

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
"Казанский (Приволжский) федеральный университет"  
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

Проф. Д. А. Таюрский



» \_\_\_\_\_ 20\_\_ г.

*подписано электронно-цифровой подписью*

## Программа дисциплины

Статистический анализ данных

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2018

## 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) Кашина О.А. ; заведующий кафедрой, д.н. (доцент) Миссаров М.Д. (кафедра анализа данных и исследования операций, отделение фундаментальной информатики и информационных технологий), Moukadas.Missarov@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:

| Шифр компетенции | Расшифровка приобретаемой компетенции  |
|------------------|--|
| ПК-1             | Способен разрабатывать и применять математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности в задачах анализа данных и машинного обучения |
| ПК-4             | Способен строить математические модели и анализировать данные, обосновывать и выбирать решение в прикладных задачах  |

Обучающийся, освоивший дисциплину (модуль):

Должен демонстрировать способность и готовность:

- применять теоретические методы статистического анализа данных;

- использовать прикладное программное обеспечение (пакет R) для моделирования данных, для визуализации и интерпретации выборочных данных и их обработки методами статистического анализа;

- анализировать результаты решения прикладных задач статистического анализа, делать содержательные выводы.

### 2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.03 Дисциплины (модули)" основной профессиональной образовательной программы 01.04.02 "Прикладная математика и информатика (Анализ данных и его приложения)" и относится к вариативной части.

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 4 зачетных(ые) единиц(ы) на 144 часа(ов).

Контактная работа - 40 часа(ов), в том числе лекции - 16 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 24 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 104 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   | Семестр | Виды и часы контактной работы, их трудоемкость (в часах) |                      |                     | Самостоятельная работа |
|----|---|---------|--|----------------------|---------------------|------------------------|
|    |   |         | Лекции   | Практические занятия | Лабораторные работы |                        |
| 1. | Тема 1. Вероятностно-статистические модели. Точечное оценивание. Моделирование и визуализация данных в среде R. | 1       | 2  | 0                    | 4                   | 18                     |
| 2. | Тема 2. Доверительное оценивание. Deskриптивные статистики. Решение задач в среде R.                            | 1       | 4  | 0                    | 6                   | 16                     |
| 3. | Тема 3. Проверка гипотез о параметрах одной выборки. Расчёты в среде R.   | 1       | 4  | 0                    | 6                   | 20                     |
| 4. | Тема 4. Критерии согласия.  | 1       | 2  | 0                    | 0                   | 16                     |
| 5. | Тема 5. Сравнительный анализ двух выборок. Решение задач в среде R.   | 1       | 2  | 0                    | 4                   | 18                     |
| 6. | Тема 6. Дисперсионный анализ. Решение задач в среде R.  | 1       | 2  | 0                    | 4                   | 16                     |
|    | Итого   |         | 16   | 0                    | 24                  | 104                    |

##### 4.2 Содержание дисциплины (модуля)

###### Тема 1. Вероятностно-статистические модели. Точечное оценивание. Моделирование и визуализация данных в среде R.

Вероятностно-статистические модели. Задача точечного оценивания.

Свойства статистических оценок: несмещённость, эффективность, состоятельность. Методы нахождения точечных оценок неизвестных параметров распределений: метод максимального правдоподобия, метод моментов. Оценивание параметров нормального распределения методом максимального правдоподобия и статистические свойства полученных оценок.

###### Тема 2. Доверительное оценивание. Deskриптивные статистики. Решение задач в среде R.

Определение доверительного интервала. Доверительный интервал для среднего нормального распределения при известной дисперсии. Доверительный интервал для дисперсии нормального распределения при известном среднем. Распределение хи-квадрат. Доверительный интервал для среднего нормального распределения при неизвестной дисперсии. Распределение Стьюдента. Доверительный интервал для дисперсии нормального распределения при неизвестном среднем. Построение доверительного интервала для среднего в случае больших выборок. Построение доверительного интервала для "доли" в случае больших выборок.

###### Тема 3. Проверка гипотез о параметрах одной выборки. Расчёты в среде R.

Основные типы гипотез и общая логическая схема статистического критерия.

Ошибки I и II рода, размер и мощность критической области. P-значение. Проверка гипотезы о среднем значении нормального распределения при известной дисперсии. Проверка гипотезы о среднем значении нормального распределения при неизвестной дисперсии. Проверка гипотезы о доле. Проверка гипотез о среднем и о доле в случае больших выборок.

###### Тема 4. Критерии согласия.

Критерий хи-квадрат Пирсона. Доказательство критерия хи-квадрат в случае двух классов. Применение критерия Хи-квадрат Пирсона для проверки гипотезы о законе распределения вероятностей случайной величины. Проверка гипотезы об отсутствии сопряженности признаков с качественной группировкой по критерию Хи-квадрат Пирсона. Эмпирическая функция распределения. Критерий Колмогорова.

###### Тема 5. Сравнительный анализ двух выборок. Решение задач в среде R.

Проверка гипотез о средних двух независимых нормальных выборок с известными дисперсиями. Проверка гипотез о средних двух независимых нормальных выборок с неизвестными дисперсиями. Проверка гипотезы о равенстве дисперсий двух нормальных выборок. Распределение Фишера- Снедекора. Проверка гипотезы о равенстве долей. Проверка гипотез в случае парных данных.

###### Тема 6. Дисперсионный анализ. Решение задач в среде R.

Основные понятия дисперсионного анализа. Однофакторный и двухфакторный дисперсионный анализ. Условия применения метода дисперсионного анализа. для сравнения математических ожиданий генеральных совокупностей (нормальность распределения, равенство дисперсий). Сравнение средних генеральных совокупностей в случае двух групп, трёх и более групп.

#### **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 образования и науки Российской Федерации от 5 апреля 2017 года №301)

Письмо Министерства образования Российской Федерации №14-55-996ин/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

CRAN - The Comprehensive R Archive Network - <https://cran.r-project.org/>

Machine Learning Repository - <http://archive.ics.uci.edu/ml/index.php>

Электронный курс "Статистический анализ данных" - <https://edu.kpfu.ru/course/view.php?id=835>

#### **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)

CRAN - The Comprehensive R Archive Network - <https://cran.r-project.org/>

Machine Learning Repository - <http://archive.ics.uci.edu/ml/index.php>

Электронный курс "Статистический анализ данных" - <https://edu.kpfu.ru/course/view.php?id=835>

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

| Вид работ              | Методические рекомендации  |
|------------------------|--|
| лекции                 | Лекции читаются в аудитории, в традиционном формате. Приветствуется активное участие студентов в лекционном занятии: ответы на вопросы преподавателя при устном опросе, обсуждение сложных моментов. Допускается использование на лекции мобильных устройств для использования материалов электронного курса "Статистический анализ данных" (авторы: М.Д.Миссаров, О.А.Кашина) в системе дистанционного обучения КФУ.  |
| лабораторные работы    | Задания на выполнение лабораторных работ, правила выполнения лабораторных работ и правила оформления и порядок представления отчётов о выполнении лабораторных работ представлены в электронного курса "Статистический анализ данных" (авторы: М.Д.Миссаров, О.А.Кашина) в системе дистанционного обучения КФУ.  |
| самостоятельная работа | Самостоятельная работа студентов включает:<br>- изучение текстов лекций, материалов из электронного курса и предлагаемых в нём источников;<br>- ознакомление с содержанием лабораторных работ;<br>- поиск реальных данных в предлагаемых репозиториях и открытых Интернет-источниках;<br>- установка, запуск, настройка пакета R;<br>- выполнение заданий на лабораторные работы (в случае нехватки времени на их выполнение в компьютерном классе);<br>- оформление отчётов;<br>- участие в форуме в электронном курсе. |
| зачет                  | Обязательным условием для получения зачёта является выполнение всех лабораторных работ и получение общей суммы баллов в размере, составляющем не менее 56% от максимально возможной. В случае выполнения данного условия на момент проведения зачёта студент может быть освобождён от необходимости ответа на теоретические вопросы.   |

### 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 01.04.02 "Прикладная математика и информатика" и магистерской программе "Анализ данных и его приложения".

Приложение 2  
к рабочей программе дисциплины (модуля)  
Б1.В.03 Статистический анализ данных

**Перечень литературы, необходимой для освоения дисциплины (модуля)**

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2018

**Основная литература:**

Бородин, А.Н. Элементарный курс теории вероятностей и математической статистики [Электронный ресурс] : учебное пособие / А.Н. Бородин. ? Электрон. дан.

? Санкт-Петербург : Лань, 2011. ? 256 с. ? Режим доступа: <https://e.lanbook.com/book/2026>

Ширяев, А.Н. Задачи по теории вероятностей [Электронный ресурс] : учебное пособие / А.Н. Ширяев. ? Электрон. дан. ? Москва : МЦНМО, 2006. ? 416 с. ? Режим доступа: <https://e.lanbook.com/book/9447>

Статистический анализ данных, моделирование и исследование вероятностных закономерностей.

Компьютерный подход / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов и др. - М.: НИЦ ИНФРА-М, 2015. - 890 с. ISBN 978-5-16-103267-1 (online). - Режим доступа:

<http://znanium.com/bookread2.php?book=515227>

**Дополнительная литература:**

Боровков, А.А. Математическая статистика [Электронный ресурс] : учебник / А.А. Боровков. ? Электрон. дан. ? Санкт-Петербург : Лань, 2010. ? 704 с.

? Режим доступа: <https://e.lanbook.com/book/3810>

Свешников, А.А. Прикладные методы теории вероятностей [Электронный ресурс] : учебник / А.А. Свешников. ? Электрон. дан. ? Санкт-Петербург : Лань, 2012.

? 480 с. ? Режим доступа: <https://e.lanbook.com/book/3184>

Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистический анализ данных в MS Excel. - М.: ИНФРА-М, 2014. -320 с. - Режим доступа:

<http://znanium.com/bookread.php?book=429722>

Электронный курс 'Статистический анализ данных' // Миссаров М.Д., Кашина О.А. - Дата начала эксплуатации : 1 сентября 2013 года Режим доступа:

<http://edu.kpfu.ru/course/view.php?id=835>



Приложение 3  
к рабочей программе дисциплины (модуля)  
Б1.В.03 Статистический анализ данных

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

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2018

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

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.