
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«24» июня 2025 г.
Протокол № 2

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

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Машинное обучение

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2025

**Москва
2025**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	5
3. Тематический план.....	7
4. Содержание дисциплины (модуля).....	7
5. Учебно-методическое обеспечение	8
6. Материально-техническое обеспечение	8
7. Методические и оценочные материалы	10

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 Машинное обучение, 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 Машинное обучение и входит в обязательную часть Блока 1.

Дисциплина (модуль) изучается на 1 курсе в 1 семестре, доступна для прохождения при условии успешного завершения дисциплины (модуля) «Основы Python».

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):

- изучить ключевые концепции теории вероятностей для моделирования и предсказания случайных процессов в различных сценариях;
- освоить техники оценки параметров распределений и построения интервальных оценок для надежного анализа выборок;
- разработать навыки применения статистических критериев для проверки гипотез и интерпретации результатов в контексте реальных данных;
- научиться анализировать зависимости между переменными с использованием мер ковариации и корреляции для выявления связей;
- приобрести умение проектировать и проводить эксперименты, такие как A/B тестирование, для оценки эффективности изменений и принятия обоснованных выводов.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- 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;
- методику 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 A/B тестирования;
- алгоритм сплитования в A/B тестах;
- дополнительные техники A/B тестирования.

уметь:

- 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;
- делать выводы о корреляции и каузации случайных величин;
- запускать, мониторить и оцифровывать результаты проведения А/В тестов;
- рассчитывать необходимое количество наблюдений для запуска А/В теста;
- делать выводы по результатам А/В теста.

владеть:

- навыками расчета вероятности события;
- навыками построения распределений и расчета их квантилей в python;
- навыками расчета и интерпретации точечных и интервальных оценок;
- навыками использования бутстрапа для построения оценок и тестирования гипотез;
- навыками проектирования и запуска А/В теста.

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)
ПК-1.	Способен определять общие формы и закономерности области машинного обучения	ПК-1.1.	Знает основные теоретические концепции и принципы, относящиеся к области машинного обучения, а также ключевые закономерности и модели, которые помогают в анализе и интерпретации данных
		ПК-1.2.	Умеет проводить систематический анализ области разработки, выявлять и формулировать общие закономерности и тенденции, а также применять методы исследования для получения новых знаний и понимания
		ПК-1.3.	Имеет практический опыт работы в области машинного обучения, включая участие в научных проектах, исследованиях или практических заданиях, где были выявлены и описаны общие формы и закономерности
ПК-2.	Способен математически корректно ставить естественнонаучные и прикладные задачи	ПК-2.1.	Знает основные методы и подходы к математическому моделированию, а также теоретические основы естественных и прикладных наук, необходимые для корректной формулировки задач
		ПК-2.2.	Умеет анализировать практические ситуации и формулировать на их основе математические модели, включая выбор адекватных методов решения и формулировку условий задачи
		ПК-2.3.	Имеет практический опыт в разработке и решении математических задач в рамках проектов или научных исследований, где были успешно поставлены и решены естественнонаучные и прикладные задачи

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

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		<i>Очная форма</i>				
		Аудиторная работа		Контроль	Самостояте льная работа	
Лекции	Семинары (практичес кие занятия)					
1	Введение в статистику	6	6		25	Домашние задания
2	Статистические оценки	6	6		25	Домашние задания
3	Гипотезы	6	6		25	Домашние задания, Контрольная работа
4	Зависимые данные	4	4		17	Домашние задания
5	А/В тесты	8	8		32	Домашние задания
	<i>Экзамен</i>			6		
	<i>Итого:</i>	30	30	6	124	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	190				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Введение в статистику	Введение в статистику Теория вероятностей и комбинаторика Распределения
2	Статистические оценки	Точечные оценки Свойства точечных оценок Интервальные оценки
3	Гипотезы	Введение в гипотезы Тестирование гипотез
4	Зависимые данные	Корреляции Линейная регрессия
5	А/В тесты	А/В тесты

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. Кремер, Н. Ш. Теория вероятностей и математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. — 5-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 538 с. — (Высшее образование). — ISBN 978-5-534-10004-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/565694>.

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. 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. — 395 с. — (Высшее образование). — ISBN 978-5-534-21643-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/581860>.

2. Статистика : учебник для вузов / под редакцией В. С. Мхитаряна. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 503 с. — (Высшее образование). — ISBN 978-5-534-18687-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/568978>.

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

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

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не 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.

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
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		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 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
Домашние задания	70%	14	Набор задач по темам недели
Контрольная работа	15%	1	Письменная работа с набором задач, которые нужно решить за ограниченное время
Экзамен	15%	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»: « $0,7 \times$ среднее за домашние задания + $0,15 \times$ контрольная работа + $0,15 \times$ экзамен».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

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 2.

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 3.

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

1 SUPER MEGA ULTRA TOP (2 балла)

В вашей компании есть база данных с 6 типами клиентов: новички (50 человек), подписчики (80 человек), продвинутые (65 человек), отказавшиеся (40 человек), VIP (25 человек) и SUPER MEGA ULTRA TOP (30 человек).

Ответьте на следующие вопросы:

1. **(0.5 балла)** Сколько существует способов составить выборку из 6 клиентов по одному из каждого типа?
2. **(0.5 балла)** Сколько существует способов расставить 4 VIP-клиентов на 4 различных sales-менеджеров?
3. **(0.5 балла)** Мы хотим определить топ-5 клиентов по метрике LTV. Сколько всего существует возможных комбинаций этого топа?
4. **(0.5 балла)** В системе происходит сбой, который удаляет случайно выбранного клиента. Какова вероятность того, что будет удалена запись о пользователе типа «отказавшийся» или «VIP»?

2 Накликай на удачу (2 балла)

Ваша команда тестирует производительность двух версий сайта. Оказалось, что время отклика страниц распределено нормально:

- Версия А: среднее время — 850 мс, стандартное отклонение — 120 мс

- Версия В: среднее время — 780 мс, стандартное отклонение — 110 мс

Ответьте на следующие вопросы:

1. **(0.5 балла)** С какой вероятностью время отклика версии А превысит 900 мс?
2. **(0.5 балла)** С какой вероятностью время отклика версии В будет менее 700 мс?
3. **(0.5 балла)** Найдите 95-й процентиль времени отклика для версии А.
4. **(0.5 балла)** Пусть X — время отклика версии А, Y — время отклика версии В. Найдите вероятность того, что разность $X - Y$ превысит 100 мс.

3 Время — деньги (2.5 балла)

В службе технической поддержки время между поступлениями заявок распределено именно так, как вы думаете. Да-да-да, именно так, как вы подумали. Оказалось, что в среднем за минуту поступает 0,15 заявок. Количество заявок в час подчиняется распределению Пуассона.

Ответьте на следующие вопросы:

1. **(0.5 балла)** Чему равно среднее время между соседними заявками?
2. **(0.75 балла)** С какой вероятностью время до следующей заявки не превысит 10 минут?
3. **(0.75 балла)** С какой вероятностью за час поступит более 12 заявок?
4. **(0.5 балла)** Постройте 90%-доверительный интервал для количества заявок в час.

4 ZIP-файл (2 балла)

Вы тестируете новый алгоритм сжатия данных. Время выполнения алгоритма (в секундах) для 12 тестовых файлов оказалось равным 2.3, 2.8, 2.1, 2.9, 2.4, 2.7, 2.2, 2.6, 2.5, 2.8, 2.4 и 2.3.

Ответьте на следующие вопросы на уровне значимости 5%:

1. **(1 балл)** Можно ли утверждать, что среднее время выполнения алгоритма статистически значимо больше 2.0 секунд? Чему равно p-value?
2. **(1 балл)** Постройте 95%-доверительный интервал для среднего времени выполнения алгоритма.

5 А что это за буква? (1.5 балла)

$\zeta_1, \zeta_2, \dots, \zeta_n$ — независимые одинаково распределённые случайные величины с $\mathbb{E}\zeta_i = 0$ и $\mathbb{D}\zeta_i = \sigma^2$.

Рассмотрим последовательность

$$\eta_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n (-1)^{k+1} \zeta_k$$

Могут быть полезны свойства независимых случайных величин и центральная предельная теорема.

1. **(0.75 балла)** Вычислите математическое ожидание и дисперсию η_n .
2. **(0.75 балла)** К какому распределению сходится η_n при $n \rightarrow \infty$? Обоснуйте ваш ответ.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Дана выборка [1, 6, 2, 7, 10, 1, 1, 7, 12, 13, 8, 0] Расположите в порядке возрастания следующие показатели выборки. Укажите цифры через тире без пробелов. 1 - Среднее арифметическое 2 - Медиана 3 - Первый квартиль 4 - Шестой дециль	3-1-2-4	ПК-1
2.	Какое из следующих свойств статистической оценки подразумевает, что математическое ожидание статистической оценки равно оцениваемому параметру? 1 - Эффективность 2 - Состоятельность 3 - Несмещённость	3	ПК-1

3.	В колоде 52 карты (по 13 карт каждой масти). Какова вероятность вытащить или пику, или туза? Ответ дайте в процентах, округлённых до целого, т.е., например, 0,6789 запишите как 68	31 / 31.0 / 31,0 / 31% / 31.0% / 31,0% / 0.31 / 0,31	ПК-1
4.	Группа студентов разбилась на 3 команды по 15 человек. Сколько всего комбинаций из трёх капитанов (по 1 на каждую команду) можно выбрать?	3375 / 3375.0 / 3375,0	ПК-2
5.	В столовой подают девять различных блюд. Сколько у вас есть вариантов различных обедов из трёх блюд (без повторов блюд)? Два обеда с одинаковым набором блюд, идущих в разном порядке, считаются разными вариантами.	504 / 504.0 / 504,0	ПК-3
6.	Агрегатор такси представлен тремя тарифами. В каждом есть своё значение конверсии из заказа в завершённую поездку: Эконом (72%), Комфорт (60%), Люкс (48%). Количество заказов в Экономе составляет 1500, в Комфорте - 950, а в Люксе - 560. Чему равна средняя конверсия из заказа в завершённую поездку по всему агрегатору такси?	64 / 64% / 64.0 / 64,0 / 64.0% / 64,0%	ПК-1
7.	Пусть случайная величина X распределена как распределение Бернулли с параметром 0.6. Чему равна дисперсия случайной величины X ?	0.24 / 0,24	ПК-2
8.	При каком значении параметра p дисперсия распределения Бернулли принимает наибольшее значение?	0.5 / 0,5	ПК-3
9.	Выберите все верные утверждения. 1. Распределение Пуассона - это непрерывное распределение 2. Экспоненциальное распределение - это непрерывное распределение 3. Распределение Пуассона - это дискретное распределение 4. Экспоненциальное распределение - это дискретное распределение	2,3 / 3,2	ПК-3
10.	Выберите верное утверждение: 1 - Поправка Бонферони при расчёте статистического теста применяется, если необходимо провести несколько попарных несовместных сравнений 2 - Поправка Бонферони при расчёте статистического теста применяется, если необходимо сравнить группы с маленьким количеством наблюдений 3 - Поправка Бонферони при расчёте статистического теста применяется, если сравниваемые выборки являются зависимыми друг от друга	1	ПК-3
11.	Чему равен коэффициент линейной корреляции Пирсона между рядами X и Y ? $X = (1,2,6,15)$ $Y = (-2,-4,-12,-30)$	-1 / -1.0 / -1,0	ПК-3
12.	Значения p -value теста на значимость каждого из коэффициентов линейной регрессии представлены в следующем списке: [0.0032, 0.154, 0.078, 0.0001, 0.97]. На 5-процентном уровне значимости сколько факторов останется в линейной регрессии после удаления незначимых факторов?	2 / два	ПК-3

13.	Вы последовательно подбрасывали монетку 500 раз и получили 300 орлов (единиц) и 200 решек (нулей). Чему равна медиана результирующей бинарной выборки?	1 / 1.0 / 1,0	ПК-1
14.	X и Y - независимые случайные величины с дисперсиями, равными 200 и 700 соответственно. Чему равно стандартное отклонение суммы случайных величин X и Y?	30 / 30.0 / 30,0	ПК-2
15.	Чему равна дисперсия биномиального распределения с параметрами $p = 0.3$ и $n = 500$?	105 / 105.0 / 105,0	ПК-3
16.	Есть пять банков, каждый из которых совершит дефолт с вероятностью 5%. Какова вероятность, что хотя бы 4 из 5 банков останутся и не совершат дефолт? Ответ округлите до процентов с нулём знаков после запятой (например, ответ 0.6478 необходимо записать как 65).	98 / 98% / 98.0 / 98,0 / 98.0% / 98,0%	ПК-1
17.	Чему равна ширина доверительного интервала для математического ожидания для выборки, в которой дисперсия составляет 144, а количество наблюдений составляет 100? В качестве квантильного значения распределения z можно взять 1.96. Ответ округлите до одного знака после запятой.	4,7 / 4.7	ПК-2
18.	Постройте доверительный интервал для пропорции: расчётное значение конверсии равно 0.56, получено данное значение было на 625 наблюдениях. Чему равна ширина доверительного интервала? В качестве квантильного значения распределения z можно взять 1.96. Ответ округлите до двух знаков после запятой.	0,08 / 0.08	ПК-3
19.	Клиенты были разделены на две группы. В первой группе конверсия в покупку составила 13%, а во второй - 14%. В первой группе было 150 клиентов, а во второй - 160 клиентов. Необходимо проверить статистическую гипотезу о равенстве конверсий в двух группах. Нулевая гипотеза (H_0) заключается в том, что конверсии равны. Альтернативная гипотеза (H_1) заключается в том, что конверсии не равны. Рассчитайте z-значение соответствующего теста и округлите ответ до двух знаков после запятой.	-0,26 / -0.26	ПК-1
20.	Есть выборки X и Y. Стандартное отклонение X составляет 15, а стандартное отклонение Y составляет 21. Ковариация двух рядов составляет 200. Чему равна дисперсия суммы X и Y?	1066 / 1066.0 / 1066,0	ПК-2
21.	Даны две выборки: X и Y. Выборочная дисперсия X равна 16, а выборочная дисперсия Y равна 25. Выборочная ковариация равна 17. Чему равен выборочный коэффициент корреляции X и Y? Ответ округлите до двух знаков после запятой.	0.85 / 0,85	ПК-1

22.	В модели линейной регрессии RSS составляет 20% от TSS. Чему равен коэффициент детерминации модели? Ответ дайте в целых процентах.	80 / 80.0 / 80,0 / 80% / 80.0% / 80,0% / 0.8 / 0,8	ПК-2
-----	--	---	------