

УТВЕРЖДЕНА

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

**Рабочая программа дисциплины (модуля)
«Математическая статистика»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Программа двух дипломов НИУ
ВШЭ и ЦУ «Прикладная математика и информатика»

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2024

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения	5
3. Тематический план	5
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) «Математическая статистика» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Математическая статистика» является ключевым инструментом в областях программирования. Освоение статистических методов позволяет эффективно обрабатывать данные, выявлять закономерности и делать прогнозы, что является важным для успешной профессиональной деятельности.

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 «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в обязательную часть Блока 1, как дисциплина по выбору.

Дисциплина (модуль) доступна к изучению на 2 курсе в 4 семестре.

Цель изучения дисциплины (модуля): в формировании навыков сбора, анализа и интерпретации статистических данных для принятия обоснованных решений.

Задачи изучения дисциплины (модуля):

— освоение основных понятий и определений: описательные статистики, графики, случайность, независимость, выборку, генеральную совокупность, произвольное дискретное распределение, вероятность, матожидание, дисперсию, биномиальное, нормальное распределения и оценки их параметров, гипотезы, тесты, уровень значимости и мощность, тесты среднего, пропорции;

— развитие умений в заданном датасете, интерпретации описательных статистик, графики, сравнить описательные статистики в выборке и в генеральной совокупности, объяснить, почему наблюдаемые числовые характеристики выборки не обязательно соответствуют характеристикам генеральной совокупности

— развитие навыков проведения стандартных тестов для среднего и пропорции, интерпретирование результатов тестов, интерпретирование p -value для заданной пары нулевой и альтернативной гипотезы.

В результате освоения дисциплины (модуля), обучающийся должен:

знать:

— основные понятия теории вероятностей и статистики: выборка, распределение, параметр, оценка, гипотеза;

— свойства и характеристики выборочных распределений, и их связь с параметрами генеральной совокупности;

— принципы построения доверительных интервалов и проверки статистических гипотез;

— методы корреляционного анализа и область его применимости;

— основы непараметрических методов и особенности работы со статистическими моделями;

уметь:

— строить и интерпретировать выборочные характеристики (средние, дисперсии, квантили и др.);

— вычислять доверительные интервалы и проводить проверку гипотез для различных типов данных;

— применять методы корреляционного анализа для исследования зависимостей;

- использовать непараметрические критерии в ситуациях, когда предпосылки параметрических методов нарушены;
- выбирать подходящие статистические инструменты для решения практической задачи и обоснованно интерпретировать полученные результаты;

владеть:

- навыками статистического анализа данных и построения обоснованных выводов;
- способностью оценивать корректность применяемых статистических методов и их ограничений;
- приёмами визуализации данных и представления статистических результатов в аналитически понятной форме;
- навыками использования программных средств для статистических расчётов;
- культурой критического анализа данных, включая проверку предпосылок, выявление выбросов и оценку устойчивости результатов.

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 консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и	ОПК-1.1.	Знает основные концепции и теории в области математического анализа и смежных дисциплин; методы и подходы, используемые в различных областях математики
		ОПК-1.2.	Умеет применять математические методы для решения профессиональных задач
		ОПК-1.3.	Имеет практический опыт разработки и реализации математических моделей в профессиональной деятельности

	математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности		
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 разработки, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Контактная работа		Контроль	Самостоятельная работа	
Лекции	Семинары					
1	Параметрическая статистика: точечное оценивание	6	6		24	Домашнее задание Контест
2	Параметрическая статистика: интервальное оценивание	6	6		25	Домашнее задание Контрольная работа
3	Параметрическая статистика: проверка гипотез	6	6		25	Домашнее задание Контест
4	Непараметрическая статистика	6	6		25	Домашнее задание
5	Специальные методы статистики	6	6		25	Контест
	<i>Экзамен</i>			6		
	<i>Итого:</i>	<i>30</i>	<i>30</i>	<i>6</i>	<i>124</i>	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	<i>190</i>				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	<i>5</i>				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Параметрическая статистика: точечное оценивание	<p>Введение в параметрическую статистику</p> <p>1. Вероятностно-статистическая модель. Понятия наблюдения и выборки.</p> <p>2. Параметрическая статистическая модель. Статистики и оценки. Примеры статистик: выборочные характеристики, порядковые статистики.</p> <p>3. Эмпирическое распределение и эмпирическая функция распределения. Теорема Гливленко-Кантелли.</p> <p>4. Основные свойства оценок (несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность) и взаимосвязи между ними.</p> <p>Сходимость случайных векторов.</p> <p>1. Виды сходимостей случайных векторов и связи между ними. Связь между сходимостью векторов и сходимостью их компонент. Теорема о наследовании сходимости.</p> <p>2. Закон больших чисел, усиленный закон больших чисел и многомерная центральная предельная теорема для случайных векторов (б/д).</p> <p>3. Лемма Слуцкого. Пример применения леммы Слуцкого и его обобщение на многомерный случай (доказательство для одномерного случая).</p> <p>Сходимость случайных векторов.</p> <p>1. Наследование состоятельности и сильной состоятельности при взятии непрерывной функции.</p>

		<p>2. Лемма о наследовании асимптотической нормальности. + сходимость по распределению. Связь слабой и по распределению. Оценки.</p> <p>1. Метод подстановки и метод моментов, их связь. Состоятельность и асимптотическая нормальность оценки метода моментов. Квантили и выборочные квантили.</p> <p>2. Сравнение оценок, функция потерь и функция риска. Подходы к сравнению оценок: равномерный, байесовский, минимаксный, асимптотический.</p> <p>Эффективные оценки.</p> <p>1. Неравенство Рао-Крамера и эффективные оценки. Критерий эффективности оценки.</p> <p>2. Экспоненциальные семейства распределений. Их связь с условием существования эффективной оценки.</p> <p>Метод максимального правдоподобия.</p>
2	Параметрическая статистика: интервальное оценивание	<p>Доверительные интервалы.</p> <p>1. Постановка задачи доверительного оценивания. Метод центральной статистики.</p> <p>2. Асимптотические доверительные интервалы.</p> <p>Классические методы доверительного оценивания</p> <p>1. Доверительный интервал для среднего</p> <p>2. Доверительный интервал для доли</p> <p>3. Доверительный интервал для разницы средних.</p> <p>Бутстреп</p>
3	Параметрическая статистика: проверка гипотез	<p>Основы проверки статистических гипотез</p> <p>1. Основные понятия: нулевая и альтернативная гипотеза.</p> <p>2. Ошибки I и II рода, уровень значимости (α), мощность критерия (β).</p> <p>3. р-значение</p> <p>Построение тестов</p> <p>1. Лемма Неймана-Пирсона.</p> <p>2. Теорема о монотонном отношении правдоподобия. Построение равномерно наиболее мощных критериев для односторонних альтернатив.</p> <p>3. Двойственность доверительного оценивания и проверки гипотез.</p> <p>Практика проверки гипотез</p> <p>1. Решение комплексных задач на проверку гипотез.</p> <p>2. Особенности проведения и интерпретации результатов:</p> <p>2.1 Кейсы из практики.</p> <p>2.2 Возможные ошибки и способы их избежания.</p>
4	Непараметрическая статистика	<p>Основы непараметрической статистики</p> <p>1. Общие принципы непараметрических методов.</p> <p>2. Ранговая статистика:</p> <p>2.1 Коэффициенты ранговой корреляции (Спирмена, Кендалла).</p> <p>2.2 Ранговые суммы Вилкоксона.</p> <p>3. Критерий знаков и критерий Манна-Уитни.</p> <p>Непараметрические критерии согласия</p> <p>1. Критерий Колмогорова-Смирнова.</p> <p>2. Хи-квадрат тест Пирсона.</p> <p>3. Критерий Андерсона-Дарлинга.</p> <p>4. Непараметрические аналоги ANOVA:</p> <p>4.1 Критерий Крускал-Уоллиса.</p> <p>4.2 Н-статистика Фридмана.</p>
5	Специальные методы статистики	<p>Метод Монте-Карло.</p> <p>1. Variance reduction</p> <p>2. Importance sampling</p>

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. Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебник для вузов / В. Е. Гмурман. — 11-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 406 с. — (Высшее образование). — ISBN 978-5-534-08389-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559583>.

Дополнительная литература:

1. Гмурман, В. Е. Теория вероятностей и математическая статистика : учебник для вузов / В. Е. Гмурман. — 12-е изд. — Москва : Издательство Юрайт, 2025. — 479 с. — (Высшее образование). — ISBN 978-5-534-00211-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559584>.

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) «Математическая статистика» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары, контесты, контрольные работы и домашние задания, 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 удастся, 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.

Контест – интерактивная платформа с заданиями разного уровня сложности и автоматической проверкой результатов.

Контест позволяет оперативно оценивать усвоение материала и выявлять пробелы в знаниях через тесты и практические задачи. Такой формат способствует регулярной самопроверке и повышает мотивацию к изучению дисциплины.

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.

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.

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) «Математическая статистика»

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:

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 исследования, необходимыми для практической деятельности, и умеет связывать
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		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) «Математическая статистика» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	15%	Работа с данными, датасетами, решение математических задач
Контесты	30%	Устные или письменные ответы на вопросы, список которых известен студенту заранее
Контрольные работы	20%	Письменная работа с набором задач, которые нужно решить за ограниченное время
Экзамен	35%	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Математическая статистика»: $\langle 0,15 \times \text{среднее за домашние задания} + 0,3 \times \text{среднее за контесты} + 0,2 \times \text{среднее за контрольные работы} + 0,35 \times \text{за экзамен} \rangle$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные задания для контестов

Анализ клиентских предпочтений с использованием методов выборки и моделирования

Описание задачи:

Ваша компания планирует запустить новый продукт на рынок, и вам необходимо провести анализ клиентских предпочтений, чтобы понять, какова вероятность успешного запуска. Для этого вы будете использовать методы статистики, включая выборки, случайность, события и метод Монте-Карло.

Задача:

- Определение генеральной совокупности и выборки:**
 - Определите, какая является генеральная совокупность (например, все потенциальные клиенты вашей компании) и выберите случайную выборку (например, 500 респондентов) для анализа.
- Сбор данных:**

- Разработайте анкету, которая будет содержать вопросы о предпочтениях клиентов относительно нового продукта. Включите как количественные, так и качественные вопросы.

3. Анализ данных:

- Проанализируйте собранные данные, используя методы описательной статистики. Определите средние значения, медианы, моду, а также распределение ответов.
- Оцените независимость событий (например, предпочтение продукта в зависимости от возраста или пола).

4. Метод Монте-Карло:

- Используйте метод Монте-Карло для моделирования различных сценариев запуска продукта. Например, смоделируйте, как различные уровни маркетинговых расходов могут повлиять на вероятность успешного запуска.
- Проведите 1000 симуляций и проанализируйте полученные результаты.

5. Выводы и рекомендации:

- На основе анализа данных и результатов моделирования, сформулируйте рекомендации по запуску нового продукта. Укажите, какие факторы наиболее критичны для успеха.

Критерии оценивания:

1. Определение генеральной совокупности и выборки (20%):

- Четкость и обоснованность определения генеральной совокупности и выбранной выборки.

2. Сбор и анализ данных (30%):

- Качество разработанной анкеты, полнота и точность собранных данных, корректность анализа.

3. Применение метода Монте-Карло (30%):

- Правильность реализации метода, количество проведенных симуляций, качество интерпретации результатов.

4. Выводы и рекомендации (20%):

- Обоснованность и практическая ценность выводов и рекомендаций, ясность изложения.

Формат представления:

- Итоговый отчет должен быть оформлен в виде документа (PDF или Word) и включать все этапы работы.

- Презентация (10-15 минут) должна содержать основные результаты, визуализации и ключевые рекомендации.

Примерные задания для контрольных работ

Контрольная работа

1. Вероятностно-статистическая модель. Наблюдение и выборка.

Пусть $(\Omega, \mathcal{F}, P, \theta)$, $\theta \in \Theta$ — семейство вероятностных территорий, $a: \Omega \rightarrow P$ — зона.

1. Дайте определение: *наблюдение*, *выборка* X_1, \dots, X_n , *статистическая модель*.

2. При вводе параметрической модели, где X_j независимые, но не такие распределенные (укажите Θ , распределение).

3. Объясните, что означает «выборка из распределения». $\Phi\theta$.

2. Статистика и рейтинг. Примеры статистик.

Для выборки X_1, \dots, X_n из непрерывного распределения:

1. показать, что $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ являются статистиками.

- ли статистической величиной $X_1 - \theta$? Обоснуйте.
- Кто из государственных статистиков является *порядковыми статистиками*?
- Эмпирическая функция распределения.**

Пусть X_1, \dots, X_n — выборка из распределения Φ .

1. Запишите определение эмпирического распределения и эмпирической функции распределения. $F_n(x)$.

2. Вычислите $F_n(x)$ для выборок $x = (2, 2, 5, 7, 7)$ (в виде кусочно-постоянной функции).

3. Найти $|F_n(x) - F(x)|$ если Φ — шире на $[0, 10]$ (для данных выборок).

4. **Теорема Гливленко–Кантелли (понимание формулировок).**

1. Сформулируйте выводы Гливленко–Кантелли.

2. Объясните, чем отличается поточечная сходимость $F_n(x) \rightarrow F(x)$ от расширенного пох.

3. Укажите, как следует наблюдать за сходством случайных крупных утверждений Совета Европы.

5. **Свойства оценок и связи между ними.**

1. Дайте определения: несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность.

2. Верно ли: сильная состоятельность \Rightarrow состоятельность?

состоятельность \Rightarrow несмещенность? Обоснуйте (можно контрпримером).

3. Приведем пример состоятельной, но смещенной оценки.

6. **Виды обнаружения случайных векторов и связей.**

Пусть $Y_n \in \mathcal{R}^d, Y \in \mathcal{R}^d$.

1. Дайте определение сходимости $Y_n \rightarrow Y$ по возможности, почти вероятно, по распределению.

2. Докажите/обосните: если $Y_n \rightarrow Y$ по возможности, то для любой непрерывной функции $g: \mathcal{R}^d \rightarrow \mathcal{R}$ выполнено $g(Y_n) \rightarrow g(Y)$ по вероятности.

3. Верно ли, что $Y_n \rightarrow Y$ по распределению \Leftrightarrow по комменно $Y_n, u \rightarrow Y$ по распределению? Уточните условия/комментарии.

7. **Наследование сходимости (теорема о наследовании).**

Пусть $Y_n \rightarrow Y$ по распределению в \mathcal{R}^d .

1. Сформулируйте выводы о наследовании (теорема о непрерывном отображении).

2. Примените ее к $g(y) = \|y\|$ и сделаю вывод о сходимости $\|Y_n\|$.

3. Примените к $g(y) = Ay + b$ где A — матрица, b — ми.

8. **Закон больших чисел (векторный).**

Пусть ξ_1, ξ_2, \dots — независимые одинаковые распределённые случайные реплики в $\mathcal{R}^d, E \|\xi_1\| < \infty$.

1. Сформулируйте многомерный закон больших чисел для $\xi_n = \frac{1}{n} \sum_{k=1}^n \xi_k$.

2. Объясните, как следует из одномерного ЗБЧ (через компоненты).

3. При каких условиях можно сформулировать усиленный ЗБЧ (п.н. сходимость)?

9. **Многомерная ЦПТ (формулировка + простое приложение).**

Пусть $\xi_k \in \mathcal{R}^d$ — iid, $E \xi_1 = \mu$, Ковариант $(\xi_1) = \Sigma$ (конечный).

1. Сформулируйте многомерную ЦПТ для $\xi_n = \frac{1}{n} \sum_{k=1}^n \xi_k$.

2. Йендер $d=2: \xi_k = (X_k, X_k^2)$ где $X_k \sim N(0, 1)$ Найти μ и Σ и запишите превышение веса $n(\xi_n - \mu)$.

10. **Лемма Слуцкого и пример применения (одномерный случай).**

1. Сформулируйте лемму Слуцкого в одномерном случае.

2. Укажите ее для случая: $X_n \Rightarrow X, Y_n \rightarrow P, \dots, X_n + Y_n \Rightarrow X + c$.

3. Примените: если $\ln(X - \mu) \Rightarrow N(0, \sigma^2)$ и $S_n^2 \rightarrow P\sigma^2$, чтобы прийти к предельному уровню доходов $\ln S_n X - \mu$.

11. Наследование состоятельности/сильной состоятельности непрерывного снижения.

Пусть $\hat{\theta}_n$ — состоятельная (или сильно состоятельная) оценка θ условна.

1. Сформулируйте и дайте заключение о наследовании состоятельности $g(\hat{\theta}_n)$ как оценки $g(\theta)$.

2. Примените к $g(\hat{\theta}_n) = m$ и $g(\theta) = m$ (укажите необходимые условия во втором случае).

12. Наследование асимптотической нормальности (дельта-метод / лемма).

Пусть $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \tau^2)$, g дифференцируемая в θ , $g'(\theta) \neq 0$.

1. Сформулируйте результат о пределе распределения $\sqrt{n}(g(\hat{\theta}_n) - g(\theta))$.

2. Примените к $g(\hat{\theta}_n) = \theta^2$ при $\theta \neq 0$.

3. Примените к $g(\hat{\theta}_n) = \ln \hat{\theta}$ при $\theta > 0$.

13. Метод подстановки и метод моментов. Квантили.

Пусть $X_1, \dots, X_n \sim \text{Эксп}(\lambda)$ с катая $f(x) = \lambda e^{-\lambda x}, x \geq 0$.

1. Найти λ методом моментов.

2. Используйте «подстановочную» меру для медианы $m(\lambda)$ через λ ММ.

3. Определите выборочную медиану (выбор квантильного уровня 0,5) и предлагаем оценку из равенства «теоретическая медиана = выборочная медиана».

14. Сравнение оценок: функциональные потери и риск.

Пусть $X_1, \dots, X_n \sim N(\theta, 1)$. две оценки: $\delta_1 = \bar{X}$ и $\delta_2 = 0$.

1. Для квадратичной потери $L(\theta, a) = (a - \theta)^2$ найти предсказуемые $R(\theta, \delta_1), R(\theta, \delta_2)$.

2. Для всех θ δ_1 лучше δ_2 по риску?

3. Сформулируйте, что означает равномерное доминирование одной оценки над другой.

15. Неравенство Рао–Крамера, эффективность, экспоненциальные компоненты и ММП.

Пусть $X_1, \dots, X_n \sim \text{Пуассона}(\lambda), \lambda > 0$.

1. Наблюдайте за максимальным правдоподобием $\hat{\lambda}_{MLE}$.

2. проверить несмещенность $\hat{\lambda}_{MLE}$ и найти ее дисперсию.

3. Вычислите информацию Фишера $I(\lambda)$ и ближайшая граница Рао – Крамера для несмещенных оценок λ .

4. эффективность $\hat{\lambda}_{MLE}$?

5. Запишите распределение Пуассона как одномерное экспоненциальное семейство и укажите, какая статистика является достаточно естественной.

Домашнее задание

Имеется информация за 10 дней в агрегаторе такси:

- Количество заказов = 8770.
- Среднее время поиска и назначения водителя на заказ составляет 2 минуты и 1 секунду (стандартное отклонение - 32 секунды).
- Конверсия из заказа в назначение водителя на заказ равна 94%
- Среднее время подачи такси после назначения водителя на заказ составляет 5 минут и 16 секунд (стандартное отклонение - 1 минута и 5 секунд).
- Конверсия из назначения в прибытие автомобиля в точку А составляет 82%.
- Среднее время поездки (от точки А до точки В) составляет 24 минуты и 36 секунд (стандартное отклонение - 7 минут и 2 секунды).
- Конверсия из прибытия автомобиля в точку А в прибытие в точку В составляет 88%.

Задача 1 (4 балла)

Вы заказываете такси.

Можно ли утверждать, что с вероятностью 95% вы будете возле точки **В** раньше чем через 35 минут?

Подсказка

Для ответа на вопрос рассчитайте границы доверительного интервала для трёх временных интервалов между этапами воронки.

Задача 2 (4 балла)

Предположите, что все заказы были равномерно распределены по 10 дням, а конверсия была неизменной в течение всего периода времени.

Средняя стоимость одного заказа составляет 750 рублей.

Какую максимальную выручку мы можем ожидать за один среднестатистический день с вероятностью 95%?

Округлите ответ до 0 знаков после запятой.

Задача 3 (2 балла)

Представьте, что все три временных интервала между этапами конверсии распределены нормально с указанными значениями параметров.

С помощью правила трёх сигм рассчитайте нижнюю границу для каждого из них, опираясь на размах в три сигмы. Предположите, что меньшее время чем нижняя граница практически невозможно.

Сколько целых минут составит самый быстрый путь от момента заказа до прибытия в точку В?

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Вычислите выборочное среднее для выбора: 2, 4, 6, 8.	5	ОПК-1
2.	Укажите значение второй порядковой статистики для выборки: 5, 1, 9, 2.	2	ОПК-1
3.	Для выбора 1, 2, 2, 4 укажите значение эмпирической функции распределения. $F_n(2)$.	0,75	ПК-2
4.	Укажите название кондуктов, утверждающих равномерную сходимость. $F_n(x)$ к $F(x)$ почти наверняка.	Гливенко–Кантелли	ПК-2
5.	Укажите пределы предельных измерений: если $X_n \Rightarrow X$, то говорят о сходимости по _____.	нейтральное	ПК-1
6.	Укажите значение $z_{0,975}$ (квантиль стандартного нормального распределения).	1.96	ПК-1
7.	Для нормальных моделей при закрытии $\sigma=2, n=25, x=10$ укажите 95%-доверительный интервал для μ в виде «левая;правая» (с расстояния до 0,01).	9.22;10.78	УК-2
8.	При взгляде на гипотезы на уровне инноваций $\alpha=0,05$ укажите решение, если р-значение равно 0,03 (одно слово).	закрываем	УК-1
9.	Укажите формулировку ошибки I рода (одно словосочетание).	ложное отклонение	УК-2
10.	Укажите название основы построения наиболее мощного критерия для простых альтернатив и простых нулевых гипотез.	Неймана–Пирсона	УК-1