

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:
Директор



А. В. Замятин

« 15 » июня 2023 г.

Рабочая программа дисциплины

Статистический анализ данных

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки :
Интеллектуальный анализ больших данных

Форма обучения

Очная

Квалификация

Магистр

Год приема

2023

Код дисциплины в учебном плане: Б1.О.02.04

СОГЛАСОВАНО:

Руководитель ОП

А.В. Замятин

Председатель УМК

С.П. Сущенко

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

- ОПК-2 – способность совершенствовать и реализовывать новые математические методы решения прикладных задач;
- ОПК-1 – способность решать актуальные задачи фундаментальной и прикладной математики.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.1 Использует результаты прикладной математики для освоения, адаптации новых методов решения задач в области своих профессиональных интересов.

ИОПК-2.2 Реализует и совершенствует новые методы, решения прикладных задач в области профессиональной деятельности.

ИОПК-2.3 Проводит качественный и количественный анализ полученного решения с целью построения оптимального варианта.

ИОПК-1.3 Решает актуальные задачи фундаментальной и прикладной математики.

2. Задачи освоения дисциплины

– Научить студентов решать задачи статистического анализа данных, начиная от их формулирования исходных задач соответствующей предметной области на языке прикладной статистики, выбора методов решения и критериев качества полученных решений и заканчивая формулировкой полученных выводов на языке предметной области.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к обязательной части образовательной программы. Дисциплина входит в модуль «Общепрофессиональные дисциплины».

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Первый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования: «Введение в интеллектуальный анализ данных».

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 16 ч.

-лабораторные: 16 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Раздел 1. Введение в статистический анализ и первичная статистическая обработка

1.1. Задачи и этапы статистического анализа

1.2. Типы и структуры данных

1.3. Предварительная обработка данных

Раздел 2. Критерии сравнения групп

2.1. Параметрические критерии

2.2. Непараметрические критерии

Раздел 3. Корреляционный анализ

3.1. Количественная корреляция

3.2. Ранговая корреляция

3.3. Корреляционный анализ количественных данных

Раздел 4. Регрессионный анализ

4.1. Парная регрессия

4.2. Множественная регрессия

Раздел 5. Дисперсионный анализ

5.1. Однофакторный дисперсионный анализ

5.2. Двухфакторный дисперсионный анализ

Раздел 6. Задачи классификации и кластеризации

6.1. Методы классификации

6.2. Методы кластеризации

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине осуществляется на основании проверки лабораторных работ, выполняемых студентами на компьютерах в течение семестра. Каждая работа оценивается на «зачет»/ «незачет». В течение семестра должны быть в срок выполнены все предложенные работы.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Промежуточная аттестация проходит в виде экзамена. Экзаменационная оценка складывается из текущего посещения (не менее 80% занятий), в срок выполненных лабораторных работ и результатов тестирования (при онлайн обучении) или письменного коллоквиума по нижепредставленным темам (при оффлайн форме).

Типовые вопросы для проведения промежуточной аттестации по дисциплине.

1. Предварительная обработка данных. Обработка пропущенных значений и выбросов.
2. Критерии проверки нормальности.
3. Параметрические критерии сравнения выборок.
4. Непараметрические критерии сравнения выборок.
5. Общая постановка МНК-оценивания параметров линейной регрессии. Оценивание дисперсии ошибок.
6. Свойства МНК-оценок параметров линейной регрессии.
7. Обобщение оценок параметров линейной регрессии для случая коррелированных гомоскедастичных наблюдений.
8. Обобщение оценок параметров линейной регрессии для случая коррелированных гетероскедастичных наблюдений.
9. Оценки параметров линейной регрессии при связывающих эти параметры ограничениях.
10. Нелинейные модели регрессии, допускающие линеаризацию. Проверка гипотезы об адекватности модели регрессии.
11. Итерационные алгоритмы оценивания параметров регрессии.
12. Доверительные интервалы для параметров регрессии. Интервалы предсказания.
13. Коэффициенты детерминации и парной корреляции, корреляционное отношение: определения и свойства.

14. Частный и множественный коэффициенты корреляции: определения и свойства.
15. Понятие ранговой корреляции. Основные типы задач анализа ранговых связей.
16. Коэффициенты ранговой корреляции Кендалла и Спирмена. Обобщенный коэффициент ранговой корреляции.
17. Проверка гипотезы о статистически зависимой ранговой связи.
18. Коэффициент конкордации и его свойства.
19. Категоризованные данные. Анализ зависимости признаков по таблицам сопряженности.
20. Общая постановка задачи дисперсионного анализа.
21. Однофакторный дисперсионный анализ. Проверка гипотезы о влиянии фактора на исследуемый объект.
22. Исследование влияния на объект уровней фактора методами множественного сравнения.
23. Общее решение задачи двухфакторного дисперсионного анализа.
24. Двухфакторный дисперсионный анализ с равным числом $r > 1$ наблюдений в ячейке.
25. Двухфакторный дисперсионный анализ с неравным числом наблюдений в ячейке.
26. Неполные сбалансированные блоки в задачах дисперсионного анализа.
27. Решение задачи трехфакторного дисперсионного анализа.
28. Общая постановка задачи дискриминантного анализа.
29. Решение задачи параметрического дискриминантного анализа. Расщепление смесей распределений.
30. Типы расстояний и мер близости между объектами и между классами.
31. Типы функционалов качества разбиения множества объектов на классы.
32. Основные типы кластер-процедур.
33. Стационарные временные ряды и их основные характеристики. Проверка гипотезы о случайности.
34. Сглаживание временного ряда методом простого скользящего среднего.
35. Сглаживание временного ряда методом взвешенного скользящего среднего. Выбор порядка аппроксимирующего полинома.
36. Точные и интервальные оценки прогноза будущих значений временного ряда.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle»

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Джеймс Г., Уиттон Д., Хасты Е., Тибширани Р Введение в статистическое обучение с примерами на языке R. – М.: ДМК Пресс, 2016. – 450 с.

– Кабанова Т. В. Применение пакета R для решения задач прикладной статистики: учебное пособие: [для студентов и аспирантов университетов]. – Томск: Издательский Дом Томского государственного университета, 2019. – 124 с.

– Марголис Н. Ю., Кабанова Т. В. Прикладная статистика: учебно-методическое пособие. Ч. 1. – Том. гос. ун-т, 2007. – 46 с.

– Марголис Н. Ю., Кабанова Т. В. Прикладная статистика: учебно-методическое пособие. Ч. 2. – Том. гос. ун-т, 2007. – 58 с.

б) дополнительная литература:

– М. Кендалл,

- А. Стьюарт. Статистические выводы и связи. – Наука. Физматлит, 1973. – 432 с.
- С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. Прикладная статистика. Классификация и снижение размерности. – Финансы и статистика, 1989. – 608 с.
- Айвазян С. А, Мхитарян В. С. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 1. – ЮНИТИ-ДАНА, 2001. – 270 с.
- Айвазян С. А, Мхитарян В. С. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 2. – ЮНИТИ-ДАНА, 2001. – 432 с.

в) ресурсы сети Интернет:

- <http://statsoft.ru/#tab-STATISTICA-link>
- <https://www.r-project.org/>
- https://rstudio-pubstatic.s3.amazonaws.com/33285_9ffa5341065d4bff9607fd6ccbba75d98.html#-
- <http://www-01.ibm.com/software/ru/analytics/spss/index.html>
- <http://itmu.vsuet.ru/Posobija/MathCAD/g113/index.htm#anc1323>
- <http://www.exponenta.ru/>

13. Перечень информационных технологий

- а) лицензионное и свободно распространяемое программное обеспечение:
- MS Windows, MS Office, Mathcad, Statistica, R, R Studio.

14. Материально-техническое обеспечение

Для реализации дисциплины необходимы лекционные аудитории и аудитории для проведения лабораторных занятий. Специальные технические средства (проектор, компьютер и т.д.) требуются для демонстрации материала в рамках изучаемых разделов. Вся основная и дополнительная литература, необходимая для самостоятельной работы и подготовки к экзамену, имеется в научной библиотеке ТГУ.

15. Информация о разработчиках

Кабанова Татьяна Валерьевна, канд. физ.-мат. наук, доцент кафедры теории вероятностей и математической статистики ИПМКН ТГУ.