

УТВЕРЖДЕНА

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

**Рабочая программа дисциплины (модуля)
«Основы статистики. Углубленный курс»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Математика и искусственный интеллект

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 | 4 |
| 3. Тематический план | 6 |
| 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. Углубленный курс» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль Математика и искусственный интеллект, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Основы статистики. Углубленный курс» позволяет принимать обоснованные решения на основе данных и повышает качество научных исследований. Это способствует развитию критического мышления и умению работать с большими и сложными информационными массивами.

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 Математика и искусственный интеллект и входит в Блок 1, часть, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 1 курсе в 1 семестре.

Цель изучения дисциплины (модуля): освоить продвинутые методы статистического анализа и интерпретации данных для решения прикладных задач в различных научных и практических областях.

Задачи изучения дисциплины (модуля) направлены на формирование у студентов следующий знаний, умений и навыков:

- знание методов очистки данных: обработка пропущенных значений, выбросов, дубликатов;
- знание формулировки неравенств Маркова и Чебышева и условия их применения;
- знание основ стратификации выборки: принципы построения страт, преимущества метода;
- знание свойства многомерных дискретных распределений и их параметры;
- знание точной формулировки асимптотических теорем (ЦПТ, ЗБЧ) и условия их применимости;
- умение проводить предварительный анализ данных: выявлять и обрабатывать аномалии;
- умение применять неравенства Маркова и Чебышева для оценки вероятностей;
- умение проектировать стратифицированные выборки и обосновывать выбор страт;
- умение вычислять ковариацию и корреляцию для нормально распределенных величин;
- умение проверять оценки на несмещенность и состоятельность экспериментальными методами;
- навык оптимизации процедур очистки данных для конкретных исследовательских задач;
- навык расчета мощности тестов и минимально детектируемого эффекта для планирования экспериментов;
- навык определения минимально необходимого размера выборки для достижения требуемой точности;
- навык пользования современными методами data manipulation (pandas, dplyr) для эффективной обработки данных;
- навык построения эффективных оценок для среднего значения с учетом особенностей данных.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

| Компетенция | Содержание компетенции | Индикатор компетенции | Перечень планируемых результатов обучения по дисциплине (модулю) |
|-------------|---|-----------------------|--|
| УК-1. | Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач | УК-1.1. | Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности. |
| | | УК-1.2. | Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем |
| | | УК-1.3. | Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации |
| 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 | ПК-1.1. | Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов. |

| | | | |
|--|--------------------------------|---------|---|
| | математики и компьютерных наук | 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 | Очистка данных | | 3 | | 3 | Домашнее задание |
| 2 | Доказательства неравенств Маркова и Чебышева для выборок | | 3 | | 3 | Домашнее задание |
| 3 | Стратификация | | 3 | | 3 | Домашнее задание Кейсы |
| 4 | Многомерное дискретное распределение | | 3 | | 3 | Домашнее задание |
| 5 | Ковариация и корреляция в нормальном распределении | | 3 | | 3 | Домашнее задание Кейсы |
| 6 | Точные формулировки асимптотических теорем | | 3 | | 3 | Домашнее задание Кейсы |
| 7 | Экспериментальная проверка несмещённости и состоятельности | | 3 | | 4 | Домашнее задание |
| 8 | Мощность теста и минимально детектируемый эффект | | 3 | | 4 | Домашнее задание |
| 9 | Минимально необходимый размер выборки в тестах | | 2 | | 4 | Домашнее задание |
| 10 | Data manipulation | | 2 | | 4 | Домашнее задание Кейсы |
| 11 | Повышение эффективности оценок для среднего | | 2 | | 4 | Домашнее задание |
| | <i>Зачет</i> | | | | 8 | |
| | Итого: | | 30 | | 8 | 38 |
| | Объем дисциплины (модуля) (в ак. ч.) | 76 | | | | |
| | Объем дисциплины (модуля) (в зач. ед.) | 2 | | | | |

4. Содержание дисциплины (модуля)

| №п/п | Наименование раздела дисциплины (модуля) | Содержание дисциплины (модуля) по темам |
|------|--|---|
| 1 | Очистка данных | Выявление и обработка пропущенных значений; удаление выбросов; нормализация и стандартизация данных |
| 2 | Доказательства неравенств Маркова и Чебышева для выборок | Формулировка неравенств; доказательство для случайных величин; применение к выборочным средним |
| 3 | Стратификация | Определение и цели стратификации; методы формирования страт; влияние стратификации на оценку параметров |
| 4 | Многомерное дискретное распределение | Определение и свойства; совместное распределение вероятностей; маргинальные и условные распределения |
| 5 | Ковариация и корреляция в нормальном распределении | Понятия ковариации и корреляции; свойства ковариационной матрицы; интерпретация корреляции в многомерном нормальном распределении |
| 6 | Точные формулировки асимптотических теорем | Закон больших чисел; центральная предельная теорема; теорема о сходимости по вероятности |
| 7 | Экспериментальная проверка несмещённости и состоятельности | Методы проверки несмещённости; проверка состоятельности оценок; анализ результатов эксперимента |
| 8 | Мощность теста и минимально детектируемый эффект | Определение мощности теста; расчет минимально детектируемого эффекта; влияние размера выборки на мощность |
| 9 | Минимально необходимый размер выборки в тестах | Критерии определения размера выборки; формулы расчета; практические примеры |
| 10 | Data manipulation | Фильтрация и сортировка данных; преобразование и агрегирование; работа с пропущенными и аномальными значениями |
| 11 | Повышение эффективности оценок для среднего | Методы уменьшения дисперсии оценки; использование стратифицированных и взвешенных средних; применение дополнительной информации |

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. Углубленный курс» в рамках текущего контроля успеваемости используются такие виды учебной работы, как 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.

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

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

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. Углубленный курс»

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*, 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 |
| 6 | Хорошо | Зачтено | |

| Десятибалльная оценка | Пятибалльная оценка | Оценка за зачет | Общая характеристика результата обучения по дисциплине (модулю) |
|-----------------------|---------------------|-----------------|--|
| | | | 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 | 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. Углубленный курс» оценивается следующим образом:

| Активность | Вес | Количество | Описание |
|------------------|-----|------------|--|
| Домашние задания | 40% | 10 | Работа с данными, датасетами, решение математических задач. |
| Кейсы | 60% | 6 | Решение бизнес-кейса с опорой на данные, подготовка презентации. |

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

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1

1. Среди всех пользователей:
 - Найди среднее время, проводимое клиентами в мобильном приложении, и смещённое стандартное отклонение.
 - Найди распределение T и покажи результат на гистограмме.
 2. Построй N простых случайных выборок размера 1 клиент. Среди них:
 - Найди среднее T и его смещённое стандартное отклонение.
 - Найди распределение T и покажи результат на гистограмме.
 3. Построй N простых случайных выборок размера 10 клиентов, посчитай по каждой такой выборке среднее T^- и среди них:
 - Найди среднее T^- и его смещённое стандартное отклонение.
 - Найди распределение T^- и покажи результат на гистограмме.
 4. Повтори предыдущий пункт для выборки размера 500 клиентов.
 5. Объясни, что показывают (приблизительно или точно) распределения, изображенные на гистограммах в предыдущих пунктах.
 6. Сравни распределения, найденные в предыдущих пунктах, и прокомментируй.
 7. Сравни средние, найденные в предыдущих пунктах.
 8. Построй N выборок, стратифицированных по новизне модели мобильного устройства, посчитай по каждой такой выборке среднее T^- и среди них:
 - Найди среднее T^- и его смещённое стандартное отклонение.
 - Найди распределение T^- и покажи результат на гистограмме.
 - Сравни с предыдущим при размере выборки 10 клиентов и 500 клиентов.
 9. Построй N выборок, стратифицированных по операционной системе мобильного устройства. Посчитай по каждой такой выборке среднее T^- и среди них:
 - Найди среднее T^- и его смещённое стандартное отклонение.
 - Найди распределение T^- и покажи результат на гистограмме.
 - Сравни с предыдущим при размере выборки 10 клиентов и 500 клиентов.
 10. Создай таблицу со сводными результатами всех вышеприведённых сравнений. Таблица должна иметь следующие столбцы:
 1. *type* — тип выборки (SRS / Strat_Model / Strat_OS).
 2. *quantity* — количество наблюдений в одной выборке.
 3. *mean* — среднее из средних N выборок.
 4. *std* — стандартное отклонение средних в N выборках.
- Сделай общие по выводы по таблице.
11. Посчитай первую квартиль по генеральной совокупности и построй гистограмму распределения первой квартили Q_1 , найденной по каждой из N выборок размера 500. Сделай выводы.
 12. Посчитай максимум по генеральной совокупности и построй гистограмму распределения максимума по выборкам размера 500. Сделай выводы.

Домашнее задание 2

В блоге планируется реклама нового продукта. Конверсия рекламы — доля купивших продукт после просмотра рекламы. Скорее верно или скорее неверно каждое из следующих утверждений?

1. Точно найти конверсию рекламы среди всех подписчиков блога можно, только показав рекламу всем подписчикам.
2. Если показать рекламу части подписчиков в блоге и посчитать конверсию, она совпадёт с конверсией всего блога.
3. Конверсия в выборке приближает конверсию в блоге. Способ получения выборки не влияет на качество приближения.
4. Конверсия в выборке приближает конверсию в блоге. Но если выборку формировать случайно, то приближению будет меньше доверия, чем при неслучайном отборе подписчиков в выборку.
5. При увеличении размера простой случайной выборки точность приближения растёт.
6. При увеличении размера выборки точность приближения растёт, независимо от способа получения выборки.
7. Точность приближения конверсии при использовании стратифицированной случайной выборки, как правило ниже, чем при простой случайной.

Примерные кейсы

Кейс, проблема, идея



Кейс

- Веб-платформа для изучения иностранного языка
- Клиенты потребляют разный контент и платят по разным тарифам
- Доходы от некоторых клиентов даже меньше, чем расходы платформы на обслуживание этих клиентов
- Компания хочет внедрить новые продукты для имеющихся клиентов, чтобы повысить среднюю прибыль с клиента
- Повышение средней прибыли с клиента на 800 рублей считается экономически значимым
- При тестировании гипотез требуется использовать уровень значимости 5%
- При определении размера выборки исходи из мощности 80%, попробуй и другие



Проблема

Нужно как-то тестировать новый продукт до запуска



Идея решения

Описать процесс тестирования продукта. Процесс должен с требуемыми вероятностями обнаруживать экономически значимую пользу от продукта

Роль, задача, результат



Роль

Аналитик отдела рекламы и маркетинга



Задача

Описать процесс тестирования продукта на слайдах



Доступные данные

Датасет [golden_mine.csv](#): прибыль по текущим клиентам, пол, возраст



Ожидаемый результат

- Презентация на слайдах в формате pdf
- Jupyter Notebook в Google Colab с расчётами

Требования к слайдам



Если слайды или Jupyter Notebook не приложен, решение кейса **оценивается в 0 баллов**

| Понятность и внешний вид | Обоснованность | Реакция заказчика |
|--|---|---|
| <ul style="list-style-type: none"> Внешний вид презентации не мешает воспринимать информацию Понятно, на какие вопросы отвечает каждый слайд Содержимое таблиц, графиков понятно из слайда без необходимости открывать исходный датасет Выводы явно сформулированы | <ul style="list-style-type: none"> Выводы основаны на таблицах, графиках, показателях, полученных из данных Таблицы и графики получены скриншотом или картинкой из Jupyter Notebook, поэтому их можно перепроверить Выводы явно сформулированы | <ul style="list-style-type: none"> 0 баллов Не принимает, ищет другого исполнителя 1 балл Частично принимает, считает необходимым отдать на доработку текущему исполнителю 2 балла Принимает, готов пересылать слайды от своего имени, под свою ответственность |
| МАКСИМУМ 5 БАЛЛОВ | МАКСИМУМ 5 БАЛЛОВ | МАКСИМУМ 2 БАЛЛА |

Максимальная оценка — 10 баллов

Бонусные баллы могут поднять основную оценку, но не выше **10 баллов**, даже если в сумме получится 11 или 12 баллов

4

Задания для промежуточной аттестации по дисциплине (модулю)

| № п/п | Задание | Ответ | Компетенция | | | | | | | | |
|-------------------|---|----------|-------------|---|---|-------------------|-----|-----|-----|---|-------|
| 1. | Нулевая гипотеза в статистике — это А. Утверждение о значениях параметров распределения генеральной совокупности (например, $p = 0.3$ или $\mu = 100$). В. Утверждение о статистических характеристиках, вычисленных по данным выборки (например, выборочное среднее равно 100). С. Утверждение о конкретном значении случайной величины в отдельном наблюдении (например, что следующий платёж превысит 100 рублей). D. Утверждение произвольного характера, не ограниченное описанием параметров генеральной совокупности. | А | УК-1 | | | | | | | | |
| 2. | Случайная величина представлена таблицей PMF <table border="1" style="margin-left: 20px;"> <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> Чему равно CDF(5) | Значение | 2 | 4 | 5 | Вероятность (PMF) | 0.2 | 0.4 | 0.4 | 1 | ОПК-1 |
| Значение | 2 | 4 | 5 | | | | | | | | |
| Вероятность (PMF) | 0.2 | 0.4 | 0.4 | | | | | | | | |
| 3. | Пусть X имеет нормальное распределение $N(a,b)$, где a – матожидание, b – дисперсия. Какое распределение имеет $2X$ при $a = 1$ и $b = 1$? (используйте символ N , скобки, запятую в качестве разделителя параметров и никаких пробелов для удобства автопроверки) | $N(2,4)$ | ПК-1 | | | | | | | | |