ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол №1

Рабочая программа дисциплины (модуля) «Современные методы статистики»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Искусственный интеллект*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B80C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2024

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 7
4. **Содержание дисциплины (модуля)** 7
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Современные методы статистики» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Искусственный интеллект», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Современные методы статистики» позволяет освоить строгие статистические методы, необходимые для корректного анализа данных в условиях реальной неопределённости, эндогенности и сложной структуры. Полученные знания и навыки формируют фундамент для проведения научных исследований, A/B тестирования, эконометрического моделирования и причинно-следственного анализа в науке и индустрии.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Искусственный интеллект» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 3 курсе в 5 семестре на выбор. Доступна к изучению после успешного освоения дисциплины (модуля): «Математическая статистика».

Цель изучения дисциплины (модуля): сформировать у обучающихся глубокое понимание современных статистических методов, включая причинно-следственный вывод, асимптотические и байесовские подходы, а также методы оценки эффектов воздействия в условиях сложных данных и эндогенности.

Задачи изучения дисциплины (модуля):

— изучить базовые и продвинутые методы математической статистики, включая асимптотику, бутстрап, методы коррекции смещения и условия применимости оценок метода наименьших квадратов;

— освоить теоретические основы линейной регрессии, включая теоремы Гаусса-Маркова и Фриша-Ву-Ловелла, и научиться интерпретировать результаты оценивания;

— научиться выявлять и устранять смещение оценок, вызванное пропущенными переменными и эндогенностью, с применением причинно-следственных моделей и метода инструментальных переменных;

— освоить современные подходы к выбору моделей и оценке эффектов воздействия, включая двойное машинное обучение и анализ на причинных графах;

— развить навыки байесовского вывода с использованием MCMC-методов, фильтров Калмана и многочастичных фильтров для работы с моделями пространства состояний.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— базовые понятия и методы статистики: асимптотика, A/B тестирование, виды бутстрапа, метод Холла;
— теоремы линейного регрессионного анализа: Фриша-Ву-Ловелла, Гаусса-Маркова, свойства оценок МНК;
— причины смещения оценок: пропущенные переменные, эндогенность;
— причинно-следственные модели: модель рубина, причинные графы, критерий обходных путей;
— подходы к выбору моделей и проблемы неравномерной инференции по Либу-Потчеру;
— основы методов двойного машинного обучения;

— основы байесовской статистики и MCMC: EM-алгоритм, сэмплер Гиббса, метод Метрополиса-Гастингса, модель стохастических блоков, фильтры Калмана и многочастичный.

уметь:

— применять асимптотические методы для построения доверительных интервалов и статистических тестов;

— использовать бутстрап для коррекции смещения и оценки неопределённости;

— оценивать простую и множественную регрессию методом наименьших квадратов и проверять условия теоремы гаусса-маркера;

— интерпретировать результаты регрессионного анализа и выявлять смещение из-за пропущенных переменных;

— применять метод инструментальных переменных (2SLS) для оценки эффектов воздействия в условиях эндогенности;

— строить байесовские A/B тесты с расчётом минимального детектируемого эффекта;

— применять EM -алгоритм для решения задач классификации и кластеризации.

владеть:

— методами бутстрапа различных видов, включая метод Холла, для коррекции смещения и построения доверительных интервалов в сложных моделях;

— техникой двойного машинного обучения (двойное LASSO) для оценки причинных эффектов при большом числе контрольных переменных;

— причинно-следственным выводом на графах с использованием критерия обходных путей и моделью рубина для формализации эффектов воздействия;

— методами MCMC, включая сэмплер Гиббса и алгоритм Метрополиса-Гастингса, для байесовского вывода в моделях пространства состояний и стохастических блоковых моделях;

— фильтром Калмана и многочастичным фильтром для анализа и прогнозирования в моделях пространства состояний.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности.
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области

		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов.
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты.
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области искусственного интеллекта, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары			
1	A/B -тестирование	4	4		16	Домашнее задание
2	Линейные модели и оценка эффектов воздействия в неэкспериментальных данных	18	16	2	78	Домашнее задание Контрольная работа
3	Байесовские методы	8	8		34	Домашнее задание
	Экзамен			2		
	Итого:	***30***	***28***	***4***	***128***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	A/B -тестирование	Базовые понятия и методы статистики. Асимптотические методы для тестов и доверительных интервалов. Эффекты воздействия, как параметры распределения: модель Рубина. Экспериментальная оценка эффектов воздействия, A/B тестирование Бутстрап для коррекции смещения оценок и инференции. Метод Холла. Применение бутстрапа при A/B тестировании
2	Линейные модели и оценка эффектов воздействия в неэкспериментальных данных	Линейные модели с стохастическими регрессорами. Отличие прогнозной и причинно-следственной задачи при исследовании данных. Оценки МНК и проекции. Теорема Фриша-Ву-Ловелла Регрессия по всей популяции. Идентификация параметров линейной модели. Наилучший линейный предиктор. L2 пространство и его свойства. Проекция в L2 Предположения линейной модели и статистические свойства МНК оценок. Теорема Гаусса-Маркова. Тесты: t,F Асимптотические свойства оценок. Тесты параметров линейной регрессии: z-тест, тест Вальда. Миссспецификация модели: смещение из-за пропущенных переменных Контрольные переменные, инструментальные переменные. Выбор модели: секвенциальное тестирование, информационные критерии Неравномерная инференция по Либу-Потчеру Использование графических моделей для выбора регрессионной модели. Критерий обходных путей LASSO, post-LASSO, двойное LASSO Двойное машинное обучение

3	Байесовские методы	Постановка задачи байесовской статистики, примеры. Условные распределения, свойства нормальных векторов Модели пространства состояний: фильтр Калмана, многочастичный фильтр Сэмплер Гиббса, алгоритм Метрополиса-Гастингса, сходимость. Модель стохастических блоков Байесовское A/B тестирование, минимальный выявляемый эффект. Байесовские алгоритмы классификации и кластеризации: EM-алгоритм

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Мэрфи, К. П. Вероятностное машинное обучение: введение : практическое руководство / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2023. - 990 с. - ISBN 978-5-93700-119-1. - Текст : электронный. - URL: https://znanium.com/catalog/product/2109489.

2. Курт, У. Байесовская статистика: Star Wars, LEGO, резиновые уточки и многое другое : практическое руководство / У. Курт. - Санкт-Петербург : Питер, 2021. - 304 с. - (Библиотека программиста). - ISBN 978-5-4461-1655-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/1817495.

Дополнительная литература:

1. Брюс, П. Практическая статистика для специалистов Data Science / П. Брюс, Э. Брюс. - 2-е изд. - Санкт-Петербург : БХВ-Петербург, 2021. - 352 с. - ISBN 978-5-9775-6705-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2142261.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование"	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое

Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Современные методы статистики» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, контрольная работа, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удается, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Контрольная работа – письменная работа с набором задач, которые нужно решить за ограниченное время.

Цель контрольной работы – получить специальные знания по одной или нескольким темам дисциплины и продемонстрировать навыки их практического применения.

Бонусные баллы — это оценки, которые студенты могут получить за выполнение дополнительных заданий.

Формат бонусных баллов позволяет студентам улучшить общую оценку по дисциплине (модулю) и стимулирует углубленное изучение материала.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Современные методы статистики»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме ***экзамена***, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	**Пятибалльная оценка**	**Общая характеристика результата обучения по дисциплине (модулю)**
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	
8	Отлично	
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Современные методы статистики» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	30%	7	Набор задач по темам недели
Контрольная работа	20%	1	Письменная работа с набором задач, которые нужно решить за ограниченное время
Экзамен	50%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

Формула расчёта итоговой оценки по дисциплине (модулю) «Современные методы статистики»: «0,3 × среднее за домашние задания + 0,2 × контрольная работа + 0,5 × экзамен».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

1. Объясните, в чём заключается разница между асимптотическими и бутстрап-подходами к построению доверительных интервалов в A/B тестировании;
2. Опишите, как модель Рубина формализует понятие эффекта воздействия и какие предположения необходимы для его идентификации;
3. Реализуйте бутстрап-оценку разности средних в A/B тесте и сравните полученный доверительный интервал с асимптотическим подходом;
4. Объясните, как метод Холла улучшает точность бутстрап-интервалов по сравнению с обычным процентильным методом;
5. Проанализируйте, в каких случаях стандартное A/B тестирование может давать смещённые оценки и как бутстрап помогает скорректировать неопределённость.

Домашнее задание 2.

1. Оцените модель линейной регрессии методом МНК и проверьте выполнение предпосылок теоремы Гаусса-Маркова на примере реальных или синтетических данных;

2. Примените теорему Фриша-Ву-Ловелла для интерпретации коэффициентов множественной регрессии как частичных эффектов;

3. Объясните, как пропущенная переменная, коррелирующая с регрессором, приводит к смещению оценки МНК, и продемонстрируйте это на примере;

4. Реализуйте оценку эффекта воздействия с помощью метода инструментальных переменных (2SLS) в случае эндогенности регрессора;

5. Примените двойное LASSO для отбора контрольных переменных и оценки причинного эффекта в условиях большого числа ковариат.

Домашнее задание 3.

1. Постройте байесовский A/B тест для биномиальных данных, используя сопряжённые априорные распределения, и рассчитайте вероятность превосходства одной группы над другой;

2. Реализуйте EM-алгоритм для задачи кластеризации методом смеси нормальных распределений и проанализируйте сходимость;

3. Опишите, как работает сэмплер Гиббса в байесовской линейной регрессии, и реализуйте его на простом примере;

4. Примените фильтр Калмана для оценки скрытого состояния в модели случайного блуждания с наблюдениями с шумом;

5. Объясните, в каких случаях предпочтительнее использовать многочастичный фильтр вместо фильтра Калмана, и опишите его основные шаги: предсказание, взвешивание, ресэмплирование.

Примерные задания для контрольной работы

1. Объясните, в чём заключается различие между прогнозной и причинно-следственной интерпретацией модели линейной регрессии, и приведите пример, когда высокая предсказательная способность не означает корректной оценки эффекта воздействия;

2. Опишите, как модель Рубина формализует понятие потенциальных исходов и какие предположения (SUTVA, случайная выборка, отсутствие скрытой эндогенности) необходимы для идентификации среднего эффекта обработки (ATE);

3. Реализуйте A/B тест на синтетических данных: симулируйте две группы, оцените разность средних, постройте доверительный интервал с использованием асимптотического подхода и сравните с результатом бутстрапа;

4. Объясните, как метод Холла улучшает точность бутстрап-доверительных интервалов за счёт учёта смещения и асимметрии распределения оценки;

5. Оцените модель линейной регрессии методом МНК на примере набора данных и проверьте выполнение ключевых предпосылок теоремы Гаусса-Маркова: линейность, случайность выборки, отсутствие мультиколлинеарности, гомоскедастичность, ненормальность остатков;

6. Примените теорему Фриша-Ву-Ловелла для интерпретации коэффициента в множественной регрессии как результата двухэтапной проекции: сначала зависимой переменной, затем регрессора на пространство остальных предикторов;

7. Объясните, как пропущенная переменная, коррелирующая с включённым регрессором, приводит к смещению оценки МНК, и продемонстрируйте этот эффект на численном примере;

8. Реализуйте оценку эффекта воздействия с помощью метода инструментальных переменных (2SLS) в случае, когда регрессор эндогенен из-за обратной связи или скрытой переменной;

9. Опишите, как критерий обходных путей (backdoor criterion) используется в причинных графах для выбора корректного набора контрольных переменных, и приведите пример графа, где контроль по одной переменной устраняет смещение;

10. Примените технику двойного машинного обучения (double machine learning) с двойным LASSO для оценки причинного эффекта в условиях большого числа потенциальных ковариат, используя модель предсказания результата и модель предсказания воздействия.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Назовите основную цель a/b тестирования в контексте причинно-следственного вывода	оценка эффекта воздействия путём сравнения контрольной и тестовой групп	УК-1
2	Укажите, как в модели рубина определяется потенциальный исход для индивида при отсутствии воздействия	y_i(0) — результат, который был бы наблюдаем, если бы индивид не получил treatment	УК-1
3	Объясните, что означает предпосылка SUTVA в модели потенциальных исходов	отсутствие взаимодействия между индивидами и однозначность воздействия	УК-1
4	Назовите условие, при котором оценка ATE является несмещённой в рандомизированном эксперименте	случайное назначение воздействия и отсутствие систематических различий между группами	УК-1
5	Назовите асимптотический тест для проверки гипотезы о разности пропорций в a/b тесте	z-тест для долей	УК-2
6	Объясните, зачем используется бутстрап при построении доверительных интервалов в условиях малой выборки	для аппроксимации распределения статистики без предположений о нормальности	УК-2
7	Укажите, в чём заключается преимущество метода Холла при построении бутстрап-интервалов	он корректирует смещение и асимметрию, обеспечивая более точные покрытия	УК-2
8	Назовите проблему, возникающую при анализе данных без учёта множественной проверки гипотез	увеличение вероятности ложноположительного результата (family-wise error rate)	УК-2

9	Назовите теорему, утверждающую, что МНК-оценки являются наилучшими линейными несмещёнными в условиях выполнения классических предпосылок	теорема Гаусса-Маркова	ОПК-1
10	Объясните, как теорема Фриша-Ву-Ловелла помогает интерпретировать коэффициенты множественной регрессии	она показывает, что коэффициент — это эффект после "очистки" от влияния других переменных	ОПК-1
11	Укажите, какое свойство оценок МНК нарушается при наличии эндогенности регрессора	несмещённость и состоятельность	ОПК-1
12	Назовите тест, используемый для проверки значимости группы коэффициентов в линейной модели	F-тест	ОПК-1
13	Назовите метод, позволяющий оценить причинный эффект в условиях эндогенности с помощью дополнительной переменной	метод инструментальных переменных (IV)	ПК-1
14	Объясните, как критерий обходных путей помогает выбрать набор контрольных переменных в причинном графе	он определяет, какие переменные нужно контролировать, чтобы закрыть все backdoor-пути	ПК-1
15	Назовите подход, позволяющий отбирать переменные и оценивать причинный эффект при большом числе ковариат	двойное машинное обучение с двойным LASSO	ПК-1
16	Назовите метод, используемый для оценки параметров в моделях с пропущенными данными или латентными переменными	EM-алгоритм	ПК-1
17	Назовите байесовский подход к A/B тестированию, основанный на апостериорном распределении вероятности превосходства	байесовское A/B тестирование с расчётом posterior probability	ПК-2
18	Укажите, как сэмплер Гиббса используется для байесовского вывода в моделях со скрытыми переменными	путём последовательной выборки из условных распределений	ПК-2
19	Назовите рекурсивный алгоритм для оценки скрытого состояния в линейной модели с гауссовским шумом	фильтр Калмана	ПК-2
20	Назовите метод, используемый для байесовского вывода в нелинейных и не гауссовских моделях пространства состояний	многочастичный фильтр / particle filter	ПК-2