

**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**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Биостатистика
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Алгоритмическая биология
	Физтех-школа Биологической и Медицинской Физики
	центр образовательных программ Физтех-школы биологической и медицинской физики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Экзамен

Аудиторных часов: 30 всего, в том числе:

лекции: 15 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Подготовка к экзамену: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составили:

Е.М. Мясникова, канд. физ.-мат. наук

Е.С. Чефанова, старший преподаватель

Программа обсуждена на заседании центра образовательных программ Физтех-школы биологической и медицинской физики 09.06.2022

## Аннотация

Дисциплина направлена на освоение студентами основ биостатистики. 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 оценки уменьшения количества переменных при анализе больших массивов данных, методы классификации данных, основы байесовского анализа данных.

### 1. Цели и задачи

#### Цель дисциплины

- дать 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 формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

### 3. 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:

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

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

##### 4.1. 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	4	4		10
2	Анализ сопряженности	4	4		10
3	Многомерные методы	4	4		5
4	Байесовская статистика	3	3		5
Итого часов		15	15		30
Подготовка к экзамену		30 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 2 (Весенний)

###### 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.

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

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

## 6. 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, - WILEY-VCH Verlag GmbH & Co. KGaA, 2008
- 2) Статистическая генетика человека. Отредактированный Робертом К. Элстоном. - Springer Science + Business Media, LLC 2012

### Дополнительная литература

Предоставляются на кафедре:

- 1) Темы по биостатистике / под редакцией Волтера Т. Амброзиуса. - New Jersey: Humana Press Inc, 2007
- 2) Агостино Ди Чаччо, Мауро Коли, Хосе Мигель Ангуло Ибакнес. Расширенные статистические методы анализа больших массивов данных. - Берлин: Springer-Verlag, 2012

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

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

## 8. 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 потребуются Zoom. Google Drive для доступа к материалам курса. Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

## 9. 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)

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Алгоритмическая биология Физтех-школа Биологической и Медицинской Физики центр образовательных программ Физтех-школы биологической и медицинской физики
<b>курс:</b>	<u>1</u>
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Экзамен

#### Разработчики:

Е.М. Мясникова, канд. физ.-мат. наук  
Е.С. Чефанова, старший преподаватель

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

## 2. 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:

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

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

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

- 1) Воспроизводимость экспериментальных результатов.
- 2) Байесовский фактор. Приоры.
- 3) Статистика в эпидемиологии. Анализ больших образцов.
- 4) Байесовские оценки частоты редких событий.

- 5) Рассчитайте медиану для набора чисел.
- 6) Рассчитайте стандартную ошибку для выборки.
- 7) Диаграммы воронок.
- 8) Множественная регрессия.
- 9) ROC анализ.
- 10) Ограничение p-значений.

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 экзамене.

#### **4. 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.
- 4) FDR-метод учета множественных сравнений.
- 5) Байесовские оценки частот редких событий.
- 6) Страта и парадокс Симпсона.
- 7) Дисперсионный анализ.
- 8) Частные корреляции и искажающие факторы.
- 9) Статистика биомаркеров.
- 10) Чувствительность и специфичность тестов.

Типовые билеты:

Билет №1

Страта и парадокс Симпсона.

Дисперсионный анализ.

Билет № 2

Частные корреляции и искажающие факторы.

Статистика биомаркеров.

Билет № 3

Чувствительность и специфичность тестов.

ROC анализ.

#### **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.

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

При проведении устного экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать одного астрономического часа.