
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол № 2

**Рабочая программа дисциплины (модуля)
«Введение в статистику»**

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль) подготовки: Бизнес и аналитика

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2026

**Москва
2026**

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) «Введение в статистику» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 38.03.05 Бизнес-информатика, профиль Бизнес и аналитика, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 838 от 29.07.2020 года.

Изучение дисциплины (модуля) «Введение в статистику» 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 38.03.05 Бизнес-информатика, профиль Бизнес и аналитика и входит в обязательную часть Блока 1.

Дисциплина (модуль) доступна к изучению на 1 или 2 курсе в 1, 2 или 3 семестре на выбор после успешного освоения дисциплин «Разработка на Python» или распределительный контекст по программированию (должен быть пройден на углубленный уровень).

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

Задачи изучения дисциплины (модуля):

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

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

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

В результате освоения дисциплины (модуля), обучающийся должен:

знать:

— описательные статистики, графики;
— случайность, независимость, выборку, генеральную совокупность;
— произвольное дискретное распределение, вероятность, матожидание, дисперсию;
— биномиальное, нормальное распределения и оценки их параметров;
— гипотезы: нулевая и альтернативная. односторонний биномиальный тест для пропорции;

уметь:

— в заданном датасете для заданных столбцов рассчитать описательные статистики в python;
— проинтерпретировать описательные статистики;
— в заданном датасете для заданных столбцов построить заданные графики в python;
— проинтерпретировать графики;
— описать способ получения случайной выборки из имеющейся генеральной совокупности;

- сгенерировать случайные выборки из заданной генеральной совокупности в python;
- сравнить описательные статистики в выборке и в генеральной совокупности, объяснить, почему наблюдаемые числовые характеристики выборки не обязательно соответствуют характеристикам генеральной совокупности;
- строить эмпирическое распределение выборочного среднего методом монте-карло;
- по заданному дискретному распределению вычислить вероятности, матожидание, дисперсию;
- по текстовому описанию задачи построить дискретное распределение;
- вычислить вероятности, матожидание и дисперсию для биномиального и нормального распределений;
- по текстовому описанию задачи построить биномиальное и нормальное распределение;
- применить центральную предельную теорему для суммы и среднего независимых одинаково дискретно распределенных случайных величин;
- определять количество наблюдений, необходимых для того, чтобы выборочное среднее оценивало матожидание в генеральной совокупности с заданной точностью и вероятностью;
- по текстовому описанию задачи формулировать гипотезу о пропорции;
- проводить биномиальный тест для пропорции, интерпретировать результат теста;
- владеть:**
 - навыками критического анализа данных и их статистических свойств;
 - языком статистического мышления: умением формулировать вопросы, поддающиеся количественной проверке, и переводить их в статистические гипотезы;
 - инструментами python (pandas, numpy, scipy.stats, matplotlib) для выполнения описательного анализа, визуализации и проверки гипотез;
 - способностью объяснять статистические выводы нетехнической аудитории, избегая распространённых интерпретационных ошибок;
 - навыками проектирования простых выборочных исследований с учётом требований к репрезентативности и точности оценок.

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-3.	Способен управлять процессами создания и использования продуктов и услуг в сфере информационно-коммуникационных технологий, в том числе разрабатывать алгоритмы и программы для их практической реализации	ОПК-3.1.	Знает принципы управления процессами разработки и внедрения продуктов и услуг в сфере информационно-коммуникационных технологий
		ОПК-3.2.	Умеет разрабатывать алгоритмы и программы, обеспечивающие эффективное создание и использование информационных продуктов
		ОПК-3.3.	Имеет практический опыт в управлении проектами в области информационно-коммуникационных технологий, включая координацию команд и ресурсов для достижения поставленных целей
ПК-1.	Способен использовать основные методы естественнонаучных, экономических и ИТ-дисциплин в профессиональной деятельности для теоретического и экспериментального исследования	ПК-1.1.	Знает ключевые методы естественнонаучных, экономических и ИТ-дисциплин, применяемые в профессиональной деятельности
		ПК-1.2.	Умеет интегрировать различные методологические подходы для проведения теоретических и экспериментальных исследований
		ПК-1.3.	Имеет практический опыт применения методов в реальных проектах для достижения научных и практических результатов
ПК-2.	Способен использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования для решения задач профессиональной деятельности	ПК-2.1.	Знает основные математические методы и инструментальные средства, применяемые для обработки и анализа информации
		ПК-2.2.	Умеет эффективно использовать математический аппарат для систематизации данных и решения профессиональных задач
		ПК-2.3.	Имеет практический опыт работы с инструментами анализа информации в рамках исследовательских проектов

3. Тематический план

№п/ п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятель ная работа	
		Лекции	Семинары			
1	Работа с датасетами и презентация результатов	6	7		31	Домашнее задание Кейсы
2	Связь выборки и теоретической модели данных	6	4	2	31	Домашнее задание Контрольная работа
3	Вероятностные модели	6	8	2	15	Домашнее задание Кейсы
4	Связь выборки и теоретической модели данных	3	7	2	33	Домашнее задание
5	Вероятностные модели	3	4		16	Домашнее задание Контрольная работа
	Экзамен			4		
Итого:		24	30	10	126	
Объем дисциплины (модуля) (в ак. ч.)		190				
Объем дисциплины (модуля) (в зач. ед.)		5				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Работа с датасетами и презентация результатов	NumPy: введение NumPy: продвинутые операции над массивами Pandas: введение Pandas: очистка и сводные таблицы Визуализация данных Контрольная работа 1
2	Связь выборки и теоретической модели данных	Модель равновероятных исходов Выборки
3	Вероятностные модели	Дискретная модель
4	Связь выборки и теоретической модели данных	Контрольная работа 2
5	Вероятностные модели	Связь случайных величин Биномиальное и непрерывные распределения Нормальное распределение Асимптотические теоремы Введение в гипотезы. Контрольная работа 3

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. 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>.

3. Энатская, Н. Ю. Математическая статистика и случайные процессы : учебник для вузов / Н. Ю. Энатская. — Москва : Издательство Юрайт, 2025. — 191 с. — (Высшее образование). — ISBN 978-5-9916-9808-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561150>.

Дополнительная литература:

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

6. Материально-техническое обеспечение

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

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

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

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

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

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

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

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

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

Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Введение в статистику» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары, кейсы, контрольные работы и домашние задания, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

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

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

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

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

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.

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

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

Домашнее задание – набор задач по темам недели.

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

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

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

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

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

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Введение в статистику»

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

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

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

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет
9	Отлично	
8	Отлично	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	
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) «Введение в статистику» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	15%	Работа с данными, датасетами, решение математических задач
Кейсы	20%	Решение бизнес-кейса с опорой на данные, подготовка презентации
Контрольные работы	30%	Письменная работа с набором задач, которые нужно решить за ограниченное время
Экзамен	35%	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

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

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

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 задания для контрольных работ

Контрольная работа

Задание 1

Что из перечисленного НЕ является корректным способом создания numpy-массива?

- а) `np.array([1, 2, 3])`
- б) `np.arange(5)`
- в) `np.list([1,2,3])`
- г) `np.zeros((3,3))`

Задание 2

Создайте numpy-массив из 10 элементов, заполненный случайными целыми числами от 1 до 100. Выведите этот массив и вычислите его среднее значение.

Задание 3

Объясните, что такое broadcasting в numpy и приведите пример его применения.

Задание 4

Дан DataFrame pandas с колонками "Возраст" и "Зарплата". Напишите код, который добавит новую колонку "Налог", равную 13% от "Зарплаты".

Задание 5

Какой метод pandas используется для объединения двух DataFrame по общему столбцу?

- а) `merge()`
- б) `concat()`
- в) `join()`
- г) `append()`

Задание 6

Используя pandas, загрузите CSV-файл с данными о продажах, выведите первые 5 строк, а затем сгруппируйте данные по столбцу "Категория" и посчитайте сумму по столбцу "Объем продаж".

Задание 7

Что из перечисленного НЕ относится к функциям seaborn?

- а) `distplot`
- б) `boxplot`
- в) `scatter_matrix`
- г) `heatmap`

Задание 8

Постройте с помощью matplotlib гистограмму распределения значений массива numpy из 1000 случайных чисел, равномерно распределенных от 0 до 1.

Задание 9

Что такое модель равновероятных исходов? Приведите пример задачи, где она применяется.

Задание 10

Дан DataFrame с результатами эксперимента, где каждый исход равновероятен. Напишите код на pandas, который проверит, что частоты исходов примерно равны.

Задание 11

Какая из функций seaborn используется для построения диаграммы рассеяния с отображением регрессии?

- а) regplot
- б) barplot
- в) violinplot
- г) pairplot

Задание 12

Используя seaborn, постройте boxplot для набора данных о росте людей по категориям "Мужчины" и "Женщины".

Задание 13

Опишите разницу между merge и join в pandas.

Задание 14

Напишите код, который создаст DataFrame с двумя столбцами: "Событие" с тремя равновероятными исходами ('A', 'B', 'C') и "Вероятность" с равными значениями вероятностей.

Задание 15

В модели равновероятных исходов с 5 исходами найдите вероятность наступления события, состоящего из двух исходов.

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 получаса.

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.	Какой метод в numpy используется для создания массива с равномерно распределенными значениями между двумя числами? a) np.linspace b) np.arange c) np.random.rand d) np.zeros	a	ПК-1
2.	Назовите функцию numpy для вычисления стандартного отклонения массива.	std	ПК-2
3.	Как называется структура данных в pandas, представляющая одномерный массив с индексами?	Series	ПК-2
4.	Укажите метод pandas для чтения CSV-файла.	read_csv	ОПК-3
5.	Назовите операцию в pandas для объединения двух DataFrame по строкам.	concat	ОПК-3
6.	Какой тип графика в seaborn лучше всего подходит для визуализации распределения одной числовой переменной? a) scatterplot b) boxplot	c	ПК-1

	c) histplot d) lineplot		
7.	Назовите функцию matplotlib для создания диаграммы рассеяния.	scatter	ОПК-3
8.	Как называется график в seaborn, показывающий взаимосвязи между несколькими переменными?	pairplot	ОПК-3
9.	Укажите понятие, описывающее модель, где все исходы имеют одинаковую вероятность.	равновероятные исходы	ПК-2
10.	Назовите метод pandas для группировки данных и применения агрегатных функций.	groupby	ПК-1
11.	Что такое генеральная совокупность? a) Часть данных из большого набора b) Полный набор объектов исследования c) Выборка с повторениями d) Среднее значение выборки	b	ПК-1
12.	Назовите тип вероятностной модели, где случайная величина принимает счетное число значений.	дискретная	ПК-2
13.	Укажите условие, при котором два события считаются независимыми.	$P(A \text{ и } B) = P(A) * P(B)$	ПК-2
14.	Назовите распределение, моделирующее число успехов в серии независимых испытаний.	биномиальное	ОПК-3
15.	Укажите пример непрерывного распределения вероятностей.	нормальное	ОПК-3
16.	Что утверждает Закон больших чисел (ЗБЧ)? a) Среднее выборки стремится к математическому ожиданию b) Дисперсия выборки уменьшается с ростом размера c) Выборка становится нормальной при большом n d) p-value всегда меньше 0.05	a	ПК-1
17.	Назовите теорему, объясняющую, почему сумма независимых случайных величин стремится к нормальному распределению.	Центральная предельная	ОПК-3
18.	Укажите значение p-value, при котором нулевая гипотеза отвергается на уровне значимости 0.05.	$< 0.05 / < 0,05$	ПК-2
19.	Назовите тип тестов, основанных на больших выборках и приближениях.	асимптотическое	ОПК-3
20.	Укажите понятие, описывающее минимальную разницу, которую тест может обнаружить.	детектируемый эффект	ОПК-3