

**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>	Дополнительные главы биостатистики
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Биоинженерия и биоинформатика Физтех-школа Биологической и Медицинской Физики кафедра биоинформатики и системной биологии
<b>курс:</b>	1
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Подготовка к экзамену: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: А.В. Рубанович, д-р биол. наук

Программа обсуждена на заседании кафедры биоинформатики и системной биологии 11.05.2020

## Аннотация

Целью дисциплины является освоение специфических методов статистической обработки биологических, генетических, медицинских и эпидемиологических данных. После освоения курса студент будет понимать основные методы оценки статистической значимости, методы учета множественности сравнений, методы мета-анализа, статистические характеристики ассоциативных тестов, ROC-анализ, методы оценки наследуемости и генетических рисков, методы сокращения числа переменных при анализе больших массивов данных, методы классификации данных, основы байесовского анализа данных.

### 1. Цели и задачи

#### Цель дисциплины

освоение специфических методов статистической обработки биологических, генетических, медицинских и эпидемиологических данных.

#### Задачи дисциплины

- формирование базовых знаний о методах статистического анализа результатов биологических экспериментов, обработка больших массивов современных генетических данных, оценка генетических рисков и показателей наследуемости, статистика биомаркеров, ROC-анализ, анализ множественности тестов, мета-анализ, построение байесовских оценок;
- практическое освоение студентами компьютерных методов проведения статистического анализа;
- формирование у студентов основных навыков статистического анализа и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в области генетики, эпидемиологии и физико-химической биологии.

### 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-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.2 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях
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-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)

### 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 мета-анализа;
- статистические характеристики ассоциативных тестов;
- ROC-анализ;
- методы оценки наследуемости и генетических рисков;
- методы сокращения числа переменных при анализе больших массивов данных;
- методы классификации данных;
- основы байесовского анализа данных.

уметь:

- пользоваться Интернетом и справочной литературой по биостатистике научного и прикладного характера для быстрого поиска необходимых данных и понятий;
- сравнивать между собой методы статистической обработки и адекватно оценивать их применимость;
- применять основные методы биостатистики в научных исследованиях;
- применять основные методы биостатистики при работе в лаборатории.

владеть:

- навыками обработки больших массивов данных;
- культурой 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 описательные статистики	12			8
2	Анализ сопряженности признаков	4			7
3	Многомерные методы статистического анализа	10			9
4	Байесовская статистика	4			6
Итого часов		30			30
Подготовка к экзамену		30 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 1 (Осенний)

###### 1. Структура биологических данных и описательные статистики

Организация файлов и управление данными в программах EXCEL, SPSS и STATISTICA. Описательные статистики. Некоторые приемы быстрых статистических вычислений. Проверка статистических гипотез. Точные и опосредованные критерии. Ошибки I и II рода. Мощность теста. Множественные сравнения. Контроль ошибок I рода. Страты и парадокс Симпсона.

Параметрические и непараметрические критерии сравнения. Дисперсионный анализ.

## 2. Анализ сопряженности признаков

Регрессионный анализ. Анализ остатков. Частные корреляции и конфаундеры. Сопряженность качественных признаков. Отношение шансов и относительный риск. Статистика биомаркеров. Оценки чувствительности и специфичности теста. ROC-анализ.

## 3. Многомерные методы статистического анализа

Множественный регрессионный анализ. Методы сокращения числа предикторов. Парадокс Фридмана. Оценки наследуемости и генетического риска. Проблема «missing heritability». Факторный анализ. Метод главных компонент. Методы классификации. Кластерный анализ. Дискриминантный анализ.

## 4. Байесовская статистика

Ограниченность концепции p-value. Анализ воспроизводимости результатов экспериментов. Байесовский фактор. Приоры. Статистика в эпидемиологии. Анализ больших выборок. Байесовские оценки частот редких событий.

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

Стандартная учебная аудитория, оснащенная мультимедийным оборудованием.

## 6. Перечень рекомендуемой литературы

### Основная литература

Предоставляется на кафедре:

- 1) Biswas, A., Datta, S., Fine, J. P., Segal, M. R. (eds.) Statistical Advances in the Biomedical Sciences Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics, 2008, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany
- 2) Statistical Human Genetics. Edited by Robert C. Elston. Springer Science+Business Media, LLC 2012

### Дополнительная литература

Предоставляется на кафедре:

- 1) Topics in Biostatistics. Edited by Walter T. Ambrosius 2007 Humana Press Inc. 999 Riverview Drive, Suite 208, Totowa, New Jersey 07512
- 2) Agostino Di Ciaccio, Mauro Coli, Jose Miguel Angulo Ibañez. Advanced Statistical Methods for the Analysis of Large Data-Sets. Springer-Verlag Berlin Heidelberg, 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 потребуется 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, прослушавший курс, должен с одной стороны, овладеть теоретическим аппаратом, а с другой стороны, должен научиться применять полученные знания на практике. Успешное освоение курса требует самостоятельной работы студента. В программе курса для самостоятельной работы студента над темой отводится минимальное время.

Самостоятельная работа включает в себя:

- проработку учебного материала,
- чтение и конспектирование дополнительной литературы,
- подготовку ответов на вопросы, предназначенных для самостоятельного изучения,
- решение задач, предлагаемых студентам на аудиторных занятиях,
- подготовку к экзамену.

Руководство и контроль самостоятельной работы студента осуществляется в форме индивидуальных консультаций. Показателем владения материалом служит умение решать задачи. Для формирования умения применять теоретические знания на практике студенту необходимо решать как можно больше задач. При решении задач каждое действие необходимо аргументировать, ссылаясь на рассмотренный ранее теоретический аппарат.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Биоинженерия и биоинформатика Физтех-школа Биологической и Медицинской Физики кафедра биоинформатики и системной биологии
<b>курс:</b>	1
<b>квалификация:</b>	магистр
Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен	
<b>Разработчик:</b>	А.В. Рубанович, д-р биол. наук

## 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-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.2 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях
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-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Дополнительные главы биостатистики» обучающийся должен:

### знать:

- основные методы оценки статистической значимости;
- методы учета множественности сравнений;
- методы мета-анализа;
- статистические характеристики ассоциативных тестов;
- ROC-анализ;
- методы оценки наследуемости и генетических рисков;
- методы сокращения числа переменных при анализе больших массивов данных;
- методы классификации данных;
- основы байесовского анализа данных.

### уметь:

- пользоваться Интернетом и справочной литературой по биостатистике научного и прикладного характера для быстрого поиска необходимых данных и понятий;
- сравнивать между собой методы статистической обработки и адекватно оценивать их применимость;
- применять основные методы биостатистики в научных исследованиях;
- применять основные методы биостатистики при работе в лаборатории.

### владеть:

- навыками обработки больших массивов данных;
- культурой 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) Отношение шансов и относительный риск
- 2) Множественная регрессия и парадокс Фридмана
- 3) Методы оценки публикационного сдвига. Графики-воронки.
- 4) FDR-метод учета множественности 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 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 публикационного сдвига. Графики-воронки.
- 4) FDR-метод учета множественности сравнений
- 5) Байесовские оценки частот редких событий

Типовые билеты:

Билет №1

Страты и парадокс Симпсона.

Билет №2

Дисперсионный анализ

Билет №3

Частные корреляции и конфаундеры.

Билет №4

Статистика биомаркеров

Билет №5

Чувствительность и специфичность тестов

Билет №8

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.