

ПРИКЛАДНОЙ АНАЛИЗ ДАННЫХ

Часть 1

Улан-Удэ • 2017

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
БУРЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ПРИКЛАДНОЙ АНАЛИЗ ДАННЫХ

Часть 1

*Рекомендовано УМС БГУ
в качестве учебно-методического пособия для обучающихся
по направлениям подготовки 02.03.03 Математическое
обеспечение и администрирование информационных систем,
02.03.01 Математика и компьютерные науки, 01.03.02 Прикладная
математика и информатика, 09.03.03 Прикладная информатика*

Улан-Удэ
Издательство Бурятского госуниверситета
2017

УДК 519.23
ББК 22.172я73
П 759

Утверждено к печати
редакционно-издательским советом
Бурятского государственного университета

Рецензенты

Т. Г. Дармаев, кандидат физико-математических наук, доцент,
заведующий лабораторией вычислительных
и геоинформационных систем Научно-образовательного
и инновационного центра системных исследований
и автоматизации Бурятского государственного университета

С. Г. Баргуев, кандидат физико-математических наук, доцент,
заведующий кафедрой высшей математики
и общеобразовательных дисциплин Бурятского института
инфокоммуникаций Сибирского государственного университета
телекоммуникаций и информатики

П 759 **Прикладной анализ данных. Часть 1:** учебно-методическое пособие / сост. А. С. Цыбиков. — Улан-Удэ : Издательство Бурятского государственного университета, 2017. — 146 с. ISBN 978-5-9793-0038-2

В учебно-методическом пособии изложены основные понятия, параметрические и непараметрические методы и алгоритмы обработки и анализа данных

Пособие предназначено для бакалавриата по направлениям подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем, 02.03.01 Математика и компьютерные науки, 01.03.02 Прикладная математика и информатика, 09.03.03 Прикладная информатика, 01.03.01 Математика

УДК 519.23
ББК 22.172я73

ISBN 978-5-9793-0038-2

© А. С. Цыбиков, составление, 2017
© Бурятский госуниверситет, 2017

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	6
ГЛАВА 1. ФУНДАМЕНТАЛЬНЫЕ ПОНЯТИЯ	9
1.1. Выборочное исследование	9
1.2. Что такое переменная?	10
1.3. Шкалы измерений	10
1.4. Простейшие описательные статистики	13
1.5. Свойства описательных статистик	14
1.6. Какие статистики выбрать?	18
1.7. Статистическая значимость и степень свободы	18
Контрольные вопросы	20
Лабораторная работа №1	20
ГЛАВА 2. РАСПРЕДЕЛЕНИЕ ПЕРЕМЕННОЙ	21
2.1. Нормальное распределение	21
2.2. Проверка нормальности распределения	23
2.3. Двумерное нормальное распределение	24
2.4. Другие виды распределения	25
2.5. Основные принципы оценивания. Метод максимального правдоподобия	28
2.6. Метод минимума Хи-квадрат	30
Контрольные вопросы	31
Лабораторная работа №2	32
ГЛАВА 3. ЗАВИСИМОСТИ МЕЖДУ ПЕРЕМЕННЫМИ. КОРРЕЛЯЦИЯ	33
3.1. Зависимые и независимые переменные	33
3.2. Понятие корреляции	34
3.3. Коэффициент корреляции г-Пирсона	37
3.4. Корреляция и регрессия	43
2.4. Частная корреляция	47
Контрольные вопросы	49
Лабораторная работа №3	49
ГЛАВА 4. НЕПАРАМЕТРИЧЕСКАЯ КОРРЕЛЯЦИЯ	51
4.1. Коэффициент корреляции г-Спирмена	51
4.2. Корреляция бинарных данных	52
Контрольные вопросы	55
Лабораторная работа №4	56

ГЛАВА 5. ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ ДВУХ ВЫБОРОК	57
5.1. Сравнение дисперсий	57
5.2. Критерий t-Стьюдента для одной выборки	59
5.3. Критерий t-Стьюдента для независимых выборок	60
5.4. Критерий t-Стьюдента для зависимых выборок	62
Контрольные вопросы	64
Лабораторная работа №5	64
ГЛАВА 6. НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ ВЫБОРОК	66
6.1. О непараметрических методах	66
6.2. Сравнение двух независимых выборок	68
6.3. Сравнение двух зависимых выборок	70
6.4. Сравнение более двух независимых выборок	73
6.5. Сравнение более двух зависимых выборок	75
Контрольные вопросы	77
Лабораторная работа №6	78
ГЛАВА 7. ДИСПЕРСИОННЫЙ АНАЛИЗ (ANOVA)	79
7.1 Назначение и общие понятия ANOVA	79
7.2. Однофакторный ANOVA	84
7.3. Множественные сравнения в ANOVA	90
Контрольные вопросы	93
Лабораторная работа №7	94
ГЛАВА 8. МНОГОФАКТОРНЫЙ ANOVA	95
8.1. Назначение и модель многофакторного ANOVA	95
8.2. Общая схема вычислений	100
Контрольные вопросы	108
Лабораторная работа №8	108
ГЛАВА 9. ANOVA С ПОВТОРНЫМИ ИЗМЕРЕНИЯМИ	110
9.1. Назначение и особенность ANOVA с повторными измерениями	110
9.2. Общая схема вычислений	113
Контрольные вопросы	119
Лабораторная работа №9	120
ЗАКЛЮЧЕНИЕ	121
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	122
ПРИЛОЖЕНИЕ	123
Приложение 1. Стандартные нормальные вероятности	124

Приложение 2. Критические значения критерия t-Стьюдента	125
Приложение 3. Критические значения критерия F-Фишера	125
Приложение 4. Критические значения критерия χ^2	129
Приложение 5. Критические значения для числа серий	131
Приложение 6. Критические значения коэффициента корреляции r-Пирсона (r-Спирмена)	133
Приложение 7. Критические значения z-преобразования Фишера для коэффициентов корреляции	135
Приложение 8. Критические значения критерия Фишера для проверки ненаправленных альтернатив	136
Приложение 9. Критические значения критерия U-Манна-Уитни	138
Приложение 10. Критические значения критерия T-Вилкоксона	140
Приложение 11. Критические значения критерия G знаков	141
Приложение 12. Критические значения критерия H-Краскала-Уолеса для трех выборок численностью $n \leq 5$	142
Приложение 13. Критические значения критерия χ^2 -Фридмана для трех выборок численностью $n < 10$	145

ПРЕДИСЛОВИЕ

Настоящее учебное издание представляет собой учебно-методическое пособие для дисциплин «Прикладной анализ данных» и «Технология обработки и анализа данных» в рамках реализации образовательной программы высшего образования по направлениям подготовки бакалавров 02.03.03 Математическое обеспечение и администрирование информационных систем, 02.03.01 Математика и компьютерные науки, 01.03.02 Прикладная математика и информатика, 01.03.01 Математика очной формы обучения и 09.03.03 Прикладная информатика очной и заочной формы обучения и подготовлено в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования.

Дисциплина «Прикладной анализа данных» относится к:

– дисциплинам по выбору вариативной части Блока 1 в структуре образовательной программы по направлениям подготовки Математическое обеспечение и администрирование информационных систем, Прикладная информатика

Дисциплина «Технология обработки и анализа данных» относится к:

– обязательным дисциплинам вариативной части Блока 1 в структуре образовательной программы по направлениям подготовки Прикладная математика и информатика, Математика и компьютерные науки, Математика;

Изучение дисциплины направлено на формирование общекультурных/ общепрофессиональных/ профессиональных компетенций:

ОПК-2.Способность анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования.

ОПК-4.Способность находить, анализировать, реализовывать программно и использовать на практике математические алгоритмы, в том числе с применением современных вычислительных систем.

ПК-1.Способность собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям.

ПК-3.Готовность к разработке моделирующих алгоритмов и реализации их на базе языков и пакетов прикладных программ моделирования.

В результате освоения дисциплины обучающийся должен:

Знать: основные понятия, методы и алгоритмы обработки и анализа данных, в т. ч. параметрические и непараметрические критерии проверки гипотез о связях, различиях и сдвигах между двумя и несколькими выборками;

Уметь: формулировать задачи анализа данных, возникающие в ходе научно-исследовательской и аналитической деятельности; проводить предобработку данных; корректно подбирать и применять соответствующие методы обработки и анализа; применять автоматизирующие и визуализирующие программные средства анализа данных; представлять результаты практической работы в виде отчетов;

Владеть: методическими и техническими способами решения стандартных задач анализа данных

Основной задачей настоящего учебно-методического пособия систематизация и практическая реализация знаний в рамках дисциплин «Прикладной анализ данных» и «Технология обработки и анализа данных».

Пособие состоит из 9 глав и приложений:

Первые две главы включают базовым понятиям и описательным статистикам.

В третьей и четвертой главах описаны параметрические и непараметрические методы исследования взаимосвязей между признаками.

В пятой и шестой главах рассматриваются основные параметрические и непараметрические методы исследования различий и сдвигов между выборками.

Последние три главы посвящены однофакторному и многофакторному дисперсионному анализу (ANOVA) – математический аппарат для сравнения нескольких выборок.

В приложении 13 таблиц критических значений необходимых для применения соответствующих статистических методов.

Желаем успехов!