

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ЧЕРЕПОВЕЦКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

Кафедра математики и информатики

УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ
«МЕТОДЫ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

Направление подготовки (специальность):

01.03.02 Прикладная математика и информатика

Образовательная программа:

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Очная форма обучения

Составители:

Парыгина С.А., доцент кафедры МиИ,
канд. психол. наук, доцент

г. Череповец - 2022

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

Основная литература

1. Павлов, Л. А. Структуры и алгоритмы обработки данных: учебник для вузов / Л. А. Павлов, Н. В. Первова. — 3-е изд., стер. — Санкт-Петербург: Лань, 2021. — 256 с. — ISBN 978-5-8114-7259-8. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/156929>
2. Кораблин, Ю. П. Структуры и алгоритмы обработки данных: учебно-методическое пособие / Ю. П. Кораблин, В. П. Сыромятников, Л. А. Скворцова. — Москва: РТУ МИРЭА, 2020. — 219 с. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/163860>
3. Статистическая обработка экспериментальных данных. Язык R в задачах планирования эксперимента: учебное пособие / А. С. Тарасов, И. А. Орлова, Е. С. Геращенко [и др.]. — Рязань: РГРТУ, 2018. — 32 с. — ISBN 978-5-6041320-6-7. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/168237>
4. Боровков, А. А. Математическая статистика: учебник для вузов / А. А. Боровков. — 5-е изд., стер. — Санкт-Петербург: Лань, 2021. — 704 с. — ISBN 978-5-8114-7677-0. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/164711>
5. Родионова, Т. Е. Информационные технологии обработки данных: учебное пособие / Т. Е. Родионова. — Ульяновск: УлГТУ, 2020. — 113 с. — ISBN 978-5-9795-2017-9. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165028>

Дополнительная литература

1. Планирование эксперимента и обработка экспериментальных данных на ЭВМ: учебное пособие. — Санкт-Петербург: СПбГУТ им. М.А. Бонч-Бруевича, 2014 — Часть 1: Обработка экспериментальных данных на ЭВМ — 2014. — 88 с. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/180145>
2. Планирование эксперимента и обработка экспериментальных данных на ЭВМ: учебное пособие. — Санкт-Петербург: СПбГУТ им. М.А. Бонч-Бруевича, 2014 — Часть 2: Планирование эксперимента — 2014. — 86 с. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/180146>
3. Методы математической статистики и информатики для обработки данных: учебное пособие для вузов / Шутикова М.И., Парыгина С.А. - Череповец: ФГБОУ ВПО ЧГУ, 2015. - 78 с. <https://edu.chsu.ru/>
4. Сидняев Н.И. Теория вероятностей и математическая статистика: учебник для бакалавров/ Сидняев Н.И. - Москва: Юрайт: ИД "Юрайт", 2014. - 219 с.: ил. + Приложения. - (Бакалавр). - Библиогр.: с. 218-219. - ISBN 978-5-9916-1379-8 (Изд-во "Юрайт").
5. Теория вероятностей и математическая статистика: учебное пособие для вузов / Бирюкова Л.Г., Бобрик Г.И., Ермаков В.И. и др.; под ред. В.И. Ермакова. - Москва: ИНФРА-М, 2012. - 287 с.: ил + Приложение. - (Высшее образование). - Библиогр.: с. 276-279. - ISBN 978-5-16-004996-0.
6. Прикладная статистика. Основы эконометрики. = ProbabilitiTheoryAndAppliedStatisnics: учебник для вузов в 2-х томах. Т. 1. Теория вероятностей и прикладная статистика / Айвазян С.А., Мхитарян В.С. - 2-е изд., испр. - Москва: ЮНИТИ-ДАНА, 2001. - 656 с.

7. Прикладная статистика. Основы эконометрики = ESSENTIALS of ECONOMETRICS: учебник для вузов в 2-х томах. Т.2. Основы эконометрики / Айвазян С.А. - 2-е изд., испр. - Москва: ЮНИТИ-ДАНА, 2001. - 432 с.

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине (модулю)

1. Айвазян, С. А. Прикладная статистика в упражнениях и задачах. / С. А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ-ДАНА, 2001.
2. Грешилов А.А. Математические методы принятия решений: учебное пособие для вузов / Грешилов А.А. - 2-е изд., испр. и доп. - Москва: Изд-во МГТУ им. Н.Э. Баумана, 2014. - 647 с.: ил. + Приложения. Предметный указатель. - Библиогр.: с. 557-562. - ISBN 978-5-7038-3910-2.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая перечень информационных справочных систем (при необходимости)

1. Электронная библиотека «Университетская библиотека online». URL: <http://biblioclub.ru/>
2. Информационная система «Единое окно доступа к образовательным ресурсам». URL: <http://window.edu.ru/>
3. Образовательный портал Череповецкого государственного университета. URL: <https://edu.chsu.ru/>
4. Образовательная платформа Открытое образование, онлайн курсы Университета ИТМО: Обработка и анализ больших данных. URL: <https://openedu.ru/course/ITMOUniversity/BIGDATA2035/>

Учебно-методические указания и рекомендации к изучению тем лекционных и практических занятий, самостоятельной работе студентов

Лекции

раздел дисциплины	Название темы лекции	час
1. Введение в прикладную статистику.	Основные подходы к интерпретации и анализу исходных экспериментальных данных. Основные этапы прикладного анализа данных.	2
2. Корреляционный анализ экспериментальных данных.	Назначение и место корреляционного анализа в прикладном статистическом исследовании. Корреляционный анализ количественных переменных: коэффициент детерминации, парный коэффициент корреляции Пирсона, корреляционное отношение. Множественный и частный коэффициенты корреляции. Корреляционный анализ порядковых переменных: понятие о ранговой корреляции, коэффициенты Спирмена и Кендалла, коэффициент конкордации. Корреляционный анализ категоризированных переменных: таблицы сопряжённости, основные меры связи. Критерий для проверки значимости выборочного коэффициента	10

	корреляции.	
3. Регрессионный анализ экспериментальных данных.	<p>Основные понятия и постановка задачи статистического исследования зависимостей. Основные типы зависимостей между количественными переменными.</p> <p>Основные этапы статистического исследования зависимостей.</p> <p>Выбор общего вида функции регрессии. Определение параметров линейной регрессии методом наименьших квадратов.</p>	8
4. Дисперсионный анализ как многомерный метод обработки экспериментальных данных.	<p>Понятие о дисперсионном анализе и его роль в обработке экспериментальных данных. Структура и математическая модель данных. Однофакторный и многофакторный дисперсионный анализ. Одинаковое и неодинаковое число испытаний на различных уровнях.</p>	6
	Итого	26

Практические занятия

№ раздела (модуля)	Тема практического занятия	Количество часов
2	<p>Корреляционный анализ количественных переменных: коэффициент детерминации, парный коэффициент корреляции Пирсона, корреляционное отношение. Множественный и частный коэффициенты корреляции.</p> <p>Корреляционный анализ порядковых переменных.</p> <p>Корреляционный анализ категоризированных переменных.</p> <p>Критерий для проверки значимости выборочного коэффициента корреляции.</p>	6
2	Контрольная работа.	2
3	Основные типы зависимостей между количественными переменными. Основные этапы статистического исследования зависимостей. Выбор общего вида функции регрессии. Определение параметров линейной регрессии методом наименьших квадратов.	4
3	Контрольная работа	2
4	Однофакторный и многофакторный дисперсионный анализ. Одинаковое и неодинаковое число испытаний на различных уровнях.	4
4	Контрольная работа	2
	Итого	20 ч.

Лабораторные работы

№ раздела (модуля)	Тема лабораторного занятия	Количество часов
2	<p>Корреляционный анализ количественных переменных: коэффициент детерминации, парный коэффициент корреляции Пирсона, корреляционное отношение. Множественный и частный коэффициенты корреляции.</p> <p>Корреляционный анализ порядковых переменных. Корреляционный анализ</p>	8

	категоризированных переменных. Критерий для проверки значимости выборочного коэффициента корреляции.	
3	Основные этапы статистического исследования зависимостей. Выбор общего вида функции регрессии. Определение параметров линейной регрессии методом наименьших квадратов.	8
4	Однофакторный и многофакторный дисперсионный анализ. Одинаковое и неодинаковое число испытаний на различных уровнях.	8
	Итого	20 ч.

Средства контроля качества обучения

Образцы индивидуальных заданий

Задание 1. Определить тип шкалы, в соответствии с которой сделаны следующие измерения, и обосновать свой выбор: а) измерение времени пробега стометровой дистанции, б) определение семейства растений, к которому относится ромашка.

Задание 2. Проверить значимость коэффициента корреляции (полученного в каком-нибудь ИЗ) с помощью соответствующего критерия на уровне значимости 0,01.

Задание 3. Наблюдения за 9 подростками по переменным «Рост» (X) и «Пол» (Y) приведены в таблице.

Номер испытуемого	Рост в см, X	Пол (1 – м., 0 – ж.), Y
1	150	1
2	170	0
3	160	1
4	165	1
5	140	0
6	183	1
7	157	0
8	152	0
9	163	1

Определить, связаны ли линейной зависимостью переменные X и Y .

Задание 4.12 учащихся ранжируются экспертом по их отношению к себе (переменная X) и к другим учащимся (переменная Y). Исходные данные представлены в таблице.

Номер учащегося	Ранги по переменной X	Ранги по переменной Y
1	2	6
2	8	5
3	12	10
4	3	7
5	1	3
6	6	4
7	7	9
8	10	8
9	4	1
10	9	11
11	11	12
12	5	2

Определить, связаны ли линейной зависимостью переменные X и Y .

Задание 5. Даны результаты 10 независимых наблюдений над системой точек (X , Y):

X	0,9	3,1	-0,8	-1,8	2,1	4,8	1,1	0,1	2,2	2,5
Y	1,4	2,2	0,8	0,6	1,8	2,8	1,5	1,2	1,8	1,9

Найти выборочный коэффициент корреляции. Проверить его на значимость (уровень значимости $\alpha = 0,05$). Предположив, что зависимость между x и y близка к линейной, найти уравнение линий регрессии y по x и x по y . Построить их.

Образцы заданий для контрольных работ

Задание 1. В выборке из 12 испытуемых изучалась связь между количеством сданных студентом в сессию дисциплин (переменная X) и количеством использованных им учебно-методических материалов (переменная Y).

По экспериментальным данным таблицы требуется:

1) построить диаграмму рассеяния и сделать предварительный вывод; 2) найти коэффициент корреляции Пирсона; 3) определить тип связи между признаками X и Y , если она существует, на основании 1-го и 2-го пунктов; 4) найти и построить уравнение регрессии в той же системе координат, что и диаграмму рассеяния; 5) сделать итоговый вывод

Номер испытуемого	X	Y
1	7	1
2	9	2
3	9	4
4	10	4
5	11	6
6	11	5
7	12	6
8	11	7
9	12	8
10	11	7
11	12	9
12	8	3

Задание 2. Произведено 4 испытания на каждом из 3 уровней. Результаты испытаний приведены в таблице. Методом дисперсионного анализа на уровне значимости 0,05 проверить нулевую гипотезу о равенстве математических ожиданий. Предполагается, что выборки извлечены из нормальных генеральных совокупностей одинаковыми дисперсиями

Номер испытания	Уровни фактора F_j		
	F_1	F_2	F_3
1	51	52	42
2	52	54	44
3	56	56	50
4	57	58	52
Групповое среднее \bar{X}_j	54	55	47

Образцы заданий для лабораторных работ

Задание 1. Исследователь изучает связь между временем, которое тратят испытуемые на просмотр телевизора (в часах) (переменная X) и количеством хронических заболеваний, диагностированных у них (переменная Y). По экспериментальным данным:

1. Построить диаграмму рассеяния.

2. Вычислить коэффициент корреляции Пирсона – 2-мя способами: 1) вручную; 2) с помощью электронного пакета.
3. Сделать вывод относительно типа линейной связи между переменными X и Y и обосновать его.

Задание 2. Исследователь изучает связь успешности рыночной деятельности бизнес-элиты с их уровнем самооценки (в баллах). Показатель успешности бизнесменов определяется путём ранжирования их по среднему уровню доходов (переменная X), показатель уровня самооценки определяется с помощью соответствующего тестирования (переменная Y). По экспериментальным данным определить:

1. Ранговые значения переменной Y .
2. Коэффициенты ранговой корреляции Спирмена и τ Кендалла – 2-мя способами: вручную и с помощью электронного пакета.
3. Сделать вывод относительно типа линейной связи между переменными X и Y и обосновать его.

Вопросы к экзамену

1. Основные подходы к интерпретации и анализу исходных экспериментальных данных.
2. Основные этапы прикладного анализа данных.
3. Основные типы зависимостей между количественными переменными.
4. Корреляционный анализ количественных признаков: коэффициент детерминации.
5. Корреляционный анализ количественных признаков: парный коэффициент корреляции Пирсона.
6. Корреляционный анализ количественных признаков: корреляционное отношение.
7. Множественный и частный коэффициенты корреляции.
8. Корреляционный анализ порядковых переменных: понятие о ранговой корреляции и коэффициент Спирмена.
9. Корреляционный анализ порядковых переменных: коэффициент Кендалла.
10. Корреляционный анализ порядковых переменных: коэффициент конкордации.
11. Корреляционный анализ категоризированных переменных: таблицы сопряжённости, меры связи.
12. Критерий для проверки значимости выборочного коэффициента корреляции.
13. Основные понятия и постановка задачи статистического исследования зависимостей.
14. Выбор общего вида функции регрессии.
15. Построение уравнения регрессии.
16. Статистические критерии проверки гипотез об общем виде функции регрессии.
17. Понятие о дисперсионном анализе.
18. Постановка задачи и модель данных однофакторного дисперсионного анализа.
19. Алгоритм однофакторного дисперсионного анализа.
20. Одинаковое и неодинаковое число испытаний на различных уровнях.
21. Многофакторный дисперсионный анализ.