

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

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:
Директор
А. В. Замятин

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
2024

СОГЛАСОВАНО:
Руководитель ОП
А.В. Замятин

Председатель УМК
С.П. Сущенко

Томск – 2024

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте;.

ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями;.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.3 Развивает и применяет математические, естественнонаучные, социально-экономические и профессиональные знания для решения задач

ИОПК-3.1 Осуществляет сбор, обработку и анализ научно-технической информации, необходимой для решения профессиональных задач

ИОПК-3.2 Умеет работать с различными видами информации с помощью различных средств информационных и коммуникационных технологий

ИОПК-3.3 Формулирует результаты, полученные в ходе решения исследовательских задач, в виде аналитических обзоров с обоснованными выводами и рекомендациями

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.
- ~ Изучить основные методы статистического анализа данных.
- ~ Сформировать навыки работы в программах статистической обработки данных.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплины (модули)».

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.

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

Тема 1. Введение в статистический анализ.

Типы данных. Графические и табличные способы представления данных.
Предварительная обработка данных.

Тема 2 .Критерии сравнения групп.

Параметрические критерии. t-критерий Стьюдента. Критерий Фишера. Дисперсионный анализ. Непараметрические критерии. Критерии Манна-Уитни, Вилкоксона, Краскала-Уолиса, Фридмана.

Тема 3. Корреляционный анализ.

Парный коэффициент корреляции Пирсона. Z-преобразование Фишера. Корреляционный анализ, Ранговая корреляция. Коэффициент Спирмена, Кендалла, конкордации Кендалла. Корреляционный анализ категоризованных данных.

Тема 4. Парная регрессия.

Определение простой регрессии. Метод наименьших квадратов оценки параметров простой регрессии. Условия Гаусса-Маркова. Теорема Гаусса-Маркова. Оценки дисперсий. Проверка качества модели регрессии, Коэффициент детерминации, его интерпретация, общая адекватность модели. Нелинейные модели и линеаризация.

Тема 5. Множественная регрессия.

Основные понятия и задачи регрессионного анализа, Общая постановка задачи множественной регрессии. Метод наименьших квадратов оценки параметров регрессии. Теорема Гаусса-Маркова. Оценка дисперсий. Проверка качества модели множественной регрессии. Фиктивные переменные. Случай коррелированных наблюдений Гетероскедастичность. Мультиколлинеарность.

Тема 6. Задача классификации.

Основные понятия и задачи классификации. Бинарная классификация и логистическая регрессия. Метрики качества. ROC-анализ.

Тема 7. Кластерный анализ.

Основные подходы в задачах кластеризации. Итерационные, плотностные, иерархические алгоритмы. Расстояния между объектами Расстояния между классами. Проверка качества кластеризации.

Тема 8. Анализ временных рядов.

Понятие временного ряда, основные модели временных рядов, задачи анализа временных рядов. Декомпозиция временных рядов. Прогнозирование во временных рядах.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине осуществляется на основании проверки практических заданий, выполняемых студентами на компьютерах в течение семестра и фиксируется в форме контрольной точки не менее одного раза в семестр.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в первом семестре проводится форме теста из 20 вопросов. Продолжительность экзамена 45 минут.

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

11. Учебно-методическое обеспечение

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

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

в) План практических занятий по дисциплине.

г) Методические указания по проведению лабораторных работ.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- ☐ 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.
- ☐ Мыльников Л.А. Статистические методы интеллектуального анализа данных. – СПб.: БХВ-Петербург, 2021. – 240 с.
- ☐ Дж. Д. Лонг, Пол Титор. Р. Книга рецептов: Проверенные рецепты для статистики, анализа и визуализации данных / пер. с англ. Д. А. Беликова. – М.: ДМК Пресс, 2020. – 510 с.

б) дополнительная литература:

- ☐ Кендалл М., Стьюарт А. Статистические выводы и связи. Наука. Физматлит. 1973. 432 с.
- ☐ Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. 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. Прикладная статистика. Основы эконометрики: Учебник для экономических специальностей вузов: В 2 т. Т. 2. ЮНИТИ-ДАНА. 2001, 432 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.
- ☐ Джеймс Г., Уиттон Д., Хастис Е., Тибширани Р. Введение в статистическое обучение с примерами на языке R. – М.: ДМК Пресс, 2016. 450 с.

в) ресурсы сети Интернет:

- ❑ Статистические методы машинного обучения
<https://moodle.ido.tsu.ru/course/view.php?id=1419>

- ❑ Статистика в Data Science – исчерпывающий гид для амбициозных практиков ML <https://habr.com/ru/company/skillfactory/blog/526972/>

- ❑ Введение в Data Science и машинное обучение <https://stepik.org/course/4852>
10 примеров использования статистических методов в проекте машинного обучения <https://www.machinelearningmastery.ru/statistical-methods-in-an-applied-machine-learning-project/>

- ❑ Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных www.machinelearning.ru/

- ❑

❑ 13. Перечень информационных технологий

- а) лицензионное и свободно распространяемое программное обеспечение:
- ~ Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- ~ публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.)
- ~ R The R Foundation, США свободно распространяемое.
- ~ RStudio RStudio, PBC, США свободно распространяемое.
- ~ JASP Амстердамский университет, Нидерланды свободно распространяемое.

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ –
<http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ –
<http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юпайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных:

- ❑ Искусственный интеллект и сферы его применения. Новости разработки квантовых компьютеров. Исследования искусственных нейронных сетей. <https://ai-news.ru>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Лаборатории, оборудованные персональными компьютерами, соответствующим необходимым программным обеспечением, выходом в интернет.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

15. Информация о разработчиках

Кабанова Татьяна Валерьевна, кандидат физ.-мат. наук, доцент, кафедра ТВиМС
ИПМКН ТГУ.