полученных теоретических частот показывает, что условие  *5* будет выполнено, если первые три промежутка объединить в один общий и последние два также в один общий. На рисунке 3.5 обозначены: nn1- наблюдаемые частоты, nt1 – теоретические частоты после объединения.

****

*Рис. 3.5. Объединение частот*

Найдем наблюдаемое значение критерия χN (рис. 3.6). Результирующий вариант имеет 6 интервалов. Квантиль порядка 1-α =1- 0,05 распределения хи-квадрат определен с помощью встроенной функции **qchisq**.

****

*3.6. Результат решения*

Наблюдаемое значение не входит в критическую область

, поэтому гипотеза о том, что СВ *Х* – масса пачки чая – подчинена нормальному закону распределения, согласуется с выборочными данными.

**Пример 3.2.** Дано следующее распределение успеваемости 125 студентов, сдавших три экзамена:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число сданных  экзаменов | 0 | 1 | 2 | 3 |
| Число студентов | 3 | 5 | 47 | 70 |

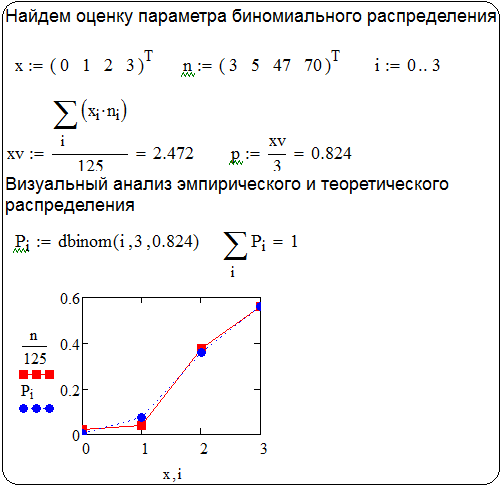
Проверить гипотезу о биномиальном распределении числа сданных экзаменов при уровне значимости 5%.

**Решение**

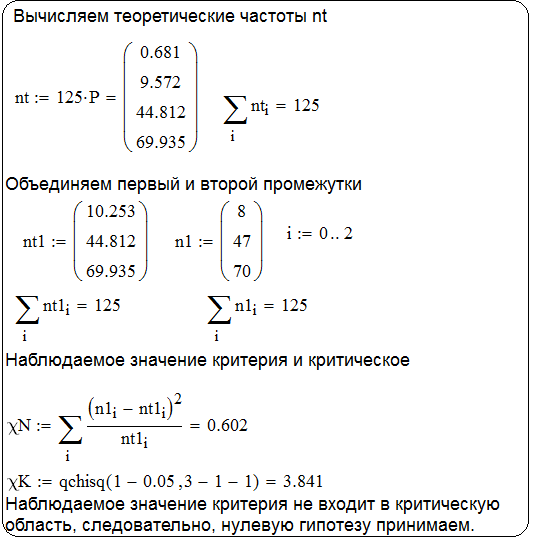
Для проверки гипотезы о биномиальном законе (рис. 3.7-3.8) необходимо найти оценку параметра выбранного закона распределения.

Биномиальный закон имеет один параметр – *p*, который найдем из условия

где *п* – число испытаний в каждом опыте (не путать с объемом выборки!).



*Рис. 3.7. Оценка параметра биномиального закона*

**

*Рис. 3.8. Проверка гипотезы о биномиальном*

*законе распределения*

**Задание для аудиторного решения**

1. Коммерсант предполагает, что объем продаж нового вида продукции в каждой из пяти торговых точек, расположенных в различных районах города, будет одинаков. Фактический объем продаж оказался разным:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Район | 1 | 2 | 3 | 4 | 5 |
| Объем продаж | 105 | 117 | 84 | 111 | 83 |

Оценить, значимы или нет различия между наблюдаемыми и ожидаемыми объемами продаж при уровне значимости 0,01.

2. При принятии на работу фирма предлагает 4 теста. СВ *Х* - число решенных тестов десятью претендентами: 0, 1, 1, 2, 2, 3, 3, 3, 4, 4. Проверить гипотезу о биномиальном распределении СВ *Х* при уровне значимости 0,05.

3. Время безотказной работы 60 элементов: 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 8, 8, 9, 9, 10, 11, 12, 13, 14, 15, 16, 16, 16, 17, 18, 20, 21, 21, 22,23. Проверить гипотезу о показательном законе распределения СВ *Х* – время безотказной работы при уровне значимости 0,05.

**Компьютерная лабораторная работа № 3**

***Проверка гипотезы о виде закона распределения СВ***

***по критерию Пирсона***

**Задания к компьютерной лабораторной работе № 3**

**Задание 1.** Используя результаты задания 2 компьютерной лабораторной работы № 2, проверить гипотезу о нормальном законе распределения СВ *Х*.

**Задание 2.** К имеющемуся эмпирическому распределению подобрать теоретический закон распределения. Осуществить проверку по критерию Пирсона.

**1.** 5,009; 10,76; 6,219; 9,973; 7,128; 5,64; 6,031; 11,92; 5,834; 5,062; 8,722; 9,212; 6,164; 8,156; 5,399; 10,483; 8,639; 11,132; 11,691; 8,775; 8,235; 11,036; 10,458; 11,978; 9,28; 6,863; 10,881; 7,631; 9,74; 5,062; 6,931; 9,115; 10,863; 10,206; 8,206; 10,211; 9,193; 10,145; 9,007; 6,061; 7,976; 8,62; 10,261; 6,183; 8,443; 9,898; 6,353; 9,095; 7,452; 8,395.

**2.** 4,493; 5,623; 6,029; 5,033; 6,566; 2,52; 2,427; 4,66; 1,78; 1,755; 8,452; 8,157; 2,819; 4,286; 6,025; 4,613; 5,783; 7,838; 5,998; 5,526; 2,474; 5,441; 2,943; 5.838; 5,677; 4.956; 6,926; 5,963; 7,436; 5,608; 8,292; 6,821; 6,342; 3,52; 3,447; 1,869; 7,81; 2,239; 1,635; 6,128; 5.361; 4,272; 4,724; 2,221; 6,904; 7,614; 7,987; 3,401; 2,018; 7,28.

**3**. 11,974; 2,301; 12,769; 2,887; 5,088; 17,559; 12,93; 9,866; 9,943; 10,167; 13,387; 11,913; 1,106; 2,811; 16,538; 14,452; 7,842; 10,951; 18,203; 4,179; 3,37; 18,131; 1,501; 2,007; 3,37; 16,558; 13,99; 1,264; 13,742; 4,916; 4,041; 7,138; 7,616; 15,443; 10,478; 15,367; 3,607; 8,243; 3,458; 2,191; 11,32; 10,865; 6,604; 15,544; 6,117; 17,114; 14,388; 7,501; 5,113; 5,13.

**4.** 0; 4; 3; 2; 4; 12; 1; 9; 4; 4; 6; 6; 8; 3; 6; 4; 5; 7; 3; 9; 5; 9; 3; 7; 3; 5; 4; 6; 2; 2; 3; 6; 2; 7; 3; 3; 3; 8; 1; 2; 5; 3; 5; 4; 2; 2; 3; 7; 4; 7.

**5.** 13,809; 9,14; 7,736; 11,415; 6,046; 27,683; 28,731; 13,034; 38,794; 39,336; 1,183; 1,856; 24,69; 14,831; 7,752; 13,249; 8,574; 2,615; 7,839; 9,439; 28,188; 9,809; 23,588; 8,393; 8,947; 11,739; 5,002; 7,957; 3,625; 9,193; 1,544; 5,299; 6,73; 19,252; 19,746; 37,005; 2,685; 31,084; 42,232; 7,412; 10,113; 14,902; 12,742; 31,326; 5,064; 3,172; 2,257; 20,062; 34,363; 4,035.

**6.** 4,35; 3,16; 3,28; 3,56; 4,04; 1,63; 4,15; 2,75; 3,03; 1,27; 4,93; 3,46; 2,47; 4,44; 1,97; 1,20; 1,69; 1,88; 4,15; 1,34; 1,55; 3,25; 3,53; 2,82; 2,34; 2,13; 4,49; 4,47; 4,18; 3,54; 3,38; 1,78; 1,93; 1,96; 2,63; 2,09; 2,52; 1,83; 3,42; 2,45; 4,90; 4,58; 4,81; 1,81; 2,42; 2,16; 1,39; 3,84; 3,69; 1,02.

**7.** 0,22; 1,51; 1,93; 0,41; 0,85; 11,25; 0,51; 0,57; 0,22; 5,52; 0,60; 4,18; 6,60; 3,20; 1,15; 3,60; 2,91; 5,29; 0,47; 4,99; 7,74; 14,96; 6,96; 9,66; 2,49; 8,13; 1,36; 2,26; 8,73; 0,19; 8,11; 15,25; 16,62; 12,48; 2,82; 7,60; 5,94; 4,44; 6,90; 3,38; 0,06; 10,53; 1,73; 6,11; 3,22; 3,30; 0,94; 11,33; 0,27; 3,73.

**8.** 1; 2; 1; 1; 2; 3; 3; 2; 4; 0; 1; 2; 3; 3; 1; 2; 6; 3; 2; 0; 3; 2; 3; 2; 4; 1; 0; 1; 4; 2; 1; 6; 2; 2; 3; 3; 1; 2; 1; 2; 4; 0; 7; 2; 2; 4; 2; 5; 2; 2.

**9.** 9,28; 6,863; 10,881; 7,631; 9,74; 5,062; 6,931; 9,115; 10,863; 10,206; 8,206; 10,211; 9,193; 10,145; 9,007; 6,061; 7,976; 8,62; 10,261; 6,183; 8,443; 9,898; 6,353; 9,095; 7,452; 8,395; **.** 5,009; 10,76; 6,219; 9,973; 7,128; 5,64; 6,031; 11,92; 5,834; 5,062; 8,722; 9,212; 6,164; 8,156; 5,399; 10,483; 8,639; 11,132; 11,691; 8,775; 8,235; 11,036; 10,458; 11,978.

**10.** 1,501; 2,007; 3,37; 16,558; 13,99; 1,264; 13,742; 4,916; 4,041; 7,138; 7,616; 15,443; 10,478; 15,367; 3,607; 8,243; 3,458; 2,191; 11,32; 10,865; 6,604; 15,544; 6,117; 17,114; 14,388; 7,501; 5,113; 5,13; 11,974; 2,301; 12,769; 2,887; 5,088; 17,559; 12,93; 9,866; 9,943; 10,167; 13,387; 11,913; 1,106; 2,811; 16,538; 14,452; 7,842; 10,951; 18,203; 4,179; 3,37; 18,131.

**11.** 40.5; 1.8; 38.2; 26.6; 70; 46,5; 33.5; 15.4; 52.5; 95.1; 70.8; 10.6; 51; 10.8; 4.9; 7.9; 5; 47; 0.3; 79.4; 56.7; 6.1; 74.8; 121.9; 25.9; 71.5; 10.1; 17.5; 9.6; 2; 12.2; 21; 106.6; 6.3; 116.1; 79.8; 103.2; 130.5; 14.2; 31.8; 16.8; 54.8; 20.7; 29.8; 10.8; 9.1; 18.5; 13.4; 47.9; 6.6.

**12.** 2; 3; 4; 5; 4; 4; 3; 4; 3; 3; 4; 4; 3; 5; 2; 5; 5; 4; 3; 3; 3; 2; 4; 4; 5; 3; 3; 3 4; 4; 5; 5; 1; 3; 3; 4; 4; 4; 2; 5; 3; 3; 3; 4; 4; 4; 3; 4; 4; 3.

**13.** 8; 2; 3; 6; 12; 2; 6; 1; 5; 1; 5; 5; 5; 4; 6; 3; 5; 6; 3; 9; 2; 7; 3; 6; 2; 7; 6; 4; 2; 3; 0; 7; 8; 4; 8; 3; 4; 3; 3; 2; 1; 3; 3; 4; 2; 2; 7; 5; 2; 4.

**14.** 1.4; 3.6; 1.8; 1.8; 4.5; 1.1; 3.8; 4.4; 2.5; 4.1; 2; 1.9; 4.8; 3.5; 4.6; 4.3; 4.8; 1.9; 2.3; 4; 1.3; 3.3; 3.6; 4.3; 2.5; 1.2; 3.8; 3.4; 4.4; 1.3; 4.6; 3.7; 4; 2.7; 4.9; 4.4; 2.7; 1.4; 1; 4.5; 4.4; 4.5; 2.2; 4.2; 4; 2.7; 4.5; 4.6; 2.3; 1.9.

**15.** 20.4; 39.4; 53.7; 16.3; 0.7; 22.1; 40.7; 5.7; 60; 49.1; 11.3; 6; 18.1; 51.8; 2.3; 19.5; 6.2; 5.6; 0.2; 22.7; 15.2; 28.8; 11.2; 9.1; 2; 37.5; 24.8; 25.5; 3.4; 7.5; 125.6; 18.6; 67.5; 10.1; 3.4; 2.9; 24.2; 1.8; 12; 13.8; 92.3; 8; 10.9; 12.9; 29.8; 8; 10.9; 12.9; 29.8; 28.4; 41.6; 8.6; 0.3; 19.8.

**3.3. Гипотезы о значениях и сравнении числовых**

**характеристик и однородности выборок**

**3.3.1. Проверка гипотез о числовых значениях параметров**

При проверке качества функционирования измерительных устройств необходимо проверить гипотезы о равенстве среднего значения *a* и дисперсии определенным числам . Если – номинальное значение измеряемого параметра и , то это означает, что прибор дает систематическую ошибку. Точность прибора определяется значением и, если , то это означает, что качество прибора не отвечает стандартным требованиям.

На практике часто требуется оценить, соответствуют ли действительности рекламные данные о параметрах того или иного товара. В этом случае возникает задача сравнения выборочной средней с анонсируемым значением этого параметра.

Или, например, что доля бракованных изделий, производимых станком, равна заданной величине . Схемы проверки гипотез о числовых значениях сведем в таблицу 3.3.

Таблица 3.3.

**Схемы проверки гипотез о числовых значениях**

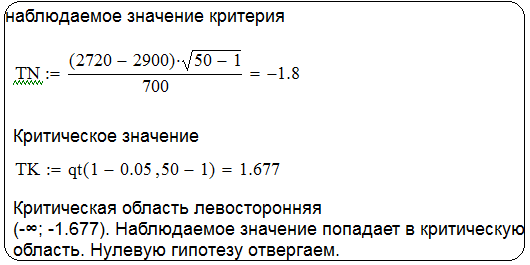
|  |  |  |  |
| --- | --- | --- | --- |
| **Гипотеза** | **Условие** | **Статистический критерий** | |
|  |  |
|  | неизвестна |  | *qt(1-)* |
|
|
|  | известна |  | (1-) |
|  | –  произвольное |  |  |
|  |  |  | (1-) |

**Пример 3.3.** Фирма-поставщик в рекламном буклете утверждает, что средний срок безотказной работы предлагаемого изделия – 2900 часов. Для выборки из 50 изделий средний срок безотказной работы оказался равным 2720 ч при выборочном среднем квадратическом отклонении 700 ч. При 5%-м уровне значимости проверить гипотезу о том, что значение 2900 ч является математическим ожиданием.

**Решение**

Предположим, что случайная величина срока безотказной работы подчинена нормальному закону распределения. Требуется проверить гипотезу о числовом значении математического ожидания (генеральной средней) при неизвестной генеральной дисперсии. Имеем: .

Другие альтернативные гипотезы нецелесообразны (потребитель обычно обеспокоен лишь тем, что срок службы изделия может оказаться меньше рекламируемого). Исходя из альтернативной гипотезы, критическая область – левосторонняя (рис. 3.9).



*Рис. 3.9. Проверка гипотезы*

Можно сделать вывод о том, что фирма в рекламе завышает срок безотказной работы изделия.

**3.3.2. Проверка гипотез о равенстве числовых характеристик двух совокупностей**

На практике часто встречаются ситуации, когда среднее значение

данных одного эксперимента отличается от среднего значения данных другого, хотя условия эксперимента являются схожими. Тогда возникает вопрос, можно ли считать это расхождение незначимым, т.е. чисто случайным, или оно вызвано существенным различием двух генеральных совокупностей. Например, такие вопросы возникают при исследовании надежности технических систем, где результаты сравниваются с предыдущими измерениями; при контроле качества изделий, изготовленных на разных предприятиях; в финансах – при сравнении уровня доходности различных активов.

Гипотезы о равенстве дисперсий возникают часто, поскольку дисперсия характеризует такие важные показатели, как точность приборов, технологических процессов, риск, связанный с отклонением доходности от заданного уровня, и т.д.

Сравнение долей признака в двух совокупностях является важной для практики задачей, так как значимое расхождение между долями признаков характеризует различие между генеральными совокупностями. Схемы проверки гипотез о равенстве числовых характеристик сведем в таблицу 3.4, в которой альтернативная гипотеза выбирается двусторонней.

Таблица 3.4

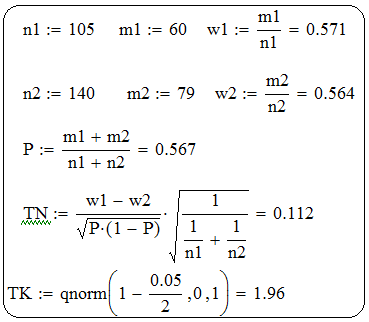
**Схемы проверки гипотез о равенстве числовых характеристик**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Гипотеза*** | ***Условие*** | ***Статистический критерий*** | |
|  |  |
|  | Дисперсии  и  известны |  | *qnorm* |
|
|
|  | Дисперсии неизвестны, но равны |  |  |
|  |  | (в числителе  должно быть большее значение) |  |
|  |  |  | *qnorm* |

**Пример 3.4.** Контрольную работу по математике по индивидуальным вариантам выполняли студенты двух групп первого курса. В первой группе было предложено 105 задач, из которых верно решено 60, во второй группе из 140 предложенных задач верно решено 80. Проверить гипотезу об отсутствии существенных различий в усвоении математики студентами обеих групп.

**Решение**

Имеем гипотезу , т.е. доли решенных задач студентами первой и второй групп равны. В качестве альтернативной возьмем гипотезу . При конкурирующей гипотезе выбираем двустороннюю критическую область (рис. 3.10).



*Рис.3.10. Проверка гипотезы*

*о равенстве долей*

Критическая область . Так как TN (наблюдаемое значение критерия) не входит в критическую область, то нулевую гипотезу принимаем.

**3.3.3. Проверка гипотез об однородности выборок**

В практике статистической обработки данных существуют задачи, связанные с необходимостью объединения нескольких массивов выборочных данных, полученных в различное время, из разнородных источников, от различных технических средств. Необходимость объединения диктуется желанием получить более полную информацию, повысить достоверность оценок.

*Необходимым условием однородности выборок служит равенство или достаточная близость математических ожиданий и дисперсий*.

***Критерий***  ***для проверки гипотезы***

***об однородности выборок***

1. Объединить выборки в один массив.

2. Для полученного массива рассчитать число интервалов, шаг, получить интервальный вариационный ряд.

3. При анализе однородности двух выборок, статистика имеет вид

где- число элементов первой и второй выборок, попавших в j-й интервал;

- общее число элементов двух выборок, попавших в j-й интервал;

*-* объемы первой и второй выборок, общий объем выборки.

4. Найти наблюдаемое значение критерия по специальной таблице или использовать встроенную функцию qchisq(1-α,k-1), k – число интервалов.

5. Задаваясь правосторонней критической областью, сделать вывод.

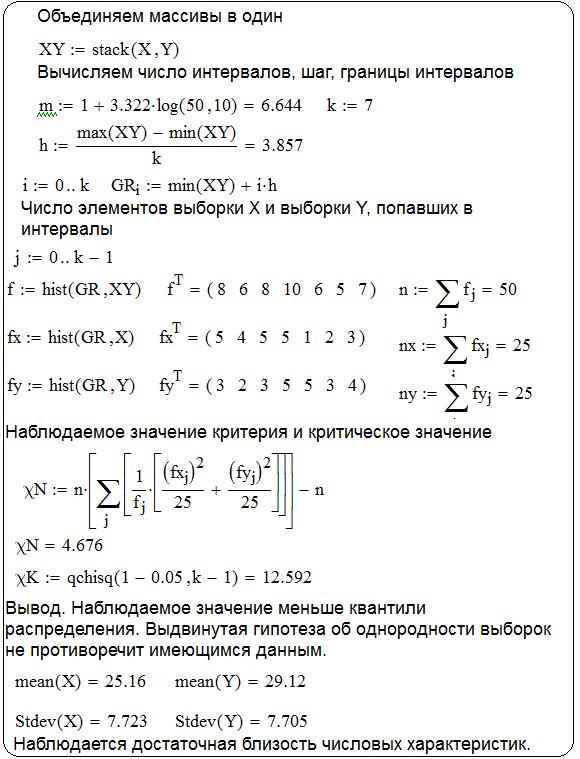
**Пример 3.5**. Из выборочных данных образованы две выборки равного объема. Проверить однородность выборок, используя критерий .

**Решение**

В среде Mathcad перечисленные пункты выполнены с применением функций stack (объединение массивов) и hist (определение частот, попавших в данный интервал). Получено 7 интервалов с границами, составляющими вектор GR.

Выборочное значение получилось меньше значения квантили распределения уровня 0,05 с k-1 степенями свободы.

Ввод выборок *X* и *Y* на рис. 3.11 не приводится.



*Рис. 3.11. Проверка однородности выборок*

*по критерию*

**Задачи для аудиторного решения**

1. Акционерное общество (АО) выпускает печенье «Русские узоры» в пачках, на которых написано: масса нетто 200 г. Осуществлена выборка для оценки средней массы печенья в пачках, выпущенных московской и санкт-петербургской фабриками АО. Результаты выборок таковы (указана масса пачек печенья «Русские узоры»):

*Московская фабрика*

201, 195, 197, 199, 202, 198, 199, 203, 195, 196, 198, 199, 194, 203, 195, 202, 197

*Санкт-петербургская фабрика*

203, 207, 191, 193, 197, 201, 196, 192, 194, 195, 198, 196.

Определить:

а) средние выборочные и исправленные средние квадратические каждой фабрики;

б) для 0,05 значимо или нет различие между средними выборочными (сначала проверить гипотезу );

в) является ли величина 200 г математическим ожиданием массы при 5%-м уровне значимости?

Можно ли считать выборки однородными?

**Компьютерная лабораторная работа № 4**

***Статистические гипотезы***

**Задание 1.**

1.1.Из большой партии ананасов одного размера случайным образом отобрано 36 штук. Выборочная средняя масса одной штуки при этом оказалась равной 930 г. Проверить гипотезу, что средняя масса одного ананаса (по утверждению поставщика) составляет 1 кг, если среднее квадратическое отклонение неизвестно, а выборочное составляет 250 г.

1.2.По утверждению руководства фирмы, средний размер дебиторского учета равен 187,5 тыс. руб. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборка равна 175 тыс. руб. при среднем квадратичном отклонении 35 тыс. руб. Может ли оказаться в действительности правильным объявленный размер дебиторского счета?

1.3.Автомат, работающий со стандартным отклонением г, фасует чай в пачки со средним весом г. В случайной выборке объема *n* = 25 пачек средний вес г. Надо ли отрегулировать автомат?

1.4. Автомат, работающий со стандартным отклонениемг, фасует чай в пачки со средним весом *a =* 80 г. В случайной выборке объема *n =* 16 пачек средний вес г. Надо ли отрегулировать автомат?

1.5.Станок, работающий со стандартным отклонением 0,5 мм, производит детали средней длины *a =* 20 мм. В случайной выборке объема *n =* 16 деталей средняя длина19,8 мм. Правильно ли настроен станок?

1.6. Станок, работающий со стандартным отклонением 0,4 мм, производит детали средней длины *a =* 30 мм. В случайной выборке объема *n =* 25 деталей средняя длина 30,1 мм. Правильно ли настроен станок?

1.7.Производитель утверждает, что средний вес пачки чая не меньше *a* = 100 г. Инспектор отобрал 10 пачек чая и взвесил. Их вес оказался 97, 102, 103, 98, 96, 105, 98, 100, 101, 99 г соответственно. Не противоречит ли это утверждению производителя? Предполагается, что вес пачек чая распределен нормально.

1.8. Производитель утверждает, что средний вес плитки шоколада не меньше *a* = 50 г. Инспектор отобрал 10 плиток шоколада и взвесил. Их вес оказался 49, 50, 51, 52, 48, 47, 49, 52, 48, 51 г соответственно. Не противоречит ли это утверждению производителя? Предполагается, что вес плитки шоколада распределен нормально.

1.9. Поставщик двигателей утверждает, что средний срок их службы равен 800 ч. Для выборки из 17 двигателей средний срок службы оказался равным 865 ч при выборочном среднем квадратическом отклонении 120 ч. Проверить нулевую гипотезу при уровне значимости 5%.

1.10.Фирма-изготовитель женских украшений, выпустив новый товар, утверждает, что 40% покупателей купят эти украшения. В ходе 10-дневной рекламной распродажи в среднем приобрели украшения 29,5% покупателей, выборочное среднее квадратическое отклонение составило 16,5%. При 5%-м уровне значимости оценить утверждение изготовителя товара.

1.11. Станок, работающий со стандартным отклонением 0,3 мм, производит детали средней длины *a =* 20 мм. В случайной выборке объема *n =* 25 деталей средняя длина19,8 мм. Правильно ли настроен станок?

1.12. Производитель утверждает, что средний вес плитки шоколада не меньше *a* = 100 г. Инспектор отобрал 10 плиток шоколада и взвесил. Их вес оказался 99, 100, 101, 102, 98, 97, 99, 102, 98, 101 г соответственно. Не противоречит ли это утверждению производителя? Предполагается, что вес плитки шоколада распределен нормально.

1.13. Производитель утверждает, что средний вес пачки чая не меньше *a* = 100 г. Инспектор отобрал 10 пачек чая и взвесил. Их вес оказался 97, 102, 103, 98, 96, 105, 98, 100, 101, 99 г соответственно. Не противоречит ли это утверждению производителя? Предполагается, что вес пачек чая распределен нормально.

1.14. Поставщик двигателей утверждает, что средний срок их службы равен 900 ч. Для выборки из 25 двигателей средний срок службы оказался равным 885 ч при выборочном среднем квадратическом отклонении 120 ч. Проверить нулевую гипотезу при уровне значимости 5%.

1.15. По утверждению руководства фирмы, средний размер дебиторского учета равен 187,5 тыс. руб. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборка равна 175 тыс. руб. при среднем квадратичном отклонении 35 тыс. руб. Может ли оказаться в действительности правильным объявленный размер дебиторского счета?

**Задание 2.**

2.1.Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать 0,15. По данным из 25 отобранных изделий вычислена несмещенная дисперсия . Выяснить, обеспечивает ли станок требуемую точность.

2.2.Партия изделий принимается в том случае, если вероятность того, что изделие окажется соответствующим стандарту, будет не менее 0,95. Среди случайно отобранных 100 изделий оказалось 98 соответствующих стандарту. Можно ли принять партию?

2.3.Производитель утверждает, что доля бракованных изделий не превосходит 3 %. В случайной выборке объема *n* = 100 изделий оказалось 5 бракованных изделий. Не противоречит ли это утверждению производителя?

2.4. Производитель утверждает, что доля бракованных изделий не превосходит 7 %. В случайной выборке объема *n =* 150 изделий оказалось 16 бракованных изделий. Не противоречит ли это утверждению производителя?

2.5.Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать 0,12. По данным из 35 отобранных изделий вычислена несмещенная дисперсия . Выяснить, обеспечивает ли станок требуемую точность.

2.6.Партия изделий принимается в том случае, если вероятность того, что изделие окажется соответствующим стандарту, будет не менее 0,92. Среди случайно отобранных 150 изделий оказалось 108 соответствующих стандарту. Можно ли принять партию?

2.7.Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать 0,14. По данным из 25 отобранных изделий вычислена несмещенная дисперсия . Выяснить, обеспечивает ли станок требуемую точность.

2.8.Партия изделий принимается в том случае, если вероятность того, что изделие окажется соответствующим стандарту, будет не менее 0,97. Среди случайно отобранных 120 изделий оказалось 112 соответствующих стандарту. Можно ли принять партию?

2.9.Производитель утверждает, что доля бракованных изделий не превосходит 2 %. В случайной выборке объема *n* = 100 изделий оказалось 4 бракованных изделий. Не противоречит ли это утверждению производителя?

2.10. Производитель утверждает, что доля бракованных изделий не превосходит 5 %. В случайной выборке объема *n =* 150 изделий оказалось 16 бракованных изделий. Не противоречит ли это утверждению производителя?

2.11. Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать 0,1. По данным из 25 отобранных изделий вычислена несмещенная дисперсия . Обеспечивает ли станок требуемую точность.

2.12. Производитель утверждает, что доля бракованных изделий не превосходит 1,5 %. В случайной выборке объема *n* = 100 изделий оказалось 3 бракованных изделий. Не противоречит ли это утверждению производителя?

2.13. Партия изделий принимается в том случае, если вероятность того, что изделие окажется соответствующим стандарту, будет не менее 0,95. Среди случайно отобранных 120 изделий оказалось 112 соответствующих стандарту. Можно ли принять партию?

2.14. Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать 0,14. По данным из 25 отобранных изделий вычислена несмещенная дисперсия . Выяснить, обеспечивает ли станок требуемую точность.

2.15. Производитель утверждает, что доля бракованных изделий не превосходит 5 %. В случайной выборке объема *n* = 100 изделий оказалось 6 бракованных изделий. Не противоречит ли это утверждению производителя?

**Задание 3.**

3.1.Было произведено измерений диаметра вала (в мм). При этом оказалось, что среднее , а стандартное среднее квадратичное отклонение . Затем вал поместили в условия с высокой температурой и провели еще измерений диаметра его оси. Среднее на этот раз оказалось равным, а стандартное отклонение . Можно ли сделать вывод, что диаметр вала существенно увеличивается при увеличении температуры?

3.2.Расходы сырья и на единицу продукции по старой и новой технологиям приведены в таблице.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | По старой технологии | | | | По новой технологии | | | | |
| Расход  сырья |  | 304 | 307 | 308 |  | 303 | 304 | 306 | 308 |
| Число  изделий |  | 1 | 4 | 4 |  | 2 | 6 | 4 | 1 |

Предполагается, что генеральные совокупности *X* и *Y* имеют нормальные распределения с одинаковыми дисперсиями и средними и . Требуется проверить гипотезу против .

3.3.Контрольную работу по высшей математике выполняли студенты двух групп. В первой группе было предложено 100 задач, из которых были правильно решены 58, во второй группе из 120 задач верно решены 65. Проверить гипотезу о том, что материал одинаково усвоен студентами обеих групп.

3.4.Инвестиция 1 рассчитана на лет, выборочная дисперсия ежегодных прибылей составляет 20 %. Инвестиция 2 рассчитана на лет, выборочная дисперсия ежегодных прибылей составляет 30 %. Предполагается, что распределение ежегодных прибылей на инвестиции подчиняется нормальному закону распределения. Равны ли риски 1 и 2?

3.5.Инвестиция 1 рассчитана на 14 лет, выборочная дисперсия ежегодных прибылей составляет 15 %. Инвестиция 2 рассчитана на =12 лет, выборочная дисперсия ежегодных прибылей сосатвляет 20 %. Предполагается, что распределение ежегодных прибылей на инвестиции подчиняется нормальному закону распределения. Раны ли риски 1 и 2?

3.6.На двух токарных станках обрабатываются втулки. Отобраны две пробы: из втулок, сделанных на первом станке, 15 шт., на втором станке -шт. По данным этих выборок рассчитаны выборочные дисперсии - 8,5 (для первого станка) и 6,3 (для второго станка). Полагая, что размеры втулок подчиняются нормальному закону распределения, выяснить, можно ли считать, что станки обладают различной точностью.

3.7.Автомат 1 и автомат 2 фасуют чай в пачки. Стандартные отклонения г и г соответственно. В случайной выборке объема пачек для автомата 1 средний вес =101 г. В случайной выборке пачек для автомата 2 средний вес =98 г. Верно ли, что оба автомата фасуют чай в пачки одинакового среднего веса?

3.8. Проводились испытания нового лекарства. В эксперименте участвовали 3000 мужчин и 3500 женщин. У 50 мужчин и 110 женщин наблюдались побочные эффекты. Можно ли утверждать, что побочные эффекты от нового лекарства у женщин возникают чаще, чем у мужчин?

3.9.Контрольную работу по математике по индивидуальным вариантам выполняли студенты двух групп первого курса. В первой группе было предложено 105 задач, из которых верно решено 60, во второй группе из 140 предложенных задач верно решено 69. Проверить гипотезу об отсутствии существенных различий в усвоении математики студентами обеих групп.

3.10**.** Проводились испытания нового лекарства. В эксперименте участвовали 3500 мужчин и 3000 женщин. У 150 мужчин и 120 женщин наблюдались побочные эффекты. Можно ли утверждать, что побочные эффекты от нового лекарства у мужчин возникают чаще, чем у женщин?

3.11. Контрольную работу по высшей математике выполняли студенты двух групп. В первой группе было предложено 10 задач, из которых были правильно решены 6, во второй группе из 12 задач верно решены 7. Проверить гипотезу о том, что материал одинаково усвоен студентами обеих групп.

3.12. Автомат 1 и автомат 2 фасуют кофе в пачки. Стандартные отклонения г и г соответственно. В случайной выборке объема пачек для автомата 1 средний вес =101 г. В случайной выборке пачек для автомата 2 средний вес =98 г. Верно ли, что оба автомата фасуют чай в пачки одинакового среднего веса?

3.13. На двух токарных станках обрабатываются втулки. Отобраны две пробы: из втулок, сделанных на первом станке, 25 шт., на втором станке -шт. По данным этих выборок рассчитаны выборочные дисперсии - 8,5 (для первого станка) и 6,3 (для второго станка). Полагая, что размеры втулок подчиняются нормальному закону распределения, выяснить, можно ли считать, что станки обладают различной точностью.

3.14. Проводились испытания нового лекарства. В эксперименте участвовали 2500 мужчин и 3000 женщин. У 50 мужчин и 110 женщин наблюдались побочные эффекты. Можно ли утверждать, что побочные эффекты от нового лекарства у женщин возникают чаще, чем у мужчин?

3.15. Инвестиция 1 рассчитана на лет, выборочная дисперсия ежегодных прибылей составляет 15 %. Инвестиция 2 рассчитана на лет, выборочная дисперсия ежегодных прибылей составляет 10 %. Предполагается, что распределение ежегодных прибылей на инвестиции подчиняется нормальному закону распределения. Равны ли риски 1 и 2?

**Глава 4. Основы регрессионного и корреляционного анализа**

*Числа не управляют миром,*

*но показывают, как управляется мир.*И. Гёте

*Качество корреляционной зависимости*

*обратно пропорционально плотности точек.*

Один из постулатов Мэрфи

**4.1. Функциональная, статистическая и**

**корреляционная зависимости**

В математике существуют различные виды зависимостей между переменными.

***Функциональная зависимость*** – это зависимость, при которой каждому значению одной (или нескольких) переменной соответствует определенное значение другой переменной (зависимой).

Пусть наблюдению подлежит случайная величина *Y*, зависящая от одной или нескольких других случайных величин *X1*, *X2*, …, *Xk*, которые называются **факторами**. В общем случае число факторов может быть неизвестно. Исследователь выбирает *k* наиболее существенных факторов. В этих условиях функциональная зависимость между *Y* и *X* недостижима, так как не учтено влияние неопределенных факторов, т.е. или , где - стохастическая переменная, включающая влияние неучтенных факторов в модели. Говорят, что между *Y* и *X* существует ***стохастическая (или статистическая, вероятностная) связь.*** Пример статистической связи – зависимость урожайности от количества внесенных удобрений и механизации предприятия.

***Статистической***называют зависимость между случайными величинами *Х* и *Y*, при которой изменение одной из случайных величин влечёт изменение распределения другой случайной величины.

***Корреляционной***называют статистическую зависимость, проявляющуюся в том, что при изменении одной из случайных величин изменяется среднее значение другой случайной величины.

Реализации двумерной случайной величины (*Х*, Y) представляются точками (*хi*, *yi*). Совокупность таких точек образует *корреляционное поле*.

Корреляционная зависимость характеризуется *линией* регрессии и *коэффициентом корреляции.*

Корреляционная зависимость может быть представлена в виде уравнения, которое называется модельным уравнением регрессии (или просто уравнением регрессии).

***Задача корреляционного анализа*** – выявление связи между случайными переменными и оценка ее тесноты.

***Задача регрессии*** - выбор модели зависимости междупеременными и определение оценок неизвестных параметров этой модели.

Выбор модели регрессионных зависимостей осуществляется исходя из теоретических представлений о возможной взаимосвязи между переменнымиили из визуального анализа графиков наблюдений.

В зависимости от количества включенных в модель факторов *Х* модели делятся на *однофакторные* (парная модель регрессии) и *многофакторные* (модель множественной регрессии).

В зависимости от вида функции модели делятся на *линейные* и *нелинейные*.

**4.2. Линейная регрессия**

При одновременном изучении двух признаков получают двумерную выборку. Пары точек наносят на координатную сетку, получая так называемое **облако точек**, которое дает предварительное представление о рассеянии и форме зависимости между признаками.

На основании регрессионного анализа наблюдаемое облако аппроксимируется уравнением регрессии. Теоретически уравнение регрессии является условным математическим ожиданием одной случайной переменной (зависимой), при условии, что вторая переменная (независимая) принимает заданные (фиксированные) значения. Таким образом, регрессия – это всегда «зависимость в среднем». Если эта зависимость задается уравнением прямой

то говорят о линейной регрессии или **линейной регрессионной модели.** Параметры и оцениваются в большинстве случаев на основе метода наименьших квадратов (МНК). Если обе переменные принадлежат двумерному нормальному распределению, то уравнение регрессии может быть записано в виде уравнения прямой линии

или

Оценки этих линейных уравнений имеют вид:

Прямые пересекаются в точке и образуют «ножницы», причем тем уже, чем больше .

При обе прямые регрессии совпадают, а при прямые регрессии перпендикулярны осям координат (переменные независимы).

Построение линейной регрессионной модели (однофакторной и множественной), проведение статистического анализа приведено в таблицах 4.1, 4.2.

Таблица 4.1

**Линейная регрессия**

|  |  |  |
| --- | --- | --- |
| ***Однофакторная регрессия***  (на наблюдаемую переменную *Y* влияет один фактор *X*)  ***Замечание***. *Уравнение называется простой линейной регрессией или парной линейной регрессией.* | | ***Множественная регрессия***  (на наблюдаемую переменную *Y* влияют несколько факторов  ) |
| ***Для оценки неизвестных параметров β применяют***  ***метод наименьших квадратов (МНК)***,  суть которого состоит в минимизации суммы квадратов отклонений фактических значений результатного признака от его расчетных значений , т.е.: | | |
| ***Алгоритм МНК в форме обобщенного обращения матрицы*** | | |
| 1. Ввести исходные данные – массивы *Y* и *X*.  2. Составить матрицу , *n* -число наблюдений, 2- число неизвестных параметров  . | | 1. Ввести исходные данные – массивы *Y* и .  2. Составить матрицу , *n* -число наблюдений, *р+1*- число неизвестных параметров  . |
| 3. Матрица неизвестных параметров | | |
| В среде Mathcad для определения параметров простой линейной регрессии можно использовать встроенные функции  1.  ;  *2. line(X,Y)* | |  |
| ***Качество уравнения регрессии*** ***определяется по величине средней***  ***ошибки аппроксимации***  (уравнение можно использовать как прогностическую модель, если  ) | | |
| ***Влияние совокупности факторов на результат Y*** | | |
| ***Выборочный линейный коэффициент корреляции*** *(характеризует степень взаимосвязи пары случайных величин, если зависимость между ними соответствует прямой линии)*  – линейная функциональная связь, ;  - *Y* и *X* некоррелированы.  В среде Mathcad используют встроенную функцию  *corr(X,Y).* | | ***Выборочный сводный коэффициент корреляции*** *(характеризует связь Y со всеми факторами, входящими в уравнение)*  - *Y* имеет функциональную связь с совокупностью факторов;  - *Y* некоррелирован ни с одним из факторов. |
| ***Проверка значимости выборочного коэффициента корреляции***  ***(t-критерий Стьюдента)***  Н0 – изучаемый фактор (факторы) не оказывает существенного влияния на результат, т.е. коэффициент корреляции генеральной совокупности равен 0.  Н1 - коэффициент корреляции генеральной совокупности отличен от 0.  *p* – число факторов, влияющих на результат,  *п* – число измерений,  *w* – критическая область двусторонняя | | |
|  | |  |
| ***Проверка значимости уравнения регрессии (F-критерий Фишера)*** – *установить, соответствует ли**математическая модель экспериментальным данным и достаточно ли включенных в уравнение факторов (одного или нескольких) для описания зависимой переменной.*  - уравнение регрессии не надежное;  - уравнение регрессии надежное  *p* – число факторов, влияющих на результат,  *п* – число измерений,  *w* – критическая область правосторонняя | | |
|  | |  |
| ***Значимость отдельных (кроме свободного члена)***  ***коэффициентов регрессии (t-критерий Стьюдента)***  – коэффициент статистически не значим;  – коэффициент статистически значим  *p* – число факторов, влияющих на результат,  *п* – число измерений,  *w* – критическая область двусторонняя.  *Если коэффициент статистически не значим, то фактор, соответствующий этому коэффициенту следует исключить из модели (при этом ее качество не ухудшится)* | | |
|  |  | |
| ***Коэффициенты эластичности и детерминации***  1***.***Коэффициент эластичности показывает, на сколько процентов в среднем изменяется результативный признак *Y* при изменении факторного признака *Xi* на 1%. Высокий уровень эластичности означает сильное влияние независимой переменной на объясняемую переменную.  2. Коэффициент детерминации ( показывает долю вариации результативного признака, объясненную вариацией факторного признака. Чаще всего, давая интерпретацию коэффициента детерминации, его выражают в процентах. | | |

Таблица 4.2

**Качественная оценка тесноты связи**

**(шкала Чеддока)**

|  |  |
| --- | --- |
| **Величина коэффициента**  **парной корреляции** | **Характеристика силы связи** |
| до 0,3 | практически отсутствует |
| 0,3-0,5 | слабая |
| 0,5-0,7 | заметная |
| 0,7-0,9 | сильная |
| 0,9-0,99 | очень сильная |

***О ложной корреляции (влияние «третьего фактора»)***

Часто корреляцию и причинную обусловленность считают синонимами, поскольку если нечто является причиной чего-либо другого, то можно говорить о связи первого и второго и, следовательно, об их коррелированности (например, действие и результат, проверка и качество, капиталовложения и прибыль). Однако корреляция может быть и без причинной обусловленности. Это можно представить так: корреляция - лишь число, которое указывает на то, что большим значениям одной переменной соответствуют большие (или меньшие) значения другой переменной. Корреляция не может объяснить, почему эти две переменные связаны между собой.

В истории статистики известен один классический пример (исследование под названием «Аисты приносят детей»). В шведской столице в течение 73 лет регистрировалось число новорожденных в год (*у*) и число аистов (*х*), которых содержало население. Данные были сведены в таблицу, по ним был рассчитан коэффициент парной корреляции. Он оказался близок к единице, экспериментальные точки «улеглись на прямую», т.е. практически указанную связь следовало толковать как функциональную.

Поскольку утверждение, содержащее в тезисе, довольно сомнительное, было решено поискать другое разумное объяснение. Оказалось, что одновременные синхронные изменения числа аистов и числа детей объясняются изменением среднего уровня жизни жителей Стокгольма. Эта переменная первоначально не являлась предметом рассмотрения, отчего и случился такой забавный курьез вследствие ложной корреляции.

***Особенности практического применения***

***линейных множественных регрессионных моделей***

Одним из условий регрессионной модели является предположение о линейной независимости объясняющих переменных, т. е., решение задачи возможно лишь тогда, когда столбцы и строки матрицы ис­ходных данных линейно независимы.

Под ***мультиколлинеарностью*** понимается высокая взаимнаякоррелированность объясняющих переменных (факторов), которая приводит к линейной зависимости нормальных уравнений. Существует несколько способов для определения наличия или отсутствия мультиколлинеарности. Один из подходов заключается в анализе коэффициентов парной корреляции.

1. Факторные признаки, у которых исключают из модели.

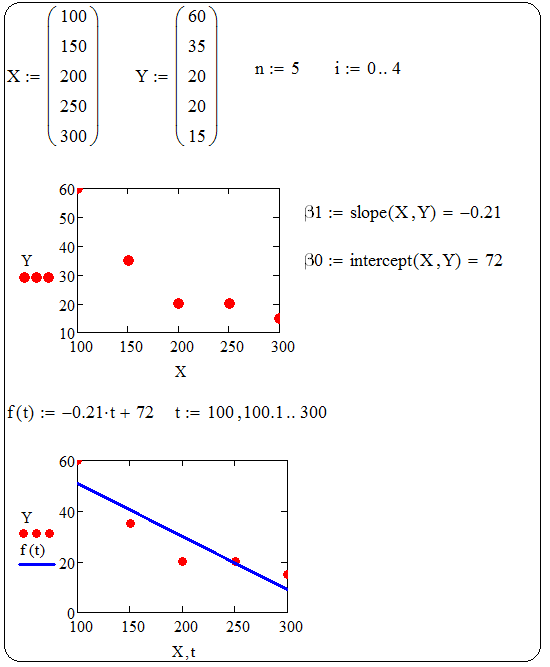
2. Считают явление мультиколлинеарности в исходных данных установленным, если коэффициент парной корреляции между двумя переменными (факторами) больше 0,8. В этом случае одну переменную исключают из рассмотрения. При этом какую пе­ременную оставить, а какую удалить из анализа, решают в первую очередь на основании экономических соображений. Если ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

**Пример 4.1.** С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах проведены измерения уровня месячной зарплаты *X* и числа уволившихся за год рабочих *Y:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *X* | 100 | 150 | 200 | 250 | 300 |
| *Y* | 60 | 35 | 20 | 20 | 15 |

Найти линейную регрессию *Y* на *X.* Провести статистический анализ.

**Решение**

****

*Рис. 4.1. Однофакторная линейная регрессия*

На рис. 4.1 решена задача простой линейной регрессии. Для нахождения коэффициентов линейной регрессии использованы встроенные функции ***slope, intercept***. Проведен сравнительный визуальный анализ расположения экспериментальных точек и полученной функциональной зависимости.

Проведем статистический анализ (рис. 4.2).

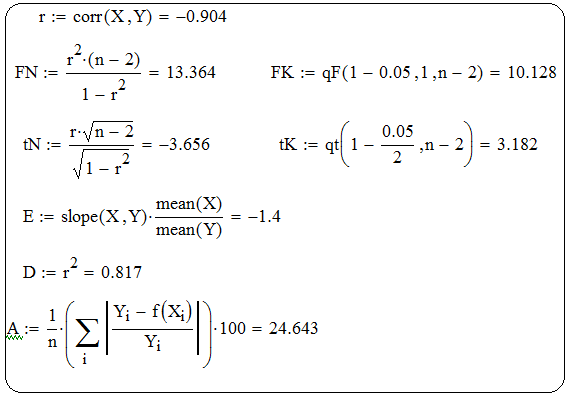
1. Коэффициент корреляции равен (-0,904), следовательно, связь тесная и обратная. Коэффициент корреляции значим, т.е. зарплата оказывает существенное влияние на текучесть рабочей силы.

2. Полученное уравнение регрессии является значимым, т.е. для описания зависимой переменной достаточно одного фактора.

3. Коэффициент эластичности равен (-1,4), т.е. при увеличении зарплаты на 1% текучесть рабочей силы уменьшается в среднем на 1,4%.

4. Коэффициент детерминации равен 0,817, т.е. на 81,7% текучесть рабочей силы зависит от заработной платы и на 18,3% от других неучтенных факторов.

5. Ошибка аппроксимации составляет 24,6%. Следовательно, полученную функциональную зависимость не следует использовать для прогнозирования. Рекомендуется подобрать другую однофакторную модель регрессии (нелинейную).

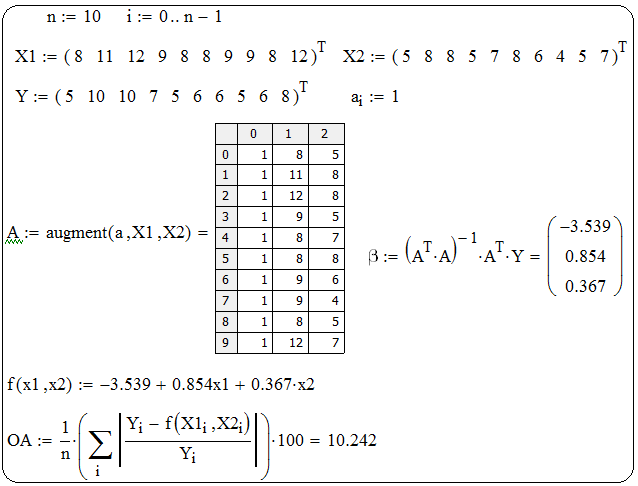


*Рис. 4.2. Статистический анализ*

**Пример 4.2.** Имеются следующие данные о сменной добыче угля на одного рабочего *Y* (т), мощности пласта *X*1(м) и уровне механизации работ *X2* (%), характеризующие процесс добычи угля в 10 шахтах. Предполагая, что между переменными *Y, X*1 и *X*2существует линейная корреляционная зависимость, найти ее аналитическое выражение (уравнение регрессии *Y* по *X*1 и *X*2*)*. Провести статистический анализ.

**Решение**

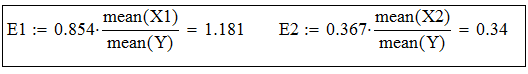
1.Матрицу *А* формируем объединением столбцов с элементами 1, *Х*1, *Х*2 с применением встроенной функции ***augment*** (рис. 4.3). По алгоритму получаем оценки параметров β, что позволяет составить уравнение двухфакторной линейной регрессии. Ошибка аппроксимации составляет 10,24%, что позволяет использовать полученное уравнение в прогностических целях.



*Рис. 4.3. Вариант двухфакторной линейной регрессии*

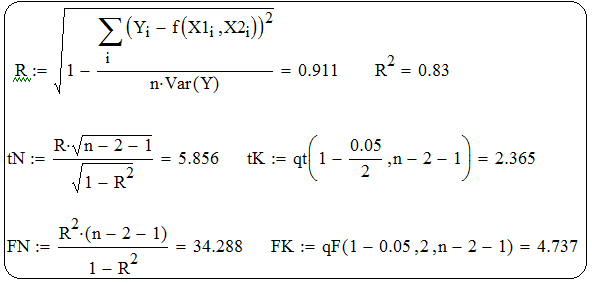
Уравнение множественной регрессии показывает, что при увеличении только мощности пласта *Х*1 (при неизменном *Х*2) на 1 м, добыча угля на одного рабочего увеличивается в среднем на 0,854 т, а при увеличении только уровня механизации работ *Х*2 (при неизменном *Х*1) – в среднем на 0,367 т.

2. Проведем сравнение раздельного влияния на добычу угля двух факторов – мощности пласта *Х*1 и уровня механизации работ *Х*2. Для этого вычислим коэффициенты эластичности (рис. 4.4).

  
*Рис. 4.4. Коэффициенты эластичности*

Увеличение факторов на 1% (от своих средних значений) приводит в среднем к росту добычи угля соответственно на 1,18% и 0,34%. Таким образом, на добычу угля большее влияние оказывает фактор *Х*1 (мощность пласта) по сравнению с фактором *Х*2 (уровень механизации).

3. Определим множественный коэффициент корреляции и проверим его значимость и значимость полученного уравнения регрессии (рис. 4.5).



*Рис. 4.5. Вычисление множественного*

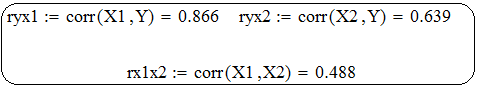
*коэффициента корреляции*

Значение , близкое к 1, указывает на тесную взаимосвязь переменной *Y* и факторов *Х*1 и *Х*2. Коэффициент детерминации свидетельствует о том, что вариация исследуемой зависимой переменной на 83% объясняется изменчивостью включенных в модель факторов.

Проверим значимость коэффициента корреляции. Наблюдаемое значение tN принадлежит критической области, следовательно, гипотезу о равенстве нулю коэффициента корреляции генеральной совокупности отвергаем.

Проверим значимость уравнения регрессии. Наблюдаемое значение FN принадлежит критической области, следовательно, гипотезу «уравнение регрессии ненадежно» отвергаем.

4. Мультиколлинеарность (рис. 4.6).



*Рис. 4.6. Парные коэффициенты корреляции*

Так как , то явление мультиколлинеарности не установлено;, следовательно, факторы не исключаются из модели.

**Задачи для аудиторного решения**

1. Дана корреляционная таблица зависимости между суточной выработкой продукции *Y* (т) и величиной основных производственных фондов *Х* (млн руб.) для 50 однотипных предприятий.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Величина ОПФ, млн руб. (*Х*)** | **Суточная выработка продукции, т (*Y*)** | | | | |
| **7-11** | **11-15** | **15-19** | **19-23** | **23-27** |
| **20-25** | 2 | 1 | - | - | - |
| **25-30** | 3 | 6 | 4 | - | - |
| **30-35** | - | 3 | 11 | 7 | - |
| **35-40** | - | 1 | 2 | 6 | 2 |
| **40-45** | - | - | - | 1 | 1 |

Найти параметры простой линейной регрессии. Провести статистический анализ.

2. В процессе тренировок многоборцев анализировались причины промахов спортсменов на этапе стрельбы. Для десяти спортсменов перед стрельбой фиксировались: сила бокового ветра (фактор *Х*1), отклонение частоты пульса спортсменов от 80 в течение двух минут (фактор *Х*2) и максимальное отклонение попаданий (см) в серии выстрелов (*Y*).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Х*1 | -1,49 | -0,45 | -2,77 | -2,2 | -2,78 | 2,18 | 2,03 | -0,6 | -2,23 | -0,6 |
| *Х*2 | -0,65 | -1,12 | 0,77 | 6,98 | -1,19 | 8,61 | 3,86 | -8,27 | -0,47 | 9,08 |
| *Y* | -0,1 | -0,33 | -0,64 | 0,42 | -0,75 | 1,57 | 0,75 | -0,7 | 0,82 | 0,84 |

Найти параметры множественной линейной регрессии. Провести статистический анализ.

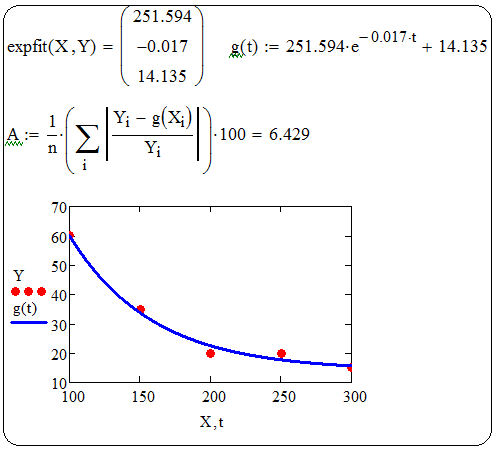
**4.3. Нелинейная однофакторная регрессия**

|  |  |
| --- | --- |
| ***Регрессии, нелинейные***  ***относительно включенных факторов, но линейные по***  ***оцениваемым параметрам*** | ***Регрессии, нелинейные по***  ***оцениваемы параметрам*** |
| Полиномы разных степеней, например,  .  ***Заменяя переменные, например, , получим уравнение множественной линейной регрессии.*** | 1.Внутренне линейные, например, функции:  степенная  показательная  экспоненциальная .  ***С помощью соответствующих преобразований приводятся к простой линейной регрессии.***  2. Внутренне нелинейные, например, функции:  ***Используют итеративные процедуры.*** |

Библиотека типовых функций регрессии Mathcad приведена в Приложении 3.

**Пример 4.5.** По условию примера 4.1 подобрать нелинейную модель регрессии, используя встроенные функции Mathcad (Приложение 3).

**Решение**



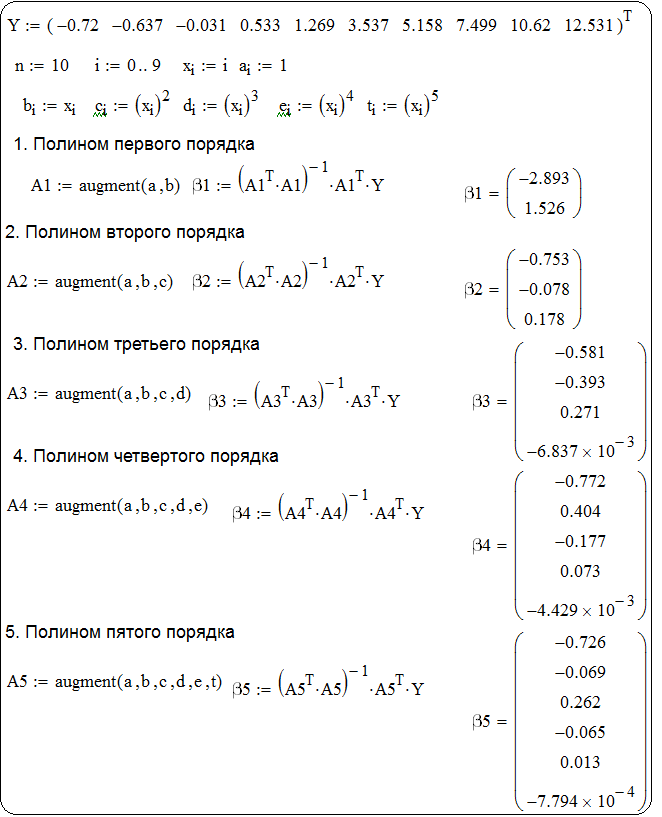
*Рис. 4.11. Решение задачи нелинейной регрессии*

**Пример 4.3.** В контрольный период проведено 10 измерений некоторой величины. Измерения проводились через фиксированный интервал времени. Результаты наблюдений *Y* сведены в таблицу. Решить задачу регрессии, используя в качестве зависимостей полиномы с 1 по 5 степеней. Сопоставить точности вариантов.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| *Y* | -0,72 | -0,637 | -0,031 | 0,533 | 1,269 | 3,537 | 5,158 | 7,499 | 10,62 | 12,531 |

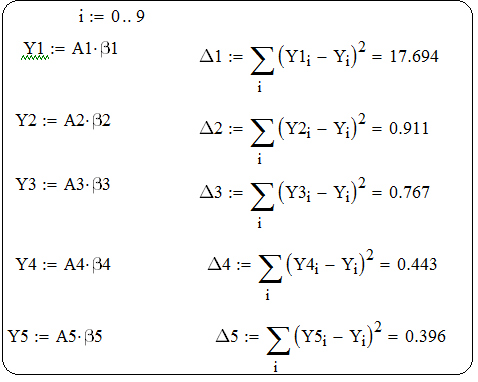
**Решение**

На рис. 4.7 приведены результаты решения в среде MathCAD. Решения для полиномов разных степеней проведены по одному алгоритму.



*Рис. 4.7. Варианты регрессии*

Для сопоставления полиномиальных регрессионных выражения (от полинома первого порядка до пятого порядка включительно) вычислены остаточные суммы квадратов, которые обозначены Δ (рис. 4.8). Проведенные вычисления подтверждают качественную зависимость остаточной суммы от порядка полинома.

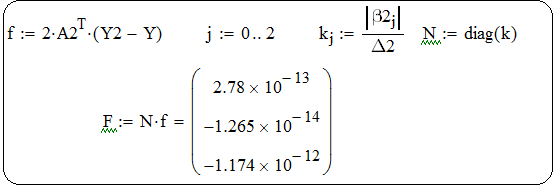
****

*Рис. 4.8. Сопоставление полиномиальных регрессий*

При выборе регрессионной модели путем сопоставления остаточных сумм квадратов (при условии их близких значений) выбирают более простой вариант модели. В качестве дополнительного упрощения выбранной регрессионной модели применяется ***функция чувствительности*** *,*с помощью которой определяют вклад оценки каждого параметра в результирующее значение.

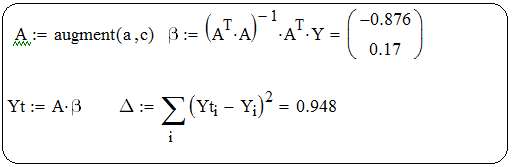
Функция чувствительности *f* представляет собой *m*-мерный вектор-столбец. Умножая вектор *f* слева на диагональную матрицу *N*, элементы диагонали которой составляют вектор , получаем вектор относительных функций чувствительности

Пусть выбранная регрессионная модель – полином второго порядка . На рис. 4.9 получен вектор относительной чувствительности для полинома второго порядка.

****

*Рис. 4.9. Получение вектора относительной чувствительности*

Анализ вектора*F* показывает, что наименьшая функция чувствительности у коэффициента , следовательно, возможно применить упрощенную регрессионную модель .

****

*Рис. 4.10. Решение упрощенной задачи регрессии*

Значение остаточной суммы квадратов**,** равное 0,948, незначительно отличается от значения 0,911. Таким образом, выбранная регрессионная модель имеет вид .

**Задачи для аудиторного решения**

1. Исследовать зависимость урожайности зерновых культур *Y* (ц/га) от количества осадков *Х* (см), выпавших в вегетационный период.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 25 | 27 | 30 | 35 | 36 | 38 | 39 | 41 | 42 | 45 | 46 | 47 | 50 | 52 | 53 |
| *Y* | 23 | 24 | 27 | 27 | 32 | 31 | 33 | 35 | 34 | 32 | 29 | 28 | 25 | 24 | 25 |

***Указание.*** *При выборе регрессионной модели можно учесть, что увеличение количества осадков приводит к увеличению урожайности до некоторого предела, после чего урожайность будет снижаться.*

2. Имеются следующие данные о выработке литья на одного работающего *Х*1 (т), браке литья *Х*2 (%) и себестоимости одной тонны литья *Y* (руб.) по 25 литейным цехам завода:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **№** | *Х*1 | *Х*2 | *Y* | **№** | *Х*1 | *Х*2 | *Y* |
| **1** | 14,6 | 4,2 | 239 | **14** | 75,8 | 3,4 | 172 |
| **2** | 13,5 | 6,7 | 254 | **15** | 27,6 | 1,1 | 201 |
| **3** | 21,5 | 5,5 | 262 | **16** | 88,4 | 0,1 | 130 |
| **4** | 17,4 | 7,7 | 251 | **17** | 16,6 | 4,1 | 251 |
| **5** | 44,8 | 1,2 | 158 | **18** | 33,4 | 2,3 | 195 |
| **6** | 111,9 | 2,2 | 101 | **19** | 17,0 | 9,3 | 282 |
| **7** | 20,1 | 8,4 | 259 | **20** | 33,1 | 3,3 | 196 |
| **8** | 28,1 | 1,4 | 186 | **21** | 30,1 | 3,5 | 186 |
| **9** | 22,3 | 4,2 | 204 | **22** | 65,2 | 1,0 | 176 |
| **10** | 25,3 | 0,9 | 1998 | **23** | 22,6 | 5,2 | 238 |
| **11** | 56,0 | 1,3 | 170 | **24** | 33,4 | 2,3 | 204 |
| **12** | 40,2 | 1,8 | 173 | **25** | 19,7 | 2.7 | 205 |
| **13** | 40,6 | 3,3 | 197 |

Найти уравнение множественной линейной регрессии *Y* по *Х*1 и *Х*2. Провести статистический анализ.

**Компьютерная лабораторная работа № 5**

***«Корреляционно-регрессионный анализ»***

**Задание 1**. По приведенным статистическим данным:

- построить поле корреляции и сформулировать гипотезу о форме связи;

- рассчитать параметры уравнения линейной регрессии;

- оценить тесноту связи с помощью показателей корреляции и детерминации;

- дать с помощью коэффициента эластичности сравнительную оценку силы связи фактора с результатом;

- с помощью *F*-критерия Фишера оценить статистическую надежность результатов регрессионного моделирования;

- с помощью t-критерия Стьюдента оценить статистическую значимость коэффициента корреляции;

- рассчитать параметры нелинейной регрессии (степенной, экспо­ненциальной);

- оценить с помощью средней ошибки аппроксимации качество уравнений.

**1.**

(4,570; 3,558), (3,017; 3,825), (3,511; 3,499), (4,393; 5,793), (5,522; 3,975), (3,066; 4,913), (4,657; 5,036), (5,143; 4,547), (3,824; 5,904), (3,248; 6,784), (3,105; 3, 708), (3,857; 5,002), (3,701; 3,124), (3,662 3,725), (5,194; 3,165), (3,190; 3,103), (2,405; 3,271), (2,807; 3,128), (3,824; 2,958), (3,631; 6,284), (4,879; 3,372), (6,959; 3,533), (4,354; 3,143), (3,651; 5,197), (5,426; 4,478), (3,229; 3,528), (3,547; 5,927), (3,296; 5,231), (4,025; 3,502), (6,285; 5,717).

**2.**

(11,49; 8,52), (10,28; 11,31), (11,65; 10,36), (11,39; 10,81), (12,15; 10,35), (9,49; 15,58), (9,92; 11,62), (11,00; 13,60), (11,78; 9,76), (12,92; 12,82), (9,76; 9,61), (12,37; 10,23), (9,46; 10,35), (10,45; 9,13), (15,72; 12,40), (12,84; 10,53), (13,00; 11,28), (12,51; 10,23), (14,07; 13,14), (10,46; 12,46), (11,75; 10,45), (12,09; 11,69), (12,72; 10,92), (15,49; 11,43), (12,14; 12,41), (11,26; 13,49), (11,81; 12,17), (9,13; 12,89), (12,24; 11,14), (13,59; 12,98), (9,55; 13,06), (15,88; 12,28), (13,65; 9,82), (9,64; 12,45), (10,18; 8,91), (11,15; 12,21), (9,98; 10,75), (9,27; 14,97), (10,75; 11,01), (12,60; 12,43).

**3.**

(3,96; 2,61), (3,55; 3,15), (3,66; 3,92), (2,93; 2,89), (4,61; 3,51), (2,58; 6,15), (2,99; 4,35), (4,40; 5,35), (4,86; 3,24), (4,27; 4,67), (3,32; 2,09), (5,38; 3,11), (1,93; 3,36), (3,67; 2,64), (6,27; 5,17), (4,14; 2,81), (5,43; 4,22), (4,31; 2,95), (5,29; 4,88), (4,20; 5,54), (3,72; 2,46), (3,99; 4,37), (4,59; 2,71), (6,17; 3,51), (4,22; 4,06), (4,86; 4,78), (3,62; 5,50), (2,60; 4,69), (4,12; 3,14), (6,03; 4,42), (2,86; 5,71), (6,28; 4,46), (5,41; 3,27), (2,82; 6,06), (2,42; 2,44), (3,39; 3,13), (2,50; 2,95), (2,91; 5,43), (2,99; 3,64), (4,20; 5,25).

**4.**

(18,20; 13,99), (16,06; 14,05), (13,97; 15,34), (15,62; 17,94), (18,36; 15,25), (14,97; 16,98), (13,11; 16,90), (16,41; 17,43), (14,44; 16,93), (16,19; 20,97), (14,90; 11,96), (16,09; 17,66), (15,27; 14,07), (14,66; 15,44), (20,00; 15,73), (11,84; 15,59), (12,52; 15,06), (17,91; 15,21), (11,12; 14,49), (17,22; 20,84), (18,12; 16,24), (19,94; 20,89), (16,29; 15,38), (17,44; 17,10), (17,48; 17,34), (15,09; 12,08), (11,76; 18,00), (15,74; 18,74), (16,03; 15,63), (19,86; 18,63).

**5.**

(264; 120), (144; 48), (48; 48), (552; 48), (72; 24), (288; 48), (240; 48), (336; 168), (24; 528), (72; 96), (72; 48), (48; 72), (168; 96), (72; 48), (96; 48), (96; 48), (24; 96), (168; 96), (48; 48), (72; 264), (72; 96), (24; 72), (48; 48), (480; 144), (24; 72), (48; 144), (96; 168), (144; 216), (336; 24), (48; 168), (456; 48), (48; 552), (96; 24), (72; 144), (192; 96), (48; 24), (24; 24), (24; 48), (24; 96), (24; 96).

**6.**

(9,0; 10,0), (5,0; 17,0), (8,0; 8,6), (6,0; 10,5), (3,0; 5,0), (5,3; 4,0), (5,0; 3,0), (4,0; 3,0), (14,0; 13,5), (5,8; 4,5), (8,5; 7,5), (5,0; 5,2), (16,0; 9,0), (19,1; 6,5), (3,9; 23,0), (6,0; 5,0), (24,0; 4,0), (22,0; 8,0), (8,0; 14,0), (4,5; 5,0), (3,0; 8, 6), (7,5; 8,0), (5,0; 1,1), (10,0; 6,5), (5,0; 7,0), (4,0; 9,3), (14,5; 4,5), (7,0; 9,0), (9,0; 7,0), (6,0; 9,0).

**7.**

(250, 530), (620, 395), (471, 25), (370, 70), (95, 0), (90, 260), (1027, 0), (695, 105), (385, 522), (260, 35), (445, 360), (125, 100), (230, 60), (275, 725), (70, 40), (970, 445), (534, 325), (100, 439), (1140, 20), (0, 690), (280, 247), (440, 91), (300, 140), (360, 320), (85, 130), (337, 1133), (1140, 0), (165, 723), (95, 240), (53, 450).

**8.**

(156; 18), (43; 29), (83; 54), (44; 58), (27; 32), (48; 81), (48; 42), (28; 91), (45; 98), (52; 49), (142; 20), (60; 54), (19; 61), (25; 156), (32; 79), (36; 80), (88; 21), (50; 19), (78; 52), (12; 118), (28; 41), (26; 48), (22; 83), (22; 30), (109; 42), (36; 35), (54; 41), (47; 69), (142; 20), (21; 14), (58; 68), (67; 31), (35; 32), (43; 17), (71; 29), (14; 34), (59; 20), (37; 20), (61; 23), (26; 24).

**9.**

(-304; -386), (35; -305), (-330; -105), (-400; -234), (-185; -160), (-160;-285), (-370; -343), (65; -35), (-51; 45), (-380; -388), (-68; 10), (48; -340), (-361;-475), (-2; -320), (-395; -240), (-356; -67), (35; -398), (-268; 70), 19; (-362; 0), (73; -10), (-192; -310), (-285; -404), (-300; 60), (-400; 5), (-349; -305), (21; -400), (-375; -80), (-365; -272), (-355; -363), (-380; -266).

**10.**

(28; -111), (115; -111), (-203; -32), (440; 98), (-353; 29), (360; 77), (79; -361), (330; -300), (-363; -105), (250; -329), (-302; 182), (-475; -322), (-276; -201), (-145; 0), (238; -115), (455; -46), (0; 0), (-109; -236), (0; 275), (86; 58), (-354; 40), (-398; 76), (-106; 95), (-185; -233), (95; 0), (-345; 0), (92; -158), (-97; -350), (200; 0), (109; -329), (254; -345), (227; -371), (370; 280), (0; -90), (95; -203), (-112; 52), (158; -70), (-142; 260), (-282; -358), (142; -299).

**11.**

(405; 142), (115; 190), (180; 90), (440; 280), (25; 382), (360; 160), (443; 270), (330; 270), (0; 360), (250; 490), (70; 395), (90; 440), (105; 50), (225; 65), (238; 273), (455; 60), (0; 545), (280; 35), (0; 180), (458; 0), (25; 260), (0; 325), (320; 0), (180; 150), (460; 275), (30; 450), (475; 440), (293; 450), (200; 475), (499; 160), (254; 0), (227; 0), (370; 220), (0; 90), (455; 0).

**12.**

(96; 216), (96; 48), (72; 72), (72; 120), (48; 96), (24; 48), (96; 144), (240; 48), (168; 72), (96; 72), (72; 48), (168; 48), (48; 120), (216; 72), (168; 96), (144; 48), (96; 192), (96; 48), (48; 144), (72; 96), (96; 120), (72; 96), (144; 72), (72; 48), (48; 168), (48; 192), (96; 216), (96; 120), (72; 48), (96; 96), (72; 144), (168; 72), (72; 120), (48; 144), (120; 72), (72; 72), (72; 48), (96; 96), (72; 96), (48; 96).

**13.**

(2,96; 3,26), (5,93; 3,09), (4,11; 2,22), (2,61; 3,95), (2,91; 3,33), (1,75; 2,79), (3,07; 3,49), (3,30; 2,10), (2,68; 1,71), (2,97; 5,22), (2,65; 2,66), (2,48; 3,55), (3,45; 3,76), (2,86; 1,77), (3,62; 3,09), (3,11; 4,10), (4,12; 4,85), (3,66; 3,10), (3,32; 2,63), (3,26; 2,13), (3,10; 3,66), (4,81; 2,59), (7,14; 3,25), (2,63; 3,32),

(2,66; 3,23), (4,13; 4,36), (3,75; 4,56), (3,91; 6,26), (3,63; 7,36), (3,66; 3,20), (2,61; 0,00), (2,86; 8,28), (3,19; 4,29), (2,85; 2,45), (1,89; 3,91).

**14.**

(13,0; 5,0), (7,0; 5,0), (7,0; 4,0), (9,0; 7,0), (14,0; 12,0), (5,0; 13,0), (6,0; 3,0), (6,0; 4,5), (6,5; 11,0), (5,5; 3,0), (5,5; 5,0), (7,0; 20,0), (6,0; 6,0), (2,5; 6,0), (6,8; 6,0), (8,0; 10,0), (4,0; 8,0), (2,0; 7,0), (7,0; 1,0), (7,0; 9,0), (5,0; 5,0), (3,0; 13,5), (7,0; 11,0), (20,0; 9,0), (3,0; 5,0), (6,0; 10,0), (5,0; 7,8), (7,0; 6,0), (2,0; 6,0), (5,5; 7,0), (5,0; 7,0), (5,0; 7,0), (2,5; 5,0), (4,8; 6,0), (7,0; 6,5), (4,0; 4,5), (2,0; 4,0), ( 3,0; 7,0), (6,0; 5,0), (4,0; 5,0).

**15.**

(10,0; 3,0), (4,0; 3,0), (19,0; 1,0), (5,0; 10,0), (8,0; 2,5), (3,0; 5,0), (3,0; 2,8), (3,0; 1,0), (10,5; 3,0), (5,0; 6,0), (5,5; 6,0), (3,8; 3,0), (4,0; 6,0), (16,5; 5,0), (16,0; 20,0), (14,0; 4,5), (4,0; 3,0), (6,0; 4,0), (3,0; 4,0), (3,5; 2,0), (5,5; 8,0), (7,0; 4,0), (15,0; 5,0), (5,0; 5,0), (4,0; 3,5), (8,0; 7,5), (5,0; 8,0), (15,5; 3,5), (2,0; 4,0), (5,0; 3,5), (3,5; 4,5), (22,8; 6,8), (4,5; 4,0), (4,0; 4,0), (7,5; 4,0), (3,0; 3,8), (4,3; 3,0), (6,0; 3,0), (5,0; 23,0), (8,8; 5,0).

**Задание 2.** Имеются статистические данные. Необходимо

- рассчитать параметры линейного уравнения множественной рег­рессии;

- дать сравнительную оценку силы связи факторов с результатом с помощью коэффициентов эластичности;

- оценить статистическую значимость параметров регрессионной модели с помощью t-критерия;

- нулевую гипотезу о значимости уравнения и показателей тесноты связи проверить с помощью F-критерия;

- оценить качество уравнения через среднюю ошибку аппроксимации;

- проверить модель на мультиколлениарность.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 |  |  |  |  | 2 |  |  |  |
| 1 | 0,5 | 24,1 | 28,0 |  | 1 | 0,6 | 26,9 | 32,2 |
| 2 | 1,3 | 8,9 | 47,4 |  | 2 | 1,4 | 9,5 | 51,2 |
| 3 | 0,4 | 2,1 | 16,5 |  | 3 | 0,5 | 2,6 | 18,1 |
| 4 | 1,2 | 7,0 | 32,7 |  | 4 | 1,3 | 8,0 | 37,3 |
| 5 | 2,0 | 14,3 | 71,9 |  | 5 | 2,2 | 16,7 | 80,9 |
| 6 | 0,9 | 10,7 | 62,7 |  | 6 | 1,2 | 11,6 | 73,5 |
| 7 | 3,3 | 87,3 | 285,0 |  | 7 | 3,7 | 98,2 | 298,4 |
| 8 | 1,2 | 13,3 | 49,0 |  | 8 | 1,3 | 14,8 | 55,6 |
| 9 | 5,5 | 110,7 | 425,7 |  | 9 | 5,9 | 124,7 | 533,6 |
| 10 | 0,2 | 1,5 | 2,3 |  | 10 | 0,2 | 1,7 | 2,8 |
| 11 | 0,9 | 4,3 | 18,8 |  | 11 | 1,1 | 4,7 | 19,4 |
| 12 | 1,3 | 21,6 | 31,5 |  | 12 | 1,5 | 23,1 | 33,2 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 |  |  |  |  | 4 |  |  |  |
| 1 | 0,7 | 27,4 | 33,8 |  | 1 | 0,9 | 28,2 | 35,4 |
| 2 | 1,5 | 10,2 | 55,6 |  | 2 | 1,6 | 10,8 | 57,9 |
| 3 | 0,6 | 3,2 | 19,8 |  | 3 | 0,6 | 3,7 | 21,9 |
| 4 | 1,4 | 8,8 | 41,0 |  | 4 | 1,5 | 9,2 | 43,1 |
| 5 | 2,3 | 17,2 | 91,0 |  | 5 | 2,4 | 18,3 | 97,1 |
| 6 | 1,2 | 12,4 | 78,8 |  | 6 | 1,2 | 13,0 | 83,4 |
| 7 | 3,6 | 105,7 | 312,3 |  | 7 | 3,8 | 115,2 | 323,8 |
| 8 | 1,4 | 15,3 | 62,8 |  | 8 | 1,5 | 16,2 | 69,9 |
| 9 | 6,3 | 133,2 | 594,0 |  | 9 | 6,5 | 141,4 | 635,8 |
| 10 | 0,3 | 1,8 | 3,0 |  | 10 | 0,3 | 1,9 | 3,4 |
| 11 | 1,1 | 5,2 | 22,1 |  | 11 | 1,2 | 5,7 | 23,8 |
| 12 | 1,7 | 24,2 | 35,0 |  | 12 | 1,8 | 25,5 | 37,2 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 |  |  |  |  | 6 |  |  |  |
| 1 | 0,9 | 28,4 | 37,2 |  | 1 | 0,9 | 31,3 | 43,0 |
| 2 | 1,6 | 11,7 | 61,2 |  | 2 | 1,7 | 13,4 | 64,7 |
| 3 | 0,6 | 3,9 | 23,2 |  | 3 | 0,7 | 4,5 | 24,0 |
| 4 | 1,6 | 9,7 | 47,2 |  | 4 | 1,7 | 10,0 | 50,2 |
| 5 | 2,5 | 19,3 | 102,1 |  | 5 | 2,6 | 20,0 | 106,0 |
| 6 | 1,2 | 14,2 | 96,9 |  | 6 | 1,3 | 15,0 | 96,6 |
| 7 | 4,0 | 121,6 | 334,5 |  | 7 | 4,1 | 137,1 | 347,0 |
| 8 | 1,5 | 17,3 | 75,6 |  | 8 | 1,6 | 17,9 | 85,6 |
| 9 | 6,8 | 152,4 | 721,0 |  | 9 | 6,9 | 165,4 | 745,0 |
| 10 | 0,4 | 1,9 | 3,6 |  | 10 | 0,4 | 2,0 | 4,1 |
| 11 | 1,2 | 6,3 | 24,8 |  | 11 | 1,3 | 6,8 | 26,8 |
| 12 | 1,8 | 26,4 | 40,1 |  | 12 | 1,9 | 27,1 | 42,7 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 7 |  |  |  |  | 8 |  |  |  |
| 1 | 1,0 | 33,4 | 45,2 |  | 1 | 1,2 | 33,8 | 45,6 |
| 2 | 1,8 | 13,9 | 68,7 |  | 2 | 1,8 | 14,1 | 69,5 |
| 3 | 0,8 | 4,8 | 23,0 |  | 3 | 0,9 | 4,9 | 25,1 |
| 4 | 1,9 | 12,5 | 50,8 |  | 4 | 2,0 | 12,9 | 62,1 |
| 5 | 2,8 | 22,1 | 112,7 |  | 5 | 2,9 | 23,0 | 121,3 |
| 6 | 1,5 | 16,7 | 94,5 |  | 6 | 1,5 | 16,8 | 98,9 |
| 7 | 4,5 | 138,1 | 351,6 |  | 7 | 4,7 | 139,0 | 366,1 |
| 8 | 1,9 | 18,4 | 88,6 |  | 8 | 2,0 | 20,1 | 91,7 |
| 9 | 7,4 | 166,9 | 721,7 |  | 9 | 7,6 | 171,2 | 747,2 |
| 10 | 0,5 | 2,2 | 4,0 |  | 10 | 0,8 | 2,7 | 5,0 |
| 11 | 1,2 | 7,4 | 27,3 |  | 11 | 1,3 | 7,3 | 27,7 |
| 12 | 2,4 | 28,9 | 48,7 |  | 12 | 2,5 | 29,3 | 49,4 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 9 |  |  |  |  | 10 |  |  |  |
| 1 | 1,3 | 34,1 | 46,1 |  | 1 | 1,7 | 35,2 | 45,1 |
| 2 | 1,9 | 14,4 | 72,5 |  | 2 | 2,5 | 15,4 | 71,0 |
| 3 | 1,1 | 5,2 | 26,0 |  | 3 | 1,2 | 5,5 | 26,5 |
| 4 | 2,1 | 13,5 | 63,0 |  | 4 | 2,0 | 13,4 | 52,1 |
| 5 | 3,0 | 23,1 | 122,5 |  | 5 | 3,2 | 24,5 | 115,8 |
| 6 | 1,8 | 17,1 | 100,2 |  | 6 | 1,9 | 17,3 | 101,8 |
| 7 | 4,8 | 142,5 | 350,1 |  | 7 | 5,5 | 151,2 | 350,3 |
| 8 | 2,1 | 22,1 | 92,4 |  | 8 | 2,5 | 24,1 | 94,8 |
| 9 | 7,7 | 173,1 | 755,8 |  | 9 | 7,9 | 174,1 | 757,2 |
| 10 | 0,8 | 2,8 | 5,1 |  | 10 | 1,1 | 3,2 | 5,6 |
| 11 | 1,5 | 7,4 | 27,9 |  | 11 | 2,0 | 9,1 | 28,1 |
| 12 | 2,6 | 29,8 | 50,8 |  | 12 | 2,7 | 30,6 | 51,4 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 |  |  |  |  | 12 |  |  |  |
| 1 | 0,5 | 24,1 | 28,0 |  | 1 | 0,6 | 26,9 | 32,2 |
| 2 | 1,3 | 8,9 | 47,4 |  | 2 | 1,4 | 9,5 | 51,2 |
| 3 | 0,4 | 2,1 | 16,5 |  | 3 | 0,5 | 2,6 | 18,1 |
| 4 | 1,2 | 7,0 | 32,7 |  | 4 | 1,3 | 8,0 | 37,3 |
| 5 | 2,0 | 14,3 | 71,9 |  | 5 | 2,2 | 16,7 | 80,9 |
| 6 | 0,9 | 10,7 | 62,7 |  | 6 | 1,2 | 11,6 | 73,5 |
| 7 | 3,3 | 87,3 | 285,0 |  | 7 | 3,7 | 98,2 | 298,4 |
| 8 | 1,2 | 13,3 | 49,0 |  | 8 | 1,3 | 14,8 | 55,6 |
| 9 | 5,5 | 110,7 | 425,7 |  | 9 | 5,9 | 124,7 | 533,6 |
| 10 | 0,2 | 1,5 | 2,3 |  | 10 | 0,2 | 1,7 | 2,8 |
| 11 | 0,9 | 4,3 | 18,8 |  | 11 | 1,1 | 4,7 | 19,4 |
| 12 | 1,3 | 21,6 | 31,5 |  | 12 | 1,5 | 23,1 | 33,2 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 13 |  |  |  |  | 14 |  |  |  |
| 1 | 1,3 | 34,1 | 46,1 |  | 1 | 1,7 | 35,2 | 45,1 |
| 2 | 1,9 | 14,4 | 72,5 |  | 2 | 2,5 | 15,4 | 71,0 |
| 3 | 1,1 | 5,2 | 26,0 |  | 3 | 1,2 | 5,5 | 26,5 |
| 4 | 2,1 | 13,5 | 63,0 |  | 4 | 2,0 | 13,4 | 52,1 |
| 5 | 3,0 | 23,1 | 122,5 |  | 5 | 3,2 | 24,5 | 115,8 |
| 6 | 1,8 | 17,1 | 100,2 |  | 6 | 1,9 | 17,3 | 101,8 |
| 7 | 4,8 | 142,5 | 350,1 |  | 7 | 5,5 | 151,2 | 350,3 |
| 8 | 2,1 | 22,1 | 92,4 |  | 8 | 2,5 | 24,1 | 94,8 |
| 9 | 7,7 | 173,1 | 755,8 |  | 9 | 7,9 | 174,1 | 757,2 |
| 10 | 0,8 | 2,8 | 5,1 |  | 10 | 1,1 | 3,2 | 5,6 |
| 11 | 1,5 | 7,4 | 27,9 |  | 11 | 2,0 | 9,1 | 28,1 |
| 12 | 2,6 | 29,8 | 50,8 |  | 12 | 2,7 | 30,6 | 51,4 |

|  |  |  |  |
| --- | --- | --- | --- |
| 15 |  |  |  |
| 1 | 0,9 | 28,4 | 37,2 |
| 2 | 1,6 | 11,7 | 61,2 |
| 3 | 0,6 | 3,9 | 23,2 |
| 4 | 1,6 | 9,7 | 47,2 |
| 5 | 2,5 | 19,3 | 102,1 |
| 6 | 1,2 | 14,2 | 96,9 |
| 7 | 4,0 | 121,6 | 334,5 |
| 8 | 1,5 | 17,3 | 75,6 |
| 9 | 6,8 | 152,4 | 721,0 |
| 10 | 0,4 | 1,9 | 3,6 |
| 11 | 1,2 | 6,3 | 24,8 |
| 12 | 1,8 | 26,4 | 40,1 |

**Глава 5. Теория систем массового обслуживания**

Теория систем массового обслуживания (СМО) начала развиваться в начале XX столетия. В 1909 г. шведский математик Эрланг применил теорию вероятностей к исследованию зависимости обслуживания телефонных вызовов от числа поступающих на телефонную станцию вызовов. В нашей стране известный математик А.Я. Хинчин систематизировал основные положения СМО в монографии «Теория очередей». Именно такое название теории СМО используется за рубежом.

В последние годы применение теории СМО в экономике приобрело особую актуальность в связи с использованием ряда ее аспектов в финансово-экономической сфере (банки различных типов, страховые организации, налоговые инспекции, аудиторские службы). Теория СМО широко применяется также в сфере обслуживания (различные системы связи, АЗС, магазины, ремонтные предприятия и т.д.) и в современных высоких технологиях (компьютерные сети, базы данных, военные системы ПВО).

**5.1. Структура и классификация**

**систем массового обслуживания**

СМО представляют собой системы специфического вида. *Системой* называется целостное множество взаимосвязанных элементов, которые нельзя разделить на независимые подмножества.

Основными элементами СМО являются: входной поток заявок; очередь; каналы обслуживания (приборы, операторы, продавцы и пр.); выходной поток заявок (обслуженные заявки). Классификация систем массового обслуживания приведена в таблице 5.1.

Таблица 5.1.

**Классификация СМО**

|  |  |
| --- | --- |
| **По числу каналов** | |
| ***одноканальные*** | ***многоканальные*** |
| **По дисциплине обслуживания** | |
| ***с отказами*** (заявка получает отказ при условии занятости каналов, например, вызовов абонента через АТС) | ***с ожиданием*** (***очередью***) (в случае занятости системы заявка поступает в очередь, например, обслуживание покупателей в магазине) |

***Показатели эффективности СМО*** описывают ее возможность справляться с потоком заявок.

|  |  |
| --- | --- |
| **Показатели эффективности** | |
| ***СМО с отказами*** | ***СМО с очередью*** |
| *A -* абсолютная пропускная способность СМО (среднее число заявок, обслуживаемых в единицу времени);  *Q –* относительная пропускная способность (средняя доля пришедших заявок, обслуживаемых СМО);  *Pserv* -вероятность обслуживания (вероятность того, что заявка будет принята на обслуживание);  *Potk* -вероятность отказа (вероятность того, что заявка покинет СМО необслуженной);  - среднее число занятых каналов (для многоканальной системы). | К показателям эффективности СМО с отказами добавляются:  *LСМО* – среднее число заявок в системе;  *Lоч* - среднее число заявок в очереди;  - среднее число заявок, находящихся по обслуживанием;  - среднее время пребывания заявки в очереди;  - среднее время пребывания заявки в СМО ;  *Рзан* -вероятность того, что канал занят(степень загрузки канала). |

**5.2. Марковский случайный процесс в СМО. Уравнения Колмогорова**

Процессы поступления и обслуживания заявок в СМО являются случайными, что обусловлено случайным характером потока заявок и длительности их обслуживания.

***Случайным процессом*** *X(t)* называется процесс, значение которого при любом значении аргумента *t* является случайной величиной. При фиксированном *t=t0 X(t0)* представляет собой обычную величину. Случайные процессы упрощают исследование СМО.

В основе СМО находится ***марковский случайный процесс*** (процесс без последствия), когда вероятность состояния СМО в будущем зависит только от ее состояния в настоящем и не зависит от прошлого (название по имени известного российского математика А.А. Маркова). Условие марковского случайного процесса: необходимо, чтобы все потоки событий, при которых система переходит из одного состояния в другое (потоки заявок, потоки обслуживания и т.д.) были пуассоновскими.

Пуассоновский поток (простейший поток) событий обладает следующими свойствами:

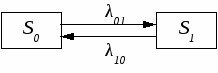
- отсутствие последствия (число событий, попавших на заданный временной интервал, не зависит от числа событий, попавших на другие интервалы);

- ординарности (вероятность попадания на элементарный временной интервал двух и более событий пренебрежимо мала по сравнению с вероятностью попадания одного события);

- стационарности (число событий, попавших на заданный временной интервал, зависит лишь от длины интервала и не зависит от числа событий, попавших на другие интервалы).

Для простейшего потока справедлив закон Пуассона. Плотность вероятности случайной величины при этом , где *λ* – интенсивность потока.

При анализе случайных процессов с дискретными состояниями пользуются *графом состояний*, где прямоугольниками изображают состояния системы, а переходы из состояния в состояние – стрелками. Если у стрелок проставлены интенсивности, то граф состояния называется *размеченным*. Переходы системы из состояния *Si* в состояние *Sj* происходят под воздействием простейших потоков событий с интенсивностями *λij.* Простейший граф состояний представлен на рисунке

5.1.

*Рис. 5.1. Граф состояний*

На рис. 5.1 изображена СМО, состоящая из одного канала обслуживания, который находится в двух возможных состояниях: либо свободен *(S0*), либо занят *(S1)*; *λ01* – интенсивность поступления заявок; *λ10* – интенсивность обслуживания заявок в единицу времени. Стрелка из *S0* в *S1* означает переход системы из состояния «канал свободен» в состояние «канал занят». Стрелка из *S1* в *S0* означает обратный переход.

Анализ состояния СМО сводится к определению вероятности, с которой система пребывает в данном состоянии. В общем случае ***вероятностью i-го состояния pi(t)*** называется вероятность того, что в момент *t* система будет находиться в состоянии *Si*. Для любого момента *t* справедливо соотношение

Определить вероятности состояний СМО можно, решив систему уравнений Колмогорова.

***Правило составления системы уравнений Колмогорова***

1. Слева в уравнениях стоит предельная вероятность данного состояния умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния.

2. Справа – сумма произведений интенсивностей всех потоков, входящих в *i*-е состояние, на вероятности тех состояний, из которых эти потоки исходят.

3. Для решения системы вводится нормировочное уравнение

Для достаточно большого значения времени *t* распределение вероятностей стабилизируется и практически не зависит от времени.

**5.3. Расчет показателей эффективности СМО с отказами**

**5.3.1. Одноканальные СМО с отказами**

Система *S* может находиться в одном из двух состояний: *S0* – канал свободен или *S1* – канал занят. Из состояния *S0* в состояние *S1* систему переводит поток входящих заявок, а из состояния *S1*в состояние *S0* – поток обслуживаний. ***Плотности вероятностей перехода*** из состояния *S0* в состояние *S1* и обратно равны соответственно *λ* и *μ*. Граф состояний СМО показан на рисунке

http://ru.convdocs.org/pars_docs/refs/105/104994/104994_html_m1ee68433.gif

*Рис. 5.2. Граф состояний*

*одноканальной СМО с отказами*

Показатели эффективности одноканальной СМО с отказами вычисляются по формулам, которые приведены в таблице 5.2.

Таблица 5.2.

**Показатели эффективности**

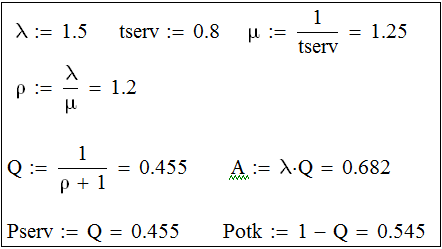
**одноканальной СМО с отказами**

|  |  |  |
| --- | --- | --- |
| **Показатели** | **Обозначение** | **Формула для вычисления** |
| Интенсивность  входящего потока  заявок | *λ* | известно из условия задачи |
| Интенсивность  выходящего потока обслуженных заявок | *μ* | известно из условия задачи; |
| Среднее время обслуживания заявки |  | известно из условия задачи; |
| Приведенная интенсивность  потока заявок | ρ |  |
| Относительная пропускная  способность СМО | Q |  |
| Абсолютная пропускная  способность СМО | А |  |
| Вероятность того, что заявка будет обслужена |  |  |
| Вероятность отказа |  |  |
| Среднее время простоя канала |  |  |
| Среднее время пребывания  заявки в системе |  |  |

**Пример 5.1.** Телефонная АТС имеет одну линию, на которую в среднем приходит 0,8 вызовов в минуту. Среднее время разговора 1,5 мин. Вызов, пришедший во время разговора, не обслуживается. Считая потоки вызовов пуассоновскими, найти абсолютную и относительную пропускную способность станции, и вероятность отказа абоненту.

**Решение**

Телефонную станцию рассматриваем как одноканальную СМО с отказами. Решение приведено на рисунке 5.3.



*Рис. 5.3. Показатели эффективности*

*одноканальной СМО с отказами*

Абсолютная пропускная способность СМО *А* оказалась почти вдвое меньше интенсивности μ потока обслуживания, что обусловлено случайным характером потока заявок.

**5.3.2. Многоканальная СМО с отказами**

Система *S* имеет следующие состояния (нумеруем по числу заявок, находящихся в системе): *S0*, *S1*,…, *Sn*, где *Sк* – состояние системы, когда в ней находится *к* заявок, т.е. занято *к* каналов.

Граф состояний СМО показан на рисунке 3.4.

S1

S2

Sk

Sn

S0

…

…

…

…

kµ

3µ

2µ

µ

λ

λ

λ

λ

λ

λ

…

(k+1)µ

nµ

*Рис. 5.4. Граф состояний*

*многоканальной СМО с отказами*

Показатели эффективности многоканальной СМО с отказами вычисляются по формулам, которые приведены в таблице 5.3.

Таблица 5.3.

**Показатели эффективности**

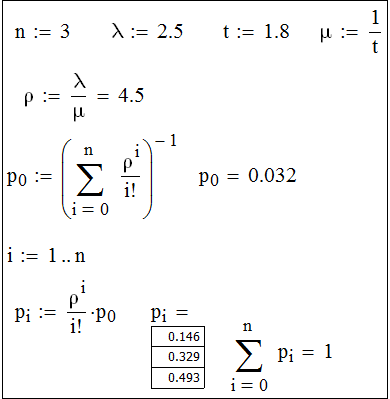
**многоканальной СМО с отказами**

|  |  |  |
| --- | --- | --- |
| **Показатели** | **Обозначение** | **Формула для вычисления** |
| Число каналов обслуживания | *n (n>1)* | известно из условия задачи |
| Интенсивность входящего  потока заявок | *λ* | известно из условия задачи |
| Интенсивность потока обслуженных заявок, выходящих *из одного канала* | *μ* | известно из условия задачи; |
| Приведенная интенсивность потока заявок | ρ |  |
| Вероятность того, что занято  0, 1, …, *п* каналов, соответственно (формулы Эрланга) |  |  |
| Вероятность отказа |  |  |
| Относительная пропускная способность СМО | Q |  |
| Абсолютная пропускная способность СМО | А |  |
| Вероятность того, что заявка будет обслужена |  |  |
| Среднее число занятых каналов |  |  |

**Пример 5.2.** Имеетсятри поста для мойки автомобилей. Автомобиль, прибывший в момент, когда все посты заняты, – получает отказ в обслуживании. Интенсивность потока автомобилей *λ* = 2,5 (автомобиля в час). Средняя продолжительность обслуживания 1,8 часа. Поток автомобилей и поток обслуживания являются простейшими. Требуется определить в установившемся режиме предельные значения вероятностей, показатели эффективности.

**Решение**

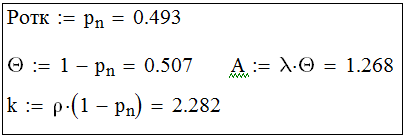
Вводим исходные данные.Вычисляем приведенную интенсивность *ρ*. По формулам Эрланга вычисляем предельные вероятности состояний (рис.5.5).

****

*Рис. 5.5.Вычисление предельных*

*вероятностей состояний*

Вычислим вероятность отказа, относительную Q и абсолютную A пропускные способности, а также среднее число занятых каналов (рис. 5.6).

****

*Рис. 5.6*. *Вычисление*

*показателей эффективности*

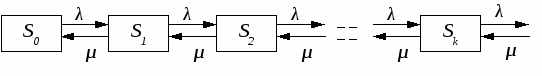
**5.4. Расчет показателей эффективности СМО**

**с ожиданием (очередью)**

**5.4.1. Одноканальная СМО с неограниченной очередью**

Имеется одноканальная СМО с очередью, на которую не наложены никакие ограничения (ни по длине очереди, ни по времени ожидания). Система *S* может находится в одном из состояний (по числу заявок, находящихся в СМО): *S0* – канал свободен; *S1* – канал занят (обслуживает одну заявку), очереди нет; *S2* – канал занят, одна заявка в очереди; …; *Sк* – канал занят, (*к*-1) заявок стоят в очереди и т.д.

Граф состояний СМО представлен на рисунке 5.7.

****

*Рис. 5.7. Граф состояний одноканальной СМО*

*с неограниченной очередью*

Показатели эффективности одноканальной СМО с очередью вычисляются по формулам, которые приведены в таблице 5.4.

Таблица 5.4.

**Показатели эффективности**

**одноканальной СМО с очередью**

|  |  |  |
| --- | --- | --- |
| **Показатели** | **Обозначение** | **Формула для**  **вычисления** |
| Длина очереди | ∞ |  |
| Интенсивность входящего  потока заявок | *λ* | известно из условия задачи |
| Интенсивность потока  обслуженных заявок,  выходящих *из одного канала* | *μ* | известно из условия задачи; |
| Приведенная интенсивность потока заявок | ρ<1 |  |
| Вероятность того, что СМО свободна и может обслужить заявку |  |  |
| Вероятность того, что СМО  занята, в очереди нет заявок |  |  |
| Вероятность того, что СМО  занята, в очереди находятся 1,…, *п,…* заявок,  соответственно |  | 2,... |
| Вероятность того, что заявка получит отказ |  |  |
| Относительная пропускная способность СМО | Q |  |
| Абсолютная пропускная  способность СМО | А |  |
| Вероятность того, что заявка будет обслужена |  |  |
| Среднее число заявок, стоящих в очереди |  |  |
| Среднее число заявок, находящихся под обслуживанием |  |  |
| Среднее число заявок в СМО (обслуживаемых и стоящих в очереди) |  |  |
| Среднее время ожидания  заявки в очереди |  |  |
| Среднее время пребывания  заявки в СМО |  |  |

***Замечание****. Прежде чем записать формулы предельных вероятностей, необходимо быть уверенным в их существовании, ведь в случае, когда время очередь может неограниченно возрастать. Если , т.е. среднее число приходящих заявок меньше среднего числа обслуженных заявок (в единицу времени), то предельные вероятности существуют. Если очередь растет до бесконечности.*

**5.4.2. Многоканальная СМО с неограниченной очередью**

Имеется *п-*канальная СМО с неограниченной очередью. Система *S* может находиться в одном из состояний (по числу заявок, находящихся в СМО): *S0* – в системе нет заявок (все каналы свободны); *S1* – занят один канал, остальные свободны; *S2* –заняты два канала, остальные свободны; …; *Sп* – заняты все *п* каналов (очереди нет); *Sп+1* - заняты все *п* каналов, в очереди одна заявка;…; *Sп+r* - заняты все *п* каналов, в очереди *r* заявок;… Граф состояний СМО представлен на рисунке 5.8.

λ

λ

Sn+1

λ

Sn

kµ

…

Sk

…

Sk

λ

λ

…

S0

S1

S0

S1

2µ

µ

λ

λ

λ

…

…

…

…

(k+1)µ

…

λ

λ

Sn+r

…

…

…

…

nµ

nµ

nµ

nµ

nµ

*Рис. 5.8. Граф состояний*

*многоканальной СМО с неограниченной очередью*

Показатели эффективности многоканальной СМО с неограниченной очередью вычисляются по формулам, которые приведены в таблице 5.5.

Таблица 5.5.

**Показатели эффективности**

**многоканальной СМО с неограниченной очередью**

|  |  |  |
| --- | --- | --- |
| **Показатели** | **Обозначение** | **Формула для вычисления** |
| Число каналов обслуживания | *n (n>1)* | известно из условия задачи |
| Длина очереди | ∞ |  |
| Интенсивность входящего потока заявок | *λ* | известно из условия задачи |
| Интенсивность потока обслуженных заявок, выходящих *из одного канала* | *μ* | известно из условия задачи; |
| Приведенная интенсивность потока заявок | ρ |  |
| Предельные вероятности состояний системы (очереди нет) |  |  |
| Предельные вероятности состояний системы (наличие в очереди 1, 2,… заявок) |  | ,…,  ,…, |
| Вероятность того, что заявка окажется в очереди |  |  |
| Вероятность того, что заявка получит отказ |  |  |
| Относительная пропускная способность СМО | Q |  |
| Абсолютная пропускная способность СМО | А |  |
| Вероятность того, что заявка будет обслужена |  |  |
| Среднее число занятых каналов |  |  |
| Среднее число заявок, стоящих в очереди |  |  |
| Среднее число заявок в СМО (обслуживаемых и стоящих в очереди) |  |  |
| Среднее время ожидания заявки в очереди |  |  |
| Среднее время пребывания заявки в СМО |  |  |

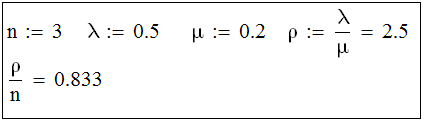
***Замечание.*** *При* *предельные вероятности существуют. Если , очередь растет до бесконечности.*

**Пример 5.3.** Система массового обслуживания с ожиданием – билетная касса с тремя окошками (с тремя кассирами). Пассажиров, желающих купить билет, приходит в среднем 5 человек за 10 мин. Поток пассажиров можно считать простейшим. Кассир в среднем обслуживает двух пассажиров за 10 минут. Определите вероятностные

характеристики СМО в стационарном режиме.

**Решение**

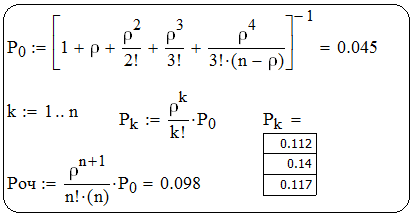
Введем начальные значения: *n* – число каналов обслуживания, λ– интенсивность потока заявок, µ– интенсивность обслуживания. Вычисляем приведенную интенсивность ρ. Проверяем условие существования предельных вероятностей (рис. 5.9).



*Рис. 5.9. Условие существования*

*предельных вероятностей*

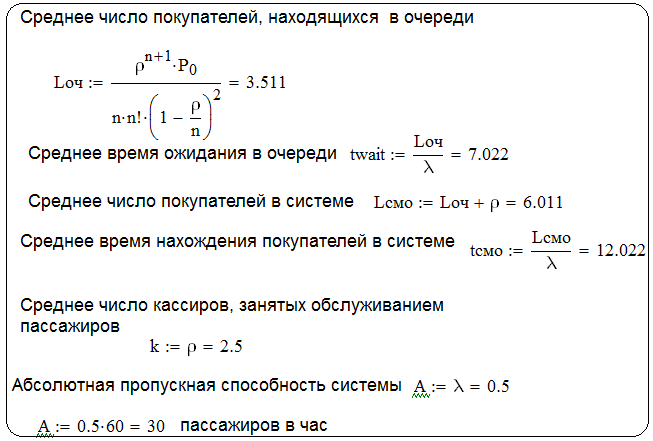
Так как , то предельные вероятности существуют. Как видно из вычислений (рисунок 5.10), в среднем 4,5% времени кассиры будут простаивать.



*Рис. 5.10. Вычисление предельных*

*вероятностей системы*

Найдем характеристики обслуживания СМО (рисунок 5.11).



*Рис. 5.11. Характеристики*

*обслуживания СМО*

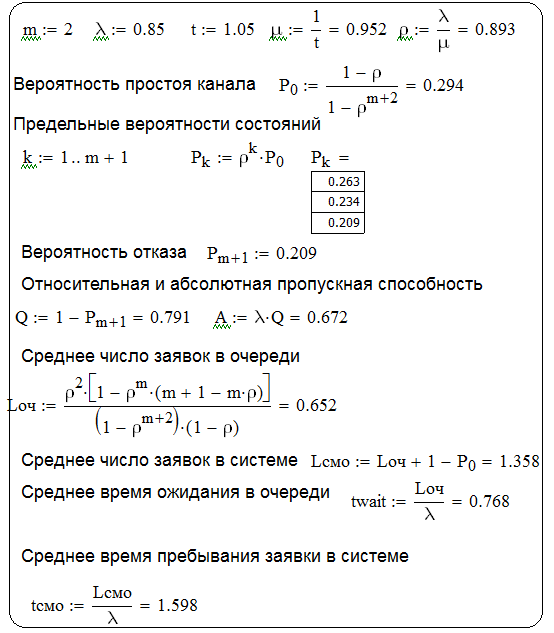
**5.4.3. СМО с ограниченной очередью**

СМО с ограниченной очередью отличаются от рассмотренных выше задач лишь тем, что число заявок в очереди ограничено (не может превосходить некоторого заданного числа *т*). Если новая заявка поступает в момент, когда все места в очереди заняты, она покидает СМО необслуженной. Формулы сведены в таблицу 5.6.

**Пример 5.4.** На вход одноканальной СМО с длиной очереди *m*= 2

поступает поток заявок с интенсивностью 0,85 заявок в час. Среднее время обслуживания одной заявки 1,05 часа. Найти основные характеристики данной СМО.

**Решение**

****

*Рис. 5.12. Расчет*

*показателей эффективности*

Таблица 5.6.

|  |  |  |
| --- | --- | --- |
| **Показатели** | **Одноканальная СМО**  **с ограниченной очередью** | **Многоканальная СМО**  **с ограниченной очередью** |
| Число каналов  Длина очереди  Приведенная интенсивность | 1  *m* | *п*  *m* |
| Предельные  вероятности |  |  |
| Вероятность  отказа |  |  |
| Относительная  пропускная  способность |  |  |
| Абсолютная  пропускная  способность |  |  |
| Среднее число  заявок, стоящих  в очереди |  |  |
| Среднее число  занятых каналов |  |  |
| Среднее число  заявок в СМО  (обслуживаемых  и стоящих в очереди) |  |  |
| Среднее время  ожидания заявки  в очереди |  |  |
| Среднее время  пребывания  заявки в СМО |  |  |

**Задачи для аудиторного решения**

1. В отделении сберегательного банка кассир обслуживает клиентов с интенсивностью 0,5 чел./мин. Среднее число клиентов, находящихся на обслуживании, равно 0,7. Предполагается, что нет ограничений на длину очереди. Определить показатели эффективности СМО и вероятность того, что ожидают своей очереди не более одного человека.

2. Рассматривается круглосуточная работа пункта проведения профилактического осмотра автомашин с одним каналом (одна группа проведения осмотра). На осмотр и выявление дефектов каждой машины затрачивается в среднем 0,5 часа. На осмотр поступает в среднем 36 машин в сутки. Если машина, прибывшая в пункт осмотра, не застает ни одного канала свободным, она покидает пункт осмотра необслуженной. Определить вероятности состояний и характеристики обслуживания пункта.

3. Решить задачу 2 для случая *п*=4.

4. Решить задачу 2 при условии, что машина, прибывшая на пункт осмотра, покидает этот пункт лишь в случае, если в очереди на осмотр стоят более 5 машин.

5. Решить задачу 3 при условии, что машина, прибывшая на пункт осмотра, покидает этот пункт лишь в случае, если в очереди на осмотр стоят более 5 машин.

**]**

**Компьютерная лабораторная работа № 6**

***Расчет показателей эффективности***

***систем массового обслуживания***

**Задание 1.** СМО с отказами представляет собой *n* диспетчеров телефонной станции. Заявка, пришедшая в момент, когда все диспетчеры заняты, получает отказ в обслуживании. Интенсивность потока заявок *λ*. Средняя продолжительность обслуживания . Поток заявок и поток обслуживания являются простейшими. Требуется определить в установившемся режиме предельные значения вероятностей, показатели эффективности.

|  |  |  |  |
| --- | --- | --- | --- |
| Вариант | *n* | *λ* |  |
| 1 | 4 | 5 | 0,9 |
| 2 | 3 | 5 | 0,7 |
| 3 | 4 | 6 | 0,8 |
| 4 | 3 | 4 | 0,9 |
| 5 | 5 | 6 | 0,6 |
| 6 | 4 | 7 | 0,7 |
| 7 | 3 | 4 | 0,8 |
| 8 | 5 | 5 | 0,9 |
| 9 | 4 | 6 | 0,8 |
| 10 | 3 | 5 | 0,7 |
| 11 | 3 | 5 | 0,7 |
| 12 | 5 | 6 | 0,6 |
| 13 | 4 | 7 | 0,7 |
| 14 | 4 | 5 | 0,9 |
| 15 | 5 | 5 | 0,9 |

**Задание 2.** СМО с ожиданием – ремонтная мастерская на *n* рабочих мест. Интенсивность потока заявок *λ*. Интенсивность обслуживания . Определите вероятностные характеристики СМО в стационарном режиме и показатели эффективности**.**

|  |  |  |  |
| --- | --- | --- | --- |
| Вариант | *n* | *λ* |  |
| 1 | 4 | 6,4 | 2 |
| 2 | 4 | 7,2 | 2 |
| 3 | 3 | 5,4 | 2 |
| 4 | 3 | 3,8 | 1,6 |
| 5 | 3 | 4,2 | 1,8 |
| 6 | 4 | 7 | 2 |
| 7 | 3 | 2,8 | 2 |
| 8 | 4 | 7,6 | 2 |
| 9 | 3 | 3,8 | 2 |
| 10 | 3 | 4,2 | 2 |
| 11 | 4 | 6,4 | 2 |
| 12 | 4 | 7,2 | 2 |
| 13 | 3 | 5,4 | 2 |
| 14 | 3 | 3,8 | 1,6 |
| 15 | 3 | 4,2 | 1,8 |

**Задание 3. Решите задачи**

**3.1.** Интенсивность потока телефонных звонков в агентство по заказу железнодорожных билетов, имеющему один телефон, составляет 16 вызовов в час. Продолжительность оформления заказа на билет равна 2,4 минуты. Определить относительную и абсолютную пропускную способность этой СМО и вероятность отказа (занятости телефона). Сколько телефонов должно быть в агентстве, чтобы относительная пропускная способность была не менее 0,75.

**3.2.** Система массового обслуживания — билетная касса с одним окошком и неограниченной очередью. Касса продает билеты в пункты А и В. Пассажиров, желающих купить билет в пункт А, приходит в среднем трое за 20 мин, в пункт В — двое за 20 мин. Поток пассажиров простейший. Кассир в среднем обслуживает трех пассажиров за 10 мин. Время обслуживания — показательное. Вычислить финальные вероятности, среднее число заявок в системе и в очереди, среднее время пребывания заявки в системе, среднее время пребывания заявки в очереди.

**3.3.** Междугородный переговорный пункт имеет четыре телефонных аппарата. В среднем за сутки поступает 320 заявок на переговоры. Средняя длительность переговоров составляет 5 мин. Длина очереди не должна превышать 6 абонентов. Потоки заявок и обслуживаний простейшие. Определить характеристики обслуживания переговорного пункта в стационарном режиме (вероятность простоя каналов, вероятность отказа, вероятность обслуживания, среднее число занятых каналов, среднее число заявок в очереди, среднее число заявок в системе, абсолютную пропускную способность, относительную пропускную способность, среднее время заявки в очереди, среднее время заявки в системе, среднее время заявки под обслуживанием).

**Список использованной литературы**

1. Гнеденко Б.В. Курс теории вероятности. – М.: ГИТТЛ, 1954.

2. Ивановский Р.И. Теория вероятностей и математическая статистика. Основы, прикладные аспекты с примерами и задачами. – СПб.: БХВ-Петербург, 2008. – 528 с.

3. Крамер Г. Математические методы статистики. / Пер. с англ. Под ред. А.Н. Колмогорова. – М.: Наука, 1993. – 900 с.

4. Климов Г.П. Стохастические системы обслуживания. – М.: Наука, 1966. – 254 с.

5. Мухина С.Н. Компьютерная математика на базе MathCAD: Учебное пособие. - Калининград: Изд-во БГАРФ, 2014. – 138 с.

6. Черняк А.А., Новиков В.А., Мельников О.И., Кузнецов А.В. Математика для экономистов на базе MathCAD: Учебное пособие. – СПб.: БХВ-Петербург, 2003. – 496 с.

Приложение 1

**Типы числовых характеристик**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **СВ *X*** | **Характеристики положения** | | | | |
| математическое  ожидание | медиана | | | мода |
| **ДСВ** |  | **\_** | | | наиболее  вероятное  значение |
| **НСВ** |  | квантиль  (делит площадь под плотностью  вероятности  пополам) | | | точка максимума  функции  плотности |
|  | **Характеристики рассеяния** | | | | |
| **D(X)**  дисперсия | среднее  квадратическое  отклонение | | | коэффициент  вариации |
| **ДСВ** |  |  | | | оценка  относительного рассеяния СВ *Х* по сравнению со  средним  значением |
| **НСВ** |  |
|  | **Теоретические моменты** | | | | |
| **Начальные моменты**  **k-го порядка** | | **Центральные моменты**  **k-го порядка** | | |
| **ДСВ** |  | |  | | |
| **НСВ** |  | |  | | |
|  | **Характеристики формы** | | | | |
| асимметрия | | | эксцесс | |
| **ДСВ** | характеризует степень  несимметричности распределения | | | характеризует степень  крутизны распределения | |
| **НСВ** |

***Замечание. Квантиль*** *порядка р – такое значение СВ Х, для которого*

Приложение 2

**Библиотека стандартных распределений**

Для работы со случайными величинами (непрерывными и дискретными) в Mathcad есть библиотека встроенных функций наиболее распространенных распределений. Каждое распределение представлено в библиотеке: функцией распределения; функцией, вычисляющей вероятность заданного значения (для дискретных распределений); плотностью вероятностей (для непрерывных распределений); квантили распределения порядка *k*; генератором случайных чисел с заданным законом распределения.

Все функции распределения вероятностей в Mathcad начинаются с буквы «**р**». Функции, имена которых начинаются с буквы «**d**», задают плотность распределения вероятностей (или вычисляют вероятность в заданной точке «k» для дискретных величин); с «**q**» – квантили и с «**r**» – генерируют вектор *m* случайных чисел с соответствующим законом распределения. Затем идет название закона, в скобках записывают (x, par), где par – параметры закона.

**Библиотека основных распределений НСВ**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Функция***  ***плотности*** | ***Функция***  ***распределения*** | | ***Числовые***  ***характеристики*** |
| **Равномерное распределение (параметры *- а, в*)** | | | |
| dunif(*x,a,b*) | рunif(*x,a,b*) | |  |
| **Экспоненциальное распределение (параметр - λ)** | | | |
| dexp(*x,λ*) | pexp(*x,λ*) | |  |
| **Нормальное распределение (параметры *- а*, σ)** | | | |
| dnorm(*x,a,σ*) | где Ф(t) – функция Лапласа  pnorm(*x,a,σ*) | |  |
| **Логнормальное распределение (параметры *- а*, σ)** | | | |
| dlnorm(*x,a,σ*) | где Ф(t) – функция Лапласа  plnorm(*x,a,σ*) |  | |
| **Распределение**  **(k - число степеней свободы)** | | | |
| dchisq(*x,k*) | pchisq(*x,k*) |  | |
| **Распределение Стьюдента (k - число степеней свободы)** | | | |
| dt(*x,k*) | pt(*x,k*) | |  |
| **Распределение Фишера (k1, k2 - две степени свободы)** | | | |
| dF(*x,k1,k2*) | pF(*x,k1,k2*) | |  |
| **Бета-распределение (параметры -** | | | |
| dbeta(x, | не выражается в элементарных функциях  pbeta(x, | |  |

**Библиотека основных распределений ДСВ**

|  |  |  |
| --- | --- | --- |
| ***Кумулятивная***  ***(накопленная)***  ***вероятность*** | ***Вероятность***  ***заданного значения*** | ***Числовые***  ***характеристики*** |
| **Биномиальный закон (параметр - *р*)** | | |
| pbinom(*k,n,p*) | dbinom(*k,n,p*) |  |
| **Закон Пуассона (параметр – λ)** | | |
| ppois(*k,**λ*) | dpois(*k,**λ*) |  |
| **Геометрическое распределение (параметр - *р*)** | | |
| pgeom(*k,p*) | dgeom(*k,p*) |  |
| **Гипергеометрическое распределение (параметры – *n1, n2, n*)** | | |
| phypergeom(*k,n1,n2,n*) | dhypergeom(*k,n1,n2,n*) |  |

Приложение 3

**Библиотека типовых функций регрессии Mathcad**

|  |  |  |
| --- | --- | --- |
| ***Регрессионная***  ***модель*** | ***Встроенная функция*** | ***Замечание*** |
|  | **expfit(X,Y,S)** | Выдаёт вектор, содержащий три коэффициента, что лучше всего аппроксимирует данные в векторах x и y. Дополнительный вектор S содержит оценки для этих трех коэффициентов. |
|  | **lgsfit(X,Y,S)** | Выдаёт вектор, содержащий эти 3 коэффициента для логистической кривой, лучше всего аппроксимируя данные в векторах x и y, используя оценки в S. |
|  | **pwrfit(X,Y,S)** | Выдаёт вектор, содержащий коэффициенты для кривой степени, что лучше всего аппроксимирует данные в векторах x и y. Вектор S содержит оценки для этих трех коэффициентов. |
|  | **sinfit(X,Y,S)** | Выдаёт вектор, содержащий коэффициенты для синусоиды, что лучше всего аппроксимирует данные в векторах x и y. Вектор S содержит оценки для этих трех коэффициентов**.** |
|  | **logfit(X,Y,S)** | Выдаёт вектор, содержащий эти три коэффициента для логарифмической кривой, что лучше всего аппроксимирует данные в векторах x и y. Вектор S содержит оценки для этих трех коэффициентов. |
|  | **lnfit(X,Y)** | Выдаёт вектор, содержащий эти 2 коэффициента для логарифмической кривой, что лучше всего аппроксимирует данные в x и y. |