

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



**УТВЕРЖДАЮ**

Проректор  
по образовательной деятельности КФУ  
Проф. Таюрский Д.А.

"\_\_" \_\_\_\_\_ 20\_\_ г.

**Программа дисциплины**

Статистические методы математической генетики Б1.В.ДВ.2

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Методы прикладной математической статистики

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Салимов Р.Ф. , Симушкин С.В.

**Рецензент(ы):**

Володин И.Н.

**СОГЛАСОВАНО:**

Заведующий(ая) кафедрой: Турилова Е. А.

Протокол заседания кафедры No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Регистрационный No

Казань  
2016

## Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) ассистент, б/с Салимов Р.Ф. кафедры математической статистики отделение прикладной математики и информатики , Rustem.Salimov@kpfu.ru ; доцент, к.н. (доцент) Симушкин С.В. кафедры математической статистики отделение прикладной математики и информатики , Sergey.Simushkin@kpfu.ru

### 1. Цели освоения дисциплины

получить современные знания в области статистических методов математической генетики; ознакомить с основными задачами и методами их решения;

### 2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.2 Дисциплины (модули)" основной образовательной программы 01.04.02 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 1 курсе, 2 семестр.

Курс "Статистические методы математической генетики" входит в число основных курсов основной образовательной программы высшего профессионального образования по направлению: 01.04.02 Прикладная математика и информатика.

В результате освоения предшествующих дисциплин студент должен:

знать:

основные понятия и методы теории вероятностей и математической статистики (базовые вероятностные распределения, теория случайных процессов);

уметь:

применять вероятностные законы к анализу реальных данных;

выбирать методику описания случайных экспериментов адекватно реальным данным;

владеть:

навыками использования математических справочников и таблиц;

### 3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1 (общекультурные компетенции)	способность к абстрактному мышлению, анализу, синтезу
ОК-2 (общекультурные компетенции)	готовность действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения
ОК-3 (общекультурные компетенции)	готовность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-4 (профессиональные компетенции)	способность использовать и применять углубленные знания в области прикладной математики и информатики
ПК-1 (профессиональные компетенции)	способность проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива
ПК-2 (профессиональные компетенции)	способность разрабатывать концептуальные и теоретические модели решаемых научных проблем и задач

В результате освоения дисциплины студент:

1. должен знать:

основные статистические методы в генетике;  
основные статистические модели в генетике;

2. должен уметь:

оценивать параметры для основных статистических моделей;  
с помощью построенных моделей строить оценки;

3. должен владеть:

навыками выбора методов построение статистических моделей;

способность и готовность применять на практике полученные знания в области статистических методов математической генетики

#### 4. Структура и содержание дисциплины/ модуля

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

Форма промежуточного контроля дисциплины экзамен во 2 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

#### 4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

##### Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Статистическая значимость в сравнении биологических последовательностей	2	1-3	3	3	0	письменное домашнее задание
2.	Тема 2. Байсовские методы в анализе биологических последовательностей	2	4-6	3	3	0	контрольная работа письменное домашнее задание

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
3.	Тема 3. Статистические подходы в предсказании эукариотических генов	2	7-10	4	4	0	письменное домашнее задание
4.	Тема 4. Анализ данных микрочипов	2	11-15	4	4	0	письменное домашнее задание
	Тема . Итоговая форма контроля	2		0	0	0	экзамен
	Итого			14	14	0	

## 4.2 Содержание дисциплины

### Тема 1. Статистическая значимость в сравнении биологических последовательностей

#### **лекционное занятие (3 часа(ов)):**

Статистическая значимость и биологическая значимость Оценка статистической значимости для локального сходства Использование статистических оценок

#### **практическое занятие (3 часа(ов)):**

Измерение сходства последовательностей Вычисление статистических оценок

### Тема 2. Байесовские методы в анализе биологических последовательностей

#### **лекционное занятие (3 часа(ов)):**

Обзор Байесовской методологии Скрытые марковские модели Парное выравнивание. Множественное выравнивание

#### **практическое занятие (3 часа(ов)):**

Построение парного Байесовского выравнивания Оценка параметров скрытой марковской цепи для выравнивания последовательностей Нахождение рекуррентных шаблонов в биологических последовательностях

### Тема 3. Статистические подходы в предсказании эукариотических генов

#### **лекционное занятие (4 часа(ов)):**

Структура эукариотических генов Методы функционального распознавание сигналов Линейный дискриминантный анализ Предсказание донора и акцептора границ сращивания Дискриминационные и вероятностные подходы для множественного предсказания генов

#### **практическое занятие (4 часа(ов)):**

Линейный дискриминантный анализ Предсказание генов с помощью SMM Предсказание генов с помощью шаблонов

### Тема 4. Анализ данных микрочипов

#### **лекционное занятие (4 часа(ов)):**

Представление данных и контроль качества Ошибки моделей, калибровка Исследование шаблонов Исследование связи фенотипа с реакцией на микрочипе Множественные испытания

#### **практическое занятие (4 часа(ов)):**

Двухгрупповое и k-групповое сравнение Уровень значимости для множественных испытаний

## 4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Статистическая значимость в сравнении биологических последовательностей	2	1-3	подготовка домашнего задания	25	домашнее задание
2.	Тема 2. Байсовские методы в анализе биологических последовательностей	2	4-6	подготовка домашнего задания	10	домашнее задание
				подготовка к контрольной работе	15	контрольная работа
3.	Тема 3. Статистические подходы в предсказании эукариотических генов	2	7-10	подготовка домашнего задания	33	домашнее задание
4.	Тема 4. Анализ данных микрочипов	2	11-15	подготовка домашнего задания	33	домашнее задание
	Итого				116	

## 5. Образовательные технологии, включая интерактивные формы обучения

Материал теоретического характера преподается в виде лекций у доски или в виде презентаций на мультимедийном экране. Предполагается использование диалоговой формы ведения занятий с постановкой и решением проблемных задач, обсуждением дискуссионных моментов.

Некоторые разделы курса студенты изучают самостоятельно по указанным методическим материалам или по Интернет-источникам с последующим докладом в виде презентаций и дискуссией.

При проведении практических занятий, на которых теоретический материал применяется к конкретным данным, студентам предлагается разработать самостоятельные вычислительные процедуры.

Перед каждым занятием, как лекционной, так и практической направленности проводится экспресс-опрос по пройденному теоретическому материалу.

## 6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

**Тема 1. Статистическая значимость в сравнении биологических последовательностей**  
домашнее задание , примерные вопросы:

Для двух биологических последовательностей оценить сходство. Построить статистические оценки. Оценить статистическую значимость.

**Тема 2. Байсовские методы в анализе биологических последовательностей**

домашнее задание , примерные вопросы:

Провести парное выравнивание с помощью байесовской методологии. Построить байесовские оценки для СММ Найти рекуррентные шаблоны в биологической последовательности

контрольная работа , примерные вопросы:

1. Однородность двух выборок 2. Построение статистических оценок 3. Построение модели марковской цепи. 4. Построение байесовских оценок для СММ

### **Тема 3. Статистические подходы в предсказании эукариотических генов**

домашнее задание , примерные вопросы:

Использование СММ для предсказания генов Использование шаблонов для предсказания генов Линейный дискриминантный анализ Нахождение псевдогенов.

### **Тема 4. Анализ данных микрочипов**

домашнее задание , примерные вопросы:

Использование критерия Стьюдента Изучение паттернов Множественное тестирование

### **Тема . Итоговая форма контроля**

Примерные вопросы к экзамену:

Вопросы к экзамену:

1. Статистическая и биологическая значимость
2. Оценка статистической значимости для локального сходства последовательностей
3. Байесовский подход
4. Скрытые марковские модели
5. Множественное выравнивание
6. Нахождение рекуррентных паттернов
7. Методы распознавания функциональных сигналов
8. Линейный дискриминантный анализ
9. Оценка донора и акцептора в сращиваниях
10. Дискриминационный и вероятностный подходы для множественного предсказания генов
11. Анализ микрочипов. Представление данных
12. Ошибки модели, калибровка.
13. Анализ микрочипов. Разделение генов
14. Исследование паттернов
15. Критерии для связи фенотипа и реакции на микрочипе
16. Множественное тестирование

#### **7.1. Основная литература:**

1. Симушкин С.В. Задачи по теории вероятностей: учеб. пособие / С.В. Симушкин, Л.Н. Пушкин. ? Казань: Казан.ун-т, 2011. ? 223 с.  
<http://kms.kpfu.ru/sites/default/files/various/%D0%AD%D0%9E%D0%A0/%D0%A1%D0%B8%D0%BC%D0%A1%D0%9E%D0%A0/%D0%A1%D0%B8%D0%BC%D0%A1%D0%9E%D0%A0/>
2. Сазанов А.А. Генетика: учеб. пособие / А.А.Сазанов - СПб: ЛГУ им А.С.Пушкина, 2011. - 264с. <http://znanium.com/bookread2.php?book=445036>
3. Григорьев Ю. Д. Методы оптимального планирования эксперимента: линейные модели: Учебное пособие. ? СПб.: Издательство "Лань", 2015. ? 320 с.: ил. ? (Учебники для вузов. Специальная литература). <http://e.lanbook.com/view/book/65949/>
4. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. ? 2-е изд., испр. ? М.: ФИЗМАТЛИТ 2012. ? 816 с.  
<http://e.lanbook.com/view/book/59747/>
5. Лагутин М. Б. Наглядная математическая статистика [Электронный ресурс] : учебное пособие / М. Б. Лагутин. ? 3-е изд. (эл.). ?М. : БИНОМ. Лаборатория знаний, 2013.? 472 с.  
<http://e.lanbook.com/view/book/56887/page473/>

#### **7.2. Дополнительная литература:**

1. В.М. Гупал. Математика и загадочный генетический код: Монография. (К 10-летию завершения программы "Геном человека") / В.М. Гупал. - М.: ИЦ РИОР: НИЦ ИНФРА-М, 2014. - 288 + II с.: 60x88 1/16. - (Научная мысль). (обложка) ISBN 978-5-369-01276-5  
<http://znanium.com/bookread2.php?book=492403>
2. Сазанов, А. А. Основы генетики [Электронный ресурс] / А. А. Сазанов. - СПб.: ЛГУ им. А. С. Пушкина, 2012. - 240 с. - ISBN 978-5-8290-1132-1.  
<http://znanium.com/bookread2.php?book=445015>
3. Ширяев А. Н. Вероятность: В 2-х кн. -- 4-е изд., переработ. и доп. -- М.: МЦНМО, 2007.  
<http://e.lanbook.com/view/book/9448/page1/>
4. Федоткин М.А. Модели в теории вероятностей. -- М.: ФИЗМАТЛИТ, 2012. -- 608с.  
<http://e.lanbook.com/view/book/59595/>

### 7.3. Интернет-ресурсы:

<http://learn.genetics.utah.edu/> - <http://learn.genetics.utah.edu/>  
<http://www.life.illinois.edu/ib/201/lectures/> - <http://www.life.illinois.edu/ib/201/lectures/>  
<http://www.stat.washington.edu/thompson/S394/genetics.html> -  
<http://www.stat.washington.edu/thompson/S394/genetics.html>  
[www.youtube.com](http://www.youtube.com) - [www.youtube.com](http://www.youtube.com)  
Википедия - [en.wikipedia.org](http://en.wikipedia.org)

### 8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Статистические методы математической генетики" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Доска, маркер, тряпка

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе Методы прикладной математической статистики .



Автор(ы):

Симушкин С.В. \_\_\_\_\_

Салимов Р.Ф. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Володин И.Н. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.