

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт математики и механики им. Н.И. Лобачевского



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

_____ Турилова Е.А.

"__" _____ 20__ г.

Программа дисциплины

Статистический анализ данных

Направление подготовки: 02.04.01 - Математика и компьютерные науки

Профиль подготовки: Статистические методы науки о данных

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2023

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): научный сотрудник, к.н. Заикин А.А. (НИЛ изучения состояния и эволюции подземных резервуаров, Научный центр мирового уровня Рациональное освоение запасов жидких углеводородов планеты (головной центр)), Kaskrin@gmail.com ; доцент, к.н. (доцент) Симушкин С.В. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), Sergey.Simushkin@kpfu.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- методы выбора оптимального статистического решения при сравнении групп по данным
- современные программные средства обработки статистических данных, основные критерии сравнения не-скольких групп данных

Должен уметь:

- формулировать и решать задачи статистического сравнения,
- находить адекватные процедуры
- составлять программы подготовки данных и их статистического анализа

Должен владеть:

- компьютерными средствами обработки данных
- основными программными средствами, применяемыми при анализе однородности

Должен демонстрировать способность и готовность:

- к применению процедур статистического сравнения
- обоснования применяемой методики

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.О.05 Дисциплины (модули)" основной профессиональной образовательной программы 02.04.01 "Математика и компьютерные науки (Статистические методы науки о данных)" и относится к обязательной части ОПОП ВО.

Осваивается на 1 курсе в 1 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 5 зачетных(ые) единиц(ы) на 180 часа(ов).

Контактная работа - 64 часа(ов), в том числе лекции - 0 часа(ов), практические занятия - 64 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 116 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 1 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

| N | Разделы дисциплины / модуля | Се-местр | Виды и часы контактной работы, их трудоемкость (в часах) | | | | | | Само-стоя-тельная ра-бота |
|----|--|----------|--|--------------------|------------------------------|---------------------------|-----------------------------|---------------------------|---------------------------|
| | | | Лекции, всего | Лекции в эл. форме | Практи-ческие занятия, всего | Практи-ческие в эл. форме | Лабора-торные работы, всего | Лабора-торные в эл. форме | |
| 1. | Тема 1. Критерии сравнения групп по качественным данным - критерии | | | | | | | | |

сопряжённости

| | | | | | | | |
|---|---|---|----|---|---|---|----|
| 1 | 0 | 0 | 16 | 0 | 0 | 0 | 30 |
|---|---|---|----|---|---|---|----|

| N | Разделы дисциплины / модуля | Се- местр | Виды и часы контактной работы, их трудоемкость (в часах) | | | | | | Само- стоя- тель- ная ра- бота |
|----|--|--------------|--|--------------------------|--|-------------------------------------|---------------------------------------|-------------------------------------|---|
| | | | Лекции, всего | Лекции в эл. форме | Практи- ческие занятия, всего | Практи- ческие в эл. форме | Лабора- торные работы, всего | Лабора- торные в эл. форме | |
| 2. | Тема 2. Критерии сравнения групп по количественным данным - дисперсионный анализ | 1 | 0 | 0 | 16 | 0 | 0 | 0 | 30 |
| 3. | Тема 3. Дискриминантный анализ - дискриминантные функции | 1 | 0 | 0 | 16 | 0 | 0 | 0 | 30 |
| 4. | Тема 4. Логит- и пробит-анализ | 1 | 0 | 0 | 16 | 0 | 0 | 0 | 26 |
| | Итого | | 0 | 0 | 64 | 0 | 0 | 0 | 116 |

4.2 Содержание дисциплины (модуля)

Тема 1. Критерии сравнения групп по качественным данным - критерии сопряжённости

Теоретические предпосылки применения статистических критериев проверки гипотезы о совпадении двух и более групп (асимптотический критерий хи-квадрат, точный критерий Фишера, таблицы сопряжённости). Реализации критериев сравнения групп в пакете программ R. Интерпретация результатов и графическое представление

Тема 2. Критерии сравнения групп по количественным данным - дисперсионный анализ

Теоретические предпосылки применения статистических критериев проверки гипотезы о совпадении двух и более групп (асимптотический критерий хи-квадрат, точный критерий Фишера, таблицы сопряжённости). Реализации критериев сравнения групп в пакете программ R. Интерпретация результатов и графическое представление

Тема 3. Дискриминантный анализ - дискриминантные функции

Задача дискриминантного анализа. Модель линейного дискриминантного анализа. Расстояния между группами данных - Евклида, Хотеллинга, Махаланобиса. Задача сокращения признаков через поиск максимума расстояния между группами. Выбор наиболее информативных признаков. Критерий отнесения данных в какую-либо группу по малому числу дискриминантных признаков

Тема 4. Логит- и пробит-анализ

Задача дискриминации для переменных отклика категориального типа. Логит-модель для зависимости альтернативных данных от значений регрессоров. Оценки максимального правдоподобия и их асимптотические свойства. Пробит-модель зависимости альтернативных данных от предикторных переменных. Оценки параметров модели. Логит-модель для категориальных данных с числом категорий, больше двух.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Естественно-научный портал - <http://en.edu.ru/>

Научная электронная библиотека - <https://elibrary.ru/defaultx.asp/>

Цифровой образовательный ресурс "Теория вероятностей" - <https://stepik.org/course/3089/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

| Вид работ | Методические рекомендации |
|------------------------|---|
| практические занятия | Практические занятия подразумевают работу с пакетами программ R или Python. Следует обратить внимание на порядок обращения к соответствующим процедурам многомерного анализа - построение регрессии, методам дисперсионного анализа. Особенно важно освоить графические процедуры и их интерпретацию. Начинать работу с данными следует с продумывания способов организация хранения |
| самостоятельная работа | Самостоятельная работа предполагает изучение дополнительной литературы. Здесь необходимо в обязательном порядке выделять ключевые моменты, которые в последствии на лекции обсудить с лектором и с сокурсниками. Кроме того, следует обратить внимание на новомодные тенденции в представлении данных для отчёта и интерпретации. Поскольку в интернет-литературе много мусора, то относится к ней нужно с большой долей критики. |
| зачет | При подготовке к зачёту следует больше внимание обратить на теоретические предпосылки применяемых процедур. Не обязательно знать досконально доказательства их корректности, но схему этого доказательства лучше иметь в виду. На зачёте каждому студенту будет предложена своя задача статистического анализа реальных данных. Окончательное решение по этим данным должно быть тщательно выверено в соответствии с изучаемыми статистическими методами. Обработка данных должна быть реализована в пакетах программ R или Python. |

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Компьютерный класс.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;

- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 02.04.01 "Математика и компьютерные науки" и магистерской программе "Статистические методы науки о данных".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.О.05 Статистический анализ данных

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 02.04.01 - Математика и компьютерные науки

Профиль подготовки: Статистические методы науки о данных

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2023

Основная литература:

1. Свешников, А. А. Прикладные методы теории вероятностей : учебник / А. А. Свешников. - Санкт-Петербург : Лань, 2022. - 480 с. - ISBN 978-5-8114-1219-8. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/210821> (дата обращения: 22.01.2022). - Режим доступа: для авториз. пользователей.
2. 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> (дата обращения: 28.03.2022).- Режим доступа: для авториз. пользователей.
3. Мельниченко, А. С. Математическая статистика и анализ данных : учебное пособие / А. С. Мельниченко. - Москва : МИСИС, 2018. - 45 с. - ISBN 978-5-906953-62-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/108035> (дата обращения: 23.03.2022).- Режим доступа: для авториз. пользователей.

Дополнительная литература:

1. Бородин, А. Н. Элементарный курс теории вероятностей и математической статистики : учебное пособие / А. Н. Бородин. - 8-е изд., стер. - Санкт-Петербург : Лань, 2022. - 256 с. - ISBN 978-5-8114-0442-1. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/210677> (дата обращения: 28.03.2022).- Режим доступа: для авториз. пользователей.
2. Соколов, Г. А. Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике : учебное пособие / Г.А. Соколов, Р.В. Сагитов. - Москва : ИНФРА-М, 2022. - 202 с. - (Высшее образование: Магистратура). - ISBN 978-5-16-003646-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1836638> (дата обращения: 19.02.2022). - Режим доступа: по подписке.
3. Мельниченко, А. С. Математическая статистика и анализ данных : учебное пособие / А. С. Мельниченко. - Москва : МИСИС, 2018. - 45 с. - ISBN 978-5-906953-62-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/108035> (дата обращения: 19.02.2022). - Режим доступа: для авториз. пользователей.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.О.05 Статистический анализ данных

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 02.04.01 - Математика и компьютерные науки

Профиль подготовки: Статистические методы науки о данных

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2023

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows