

М.М. Герасимова

МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ЗАДАЧ НА ЭВМ



Красноярск 2011

Министерство образования и науки Российской Федерации

ГОУ ВПО «СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Лесосибирский филиал

М.М. Герасимова

Методы обработки экспериментальных задач на ЭВМ

Утверждено редакционно-издательским советом СибГТУ в качестве
лабораторного практикума для студентов специальности 250401
Лесоинженерное дело направления 250400 Технология
лесозаготовительных и деревообрабатывающих производств очной формы
обучения

Красноярск 2011

Герасимова, М.М. Методы обработки экспериментальных задач на ЭВМ [Текст]: лабораторный практикум для студентов специальности 250401 Лесоинженерное дело направления 250400 Технология лесозаготовительных и деревообрабатывающих производств очной формы обучения / М.М. Герасимова. - Красноярск: СибГТУ, 2011. - 104 с.

Рецензенты: канд. физ. - мат. наук Е.Н. Яковлева (СФУ);

доц. Н.Г. Черноусова (научно-методический совет СибГТУ).

Лабораторный практикум предназначен для выполнения лабораторных работ по дисциплине «Методы обработки экспериментальных задач на ЭВМ». Он содержит теоретическую, практическую части и контрольные вопросы по основным темам курса.

© М.М. Герасимова, 2011

© ГОУ ВПО «Сибирский государственный
технологический университет»,
Лесосибирский филиал, 2011

СОДЕРЖАНИЕ

Введение.....	4
Раздел 1 Статистическая обработка результатов эксперимента	6
1.1 Статистический анализ вариационных рядов.....	6
Лабораторная работа 1 Получение и статистический анализ вариационных рядов (4 ч).....	6
1.2 Статистическая проверка статистических гипотез.....	31
Лабораторная работа 2 Проверка гипотезы о нормальном распределении генеральной совокупности (2 ч).....	31
Раздел 2 Дисперсионный анализ	39
2.1 Однофакторный дисперсионный анализ	39
Лабораторная работа 3 Однофакторный дисперсионный анализ (2 ч)	39
Раздел 3 Корреляционно-регрессионный анализ.....	51
3.1 Парная регрессия и корреляция.....	51
Лабораторная работа 4 Парный регрессионный анализ (4 ч).....	51
Раздел 4 Метод полного факторного эксперимента.....	73
4.1 Сущность метода полного факторного эксперимента	73
Лабораторная работа 5 Расчет реализованного плана полного факторного эксперимента (4 ч).....	73
Библиографический список.....	93
Приложение А (справочное) Статистико-математические таблицы.....	94

Введение

Успешное решение задач, стоящих перед лесной промышленностью, связано с созданием гибких производственных процессов, малоотходных и ресурсосберегающих технологий, компьютеризацией и автоматизацией управления производством. Эффективное использование вычислительных средств с целью принятия научно обоснованных решений в процессе планирования и управления возможно только при знании соответствующих методов и программ, а также умении самостоятельно ставить задачи, решать их с помощью ЭВМ и грамотно анализировать результаты. Поэтому возникла необходимость в подготовке инженеров, обладающих соответствующими знаниями математических методов и моделей, составляющих основу математического обеспечения ЭВМ.

Цель данного лабораторного практикума - научить студентов современным способам обработки информации, необходимым знаниям математических методов, которые используются в лесоинженерном деле, и умению на основе результатов сделать профессиональный вывод.

Лабораторный практикум построен на основе рабочего учебного плана специальности 250401.65 Лесоинженерное дело направления 250400 Технология лесозаготовительных и деревообрабатывающих производств.

Курс «Методы обработки экспериментальных задач на ЭВМ» общим объемом 60 часов изучается в течение VI семестра, из них лекций –18 часов, лабораторных занятий – 16 часов, и 26 часов дается студенту на самостоятельную работу. Изучение курса завершается зачетом.

Лабораторный практикум содержит теоретическую и практическую части по основным темам курса. В ходе выполнения лабораторной работы студент получает навыки автоматической обработки данных с помощью табличного процессора Microsoft Excel, выполняет расчёты и делает вывод по обработке статистической информации.

Прежде чем приступить к выполнению лабораторной работы, студент должен понять задание, ознакомиться с технологией выполнения работы и примером решения аналогичной задачи. Выполнив работу, студент предъявляет отчет, который должен содержать следующие сведения:

- дату выполнения;
- номер и название работы;
- исходные данные варианта;
- результаты и анализ решения задачи.

Обязательной является защита лабораторной работы. Требования к защите: студент должен пояснить применение метода решения задачи, объяснить основные этапы ее решения, сделать выводы и ответить на контрольные вопросы.

Раздел 1 Статистическая обработка результатов эксперимента

1.1 Статистический анализ вариационных рядов

Лабораторная работа 1

Получение и статистический анализ вариационных рядов (4 ч)

Цель работы: приобретение навыков получения вариационных рядов, их графического представления и вычисления оценок основных числовых характеристик выборки.

Теоретическая часть. При проведении наблюдений и экспериментов получают количественную информацию по какому-либо признаку X . Отдельные значения (варианты) признака x_1, x_2, \dots, x_n образуют непрерывный ряд. Последовательность вариантов, записанных в возрастающем порядке, называется *вариационным рядом*.

Любой вариационный ряд можно преобразовать в дискретный (интервальный) ряд. Интервальный вариационный ряд получается группировкой данных, в результате исходный ряд разбивается на определённое количество интервалов. Каждый интервал характеризуют его нижняя и верхняя границы, середина интервала и количество значений признака, которые в него попадают (частота).

Группировка данных выполняется в следующем порядке:

1. Определяется количество интервалов. При производственных исследованиях исходят из того, что количество интервалов k при малой выборке ($n < 30$) должно быть $k = 5 - 6$, при большой выборке ($n > 30$) $k = 8 - 10$.

В научных исследованиях количество интервалов рассчитывается по формуле:

$$k = 1 + 3,32 \lg n, \quad (1)$$

где n – объём выборки.

Вычисляется длина интервала h по формуле:

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{x_{\max} - x_{\min}}{1 + 3,321 \cdot \lg n}, \quad (2)$$

где x_{\max} , x_{\min} – максимальное и минимальное значения признака.

2. Устанавливаются нижняя и верхняя границы интервалов. Нижняя граница первого интервала определяется следующим образом:

$$a_0 = x_{\min} - \frac{h}{2}. \quad (3)$$

Верхняя граница интервала с номером i определяется по формуле

$$a_i = a_{i-1} + h, \quad (4)$$

где a_{i-1} - верхняя граница $(i-1)$ -го интервала (нижняя граница i -го интервала);

$$i = 1, 2, 3, \dots$$

4. Составляются интервальный и дискретный ряды распределения случайной величины X (таблицы 1, 2).

Таблица 1 – Форма табличного представления интервального ряда распределения

Границы интервала	Частота n_i	Относительная частота w_i

Таблица 2 – Форма табличного представления дискретного ряда распределения

Середина интервала x_i	Частота n_i	Относительная частота w_i

Серединой интервала является среднее значение указанного интервала, которое для i -го интервала вычисляется по формуле

$$x_i = \frac{a_{i-1} + a_i}{2}. \quad (5)$$

Ряд делится на интервалы до тех пор, пока максимальное значение x_{\max} не попадёт в последний интервал.

Частота – это количество значений признака, которые попадают в данный интервал.

Относительная частота вычисляется по формуле

$$w_i = \frac{n_i}{n}, \quad (6)$$

где w_i - относительная частота, соответствующая i -му интервалу;

n_i - частота, соответствующая i -му интервалу.

Сгруппированный ряд распределения может быть представлен как в табличной форме (таблицы 1, 2), так и в графическом виде. Существует несколько графических способов изображения данных. Так, ряд распределения может быть представлен гистограммой и полигоном распределения.

Гистограмма частот (относительных частот) строится для изображения ряда распределения в виде прямоугольников. На оси абсцисс откладываются верхние и нижние границы интервалов, на оси ординат откладывают частоту (относительную частоту). Граничные линии проводятся перпендикулярно оси абсцисс (рисунок 1).

Полигон частот (относительных частот) – это ломаная, отрезки которой соединяют точки $(x_i; n_i)$ ($(x_i; w_i)$). На оси абсцисс откладывают середины интервалов, а на оси ординат – частоту (относительную частоту) (рисунок 2).

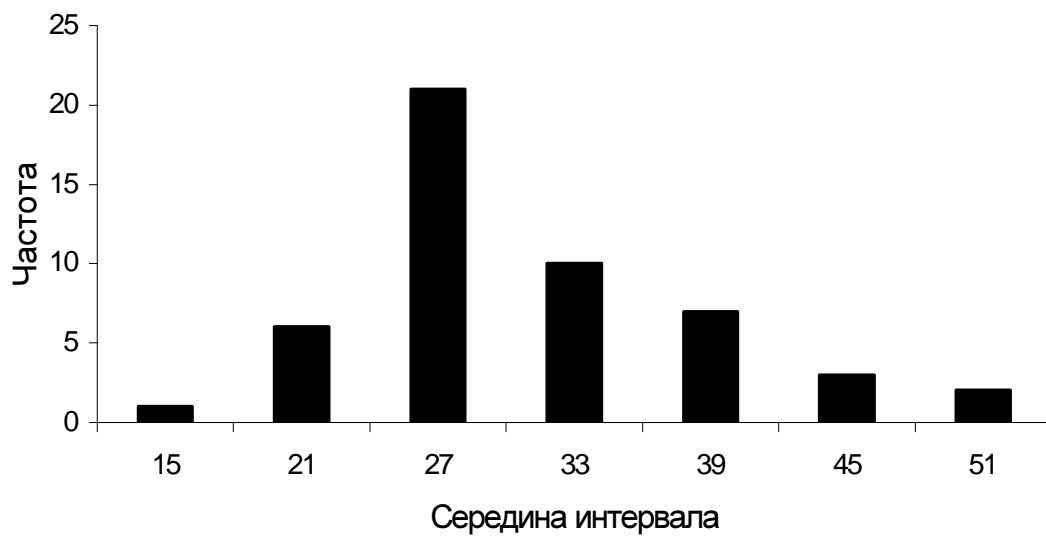


Рисунок 1 – Гистограмма частот

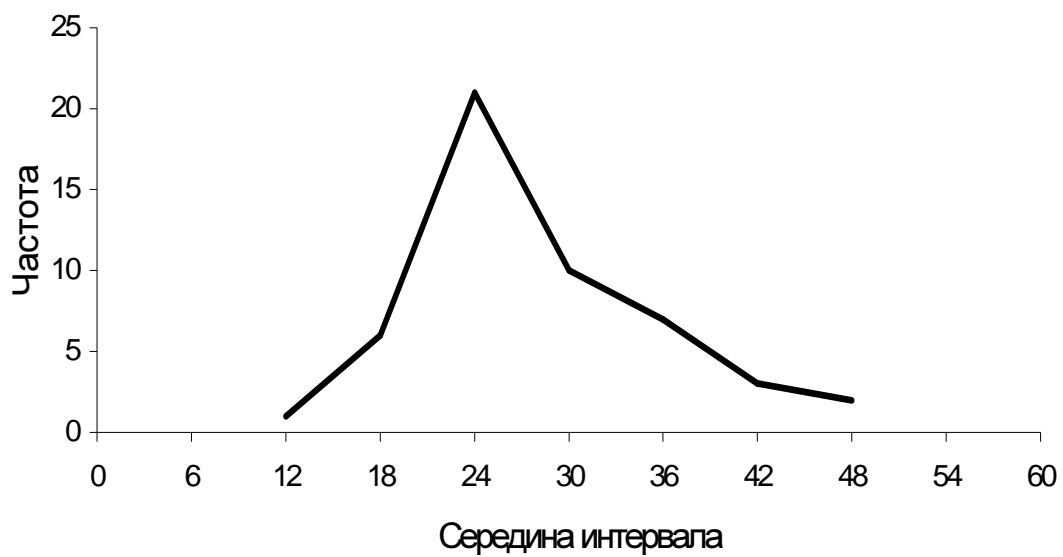


Рисунок 2 – Полигон частот

Основными статистическими показателями ряда распределения являются средние значения и показатели вариации признака.

Выборочная средняя для сгруппированного ряда распределения вычисляется по формуле

$$\bar{x} = \frac{\sum_{i=1}^n x_i n_i}{n}, \quad (7)$$

где \bar{x} - выборочная средняя;

x_i – середина i –го интервала;

n_i – частота, соответствующая i – му интервалу;

n – объём выборки.

Центр ряда распределения, кроме среднего значения, характеризуют непараметрические показатели – медиана и мода.

Медиана – это значение вариационного ряда, которое делит его на две равные совокупности.

Мода – это наиболее часто встречающееся значение вариационного ряда.

Для характеристики варьирования признака вычисляют следующие статистики: выборочная дисперсия, выборочное среднее квадратическое отклонение, коэффициент вариации.

Выборочная дисперсия характеризует разброс значений признака X относительно выборочной средней:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 n_i}{n}. \quad (8)$$

Эта величина показывает среднюю изменчивость вариантов в ряду распределения, но так как дисперсия - величина квадратичная, это не позволяет использовать данный статистический показатель в сравнительном анализе.

Выборочное среднее квадратическое отклонение показывает среднее отклонение вариантов от центра ряда распределения:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 n_i}{n}}. \quad (9)$$

Среднее квадратическое отклонение – величина линейная, что позволяет характеризовать выборочную совокупность, показывая меру рассеяния вариант по отношению к центру распределения. Но этот статистический показатель не может использоваться для сравнения изменчивости разных признаков.

Коэффициент вариации характеризует изменчивость признаков в сопоставимых единицах (процентах):

$$V = \frac{S}{\bar{x}} \cdot 100\% , \quad (10)$$

где \bar{x} - выборочная средняя;

S – выборочное среднее квадратическое отклонение.

По величине коэффициента вариации устанавливают меру изменчивости признака (таблица 3).

Таблица 3 – Соответствие между коэффициентом вариации и изменчивостью признака

Коэффициент вариации	Изменчивость признака
до 5%	слабая
6 – 10%	умеренная
11 – 20%	значительная
21 – 50%	большая
51% и более	очень большая

Достоинство этого показателя в том, что он может использоваться для сравнения изменчивости признаков, имеющих разную размерность, недостаток заключается в том, что коэффициент вариации зависит от среднего значения, поэтому при небольших значениях средних величин наблюдаются высокие величины коэффициентов вариации.

Выборочный коэффициент асимметрии характеризует асимметрию распределения:

$$A^* = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 n_i}{n \cdot S^3}. \quad (11)$$

Выборочный коэффициент эксцесса характеризует островершинность распределения

$$\mathcal{E}^* = \frac{\sum_{i=1}^n (x_i - \bar{x})^4 n_i}{n \cdot S^4} - 3. \quad (12)$$

Эти числовые характеристики служат для сравнения эмпирического (наблюдаемого) распределения с нормальным, для которого оба этих коэффициента равны нулю. Для предварительного выбора закона распределения вычисляют средние квадратические ошибки определения коэффициентов асимметрии и эксцесса:

$$S_A = \sqrt{\frac{6(n-1)}{(n+1) \cdot (n+3)}}, \quad (13)$$

$$S_{\mathcal{E}} = \sqrt{\frac{24n(n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}}. \quad (14)$$

Если выборочные коэффициенты асимметрии и эксцесса отличаются от нуля не более чем на удвоенные средние квадратические ошибки их определения, то можно сделать предположение, что эмпирическое распределение случайной величины X является нормальным.

Пример выполнения работы

Выполним статистический анализ результатов наблюдений, представленных в таблице 4.

Таблица 4 – Исходные данные

5,2	4,8	3,8	4,4	4	6	6	7,6	5	4,8
7,8	9,1	7	6,2	7	5	5	11	6,4	5,6
6,2	4,4	7,2	5,8	6,8	6	6,8	6,6	5,8	7,8
4,2	9	8,2	6,6	9,4	5,6	6,6	6,2	5,2	6,4
8,6	9,3	7	6,2	5,2	5	7	7,8	5,8	5
7,4	6,2	2,8	5	4,8	4,8	6,8	7	6,8	4,4
8	7,6	5,6	4,8	4,8	8,4	7,2	4,8	4,6	9
7	8	5,2	5,6	4,8	6,6	6,4	9,2	7,2	7,6
5,4	10,3	2,4	6	5,6	5	5,6	5,4	7,6	5,4
6,2	6	5,4	2,8	4,4	6,4	4,6	4,6	7,8	6,4

1. Для удобства обработки данных их надо сгруппировать. С этой целью найдем наибольшее и наименьшее наблюдаемые значения исследуемой случайной величины X : $x_{\max} = 11$, $x_{\min} = 2,4$. Вычисление этих значений выполняется в Excel с помощью функций МАКС и МИН.

Все остальные наблюдаемые значения находятся в промежутке $[x_{\min}, x_{\max}]$. Разобьем этот отрезок на интервалы. По формуле (2) определим длину частичного интервала h :

$$h = \frac{11 - 2,4}{1 + 3,321 \cdot \lg 100} = 1,1.$$

Начало первого интервала рассчитаем по формуле (3):

$$a_0 = 2,4 - \frac{1,1}{2} = 1,85.$$

Таким образом, минимальное значение случайной величины X (СВХ) попадает в середину первого частичного интервала. Для определения конца интервала с номером i будем использовать формулу (4), где $i = 1, 2, 3, \dots$. Процесс вычислений прекращаем, когда при каком -

либо значения i a_i станет больше или равно x_{\max} . Полученные интервалы будем считать закрытыми справа.

$$a_0 = 1,85,$$

$$a_1 = 1,85 + 1,1 = 2,95,$$

$$a_2 = 2,95 + 1,1 = 4,05,$$

$$a_3 = 4,05 + 1,1 = 5,15,$$

$$a_4 = 5,15 + 1,1 = 6,25,$$

$$a_5 = 6,25 + 1,1 = 7,35,$$

$$a_6 = 7,35 + 1,1 = 8,45,$$

$$a_7 = 8,45 + 1,1 = 9,55,$$

$$a_8 = 9,55 + 1,1 = 10,65,$$

$$a_9 = 10,65 + 1,1 = 11,75.$$

$15,32 > x_{\max}$, поэтому дальнейшее вычисление производить не нужно.

После разбивки всего диапазона изменений значений CBX на 9 частичных интервалов необходимо подсчитать число значений, попавших в каждый из них, т.е. частоту n_i , соответствующую данному интервалу. Частоту можно найти в табличном редакторе Microsoft Excel при помощи функции ЧАСТОТА. Ее аргументами являются массив данных и массив интервалов.

Массивом данных являются значения случайной величины X из таблицы 1. В Excel они введены в ячейки A1:A100. Массив интервалов -

значения границ интервалов, в нашем примере это значения a_0, \dots, a_9 , в Excel - массив C2:C11 (рисунок 3). Формулу для подсчета частот необходимо ввести как формулу массива. Сначала формула вводится в одну ячейку, например, D2. Затем выделяем диапазон D2:D11, нажимаем клавишу F2, а затем - клавиши CTRL+SHIFT+ENTER.

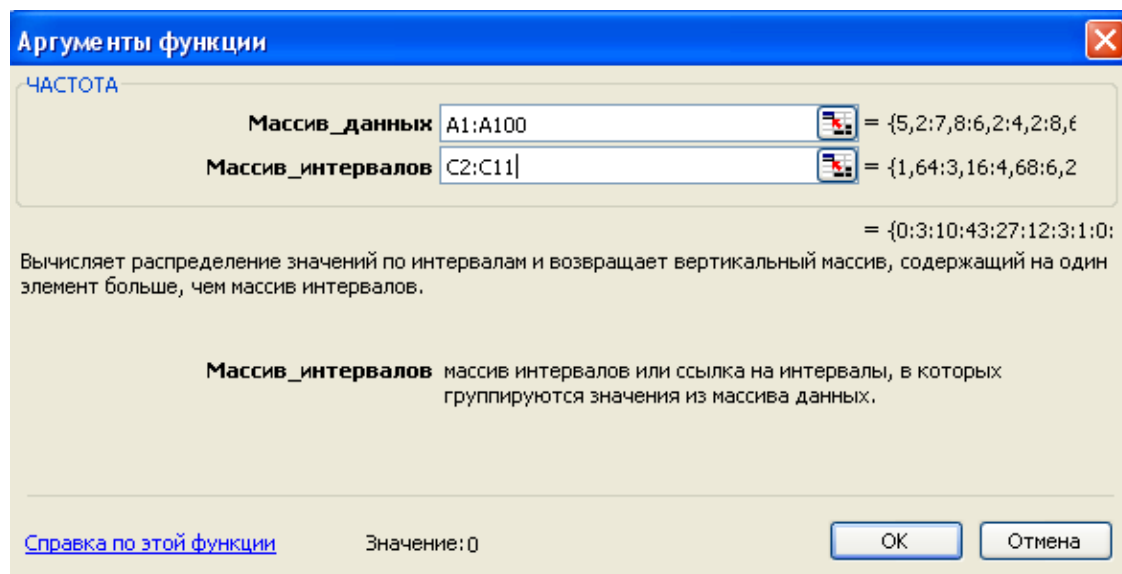


Рисунок 3 – Окно *Аргументы функции* ЧАСТОТА

В ячейках D2:D11 выводится количество значений случайной величины X , находящихся в интервале, правой границей которого является соответствующее значение в массиве C2:C11. Так, например, значение в ячейке D3, равное 3, есть частота, соответствующая интервалу $(1,85;2,95]$. Так как в массиве данных нет значений, меньших, чем нижняя граница первого интервала, то в ячейке D2 выводится значение, равное нулю (рисунок 4).

На основании результатов вычислений по формуле (6) определим относительные частоты w_i , по формуле (5) - середины частичных интервалов и составим интервальный и дискретный ряды распределения случайной величины X (таблицы 5, 6).