

УТВЕРЖДЕНА

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

**Рабочая программа дисциплины (модуля)
«Основы статистики»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Программа двух дипломов НИУ
ВШЭ и ЦУ «Искусственный интеллект»

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2024

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	4
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) «Основы статистики» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 02.04.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Искусственный интеллект», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Основы статистики» является важным для формирования навыков анализа и интерпретации данных, что позволяет принимать обоснованные решения на основе количественной информации. Статистика предоставляет инструменты для выявления закономерностей, оценки рисков и проверки гипотез, что критически важно в различных профессиональных областях.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Искусственный интеллект» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 1 или 2 курсе в 1, 2, 3 или 4 семестре на выбор.

Цель изучения дисциплины (модуля): формирование у студентов фундаментальных знаний и навыков, необходимых для анализа данных, интерпретации результатов и принятия обоснованных решений на основе статистической информации.

Задачи изучения дисциплины (модуля):

— формирование знаний и развитие понимания по темам: основные понятия, операторы и архетипы задач из теории вероятностей, формулы для различных типов комбинаторных вычислений, определения, параметры и вид основных распределений, основные точечные оценки и их влияние на распределения, закон больших чисел и ЦПТ, методика построения доверительных интервалов, бутстрап для различных задач, понятия и алгоритмы, связанные с тестированием гипотез, понятие ковариации и различные виды корреляции, а также значимость корреляций, основные понятия линейной регрессии, этапы А/В тестирования, алгоритм сплитования в А/В тестах, дополнительные техники А/В тестирования;

— освоение умений рассчитывать вероятность с помощью классических, геометрических и комбинаторных техник, визуализировать и рассчитывать плотность различных распределений, рассчитывать и интерпретировать точечные оценки на выборке, совершать переход от закона больших чисел и центральной предельной теоремы к доверительным интервалам и проверке гипотез, строить доверительный интервал, рассчитывать различные критерии: z , t , χ^2 , F , рассчитывать p -value и с его помощью делать выводы, оцифровывать зависимость между случайными величинами, рассчитывать необходимое количество наблюдений для запуска А/В-теста, проектировать и запускать А/В-тест, делать выводы по результатам А/В-теста;

— формирование навыков расчета вероятности события, построение распределений и расчета их квантили в `python`, расчета и интерпретации точечных и интервальных оценок, использования бутстрапа для построения оценок и тестирования гипотез, подведение итогов о корреляции и каузации случайных величин, запуска, мониторинга и оцифровки результатов проведения А/В-тестов.

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-6.	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1.	Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания
		УК-6.2.	Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки
		УК-6.3.	Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста
ОПК-2.	Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы	ОПК-2.1.	Знает основные математические модели и методы, используемые в естественных науках, включая статистическое моделирование, дифференциальные уравнения и численные методы, а также современные подходы к исследованию и анализу данных
		ОПК-2.2.	Умеет разрабатывать и адаптировать математические модели для решения конкретных проблем в естественных

			науках, проводить их анализ и верификацию, а также интерпретировать полученные результаты в контексте научных исследований
		ОПК-2.3.	Имеет практический опыт создания и исследования математических моделей в рамках научных проектов или исследований, включая участие в публикациях, конференциях или коллаборациях, где были разработаны и апробированы новые концепции и методы
ПК-3.	Способен решать задачи профессиональной деятельности, формулировать результат, увидеть следствия полученного результата	ПК-3.1.	Знает основные принципы и методы решения задач профессиональной деятельности, а также способы формулирования и представления результатов, включая анализ последствий и их значимость в контексте проекта
		ПК-3.2.	Умеет применять математические и компьютерные методы для решения конкретных задач, формулировать четкие и обоснованные результаты, а также анализировать их последствия для дальнейших действий и решений
		ПК-3.3.	Имеет практический опыт в решении профессиональных задач, включая участие в проектах, где были получены результаты и проанализированы их следствия, что способствовало принятию обоснованных решений
ПК-4.	Способен публично представлять собственные и известные научные	ПК-4.1.	Знает основные принципы эффективного публичного выступления, методы

	результаты		визуализации данных и основные требования к научным презентациям, включая структуру и содержание
		ПК-4.2.	Умеет четко и логично формулировать свои научные результаты, адаптируя их для различных аудиторий, а также использовать визуальные средства для улучшения восприятия информации
		ПК-4.3.	Имеет практический опыт участия в научных конференциях, семинарах или других мероприятиях, где успешно представлял свои и известные научные результаты, получая обратную связь и взаимодействуя с аудиторией

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 (практические занятия)					
1	Теория вероятностей	5	5		34	Домашние задания
2	Распределения и оценки	6	5		34	Домашние задания
3	Статистические гипотезы	6	7		34	Домашние задания, Контрольная работа
4	Зависимые данные	5	5		34	Домашние задания
5	A/B Тесты	6	6		34	Домашние задания
	<i>Экзамен</i>			2		
	Итого:	28	28	2	170	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	228				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	6				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Теория вероятностей	<ul style="list-style-type: none"> - Актуальность, состав и расписание курса - Общие понятия из теории вероятностей - Геометрическая интерпретация вероятности - Основные понятия комбинаторики (перестановки / размещения / сочетания с повторениями и без) - Задачи "на смекалку" с собеседований
2	Распределения и оценки	<ul style="list-style-type: none"> - Примеры распределений: нормальное (+правило трёх сигм), нормальное стандартное, Бернулли, биномиальное, равномерное, Пуассон, экспоненциальное, хи-квадрат - Понятия функции распределения и функции плотности - Работа с распределениями и их плотностью в python - Точечные оценки: среднее, дисперсия, стандартное отклонение, квантили, максимум, минимум, асимметрия, эксцесс - Свойства точечных оценок: несмещённость и состоятельность - Закон Больших Чисел и Неравенство Чебышева - Центрально-Предельная Теорема - Интервальные оценки: построение доверительных интервалов для среднего, доли и дисперсии - Метод бутстрап: построение доверительных интервалов с помощью бутстрапа
3	Статистические гипотезы	<ul style="list-style-type: none"> - Понятие нулевой и альтернативной гипотезы - Ошибки первого и второго рода - Зона принятия гипотезы и критическая зона

		<ul style="list-style-type: none"> - Односторонний и двусторонний критерий и его зависимость от альтернативной гипотезы - Понятие p_value - Одновыборочный и двухвыборочный Z-критерий - Распределение Стьюдента - Переход от Z- к t-критерию - Распределение Фишера - Одновыборочный и двухвыборочный критерий на дисперсии
4	Зависимые данные	<ul style="list-style-type: none"> - Понятие зависимости событий и случайных величин - Понятия априорной и апостериорной вероятности - Формула Байеса для расчёта условной вероятности - Понятие ковариации и корреляции - Корреляция Пирсона, Спирмена, Мэтьюса и Крамера - Значимость и интерпретация корреляции - Нетождественность понятий "корреляция" и "каузальность" - Модель линейной регрессии и ее качество (R-квадрат, TSS, ESS, RSS) - Теорема Гаусса-Маркова: понятия мультиколлинеарности, гетероскедастичности и автокорреляции - Корреляция признаков и VIF - Интерпретация коэффициентов линейной регрессии и их значимость - Нормализация и стандартизация признаков в линейной модели
5	A/B Тесты	<ul style="list-style-type: none"> - Определение, смысл и ситуации использования A/B тестов - Понятие таргет- и гард-метрики + карточка эксперимента - Финансы: альтернативная и реальная стоимость эксперимента - MDE и расчёт необходимого количества наблюдения - Сплит-система с хешированием и солью - A/A-тест и A/A/B-тест - Система бакетов и мастер-бакетов - Поправки на множественное сравнение Бонферонни и Холма - Множественное тестирование с помощью ANOVA и F-теста - Непараметрические критерии Манна-Уитни и Колмогорова-Смирнова - Хи-квадрат критерий и его связь с t-критерием - p-hacking и проблема подсматривания в эксперименте - Линеаризация в A/B тесте - Методы понижения дисперсии (в частности, стратификация или смена типа метрики) - Сетевой эффект и Switchback-эксперименты

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>.

2. 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>.

Дополнительная литература:

1. McKinney, W. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter / Wes McKinney. — Boston : O'Reilly Media, 2022. — 579 p. — URL: <https://wesmckinney.com/book/>.

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;
- 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 к 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) «Основы статистики» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, практические занятия, 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 удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (практическом занятии) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

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 (модуля) и продемонстрировать навыки их практического применения.

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) «Основы 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:

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 (модуль). 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
6	Хорошо	

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

Активность	Вес	Количество	Описание
Накопительная оценка			
Домашние задания	60%	10	Набор задач по темам недели
Контрольная работа		1	Письменная работа с набором задач, которые нужно решить за ограниченное время
Промежуточная аттестация			
Экзамен	40%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Основы статистики»: $\langle 0,6 \times \text{накопительная оценка} (0,7 \times \text{среднее за домашние задания} + 0,3 \times \text{контрольная работа}) + 0,4 \times \text{экзамен} \rangle$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание: Вероятность и комбинаторика

Задача 1.

Звонящий забыл последнюю цифру номера телефона и поэтому набирает её наугад. Определите вероятность того, что ему придётся звонить не более чем в 4 места.

Задача 2.

Вы хотите посмотреть звездопад, который начнётся в случайный момент времени между 0.00 (полуночью) и 1.00 (часом ночи) и который продлится десять минут.

Вы попадёте на точку, удобную для наблюдения за звёздами, в случайный момент времени между 0.00 и 1.00. Вы готовы просидеть там не больше получаса.

Какова вероятность, что вы увидите хотя бы один миг звездопада?

Задача 3.

Вы работаете аналитиком в онлайн-кинотеатре, и ваша задача — разработать стратегию линеек с контентом.

Линейка — это последовательность тайтлов (т. е. фильмов/мультфильмов/сериалов и т. д.), которая отображается на стартовой странице онлайн-кинотеатра с целью привлечь клиентов к просмотру.

Определите, сколько потенциальных вариантов линеек у нас есть. Создайте функцию, которая будет брать на вход:

- Количество тайтлов всего в онлайн-кинотеатре;
- Желаемая длина линейки;
- Параметр, который принимает значения 0 или 1, и в зависимости от его значения мы рассчитываем или количество размещений, или количество сочетаний.

Этот параметр контролирует «важность» последовательности тайтлов в линейке. Важно ли, что в данной линейке именно «Хочу замуж» идёт первым фильмом, «Подельники» — вторым? Или мы не различаем, какой фильм из этих пяти на каком месте стоит в линейке?

- Если «последовательность не важна» ($=0$), то нам подойдут сочетания.
- Если «последовательность важна» ($=1$), то нам подойдут размещения.
- Если в качестве аргумента передаётся что-то отличное от 0 или 1, функция должна вернуть принт с требованием передать 0 или 1.

Создайте функцию и протестируйте её на выборке из 58 топ-тайтлов при условии, что длина линейки равна пяти.

А также ответьте на вопрос: «Во сколько раз количество потенциальных размещений превышает количество сочетаний?»

Задача 4.

Рассмотрим строку «I love Python!»

Рассчитайте количество различных строк, которые могут получиться при перестановке символов в этой строке?

Обратите внимание, что среди символов есть не только буквы.

Домашнее задание: Биномиальное распределение. Распределение Пуассона. Экспоненциальное распределение.

Задача 1. Биномиальное распределение

На онлайн-площадку заходят пользователи, каждый из которых купит продукт с вероятностью 10%. Цена продукта — 1200 рублей.

Чтобы площадка окупалась, необходимо заработать минимум 500 000 рублей.

Сколько пользователей (N) нужно пригласить, чтобы заработать 500 000 рублей с вероятностью как минимум 95%?

Подсказка. На вопрос можно ответить с помощью:

1. цикла `for`, запущенного по сетке различных N ;
2. цикла `while` с условием достижения вероятности порога в 95%.

Задача 2. Распределение Пуассона

Каждому дню недели соответствует цифра.

Каждая цифра показывает, сколько пользователей в среднем за минуту заходили к нам в приложение в этот день недели.

Одна цифра представляет свой день недели - от понедельника до воскресенья.

Например, во вторник средний пользовательский поток был 125 пользователей в минуту.

Также известен средний пользовательский поток за последний год:

- 125 пользователей в дни с понедельника по четверг
- 175 пользователей в пятницу
- 200 пользователей в выходные

Для каждого дня недели рассчитай вероятность получить на следующей неделе такой же или более интенсивный поток, чем в среднем за год.

Задача 3. Экспоненциальное распределение

Аналитики нашего магазина разделили пользователей на две категории - "горячие" и "холодные".

Горячим пользователем требуется в среднем 8 минут на принятие решения о покупке, а холодным требуется в среднем 17 минут на принятие решения. Время, необходимое для принятия решения, у обоих распределено экспоненциально.

В течение последней минуты к нам зашло 4 пользователя: два горячих и два холодных. Какова вероятность, что за 15 минут мы не получим от них ни одной покупки?

Домашнее задание: Доверительные интервалы

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.

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.

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.

Подсказка:

Рассчитайте границы доверительного интервала для каждого звена конверсии и рассчитайте, сколько получится поездок за один день, если пройти по верхней границе всех трёх конверсий.

Задача 3.

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?

Примерные задания для контрольной работы

Задание 1.

Задача: В группе из 10 человек нужно выбрать трёх для участия в проекте. Сколько существует различных способов выбора, если:

- порядок выбора не важен;
- порядок выбора важен;
- можно выбирать одного и того же человека несколько раз?

Объясните, к каким видам комбинаторики относятся эти задачи.

Задание 2.

Рассмотрите квадрат со стороной 1. Точка выбирается случайно внутри квадрата. Найдите вероятность того, что точка попадёт в круг радиуса 0.5, вписанный в квадрат.

Задание 3.

Определите функцию плотности и функцию распределения для равномерного распределения на отрезке $[0,1]$. Постройте графики этих функций.

Задание 4.

Напишите код на Python, который генерирует 1000 случайных чисел из нормального распределения с параметрами $\mu=0$, $\sigma=1$ и строит гистограмму с наложенной теоретической кривой плотности.

Задание 5.

Дана выборка: 5, 7, 8, 9, 10. Найдите среднее, дисперсию, стандартное отклонение. Объясните, что означает несмещённость оценки среднего.

Задание 6.

Объясните суть Закона Больших Чисел и Центрально-Предельной Теоремы. Приведите пример, иллюстрирующий их применение.

Задание 7.

В эксперименте измеряется среднее время реакции. Нулевая гипотеза: среднее время равно 0.5 секунды. Альтернативная: среднее время меньше 0.5 секунды. Объясните, что такое p-value и как его использовать для принятия решения.

Задание 8.

В городе 1% населения болеет редкой болезнью. Тест на болезнь имеет 95% точность при выявлении больных и 90% точность при выявлении здоровых. Найдите вероятность того, что человек болен, если тест показал положительный результат.

Задание 9.

Объясните разницу между корреляцией и каузальностью. Опишите, что такое коэффициент детерминации (R^2) в линейной регрессии и как он интерпретируется.

Задание 10.

В компании проводится A/B тест с двумя вариантами сайта. Объясните, что такое ошибка первого рода и как поправки Бонферони помогают контролировать множественное тестирование при анализе нескольких метрик.

Примерный перечень вопросов к экзамену

- Что такое вероятность события и как она интерпретируется в геометрическом смысле?
- Опишите основные понятия комбинаторики и приведите примеры перестановок, размещений и сочетаний.
- Каковы основные правила сложения и умножения вероятностей?
- Приведите пример задачи "на смекалку", связанной с теорией вероятностей, и её решение.
- Какова роль теории вероятностей в анализе данных и принятии решений?
- Опишите нормальное распределение и правило трёх сигм.
- Каковы основные характеристики биномиального распределения?

8. Что такое функция распределения и функция плотности случайной величины?
9. Как в Python можно работать с различными распределениями и их плотностями?
10. Объясните понятия точечной оценки, несмещённости и состоятельности.
11. Каков смысл Закона Больших Чисел и Неравенства Чебышева?
12. Что такое Центральная-Предельная Теорема и как она применяется в статистике?
13. Как строятся доверительные интервалы для среднего на основе выборки?
14. Каков принцип работы метода бутстрап для построения доверительных интервалов?
15. Приведите примеры применения экспоненциального и Пуассоновского распределений.
16. Что такое нулевая и альтернативная гипотеза?
17. Опишите ошибки первого и второго рода в контексте статистических тестов.
18. Как определяется зона принятия гипотезы и критическая зона?
19. В чем разница между односторонним и двусторонним критерием?
20. Как интерпретируется p-value в контексте тестирования гипотез?
21. Каковы условия применения одновыборочного и двухвыборочного Z-критерия?
22. Что такое распределение Стьюдента и когда оно используется?
23. Объясните переход от Z- к t-критерию.
24. Каковы основные принципы применения распределения Фишера?
25. В чем разница между одновыборочным и двухвыборочным критерием на дисперсии?
26. Каковы основные понятия зависимости событий и случайных величин?
27. Объясните формулу Байеса и её применение для расчёта условной вероятности.
28. Что такое ковариация и корреляция, и как они различаются?
29. Как интерпретируются коэффициенты корреляции Пирсона и Спирмена?
30. Какова роль теоремы Гаусса-Маркова в линейной регрессии и как определяются её качества?

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Вы получили $p\text{-value} = 0.12$ при уровне значимости 0.05 ; укажите статистическое решение по нулевой гипотезе.	не отклоняется	УК-6
2	Вы рассчитали доверительный интервал для среднего и получили слишком широкие границы; укажите способ повышения точности оценки.	увеличение выборки	УК-6
3	Вы обнаружили расхождение между априорной и апостериорной вероятностями; укажите формулу для пересчёта вероятностей.	формула Байеса	УК-6
4	Вы применили несколько критериев без поправки и получили значимые различия; укажите возникающую проблему.	множественное тестирование	УК-6
5	Вы анализируете эксперимент и замечаете преждевременную остановку после достижения значимости; укажите проблему дизайна.	p-hacking	УК-6
6	Случайная величина имеет постоянную плотность на отрезке $[a,b]$; укажите тип распределения.	равномерное распределение	ОПК-2
7	Для модели линейной регрессии при наличии гетероскедастичности укажите нарушаемое условие теоремы Гаусса-Маркова.	гомоскедастичность	ОПК-2
8	Если ковариация двух случайных величин равна нулю при нормальности распределения; укажите тип зависимости.	независимость	ОПК-2

9	Для оценки доли в большой выборке используется нормальная аппроксимация; укажите предельную теорему.	центральная предельная теорема	ОПК-2
10	При сравнении более двух средних используется F-статистика; укажите название метода.	ANOVA	ОПК-2
11	Если p-value меньше уровня значимости; укажите статистическое действие.	отклонение гипотезы	ПК-3
12	Коэффициент корреляции Спирмена применяется при нарушении нормальности; укажите тип зависимости, который он измеряет.	ранговая корреляция	ПК-3
13	При значимом коэффициенте регрессии и положительном знаке β ; укажите направление влияния признака.	положительное влияние	ПК-3
14	При высоком значении VIF укажите проблему модели регрессии.	мультиколлинеарность	ПК-3
15	В A/B тесте при бинарной метрике для сравнения долей используется нормальная аппроксимация; укажите критерий.	Z-критерий	ПК-3
16	При публичной презентации результатов вы сообщаете 95% доверительный интервал; укажите интерпретацию уровня 0.95.	уровень доверия	ПК-4
17	При сравнении эмпирической и теоретической функций распределения используется непараметрический критерий; укажите его название.	критерий Колмогорова-Смирнова	ПК-4
18	Если остатки регрессии автокоррелированы; укажите нарушаемое предположение классической линейной модели.	отсутствие автокорреляции	ПК-4
19	При анализе категориальных признаков в таблице сопряженности используется χ^2 -статистика; укажите критерий.	критерий хи-квадрат	ПК-4
20	В эксперименте с сетевым эффектом применяется дизайн с чередованием периодов; укажите тип эксперимента.	switchback-эксперимент	ПК-4