

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

**УТВЕРЖДЕНО**

**Директор физтех-школы  
прикладной математики и  
информатики**

**А.М. Райгородский**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Statistical Data Analysis/Математическая статистика и анализ данных
<b>по направлению:</b>	Фотоника и оптоинформатика
<b>профиль подготовки:</b>	Фотоника, квантовые технологии и двумерные материалы Физтех-школа физики и исследований им. Ландау кафедра машинного обучения и цифровой гуманитаристики
<b>курс:</b>	1
<b>квалификация:</b>	магистр

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

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составили:

Р.Г. Нейчев, старший преподаватель

А.М. Райгородский, д-р физ.-мат. наук, доцент, главный научный сотрудник

Программа обсуждена на заседании кафедры машинного обучения и цифровой гуманитаристики 05.03.2020

## Аннотация

В данном курсе математическая статистика рассматривается как научная дисциплина, изучающая теоретические основы и методы извлечения и обработки эмпирических данных о массовых явлениях, представляемых в виде стохастических моделей. К задачам математической статистики относят также построение и исследование процедур принятия решений в условиях «стохастической неопределенности».

Курс содержит начальные сведения из математической статистики, используемые и дополняемые в дальнейшем. Рассматриваются классические вероятностные модели принятия решений о классах наблюдаемых объектов по значениям их признаков (модели классификации или выбора гипотез); при этом предполагается, что распределения признака для каждого класса объектов известны точно или с точностью до типа. Обсуждаются критерии согласия как инструмент проверки достоверности гипотез и задачи оценивания распределений (в частности, задачи т.н. параметрического оценивания).

Отдельно обсуждаются методы непараметрического оценивания распределений, предъявляющие к априорным сведениям об их свойствах гораздо меньшие требования. Курс также содержит основные сведения о регрессионном анализе, служащем для выявления и оценки вероятностных связей между изучаемыми случайными величинами.

## 1. Цели и задачи

### Цель дисциплины

изучение математических и теоретических основ современного статистического анализа, а также подготовка слушателей к дальнейшей самостоятельной работе в области анализа статистических задач прикладной математики, физики и экономики.

### Задачи дисциплины

изучение математических основ математической статистики;  
приобретение слушателями теоретических знаний в области современного статистического анализа.

## 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач в области фотоники и оптоинформатики
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области профессиональной деятельности, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-3.2 Способен использовать исследовательские методы при решении новых задач, применяя знания в различных областях науки (техники)
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений
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
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Способен самостоятельно или в составе научного коллектива планировать и проводить научные исследования в области фотоники и оптоинформатики
ПК-3 Способен профессионально работать с	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)

исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	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 понятия математической статистики;
- основные подходы к сравнению оценок параметров неизвестного распределения;
- асимптотические и неасимптотические свойства оценок параметров неизвестного распределения;
- основные методы построения оценок с хорошими асимптотическими свойствами: метод моментов, метод максимального правдоподобия, метод выборочных квантилей;
- понятие эффективных оценок и неравенство информации Рао-Крамера;
- определение и главные свойства условного математического ожидания случайной величины относительно сигма-алгебры или другой случайной величины;
- определение общей линейной регрессионной модели и метод наименьших квадратов;
- многомерное нормальное распределение и его основные свойства;
- базовые понятия теории проверки статистических гипотез;
- лемму Неймана – Пирсона и теорему о монотонном отношении правдоподобия;
- критерий хи-квадрат Пирсона для проверки простых гипотез в схеме Бернулли.

уметь:

- обосновывать асимптотические свойства оценок с помощью применения предельных теорем теории вероятностей;
- строить оценки с хорошими асимптотическими свойствами для параметров неизвестного распределения по заданной выборке из него;
- находить байесовские оценки по заданному априорному распределению;
- вычислять условные математические ожидания с помощью условных распределений;
- находить оптимальные оценки с помощью полных достаточных статистик;
- строить точные и асимптотические доверительные интервалы и области для параметров неизвестного распределения;
- находить оптимальные оценки и доверительные области в гауссовской линейной модели;
- строить равномерно наиболее мощные критерии в случае параметрического семейства с монотонным отношением правдоподобия;
- строить F-критерий для проверки линейных гипотез в линейной гауссовской модели.

владеть:

- основными методами математической статистики построения точечных и доверительных оценок: методом моментов, выборочных квантилей, максимального правдоподобия, методом наименьших квадратов, методом центральной статистики.
- навыками асимптотического анализа статистических критериев;
- навыками применения теорем математической статистики в прикладных задачах физики и экономики.

### 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	Ф-критерий для проверки линейных гипотез в гауссовской линейной модели.	10	10		15
2	Вероятностно-статистическая модель.	10	10		15
3	Основная задача математической статистики.	10	10		15
Итого часов		30	30		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 2 (Весенний)

1. Ф-критерий для проверки линейных гипотез в гауссовской линейной модели.

Ф-критерий для проверки линейных гипотез в гауссовской линейной модели. Пример с двумя гауссовскими выборками, отличающимися сдвигом: проверка гипотезы об их однородности.

2. Вероятностно-статистическая модель.

Вероятностно-статистическая модель. Понятия наблюдения и выборки. Параметрическая статистическая модель. Моделирование выборки из неизвестного распределения, принадлежащему параметрическому семейству.

3. Основная задача математической статистики.

Основная задача математической статистики. Примеры: выборка и линейная модель.

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

Стандартная учебная аудитория

#### 6. Перечень рекомендуемой литературы

##### Основная литература

1. Математическая статистика [Текст] : учеб. пособие для вузов / А. А. Натан, О. Г. Горбачев, С. А. Гуз ; Моск. физико-техн. ин-т (гос. ун-т). — М : МЗ Пресс, 2004, 2005. — 160 с.
2. Математическая статистика [Текст] : [учебник для вузов] / А. А. Боровков. — [3-е изд., испр.] — М. : Физматлит, 2007. — 704 с.
3. Введение в математическую статистику [Текст] : [учебник для вузов] / Г. И. Ивченко, Ю. И. Медведев. — М. : ЛКИ, 2010, 2014, 2015. — 600 с.
4. Теория вероятностей и математическая статистика [Текст] : учеб. пособие для вузов / П. П. Бочаров, А. В. Печинкин. — М. : Физматлит, 2005. — 295 с. : ил. + pdf-версия. - Библиогр.: с. 292. - ISBN 5-9221-0633-3. — Полный текст (Доступ из сети МФТИ / Удаленный доступ).

##### Дополнительная литература

1. Наглядная математическая статистика [Текст] : учеб. пособие для вузов / М. Б. Лагутин. — 7-е изд. — М. : БИНОМ. Лаб. знаний, 2019. — 472 с. : ил. - Библиогр.: с. 456-459. - Предм. указ.: с. 462-466. - ISBN 978-5-00101-105-7 (в пер.). — Полный текст (Режим доступа : доступ из сети МФТИ).

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

<http://dm.fizteh.ru/>

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

Мультимедийные технологии можно использовать на лекциях и практических занятиях, в том числе на презентациях.

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

1. Рекомендуется успешно сдавать контрольные работы, так как это упрощает итоговую аттестацию по предмету.
2. Для подготовки к итоговой аттестации по предмету лучше всего пользоваться материалами лекций.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Фотоника и оптоинформатика
<b>профиль подготовки:</b>	Фотоника, квантовые технологии и двумерные материалы Физтех-школа физики и исследований им. Ландау кафедра машинного обучения и цифровой гуманитаристики
<b>курс:</b>	<u>1</u>
<b>квалификация:</b>	магистр

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

**Разработчики:**

Р.Г. Нейчев, старший преподаватель

А.М. Райгородский, д-р физ.-мат. наук, доцент, главный научный сотрудник

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач в области фотоники и оптоинформатики
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области профессиональной деятельности, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-3.2 Способен использовать исследовательские методы при решении новых задач, применяя знания в различных областях науки (техники)
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений
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
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
ПК-2 Способен самостоятельно или в качестве члена (руководителя) малого коллектива организовывать и проводить научные исследования и их апробацию	ПК-2.1 Способен самостоятельно или в составе научного коллектива планировать и проводить научные исследования в области фотоники и оптоинформатики
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 «Statistical Data Analysis/Математическая статистика и анализ данных» обучающийся должен:

### знать:

- основные понятия математической статистики;
- основные подходы к сравнению оценок параметров неизвестного распределения;
- асимптотические и неасимптотические свойства оценок параметров неизвестного распределения;
- основные методы построения оценок с хорошими асимптотическими свойствами: метод моментов, метод максимального правдоподобия, метод выборочных квантилей;
- понятие эффективных оценок и неравенство информации Рао-Крамера;
- определение и главные свойства условного математического ожидания случайной величины относительно сигма-алгебры или другой случайной величины;
- определение общей линейной регрессионной модели и метод наименьших квадратов;
- многомерное нормальное распределение и его основные свойства;
- базовые понятия теории проверки статистических гипотез;
- лемму Неймана – Пирсона и теорему о монотонном отношении правдоподобия;
- критерий хи-квадрат Пирсона для проверки простых гипотез в схеме Бернулли.

### уметь:

обосновывать асимптотические свойства оценок с помощью применения предельных теорем теории вероятностей;

строить оценки с хорошими асимптотическими свойствами для параметров неизвестного распределения по заданной выборке из него;

находить байесовские оценки по заданному априорному распределению;

вычислять условные математические ожидания с помощью условных распределений;

находить оптимальные оценки с помощью полных достаточных статистик;

строить точные и асимптотические доверительные интервалы и области для параметров неизвестного распределения;

находить оптимальные оценки и доверительные области в гауссовской линейной модели;

строить равномерно наиболее мощные критерии в случае параметрического семейства с монотонным отношением правдоподобия;

строить F-критерий для проверки линейных гипотез в линейной гауссовской модели.

#### **владеть:**

основными методами математической статистики построения точечных и доверительных оценок: методом моментов, выборочных квантилей, максимального правдоподобия, методом наименьших квадратов, методом центральной статистики.

навыками асимптотического анализа статистических критериев;

навыками применения теорем математической статистики в прикладных задачах физики и экономики.

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

Примеры домашних заданий:

1. Найдите оптимальную оценку параметра  $\theta > 0$  по выборке из распределения: а)  $N(\theta, 1)$ , б)  $R(0, \theta)$ , в)  $Pois(\theta)$ , г)  $Bin(1, \theta)$  (здесь  $(0, 1)$ ). 6. Пусть  $X_1, \dots, X_n$  выборка из равномерного распределения на отрезке  $[0, \theta]$ ,  $\theta > 0$ . Постройте доверительный интервал для уровня доверия  $\gamma$ , используя статистику а)  $X$ , б)  $X(1)$ , в)  $X(n)$ .

2.. Пусть  $X_1, \dots, X_n$  выборка из нормального распределения с параметрами  $(\theta, 1)$ . Найдите байесовскую оценку параметра  $\theta$ , если априорное распределение есть  $Bin(1, p)$ . Будет ли полученная оценка состоятельной оценкой параметра?

3. Имеется 2 объекта с весами  $a$  и  $b$ . Мы взвесили с ошибками первый, второй и оба объекта вместе, причём дисперсия ошибки в последнем случае была в 4 раза больше. Свести задачу к линейной регрессионной модели и найти оценки наименьших квадратов для  $a$  и  $b$ .

4.  $X_1, \dots, X_n$  выборка из экспоненциального распределения с параметром  $\theta$ . Постройте равномерно наиболее мощный критерий уровня значимости проверки гипотезы  $H_0: \theta = 0$  против альтернативы а)  $H_1: \theta > 0$ , б)  $H_1: \theta < 0$ .

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

1. Виды сходимостей случайных векторов: с вероятностью 1, по вероятности, по распределению. Взаимоотношения между различными видами сходимостей. Усиленный закон больших чисел для случайных векторов. Многомерная центральная предельная теорема.

2. Теорема о наследовании сходимости и лемма Слуцкого. Пример их применения.

3. Гауссовские случайные векторы (многомерное нормальное распределение). Теорема о трех эквивалентных определениях. Смысл параметров гауссовского вектора.

4. Основные свойства гауссовских векторов: линейные преобразования и критерий независимости компонент. Теорема об ортогональном разложении гауссовского вектора.

5. Вероятностно-статистическая модель, параметрическая модель. Выборка, эмпирическое распределение. Теорема Гливенко-Кантелли.

6. Статистики и оценки. Несмещенность, состоятельность, сильная состоятельность и асимптотическая нормальность. Лемма о наследовании асимптотической нормальности.

7. Оценка параметра по методу подстановки. Оценка параметра по методу моментов. Теорема о состоятельности оценки метода моментов.

8. Квантили распределения, выборочные квантили, выборочная медиана. Теорема об асимптотической нормальности выборочной  $r$ -квантили.

9. Функция потерь и функция риска. Равномерный подход к сравнению оценок. Байесовский, минимаксный и асимптотический подходы к сравнению оценок.



10. Информация Фишера и вклад наблюдения. Неравенство Рао-Крамера. Эффективные оценки и критерий эффективности.
11. Понятие плотности для дискретной случайной величины. Доминируемое семейство распределений. Функция правдоподобия и оценка максимального правдоподобия. Теорема об экстремальном свойстве правдоподобия.
12. Теорема о существовании состоятельного решения уравнения правдоподобия. Состоятельность оценки максимального правдоподобия. Теорема об асимптотической нормальности решения уравнения правдоподобия.
13. Теорема Бахадура. Асимптотически эффективные оценки. Асимптотическая эффективность и эффективность оценки максимального правдоподобия.
14. Условное математическое ожидание случайной величины относительно сигма-алгебры. Заряд на вероятностном пространстве. Теорема Радона-Никодима.
15. Свойства условного математического ожидания (9 штук).
16. Условное математическое ожидание. Условное распределение и условная плотность одной случайной величины относительно другой. Теорема о вычислении условного математического ожидания с помощью условной плотности. Теорема о достаточном условии существования условной плотности.
17. Достаточные статистики. Критерий факторизации Неймана-Фишера. Теорема Колмогорова-Блекуэлла-Рао и следствие из нее.
18. Полные статистики. Теорема об оптимальной оценке. Экспоненциальное семейство распределений. Теорема о полной достаточной статистике в экспоненциальном семействе.
19. Доверительные интервалы и доверительные области. Понятие центральной статистики и метод построения доверительных областей с ее помощью. Асимптотические доверительные интервалы.
20. Линейная регрессионная модель. Оценка наименьших квадратов, формула ее вычисления. Несмещенные оценки для параметров линейной регрессионной модели.
21. Проверка статистических гипотез: гипотеза и альтернатива, критерий проверки гипотезы, ошибки первого

#### Экзаменационный билет

1. Информация и наблюдения Фишера. Неравенство Рао-Крамера. Эффективные оценки и критерии эффективности.
2. Типы сходимости случайных векторов: с вероятностью 1, по вероятности, по распределению. Взаимоотношения между разными типами конвергенции. Сильный закон больших чисел для случайных векторов. Многомерная центральная предельная теорема.

#### Критерии оценивания

Оценка «отлично (10)» выставляется студенту, проявившему всестороннее, систематическое и глубокое знание материала образовательной программы, самостоятельно выполнившего все задания, предусмотренные программой, глубоко изучившему основную и дополнительную литературу, рекомендованную программой. , активно работает в классе и понимает основные научные концепции по изучаемой дисциплине, проявил творческий подход и научный подход в понимании и представлении материала образовательной программы, ответ на который характеризуется использованием богатых и адекватных терминов, а также последовательным и логичным изложением материала;

Оценка «отлично (9)» выставляется студенту, который продемонстрировал всестороннее систематическое знание материала образовательной программы, самостоятельно выполнил все задачи, предусмотренные программой, глубоко усвоил основную литературу и знаком с рекомендуемой дополнительной литературой. по программе, активно проработал на занятиях, показал системность знаний по дисциплине, достаточную для дальнейшего изучения, а также умение самостоятельно расширять ее, ответ которой отличается точностью используемых терминов, а изложение материала в нем последовательное и логичное;

Оценка «отлично (8)» выставляется студенту, который проявил полное знание материала образовательной программы, не допускает существенных неточностей в своем ответе, самостоятельно выполнил все задания, предусмотренные программой, изучил основную литературу, рекомендованную учебной программой. программа, активно проработанная на занятиях, показала системность его знаний по дисциплине, достаточных для дальнейшего изучения, а также способность самостоятельно их усиливать;

Оценка «хорошо (7)» выставляется студенту, который проявил достаточно полное знание материала образовательной программы, не допускает существенных неточностей в ответе, самостоятельно выполнил все задания, предусмотренные программой, изучил основную рекомендованную литературу по программе, активно работал на занятиях, проявил системность своих знаний по дисциплине, достаточных для дальнейшего изучения, а также способность самостоятельно их усиливать;

Оценка «хорошо (6)» выставляется студенту, который проявил достаточно полное знание материала образовательной программы, не допускает существенных неточностей в своем ответе, самостоятельно выполнил основные задачи, предусмотренные программой, изучил основную литературу, рекомендованную программой, показал систематичность своих знаний по дисциплине, достаточную для дальнейшего изучения;

Оценка «хорошо (5)» дается студенту, продемонстрировавшему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, который, не проявляя достаточной активности на уроках, тем не менее самостоятельно выполнял, овладел основными задачами, предусмотренными программой, освоил основную литературу, рекомендованную программой, допустил ошибки в их выполнении и ответе во время тестирования, но имеет необходимые знания для исправления этих ошибок самостоятельно;

Оценка «удовлетворительно (4)» дается студенту, обнаружившему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, который, не проявляя достаточной активности на уроках, тем не менее самостоятельно выполнял, выполнил основные задачи, предусмотренные программой, изучил основную литературу, но допустил ошибки в их выполнении и в своем ответе во время теста, но имеет необходимые знания для исправления этих ошибок под руководством преподавателя;

Оценка «удовлетворительно (3)» выставляется обучающемуся, проявившему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, не проявившего активности на занятиях, самостоятельно выполнившего основные задания, предусмотренные законодательством. программа, но допускающая ошибки в их выполнении и в ответе при тестировании, но обладающая необходимыми знаниями для устранения под руководством преподавателя наиболее существенных ошибок;

Оценка «неудовлетворительно (2)» выставляется студенту, который показал пробелы в знаниях или недостаток знаний по значительной части материала основной образовательной программы, не выполнил самостоятельно основные задания, требуемые программой, допустил принципиальные ошибки в выполнении предусмотренных программой задач, не имеющего возможности продолжить учебу или начать профессиональную деятельность без дополнительной подготовки по данной дисциплине;

Оценка «неудовлетворительно (1)» ставится студенту при отсутствии ответа (отказ от ответа) или когда представленный ответ не соответствует сути вопросов, содержащихся в задании.

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

Во время экзамена студенту разрешается использовать программу дисциплины.