

УТВЕРЖДЕНА

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

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

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	6
4. Содержание дисциплины (модуля)	6
5. Учебно-методическое обеспечение	7
6. Материально-техническое обеспечение	7
7. Методические и оценочные материалы	9

1. Краткая характеристика дисциплины (модуля)

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

Дисциплина (модуль) изучается на 1 курсе в 1 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 для заданных столбцов рассчитать описательные статистики в Python;

— развитие умений интерпретации описательных статистик, графики;

— развитие умений в заданном датасете для заданных столбцов построить заданные графики в Python;

— развитие умений описания способов получения случайной выборки из имеющейся генеральной совокупности;

— развитие умений генерирования случайных выборок из заданной генеральной совокупности в Python;

— развитие умений сравнить описательные статистики в выборке и в генеральной совокупности, объяснить, почему наблюдаемые числовые характеристики выборки не обязательно соответствуют характеристикам генеральной совокупности

— развитие умений объяснения важности случайности выборки, размер выборки, почему про генеральную совокупность/распределение часто делают предположения, строят разные модели;

— развитие умений по заданному дискретному распределению вычислить вероятности, матожидание, дисперсию, по текстовому описанию задачи построить дискретное распределение, биномиальное и нормальное распределение;

— развитие навыков вычисления вероятности, матожидание и дисперсию для биномиального и нормального распределений

— умение применять центральную предельную теорему для суммы и среднего независимых одинаково дискретно распределенных случайных величин, для выборочной пропорции

— умение определять количество наблюдений, необходимых для того, чтобы

выборочная пропорция оценивала пропорцию в генеральной совокупности с заданной точностью;

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

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)
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 математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности	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
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

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 (Практические занятия)					
1	Работа с датасетами и презентация результатов	7	7		20	Домашнее задание Кейсы
2	Связь выборки и теоретической модели данных	8	8		22	Домашнее задание Кейсы
3	Вероятностные модели	8	8		22	Домашнее задание Кейсы
4	Калибровка и отбор вероятностных моделей	7	7		22	Домашнее задание Кейсы
	<i>Зачет с оценкой</i>			6		
	<i>Итого:</i>	<i>30</i>	<i>30</i>	<i>6</i>	<i>86</i>	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	<i>152</i>				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	<i>4</i>				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Работа с датасетами и презентация результатов	Описательные статистики. Сводные таблицы. Графики и презентации
2	Связь выборки и теоретической модели данных	Выборки и генеральные совокупности. Случайность. События. Независимость. Случайная выборка. Связь характеристик случайной выборки и генеральной совокупности. Метод Монте-Карло
3	Вероятностные модели	Дискретные распределения. Матожидание, дисперсия. Биномиальное распределение и введение в непрерывные распределения. Нормальное распределение, его свойства. Центральная предельная теорема
4	Калибровка и отбор вероятностных моделей	Оценки параметров: среднее и пропорция. Статистические гипотезы и тестирование. P-value, значимость и мощность. Стандартные односторонние тесты. Стандартные двусторонние тесты. Предпосылки к тестам

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;
- 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 лекция, практические занятия, кейсы 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.

Кейсы – решение бизнес-кейса с опорой на данные, подготовка презентации.

При подготовке решения бизнес-кейса с опорой на данные, важно начать с четкого определения проблемы и целей анализа. Затем соберите и проанализируйте релевантные данные, используя подходящие методы визуализации для наглядного представления результатов. В презентации акцентируйте внимание на ключевых выводах и рекомендациях, подкрепляя их данными и графиками, чтобы убедительно донести свою позицию до аудитории.

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) «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 *Зачета с оценкой*, 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	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	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	Зачтено	
8	Отлично	Зачтено	
7	Хорошо	Зачтено	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
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			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 теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	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	Зачтено	
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Основы статистики» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	40%	14	Работа с данными, датасетами, решение математических задач.
Кейсы	60%	4	Решение бизнес-кейса с опорой на данные, подготовка презентации.

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

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные задания кейсов

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:

1. **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.
2. **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 домашние задания

Домашнее задание по теме: «Вероятность и комбинаторика»

Задача 1 (2 балла)

Звонящий забыл последнюю цифру номера телефона и поэтому набирает её наугад.

Задание: определить вероятность того, что ему придётся звонить *не более чем в 4 места*.

Задача 2 (3 балла)

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 полчаса.

Задание: какова вероятность, что вы увидите хотя бы один миг звездопада?

Задача 3 (3 балла)

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.

Необходимо определить, 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.

Задание: 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 (2 балла)

Рассмотрим строку «I love Python!»

Задание: рассчитайте количество различных строк, которые могут получиться при перестановке символов в этой строке?

Обрати внимание, что среди символов есть не только буквы.

Домашнее задание по теме: «Распределения»

Задача 1. Биномиальное распределение (3 балла)

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 (3 балла)

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 (4 балла)

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 по теме: «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 (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.	<p>Сайт размещает объявления о продаже квартир в Москве. В каждом объявлении указывается информация о квартире, в том числе S = площадь квартиры, M = название ближайшей к дому станция метро. Определите типы значений для S и M.</p> <p>A. S — числовой, M — числовой B. S — числовой, M — упорядоченные категории C. S — числовой, M — неупорядоченные категории D. S — неупорядоченные категории, M — неупорядоченные категории</p>	С	ПК-1
2.	<p>Медианная зарплата в фирме равна 100 тысяч рублей в месяц означает:</p> <p>A. Большинство сотрудников получают 100 тысяч рублей в месяц. B. Примерно половина сотрудников получает меньше 100 тысяч рублей в месяц, а примерно половина сотрудников — больше. C. В среднем все сотрудники получают 100 тысяч рублей в месяц. D. Каждый сотрудник получает 100 тысяч рублей в месяц.</p>	В	ПК-1
3.	<p>Датафрейм <code>df_users</code> содержит данные о пользователях. Столбец <code>Age</code> содержит возраст в годах в числовом формате. Команда <code>df_users['Age'].mean()</code> покажет</p> <p>A. Сумму возрастов пользователей датафрейма B. Количество пользователей датафрейма C. Средний возраст пользователей датафрейма D. Смысл возраста пользователей датафрейма</p>	С	ПК-1
4.	<p>Датафрейм <code>df_users</code> содержит данные о пользователях. Столбец <code>Age</code> должен содержать возраст в годах в числовом формате.</p> <p>Есть предположение, что у одного пользователя в столбце <code>Age</code> вместо возраста указано некорректное значение, большее 200, так называемый выброс, а для остальных пользователей данные в столбце <code>Age</code> корректны.</p> <p>Какая команда НЕ ПОЗВОЛИТ проверить предположение: выявить наличие такого выброса?</p> <p>A. <code>df_users.sort_values(by='Age', ascending=True).tail(5)</code> B. <code>df_users.sort_values(by='Age', ascending=False).head(5)</code> C. <code>df_users[df_users['Age'] > 200]</code> D. <code>df_users['Age'].median()</code></p>	D	ПК-1

5.	<p>Пусть есть датафрейм <code>df_users</code> с данными о пользователях и столбец <code>Age</code> содержит возраст в числовом формате. Команда <code>df_users[df_users['Age'] > 18]</code> покажет</p> <p>А. Датафрейм, в котором останутся только те пользователи из <code>df_users</code>, у которых возраст больше 18 лет.</p> <p>В. Средний возраст пользователей, которые старше 18 лет</p> <p>С. Количество пользователей, у которых возраст больше 18 лет</p> <p>Д. Количество пользователей, кроме тех, у которых возраст больше 18 лет</p>	А	ПК-1								
6.	<p>Что из перечисленного является генеральной совокупностью?</p> <p>А. 100 школьников, опрошенных о любимом предмете.</p> <p>В. Все школьники в школе.</p> <p>С. 30 студентов, участвующих в исследовании.</p> <p>Д. 1,000 опрошенных жителей города.</p>	В	ОПК-1								
7.	Чему равна медиана для выборки значений: 1,2,3,8,9	3	ОПК-1								
8.	<p>Чему равно матожидание случайной величины</p> <table border="1" data-bbox="316 862 890 1025"> <tr> <td>Значение</td> <td>2</td> <td>4</td> <td>5</td> </tr> <tr> <td>Вероятность (PMF)</td> <td>0.2</td> <td>0.4</td> <td>0.4</td> </tr> </table>	Значение	2	4	5	Вероятность (PMF)	0.2	0.4	0.4	4	ОПК-1
Значение	2	4	5								
Вероятность (PMF)	0.2	0.4	0.4								
9.	Вероятность ошибки первого рода — это теста. Укажите пропущенный термин (2 слова)	уровень значимости / Уровень значимости	ОПК-1								
10.	Укажите пропущенное одно слово в следующем предложении. случайной величины показывает возможные значения этой величины и их вероятности.	Распределение / распределение	ОПК-1								
11.	Среднее по выборке из четырех наблюдений равно 10. В выборку добавили пятое наблюдение, равное нулю. Чему теперь равно среднее по выборке (уже от пяти наблюдений)	8	ОПК-1								
12.	<p>Как называется метод приближительного нахождения вероятности, который реализован в следующем коде</p> <pre>import scipy.stats as sps total = 999999 counter = 0 for _ in range(total): x = sps.norm(2,3).rvs(1) y = sps.norm(3,4).rvs(1) if x < y: counter = counter + 1 probability_x_less_than_y = counter / total</pre>	метод Монте-Карло / метод монте-карло	ОПК-1								