

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



Федеральное государственное автономное образовательное учреждение
высшего образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

УТВЕРЖДАЮ

Директор ИНК

В.Н. Бориков

_____ «__» _____ 2014г.

**Лабораторная работа №2
Элементарные задачи математической
статистики.**

Томск - 2014

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

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

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

1. Ввод и вывод файлов данных

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

Самый простой из них — использовать имеющееся семейство встроенных функций.

- READPRN (“file”) – чтение данных в матрицу из текстового файла;
- WRITEPRN (“file”) – запись данных в текстовый файл;
- APPENDPRN (“file”) – дозапись данных в существующий текстовый файл;
- file — путь к файлу.

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

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

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

Пример 1: Запись данных в файл "data.txt"

$x := rpost(N, a, \sigma)$	создание выборки случайных величин распределенных по нормальном закону;
$CWD := "D:\tmp\"$	устанавливается текущий рабочий каталог.
$WRITEPRN("data.txt") := x$	запись в файл «data» созданной ранее выборки x .

2. Моделирование выборок из стандартных распределений

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

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

Для этого необходимо выполнить следующие действия:

1. Установить курсор на место вставки функции в документе.
2. Вызвать диалоговое окно **Insert Function** нажатием кнопки $f(x)$ на стандартной панели инструментов или командой меню **Insert / Function** (Вставка / Функция), или нажатием клавиш $\langle \text{Ctrl} \rangle + \langle \text{E} \rangle$.

3. Выбрать в списке **Function Category** (Категория функции) выберите одну из категорий статистических функций. Категория **Probability Density** (Плотность вероятности) содержит встроенные функции для плотности вероятности, Категория **Probability Distribution** (Функция распределения) — для вставки функций или квантилей распределения, Категория **Random Numbers** (Случайные числа) — для вставки функции генерации случайных чисел.

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

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

3. Функции Mathcad для расчета численных характеристик

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

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

4. Построение гистограмм

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

4.1 Гистограмма с произвольными сегментами разбиения

$\text{hist}, (\text{intvls}, x)$ – вектор частоты попадания данных в интервалы гистограммы;

- intvls – вектор, элементы которого задают сегменты построения гистограммы в порядке возрастания $a \leq \text{intvls}_i < b$;
- x – вектор случайных данных.

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

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

$\text{sort}(x)$ – сортировка выборки в порядке возрастания;

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

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

Пример 2. Построение гистограммы

$N:=1000$

$x:= \text{binom}(N, \Delta, 0,5)$

$bin:=30$

; кол-во равных сегментов, на кот. разбивается весь диапазон

; определение границы интервала построения гистограммы

$lower:= \text{floor}(\min(x))$

; наибольшее цело число $\leq \min(x)$

$upper:= \text{ceil}(\max(x))$

; наименьшее цело число $\geq \max(x)$

$h:= \frac{upper - lower}{bin}$

; размер сегмента

$j:= 0 .. bin$

; счетчик сегментов

$int_j := lower + h \cdot j$

$f := \frac{1}{N \cdot h} \cdot \text{hist}(int, x)$

; массив начальных точек каждого сегмента

$int := int + 0.5h$

; от левой границы каждого сегмента к его центру

; нормирование гистограммы для удобства отображения на одном графике вместе с плотностью распределения

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

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

4.2 Гистограмма с разбиением на равные сегменты

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

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

Пример 3. Построение гистограммы (упрощенный вариант)

$N:=100$

; созданы выборки сл. величины

$x:= \text{exp}(N, 1)$

$bin:=30$

; кол-во равных сегментов, на кот. разбивается весь диапазон

$f := \text{histogram}(bin, x)$


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

4.3 Создание графика гистограммы

Для того чтобы создать график в виде гистограммы необходимо:

1. Построить двумерный график, по оси x откладываются границы интегралов, по оси y частота(процент) попадания значения сл. величины в заданные интервалы.
2. Перейти в диалоговом окне **Formatting Currently Selected Graph** (Форматирование) выбранного графика (например, двойным щелчком мыши) в раздел **Traces** (Графики). Установить в поле **Type** (Тип) элемент списка **bar** (столбцы) или **solidbar** (гистограмма). Тип **solidbar** специально предназначен для гистограмм.
3. Нажать кнопку ОК.

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

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

5. Полигон частот

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

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

Порядок выполнения работы .

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

Элементарные задачи математической статистики

10. Понаблюдайте, как изменится внешний вид гистограммы, если изменить количество интервалов разбиения выборки. Сделать выводы

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

Литература.

1. Васильев А.В. Mathcad 13 на примерах – СПб.: БХВ - Петербург, 2006. – 528с.
2. Плис А.И., Сливина Н.А. Mathcad 2000. Математический практикум для экономистов и инженеров: Учеб. Пособие. – М.: Финансы и статистика, 2000. -656с.: ил.
3. Кирьянов Д. В. Самоучитель Mathcad 11. - СПб.: БХВ-Петербург, 2003. - 560 с: ил.