

**Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Московский физико-технический институт  
(национальный исследовательский университет)»**

**УТВЕРЖДЕНО**

**Директор физтех-школы  
биологической и медицинской  
физики**

**Д.В. Кузьмин**

**по дисциплине:** **Рабочая программа дисциплины (модуля)**  
**Биостатистика**

**программа аспирантуры:** Компьютерные науки и информатика

**курс:** **центр образовательных программ Физтех-школы биологической и медицинской физики**  
**1**

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 48 час.

Всего часов: 78, всего зач. ед.: 2

Количество контрольных работ, заданий: 2

Программу составил: С.С. Отставнов, канд. экон. наук

Программа обсуждена на заседании центра образовательных программ Физтех-школы биологической и медицинской физики 05.06.2023

## Аннотация

Цель курса - дать представление о механизмах реализации генетической информации, биостатистики и практических навыков применения биоинформатики и статистических методов для анализа и интерпретации биологических данных. После завершения курса студент поймет основные методы оценки статистической значимости, методы учета множественности сравнений, методы метаанализа, статистические характеристики ассоциативных тестов, ROC-анализ, методы оценки наследственности и генетических рисков, методы уменьшения количества переменных при анализе больших массивов данных, методы классификации данных, основы байесовского анализа данных.

### 1. Цели и задачи

#### Цель дисциплины

- дайте представление о механизмах реализации генетической информации, биостатистике и практических навыках применения биоинформатики и статистических методов для анализа и интерпретации биологических данных.

#### Задачи дисциплины

- дать представление об основных методах статистического анализа биологических данных;
- познакомить студентов с современным пониманием статистических исследований населения;
- научить пользоваться основными базами данных в полевых условиях;
- представить базовые алгоритмы и форматы данных для статистической генетики и биостатистики.

### 2. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные методы оценки статистической значимости;
- методы учета множественности сравнений;
- методы метаанализа;
- статистические характеристики ассоциативных тестов;
- ROC-анализ;
- методы оценки наследственности и генетических рисков;
- методы уменьшения количества переменных при анализе больших массивов данных;
- методы классификации данных;
- основы байесовского анализа данных.

уметь:

- использовать Интернет и справочники по научной и прикладной биостатистике, чтобы быстро найти необходимые данные и понятия;
- сравнивать методы статистической обработки и адекватно оценивать их применимость;
- применять в научных исследованиях основные методы биостатистики;
- применять основные методы биостатистики при работе в лаборатории.

владеть:

- ведения крупномасштабных массивов данных;
- компьютерного анализа статистической значимости результатов генетических и медико-биологических экспериментов.

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

#### 3.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа

1	Структура биологических данных и описательная статистика	8			12
2	Анализ сопряженности	8			12
3	Многомерные методы	8			12
4	Байесовская статистика	6			12
Итого часов		30			48
Подготовка к экзамену		0 час.			
Общая трудоёмкость		78 час., 2 зач.ед.			

### 3.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

#### Семестр: 1 (Осенний)

##### 1. Структура биологических данных и описательная статистика

Организация файлов и управление данными в EXCEL, SPSS и STATISTICA. Описательная статистика. Некоторые хитрости быстрых статистических расчетов. Статистическая проверка гипотез. Точные и опосредованные критерии. Эрооры I и II типа. Множественное сравнительное тестирование. Контроль ошибок типа I. Группировка и парадокс Симпсона. Параметрические и непараметрические критерии сравнения. Анализ отклонений.

##### 2. Анализ сопряженности

Регрессивный анализ. Остаточный анализ. Частичные корреляции и искажающие факторы. Анализ непредвиденных обстоятельств качественных характеристик. Отношение шансов и относительный риск. Статистика биомаркеров. Оценка чувствительности и специфичности теста. ROC-анализ.

##### 3. Многомерные методы

Множественный регрессионный анализ. Способы уменьшения количества предикторов. Парадокс Фридмана. Оценка наследственности и генетического риска. Проблема «недостающей наследственности». Факторный анализ. Метод главных компонент. Методы классификации. Кластерный анализ. Дискриминантный анализ.

##### 4. Байесовская статистика

Ограничение p-значений. Воспроизводимость экспериментальных результатов. Байесовский фактор. Приоры. Статистика в эпидемиологии. Анализ больших образцов. Байесовские оценки частоты редких событий.

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

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

### 5.Перечень рекомендуемой литературы

Основная литература

Предоставляются на кафедре:

- 1) Бисвас, А., Датта, С., Файн, Дж. П., Сигал, М. Р. (ред.) Статистические достижения в биомедицинских науках, клинические испытания, эпидемиология, анализ выживаемости и биоинформатика, 2008 г., WILEY-VCH Verlag GmbH & Co. KGaA, Вайнхайм, Германия
- 2) Статистическая генетика человека. Отредактированный Робертом К. Элстоном. Springer Science + Business Media, LLC 2012 г.

Дополнительная литература

Предоставляются на кафедре:

- 1) Темы по биостатистике. Под редакцией Волтера Т. Амброзиуса 2007 Humana Press Inc. 999 Riverview Drive, Suite 208, Totowa, New Jersey 07512
- 2) Агостино Ди Чаччо, Мауро Коли, Хосе Мигель Ангуло Ибакнес. Расширенные статистические методы анализа больших массивов данных. Springer-Verlag, Берлин, Гейдельберг, 2012 г.

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

- 1) [www.ncbi.nih.nlm.gov](http://www.ncbi.nih.nlm.gov)
- 2) [www.scopus.com](http://www.scopus.com)

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

Для части занятий потребуется Zoom. Google Drive для доступа к материалам курса.

Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

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

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

## ПРИЛОЖЕНИЕ

### ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

**программа аспирантуры:** Компьютерные науки и информатика  
Физтех-школа Прикладной Математики и Информатики  
центр образовательных программ Физтех-школы биологической и медицинской  
физики

**курс:** 1

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

**Разработчик:** С.С. Отставнов, канд. экон. наук

## 1. Показатели оценивания компетенций

В результате изучения дисциплины «Биостатистика» обучающийся должен:

### знать:

- основные методы оценки статистической значимости;
- методы учета множественности сравнений;
- методы метаанализа;
- статистические характеристики ассоциативных тестов;
- ROC-анализ;
- методы оценки наследственности и генетических рисков;
- методы уменьшения количества переменных при анализе больших массивов данных;
- методы классификации данных;
- основы байесовского анализа данных.

### уметь:

- использовать Интернет и справочники по научной и прикладной биостатистике, чтобы быстро найти необходимые данные и понятия;
- сравнивать методы статистической обработки и адекватно оценивать их применимость;
- применять в научных исследованиях основные методы биостатистики;
- применять основные методы биостатистики при работе в лаборатории.

### владеть:

- ведения крупномасштабных массивов данных;
- компьютерного анализа статистической значимости результатов генетических и медико-биологических экспериментов.

## 2. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

- 1) Отношение шансов и относительный риск.
- 2) Множественная регрессия и парадокс Фридмана.
- 3) Методы оценки сдвига публикации. Диаграммы воронок.
- 4) FDR-метод учета множественных сравнений.
- 5) Байесовские оценки частот редких событий.

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Возможно выполнение патентного поиска в качестве самостоятельной задачи. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на дифференцированном зачете.

## 3. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

- 1) Отношение шансов и относительный риск
- 2) Множественная регрессия и парадокс Фридмана
- 3) Методы оценки сдвига публикации. Диаграммы воронок.
- 4) FDR-метод учета множественных сравнений
- 5) Байесовские оценки частот редких событий

Типовые билеты:

Билет №1

Страта и парадокс Симпсона.

Билет №2

Дисперсионный анализ

Билет № 3

Частные корреляции и искажающие факторы.

Билет № 4

#### Критерии оценивания

Оценка отлично (10 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

#### **4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

Оценка знаний, умений и опыта навыков деятельности осуществляется в период промежуточной аттестации в форме дифференцированного зачета. Дифференцированный зачет проходит в форме "вопрос-ответ", где вопросы задаются из заранее утвержденного перечня вопросов.

По решению кафедры оценка знаний навыков и умений и опыта деятельности может осуществляться в ходе текущего контроля, во время семинаров, практических занятий, либо по результатам самостоятельной работы студентов с использованием фондов оценочных средств, дифференцированных по показателям обучения по дисциплине.

При проведении устного дифференцированного зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном дифференцированном зачете не должен превышать одного астрономического часа.