

**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>	Методы прикладной статистики
<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Физтех-школа Прикладной Математики и Информатики кафедра дискретной математики
<b>курс:</b>	3
<b>квалификация:</b>	бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

Аудиторных часов: 30 всего, в том числе:

лекции: 15 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 15 час.

Всего часов: 45, всего зач. ед.: 1

Количество контрольных работ, заданий: 1

Программу составил: Д.А. Шабанов, д-р физ.-мат. наук, доцент, доцент

Программа обсуждена на заседании кафедры дискретной математики 04.03.2020

## Аннотация

Курс методов современной прикладной статистики изучается студентами группы Анализа данных ФПМИ в шестом семестре после того, как они овладели основными математической статистики. Курс имеет прикладную направленность. В течение семестра студенты учатся анализировать данные и обучать статистические модели, в том числе на реальных данных. В рамках данной дисциплины студенты проходят следующие темы: критерии согласия и однородности, проверка нормальности, множественная проверка гипотез, обнаружение выбросов бутстрепные методы, корреляционный анализ, дисперсионный анализ, регрессионный анализ, ядерные оценки плотности и непараметрическая регрессия, методы понижения размерности, анализ временных рядов.

### 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
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.3 Знает основные требования информационной безопасности
ОПК-5 Способен участвовать в проведении фундаментальных и прикладных исследований и разработок, самостоятельно осваивать новые теоретические, в том числе, математические методы исследований и работать на современной экспериментальной научно-исследовательской, измерительно-аналитической и технологической аппаратуре)	ОПК-5.3 Способен к профессиональной эксплуатации современной экспериментальной научно-исследовательской (измерительно-аналитической и технологической) аппаратуры
	ОПК-5.1 Способен решать поставленные задачи в области теоретических и экспериментальных исследований и разработок
	ОПК-5.2 Обладает способностью к освоению новых знаний на основе изучения литературы, научных статей и других источников

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, оценить 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-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива
	ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

### 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	Основная задача математической статистики.	2	2		
2	Различные виды сходимостей случайных векторов	2	2		
3	Эмпирическое распределение и эмпирическая функция распределения	2	2		
4	Вероятностно статистическая модель.	3	3		
5	Статистики и оценки.	3	3		
6	F критерий для проверки линейных гипотез в гауссовской линейной модели.	3	3		15
Итого часов		15	15		15
Подготовка к экзамену		0 час.			
Общая трудоёмкость		45 час., 1 зач.ед.			

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

Семестр: 6 (Весенний)

###### 1. Основная задача математической статистики.

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

###### 2. Различные виды сходимостей случайных векторов

Различные виды сходимостей случайных векторов: с вероятностью 1, по вероятности, по распределению. Три знаменитых теоремы: закон больших чисел, усиленный закон больших чисел, центральная предельная теорема. Теорема о наследовании сходимости и лемма Слуцкого. Пример применения леммы Слуцкого.

###### 3. Эмпирическое распределение и эмпирическая функция распределения

Эмпирическое распределение и эмпирическая функция распределения. Обоснованность основной задачи математической статистики и теорема Гливенко-Кантелли.

###### 4. Вероятностно статистическая модель.

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

## 5. Статистики и оценки.

Статистики и оценки. Примеры статистик: выборочные характеристики, порядковые статистики. Основные свойства оценок: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность. Примеры. Наследование состоятельности и сильной состоятельности при взятии непрерывной функции. Лемма о наследовании асимптотической нормальности.

## 6. F критерий для проверки линейных гипотез в гауссовской линейной модели.

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

## 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] : [учебник для вузов] / А. А. Боровков .— [3-е изд., испр.] .— М. : Физматлит, 2007 .— 704 с.
2. Введение в математическую статистику [Текст] : [учебник для вузов] / Г. И. Ивченко, Ю. И. Медведев .— М. : ЛКИ, 2010, 2014, 2015 .— 600 с.
3. Наглядная математическая статистика [Текст] : учеб. пособие для вузов / М. Б. Лагутин .— 2-е изд., испр. — М. : Бинوم. Лаб. знаний, 2009 .— 472 с.
4. Курс теории вероятностей и математической статистики [Текст] : [учеб. пособие для вузов] / Б. А. Севастьянов .— М. ; Ижевск : Ин-т компьютерных исследований, 2004 .— 272 с.
5. Вероятность [Текст] : в 2 т. : учебник для вузов / А. Н. Ширяев .— 3-е изд., перераб. и доп. — М. : Изд-во МЦНМО, 2004 .— Т. 1 : Элементарная теория вероятностей. Математические основания. Предельные теоремы. - 2004. - 520 с.
6. Вероятность [Текст] : в 2 т. : учебник для вузов / А. Н. Ширяев ; доп. М-вом образования России .— 3-е изд., перераб. и доп. — М. : Изд-во МЦНМО, 2004 .— Т. 2 : Суммы и последовательности случайных величин - стационарные, мартингалы, марковские цепи. - 2004. - 408 с.

### Дополнительная литература

1. Линейная алгебра [Текст] : учебник для вузов / В. А. Ильин, Э. Г. Позняк .— 3 - е изд., доп. — М. : Наука, 1984 .— 295 с.

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

Не используются

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

Учебная аудитория, экран и проектор.

## 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)**

**по направлению:** Информатика и вычислительная техника

**профиль подготовки:** Физтех-школа Прикладной Математики и Информатики  
кафедра дискретной математики

**курс:** 3

**квалификация:** бакалавр

Семестр, формы промежуточной аттестации: 6 (весенний) - Дифференцированный зачет

**Разработчик:** Д.А. Шабанов, д-р физ.-мат. наук, доцент, доцент

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.3 Знает основные требования информационной безопасности
ОПК-5 Способен участвовать в проведении фундаментальных и прикладных исследований и разработок, самостоятельно осваивать новые теоретические, в том числе, математические методы исследований и работать на современной экспериментальной научно-исследовательской, измерительно-аналитической и технологической аппаратуре)	ОПК-5.3 Способен к профессиональной эксплуатации современной экспериментальной научно-исследовательской (измерительно-аналитической и технологической) аппаратуры
	ОПК-5.1 Способен решать поставленные задачи в области теоретических и экспериментальных исследований и разработок
	ОПК-5.2 Обладает способностью к освоению новых знаний на основе изучения литературы, научных статей и других источников
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, оценить 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-2.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации
	ПК-2.2 Способен планировать и проводить научные исследования самостоятельно или в качестве члена (руководителя) малого научного коллектива
	ПК-2.3 Способен проводить апробацию результатов научно-исследовательской работы посредством публикации научных статей и участия в конференциях

## 2. Показатели оценивания компетенций



В результате изучения дисциплины «Методы прикладной статистики» 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**

### **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 и второго родов, уровень значимости и размер критерия, функция мощности. Несмещенность и состоятельность статистического критерия.
22. Сравнения критериев, равномерно наиболее мощные критерии. Лемма Неймана-Пирсона о проверке простых гипотез. Следствие из нее. Параметрические семейства распределений с монотонным отношением правдоподобия. Теорема о семействе с монотонным отношением правдоподобия.
23. Линейная гауссовская модель, достаточная статистика в линейной гауссовской модели. Линейные гипотезы,  $F$ -критерий для проверки линейной гипотезы в гауссовской линейной модели.
24. Статистика хи-квадрат Пирсона в схеме Бернулли с  $m$  исходами. Теорема Пирсона. Критерий согласия Пирсона.

#### Критерии оценивания

- оценка «отлично (10)» 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 (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений

- оценка «хорошо (7)» 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 (6)» выставляется студенту, если он 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)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины

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

Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины.

# Методы современной прикладной статистики.

Лектор – Родионов И.В.

## 1 Аннотация курса

Курс методов современной прикладной статистики изучается студентами группы Анализа данных ФПМИ в шестом семестре после того, как они овладели основами математической статистики. Курс имеет прикладную направленность. В течение семестра студенты учатся анализировать данные и обучать статистические модели, в том числе на реальных данных. В рамках данной дисциплины студенты проходят следующие темы: критерии согласия и однородности, проверка нормальности, множественная проверка гипотез, обнаружение выбросов бутстрепные методы, корреляционный анализ, дисперсионный анализ, регрессионный анализ, ядерные оценки плотности и непараметрическая регрессия, методы понижения размерности, анализ временных рядов.

## 2 Типовые задания домашних работ.

1. Пусть  $F(x) = 1 - \exp(-(\beta x)^\alpha)$ ,  $x > 0, \alpha, \beta > 0$  – функция распределения Вейбулла. Смоделировать выборку из этого распределения размера 100. Реализовать метод спейсингов и найти этим методом оценки параметров  $\alpha$  и  $\beta$ .
2. Выдана выборка  $X = (X_1, \dots, X_n)$  из неизвестного распределения  $Q$ . Рассмотрим гипотезу  $H_0 : Q \in \mathcal{P} = \{P_\theta, \theta \in \Theta\}$ , где  $\mathcal{P}$  – некое (заданное семинаристом) семейство распределений. С помощью разделения (возможно, многократного) выборки на 2 части и построения по первой части оценки параметра  $\theta$  (можно брать функции от  $\theta$ ) проверить гипотезу  $H_0$  и объяснить, почему критерий работает. Критерий должен правильно работать на достаточно большом количестве распределений.
3. Для критерия Андерсона-Дарлинга  $\{n\Omega_n^2 > u_{1-\alpha}\}$  с помощью моделирования найти критические значения  $u_{1-\alpha}$  при  $n = 25, 100, 500, 2000$  и  $\alpha = 0.01, 0.05, 0.1$ . Пользуясь полученными критическими значениями, сравнить мощности критериев Андерсона-Дарлинга и Колмогорова при проверке гипотезы  $H_0 : P = N(0, 1)$  против альтернативы  $H_1 : P = T(10)$ , где  $T(10)$  – распределение Стьюдента с 10 степенями свободы.
4. Выданы наблюдения  $X_1, \dots, X_n$ . Проверить гипотезу о нормальности распределения наблюдений с помощью статистической процедуры, контролирующей FWER на уровне  $\alpha$ . Объяснить использование именно этой процедуры. Использовать хотя бы 5 критериев проверки нормальности.
5. Выдано  $k$  выборок  $\{X_i^{(1)}\}_{i=1}^n, \dots, \{X_i^{(k)}\}_{i=1}^n$ . Проверить гипотезу о независимости выборок методами корреляционного анализа с помощью статистической процедуры, контролирующей FWER на уровне  $\alpha = 0.05$ . Являются ли выборки независимыми в совокупности?

Если нет, укажите пары выборок, которые в результате проведённой статистической процедуры признаны зависимыми.

6. В исследовании оценивается эффективность поведенческой терапии для лечения анорексии. Для 50 пациентов известен вес до начала терапии и по её окончании. Проверить, была ли терапия эффективной, с помощью статистической процедуры, контролирующей FWER на уровне 0.05.
7. Пусть данные  $\{X_{ij}\}$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, n$ , подчиняются двухфакторной модели  $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ , где  $\{\varepsilon_{ij}\}$  – н.о.р. случайные ошибки. Предложить методику (не обязательно точную) проверки  $\{\varepsilon_{ij}\}$  на нормальность.
8. Выданы данные  $\{(y_i, x_{ij}), i = 1, \dots, n+q, j = 1, \dots, k\}$ , причем  $y_{n+1}, \dots, y_{n+q}$  неизвестны. Используя пройденные методы регрессионного анализа, в рамках линейной регрессионной модели произвести отбор признаков по первым  $n$  объектам и предсказать значения откликов объектов с номерами  $n + 1, \dots, n + q$ . Описать и объяснить проделанные процедуры.
9. Выданы данные  $\{(X_i, Y_i), i = 1, \dots, n\}$ . Рассмотрим задачу непараметрической регрессии  $Y_i = m(X_i) + \varepsilon_i$ . Построить непараметрическую регрессию с помощью оценки Надарая-Ватсона и методом сглаживающего сплайна (функция `smooth.spline` в R):

$$SS(h) = \sum_{i=1}^n (Y_i - m(X_i))^2 + h \int_{X_{(1)}}^{X_{(n)}} [m''(x)]^2 dx \longrightarrow \min_m$$

и вывести графики получившихся приближений. Зачем, на ваш взгляд, добавлять в выражение суммы квадратов интеграл от квадрата второй производной функции  $m$ ? Какой выбор  $h$ , на ваш взгляд, является оптимальным? Где отличаются графики оценки Надарая-Ватсона и сглаживающего сплайна и как вы это объясните?

10. Выданы данные  $\{(y_i, x_{ij}), i = 1, \dots, n+q, j = 1, \dots, k\}$ , причем  $y_{n+1}, \dots, y_{n+q}$  неизвестны. Используя пройденные методы регрессионного анализа и РСА, по первым  $n$  объектам в рамках линейной регрессионной модели предсказать значения откликов объектов с номерами  $n + 1, \dots, n + q$ . Описать и объяснить проделанные процедуры.
11. Выдан временной ряд  $\{X_i\}_{i=1}^{n+k}$ . С помощью модели ARIMA сделать прогноз значений ряда  $(\hat{X}_{n+1}, \dots, \hat{X}_{n+2k})$  по первым  $n$  наблюдениям. Визуализировать прогноз на графике в сравнении с истинными значениями ряда.

### 3 Критерии оценивания

Система оценивания курса следующая: в курсе 12 домашних заданий, каждое из которых состоит из 6 задач и оценивается суммарно в 12 баллов. По курсу также существует необязательный письменный экзамен, на котором можно набрать суммарно 50 баллов. Итоговая оценка за курс выставляется по формуле  $\min([A + B]/12, 10)$ , где  $A$  – сумма баллов за домашние задания,  $B$  – сумма баллов за письменный экзамен.