

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

Проверка гипотез и множественное тестирование

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) Кареев И.А. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), IAKareev@kpfu.ru ; доцент, к.н. Салимов Р.Ф. (кафедра математической статистики, Институт математики и механики им. Н.И. Лобачевского), Rustem.Salimov@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:

Основы статистики: понимание понятий вероятности, выборки, среднего, дисперсии, стандартного отклонения и т.д.

Гипотезы и их проверка: понимание основных понятий, таких как нулевая и альтернативная гипотезы, уровень значимости, критическая область, р-значение и т.д.

Различные статистические тесты: знание различных тестов, таких как t-тесты (одновыборочный, двухвыборочный, зависимый), анализ дисперсии (ANOVA), корреляционный анализ и др.

Множественное тестирование: понимание проблем множественного сравнения и методов коррекции уровня значимости, таких как метод Бонферрони, метод Холма и др.

Должен уметь:

Использовать статистических программ: умение работать с программами для статистического анализа данных, такими как R, SPSS, Excel и др. в задачах множественного тестирования

Должен владеть:

Навыками коммуникации: умение ясно и точно представлять результаты статистического анализа и объяснять их значение для конкретной проблемы или исследования.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.03.02 Дисциплины (модули)" основной профессиональной образовательной программы 02.04.01 "Математика и компьютерные науки (Статистические методы науки о данных)" и относится к дисциплинам по выбору части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2 курсе в 3 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 4 зачетных(ые) единиц(ы) на 144 часа(ов).

Контактная работа - 34 часа(ов), в том числе лекции - 0 часа(ов), практические занятия - 34 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 74 часа(ов).

Контроль (зачёт / экзамен) - 36 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 3 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- 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	
1.	Тема 1. Введение в проверку гипотез и								

множественное тестирование

3	0	0	11	0	0	0	24
---	---	---	----	---	---	---	----

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. Обзор процедур множественного тестирования	3	0	0	11	0	0	0	25
3.	Тема 3. Применение к задачам генетики	3	0	0	12	0	0	0	25
	Итого		0	0	34	0	0	0	74

4.2 Содержание дисциплины (модуля)

Тема 1. Введение в проверку гипотез и множественное тестирование

общая статистическая основа для множественной проверки гипотез, основные составляющие проблемы множественной проверки, включая: распределение, генерирующее данные; интересующие параметры; нулевая и альтернативная гипотезы; статистика испытаний; несколько процедур тестирования; области отбраковки для тестовой статистики; ошибки при проверке множественных гипотез: ошибки типа I, типа II и типа III; Частота ошибок типа I; мощность, нулевое распределение тестовой статистики, используемое для получения областей отклонения для тестовой статистики, областей достоверности для интересующих параметров и скорректированных р-значений,

Тема 2. Обзор процедур множественного тестирования

Контроль ошибок типа I и выбор нулевого распределения тестовой статистики, Процедуры предельного множественного тестирования, Совместные процедуры множественного тестирования, Процедуры множественного тестирования для контроля количества ошибок типа I: FWER, Контроль количества ошибок типа I, FWER-контролирующий одиночные пошаговые процедуры, пошаговые процедуры, контролирующие FWER, пошаговые процедуры, контролирующие FWER, множественные процедуры тестирования для контроля количества ошибок I рода: gFWER, gFWER-контролирующие одношаговые и понижающие процедуры Лемана и Романо, gFWER-контролирующие одношаговые процедуры общего отсека и общего квантиля, gFWER-контролирующие процедуры многократного тестирования увеличения, gFWER-контролирующие эмпирические байесовские процедуры на основе повторной выборки, Другие gFWER-контролирующие процедуры, Сравнение контролирующих gFWER процедур, Множественное тестирование процедуры контроля доли ошибок I рода среди отвергнутых гипотез: FDR, Контроль количества и доли ошибок I рода

ошибки, усиление контроля FDR Бенджамини и Хохберга

Процедура, FDR-контролирующая пошаговая процедура Бенджамини и Йекутиэли, FDR-контролирующая основанная на повторной выборке эмпирическая процедура Байеса, Другие FDR-контролирующие процедуры, Множественные процедуры тестирования для контроля доли ошибок типа I среди отвергнутых гипотез: TPPFP, Контроль ожидаемого значение по сравнению с вероятностью хвоста для доли ошибок типа I, понижающие процедуры Лемана и Романо, контролирующие TPPFP, процедуры многократного тестирования с увеличением, контролирующие TPPFP, эмпирические процедуры Байеса, основанные на повторной выборке, контролирующие TPPFP, сравнение процедур, контролирующих TPPFP.

Тема 3. Применение к задачам генетики

Мотивация, Сравнение с другими подходами, Схема, Статистическая основа для множественных тестов ассоциации с метаданными биологических аннотаций, Профили генов аннотаций, Профили генов параметров, Меры ассоциации для профилей генов аннотаций и генов параметров, Проверка множественных гипотез, Генная онтология, Обзор онтологии генов, Обзор программного обеспечения R и Bioconductor для анализа метагенных аннотаций GO, Пакет метагенных аннотаций GO, Пакеты метагенных аннотаций для конкретных чипов Affymetrix: пакет hgu95av2, Сборка матрицы аннотаций генов GO, Тесты ассоциации между GO аннотация и дифференциальная экспрессия генов при ALL, Исследование острого лимфобластного лейкоза, проведенное Chiarotti et al., Система проверки множественных гипотез

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)

Multiple Hypothesis Testing - <https://multithreaded.stitchfix.com/blog/2015/10/15/multiple-hypothesis-testing/>

Multiple Testing - <https://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf>

Multiple testing in adaptive designs - A review -

https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0CAIQw7AJahcKEwjw-uXoub_AhUA

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

Вид работ	Методические рекомендации
практические занятия	В ходе прохождения цикла занятий практического типа по дисциплине обучающемуся слушателю для лучшего и полноценного усвоения осваиваемого материала и теории необходимо усердно и с инициативным рвением выполнять все задания для выполнения на практических занятиях, анализировать соответствие выполненных работ с заданием и теорией.
самостоятельная работа	В ходе выполнения цикла самостоятельных работ по дисциплине обучающемуся слушателю курса рекомендуется с целью лучшего и более полного усвоения осваиваемого материала и теории выполнять все работы для домашнего исполнения, изучать дополнительную литературу, формулировать вопросы на не полностью освоенные части курса.
экзамен	В ходе подготовки к экзамену по дисциплине обучающемуся слушателю курса рекомендуется с целью повышения его возможностей по успешному прохождению экзамена повторить весь ранее изученный материал, как теоретического характера, так и практические и самостоятельные работы, определить возможные проблемные места усвоения материала и провести дополнительные образовательные действия для разрешения выявленных ранее проблемных и неосвоенных участков курса.

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.

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.В.ДВ.03.02 Проверка гипотез и множественное тестирование

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:

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: 28.06.2023). - Режим доступа: для авториз. пользователей.

Лагутин, М. Б. Наглядная математическая статистика : учебное пособие / М. Б. Лагутин. - 9-е изд. (эл.). - Москва : Лаборатория знаний, 2023. - 475 с. - ISBN 978-5-93208-651-3. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/319217> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.

Хуснутдинов, Р. Ш. Сборник задач по курсу теории вероятностей и математической статистики : учебное пособие / Р. Ш. Хуснутдинов. - 2-е изд., испр. - Санкт-Петербург : Лань, 2022. - 320 с. - ISBN 978-5-8114-1668-4. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/211733> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.

Дополнительная литература:

Дерр, В. Я. Теория вероятностей и математическая статистика : учебное пособие для вузов / В. Я. Дерр. - Санкт-Петербург : Лань, 2021. - 596 с. - ISBN 978-5-8114-6515-6. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/159475> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.

Пугачев, В. С. Теория вероятностей и математическая статистика : учебное пособие / В. С. Пугачев. - 2-е изд. - Москва : ФИЗМАТЛИТ, 2002. - 496 с. - ISBN 5-9221-0254-0. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/48170> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.

Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников : учебное пособие / А. И. Кобзарь. - 2-е изд. - Москва : ФИЗМАТЛИТ, 2012. - 816 с. - ISBN 978-5-9221-1375-5. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/59747> (дата обращения: 28.06.2023). - Режим доступа: для авториз. пользователей.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.03.02 Проверка гипотез и множественное тестирование

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