

**Г О У В П О Р О С С И Й С К О - А Р М Я Н С К И Й
(С Л А В Я Н С К И Й) У Н И В Е Р С И Т Е Т**

Составлен в соответствии с государственными требованиями к минимуму содержания и уровню подготовки выпускников по направлению 11.04.04 Электроника и нанoeлектроника и Положением «Об УМКД РАУ».

УТВЕРЖДАЮ:

Директор А. А. Саркисян
«21» июля 2023г.

Инженерно-физический Институт

Кафедра: Общей физики и квантовых наноструктур

Автор(ы): к.ф.-м.н., старший преподаватель Костанян Артавазд Арамович

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС

Дисциплина: Б1.В.01 «Статистический анализ данных»

Направление: 11.04.04 «Электроника и нанoeлектроника»

**Основная образовательная программа магистратуры:
«Квантовая и оптическая электроника»**

1. Аннотация

Курс знакомит студентов с основными задачами и методами статистики анализа данных.

Цель курса: связать теорию и практику, научить студентов «видеть» статистические задачи в различных предметных областях и правильно применять современные методы прикладной статистики, показать на примерах возможности и ограничения статистических методов. Курс имеет методологическую направленность и не содержит доказательств теорем.

Каждый метод описывается по единой схеме: постановка задачи; примеры прикладных задач из области физики и техники, биологии, медицины, производства; базовые предположения и границы применимости; описание метода (для методов проверки статистических гипотез: нулевая гипотеза и альтернативы, статистика, её нулевое распределение); достоинства, недостатки, ограничения, «подводные камни»; сравнение с другими методами.

Учебная задача: Задачи курса состоят в изложении современного анализа данных с использованием методов статистики, а также в усвоении практических навыков в области применения.

Основные методы проведения занятий, лекции, практические занятия, самостоятельная работа.

Список литературы: содержит 3 наименований книг зарубежных авторов; этот список поможет студентам освоить и создать свой профессиональный исследовательский инструментарий, обеспечить целостность обучения.

Краткое содержание курса: введение в анализ данных, вероятностные распределения, визуализация данных, основные методы предиктивного анализа данных, построение предсказательных моделей, факторный анализ, метод главных компонент, регрессия на главных компонентах, нелинейное оценивание, методы классификации данных: кластерный анализ, деревья классификации, методы прогнозирования: выделение тренда, сезонных составляющих, разведочный анализ данных

Взаимосвязь с другими дисциплинами специальности: компьютерные технологии в физике, методы математического моделирования, линейная алгебра и аналитическая геометрия, теория вероятностей.

2. Требования к исходным уровням знаний и умений студентов

Студент должен знать и владеть вузовским курсом математического анализа, линейной алгебры, численных методов в рамках курса математического моделирования, статистической физики в объемах курса "Физика конденсированных сред"

Знать:

Студенты должны знать основы линейной алгебры и аналитической геометрии

Уметь:

Программировать как минимум на одном современном языке программирования, пользоваться средой программирования

Владеть:

Навыками решения вычислительных задач с использованием Wolfram Mathematica, Python или на каком-либо другом языке.

3. Цель и задачи дисциплины

Основная цель изучаемой дисциплины — ознакомить студента с современными методами анализа данных с использованием статистики, а также практическими навыками анализа данных с использованием современных библиотек языка Python.

4. Требования к уровню освоения содержания дисциплины

В результате изучения дисциплины студент должен:

Знать:

Несколько методов статистики анализа данных

Уметь:

Уметь классифицировать используемый метод для анализа данных, интерпретировать результаты их обработки, а также перечислить используемых для этого программных средств.

Владеть:

Иметь навыки для реализации задачи анализа данных “с нуля”.

5. Трудоемкости дисциплины и видов учебной работы по учебному плану

Виды учебной работы	Всего (ак. час)
<i>Общая трудоемкость изучения дисциплины, в т.ч.:</i>	72 (2 кр.)
1. Аудиторные занятия, в т. ч.:	34
1.1. Лекционные занятия	18
1.2. Семинарские занятия	-
1.3. Практические занятия	16
1.4. Лабораторные работы	-
2. Самостоятельная работа, в т. ч.:	38
2.1. Контактная самостоятельная работа	-
2.2. Бесконтактная самостоятельная работа	38
<i>Итоговый контроль</i>	<i>Зачет</i>

6. Распределение весов по формам контроля

Веса и формы контролей	Вес форм текущих контролей в результирующей оценке текущего контроля			Вес форм промежуточных контролей и результирующей оценки текущего контроля в итоговой оценке промежуточного контроля			Вес итоговых оценок промежуточных контролей в результирующей оценке промежуточного контроля	Вес результирующей оценки промежуточных контролей и оценки итогового контроля в результирующей оценке итогового контроля
	M1	M2	M3	M1	M2	M3		
Вид учебной работы/ контроля								
Контрольная работа				0	0	1		
Тест								
Курсовая работа								
Лабораторные работы								
Письменные домашние задания	0	0	0					
Эссе								
Семинар	0	0	0					
Веса результирующих оценок текущих контролей в итоговых оценках соответствующих промежуточных контролей				0	0	1		
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей							0	
Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей							0	
Вес итоговой оценки 3-го промежуточного контроля в результирующей оценке промежуточных контролей							1	
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля								1
Вес оценки экзамена/зачета в результирующей оценке итогового контроля								0
	$\Sigma=0$	$\Sigma=0$	$\Sigma=0$	$\Sigma=0$	$\Sigma=0$	$\Sigma=1$	$\Sigma=1$	$\Sigma=1$

7. Содержание дисциплины

7.1 Тематический план и трудоемкости аудиторных занятий

Разделы и темы дисциплины	Всего (ак. часов)	Лекционные занятия (ак. часов)	Семинарские занятия (ак. часов)	Практические занятия (ак. часов)	Лабораторные работы (ак. часов)
1	2	3	4	5	6
МОДУЛЬ 1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ					-
<u>1. Введение в анализ данных</u>					-
1. Основная парадигма анализа данных		1			-
2. Сбор и организация данных, типы переменных, структура данных		1		1	-
3. Элементарные понятия анализа данных и статистики: переменные, описательные статистики, значимость, группировка, связи, корреляции, эффект и др.		1		1	
4. Основные статистические критерии: критерий хи-квадрат, критерий Стьюдента, критерий Фишера		1		1	-
5. Сравнение средних в двух группах, сравнение средних в более чем двух группах		1		1	-
6. Построение и анализ таблиц		1			
11. Основные статистические критерии: критерий хи-квадрат, критерий Стьюдента, критерий Фишера		1		1	-
<u>2. Вероятностные распределения</u>					-
1. Вероятностные распределения и их свойства		1		1	-
2. Подгонка вероятностных распределений к реальным данным		1		1	-
<u>3. Визуализация данных</u>					-
1. Визуальный анализ данных: одномерный визуальный		1		1	-
2. Интерактивный визуальный анализ		1		1	-
<u>4. Основные методы предиктивного анализа данных</u>					-
1. Построение предсказательных моделей:		1		1	-
• парная регрессия					-
• множественная регрессия					-
• логит регрессия					-
• пробит регрессия					-
2. Факторный анализ, метод главных компонент		1		1	-
3. Регрессия на главных компонентах		1		1	-
4. Нелинейное оценивание		1			
5. Методы классификации данных: кластерный анализ, деревья классификации		1		1	-
6. Методы прогнозирования: выделение тренда, сезонных		1		1	-
<u>5. Разведочный анализ данных</u>		1		1	-
ИТОГО	34	18		16	-

7.2 Содержание разделов и тем дисциплины

МОДУЛЬ 1.

СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

1. Введение в анализ данных
 1. *Основная парадигма анализа данных*
 2. *Сбор и организация данных, типы переменных, структура данных*
 3. *Элементарные понятия анализа данных и статистики: переменные, описательные статистики, значимость, группировка, связи, корреляции, эффект и др.*
 4. *Основные статистические критерии: критерий хи-квадрат, критерий Стьюдента, критерий Фишера*
 5. *Сравнение средних в двух группах, сравнение средних в более чем двух группах*
 6. *Построение и анализ таблиц*
2. Вероятностные распределения
 1. *Вероятностные распределения и их свойства*
 2. *Подгонка вероятностных распределений к реальным данным*
3. Визуализация данных
 1. *Визуальный анализ данных: одномерный визуальный анализ, двумерный визуальный анализ*
 2. *Интерактивный визуальный анализ*
4. Основные методы предиктивного анализа данных
 1. *Построение предсказательных моделей:*
 - *парная регрессия*
 - *множественная регрессия*
 - *логит регрессия*
 - *пробит регрессия*
 2. *Факторный анализ, метод главных компонент*
 3. *Регрессия на главных компонентах*
 4. *Нелинейное оценивание*
 5. *Методы классификации данных: кластерный анализ, деревья классификации*
 6. *Методы прогнозирования: выделение тренда, сезонных составляющих*
5. Разведочный анализ данных

7.3 Вопросы

1. *Основная парадигма анализа данных*
2. *Сбор и организация данных, типы переменных, структура данных*
3. *Элементарные понятия анализа данных и статистики: переменные, описательные статистики, значимость, группировка, связи, корреляции, эффект и др.*
4. *Основные статистические критерии: критерий хи-квадрат, критерий Стьюдента, критерий Фишера*
5. *Сравнение средних в двух группах, сравнение средних в более чем двух группах*
6. *Построение и анализ таблиц*
7. *Вероятностные распределения и их свойства*
8. *Подгонка вероятностных распределений к реальным данным*
9. *Визуальный анализ данных: одномерный визуальный анализ, двумерный визуальный анализ*
10. *Интерактивный визуальный анализ*
11. *Построение предсказательных моделей:*
12. *Парная регрессия, множественная регрессия*
13. *Логит регрессия, пробит регрессия*
14. *Факторный анализ, метод главных компонент*
15. *Регрессия на главных компонентах*
16. *Нелинейное оценивание*
17. *Методы классификации данных: кластерный анализ, деревья классификации*
18. *Методы прогнозирования: выделение тренда, сезонных составляющих*
19. *Разведочный анализ данных*

8. Учебно-методическое обеспечение дисциплины

8.1. Рекомендуемая литература

а) Основная литература

1. Jay L. Devore, Probability and Statistics for Engineering and the Sciences, Eighth Edition, 2010
2. Peter H. Westfall, Kevin S. S. Henning, Understanding Advanced Statistical Methods, 2013

б) Дополнительная литература

1. Wes McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython, Second Edition, 2018

8.2. Программные средства освоения дисциплины

Python 3.6 и выше, библиотеки Pandas, NumPy, Scikit-learn.

8.3. Материально-техническое обеспечение дисциплины

Компьютерная аудитория , проектор