

**УТВЕРЖДЕНА**

Приказом Ректора АНО ВО  
«Центральный университет»  
Ивашкевич Е.В.  
от «19» января 2024 г. № 0119.37

**Рабочая программа дисциплины (модуля)  
«Основы статистики»  
дополнительной профессиональной программы – программы  
профессиональной переподготовки «Академия data science»**

**Траектория: Продуктовая аналитика**

**Москва  
2024**

## Содержание

|                                                            |          |
|------------------------------------------------------------|----------|
| <b>1. Краткая характеристика дисциплины (модуля) .....</b> | <b>3</b> |
| <b>2. Тематический план .....</b>                          | <b>4</b> |
| <b>3. Содержание дисциплины (модуля) .....</b>             | <b>4</b> |
| <b>4. Учебно-методическое обеспечение .....</b>            | <b>6</b> |
| <b>5. Материально-техническое обеспечение .....</b>        | <b>6</b> |
| <b>6. Методические и оценочные материалы .....</b>         | <b>8</b> |

## 1. Краткая характеристика дисциплины (модуля)

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.

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

### **Задачи изучения дисциплины (модуля):**

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

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

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

## 2. Тематический план

| №<br>п/п | Наименование<br>раздела дисциплины<br>(модуля)  | Трудоемкость, академические часы |           |          |                               | ТКУ<br>(текущий<br>контроль<br>успеваемости) |
|----------|-------------------------------------------------|----------------------------------|-----------|----------|-------------------------------|----------------------------------------------|
|          |                                                 | Очная форма                      |           |          |                               |                                              |
|          |                                                 | Аудиторная работа                |           | Контроль | Самостоя<br>тельная<br>работа |                                              |
| Лекции   | Семинары<br>(практичес<br>кие<br>занятия)       |                                  |           |          |                               |                                              |
| 1        | Теория вероятностей                             | 6                                | 6         |          | 18                            | Домашние задания                             |
| 2        | Распределения и оценки                          | 6                                | 7         |          | 18                            | Домашние задания                             |
| 3        | Статистические гипотезы                         | 7                                | 7         |          | 18                            | Домашние задания,<br>Контрольная работа      |
| 4        | Зависимые данные                                | 7                                | 7         |          | 20                            | Домашние задания                             |
| 5        | A/B Тесты                                       | 7                                | 7         |          | 20                            | Домашние задания                             |
|          | <i>Экзамен</i>                                  |                                  |           | 4        |                               |                                              |
|          | <b>Итого:</b>                                   | <b>33</b>                        | <b>34</b> | <b>4</b> | <b>94</b>                     |                                              |
|          | <b>Объем дисциплины (модуля)<br/>(в ак. ч.)</b> | <b>165</b>                       |           |          |                               |                                              |

## 3. Содержание дисциплины (модуля)

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

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

#### 4. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый слушатель 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.

Слушателям обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

##### *Основная литература:*

1. Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебник для вузов / В. Е. Гмурман. — 11-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 406 с. — (Высшее образование). — ISBN 978-5-534-08389-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/559583>.

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

##### *Дополнительная литература:*

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

#### 5. 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 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<br>(издания, курса, документа)                            | Ссылка                                                                          |
|----|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 1. | Научная электронная библиотека elibrary.ru библиотека                          | <a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a> |
| 2. | База данных для IT-специалистов                                                | <a href="https://habr.com">https://habr.com</a>                                 |
| 3. | База данных ScienceDirect                                                      | <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>       |
| 4. | Официальный сайт Министерства науки и высшего образования Российской Федерации | <a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>           |
| 5. | Федеральный портал «Российское образование»                                    | <a href="https://www.edu.ru/">https://www.edu.ru/</a>                           |
| 6. | Информационная система "Единое окно доступа к образовательным ресурсам"        | <a href="http://window.edu.ru/">http://window.edu.ru/</a>                       |
| 7. | Единая коллекция цифровых образовательных ресурсов                             | <a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a> |
| 8. | Федеральный центр информационно - образовательных ресурсов                     | <a href="http://fcior.edu.ru/">http://fcior.edu.ru/</a>                         |

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

|                                                          |            |                           |
|----------------------------------------------------------|------------|---------------------------|
| <b>Среды разработки:</b>                                 |            |                           |
| 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 |
| <b>Пакеты программных средств и библиотек:</b>           |            |                           |
| AutoPsy                                                  | зарубежное | свободно распространяемое |
| Interactive Disassembler (IDA)                           | зарубежное | свободно распространяемое |
| <b>Системы управления библиографической информацией:</b> |            |                           |
| Zotero                                                   | зарубежное | свободно распространяемое |
| <b>Сервисы и службы:</b>                                 |            |                           |
| Bind                                                     | зарубежное | свободно распространяемое |
| Docker                                                   | зарубежное | свободно распространяемое |

## 6. Методические и оценочные материалы

### Методические указания для обучающихся по освоению дисциплины (модуля)

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

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

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

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

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

*Домашнее задание* – набор задач по темам недели.

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

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

Цель контрольной работы – получить специальные знания по одной или нескольким

темам дисциплины (модуля) и продемонстрировать навыки их практического применения.

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

В процессе самостоятельной работы слушатели взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи слушателей 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)**

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**.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

| <b>Десятибалльная оценка</b> | <b>Пятибалльная оценка</b> | <b>Оценка за зачет</b> | <b>Общая характеристика результата обучения по дисциплине (модулю)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------|----------------------------|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10                           | Отлично                    | Зачтено                | Слушатель 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. Слушатель 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                | Слушатель обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 6                            | Хорошо                     | Зачтено                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

| Десятибалльная оценка | Пятибалльная оценка | Оценка за зачет | Общая характеристика результата обучения по дисциплине (модулю)                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----------------------|---------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                       |                     |                 | 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. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами. |
| 5                     | Удовлетворительно   | Зачтено         | Слушатель 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. Слушатель способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.                                                                                                                                                                                                             |
| 4                     | Удовлетворительно   | Зачтено         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 3                     | Не сдан             | Не зачтено      | Слушатель не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.                                                                                                                                                                                                                                                                                                                                                                               |
| 2                     | Не сдан             | Не зачтено      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 1                     | Не сдан             | Не зачтено      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

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

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

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

## Текущий контроль успеваемости обучающихся по дисциплине (модулю)

### Примерные домашние задания

#### Домашнее задание: Вероятность и комбинаторика

##### Задача 1.

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

##### Задача 2.

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

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

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

##### Задача 3.

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

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

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

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

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

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

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

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

##### Задача 4.

Рассмотрим строку «I love Python!»

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

Обратите внимание, что среди символов есть не только буквы.

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

##### Задача 1. Биномиальное распределение

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

Чтобы площадка окупалась, необходимо заработать минимум 500 000 рублей.

Сколько пользователей ( $N$ ) нужно пригласить, чтобы заработать 500 000 рублей с

вероятностью как минимум 95%?

Подсказка. На вопрос можно ответить с помощью:

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

### **Задача 2. Распределение Пуассона**

Каждому дню недели соответствует цифра.

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

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

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

Также известен средний пользовательский поток за последний год:

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

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

### **Задача 3. Экспоненциальное распределение**

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

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

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

## **Домашнее задание: Доверительные интервалы**

Имеется информация за 10 дней в агрегаторе такси:

- Количество заказов = 8770.
- Среднее время поиска и назначения водителя на заказ составляет 2 минуты и 1 секунду (стандартное отклонение - 32 секунды).
- Конверсия из заказа в назначение водителя на заказ равна 94%
- Среднее время подачи такси после назначения водителя на заказ составляет 5 минут и 16 секунд (стандартное отклонение - 1 минута и 5 секунд).
- Конверсия из назначения в прибытие автомобиля в точку А составляет 82%.
- Среднее время поездки (от точки А до точки В) составляет 24 минуты и 36 секунд (стандартное отклонение - 7 минут и 2 секунды).
- Конверсия из прибытия автомобиля в точку А в прибытие в точку В составляет 88%.

### **Задача 1.**

Вы заказываете такси.

Можно ли утверждать, что с вероятностью 95% вы будете возле точки В раньше чем через 35 минут?

Подсказка:

Для ответа на вопрос рассчитайте границы доверительного интервала для трёх временных интервалов между этапами воронки.

### **Задача 2.**

Предположите, что все заказы были равномерно распределены по 10 дням, а конверсия была неизменной в течение всего периода времени.

Средняя стоимость одного заказа составляет 750 рублей.

Какую максимальную выручку мы можем ожидать за один среднестатистический день с вероятностью 95%?

Округлите ответ до 0 знаков после запятой.

Подсказка:

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

### **Задача 3.**

Представьте, что все три временных интервала между этапами конверсии распределены нормально с указанными значениями параметров.

С помощью правила трёх сигм рассчитайте нижнюю границу для каждого из них, опираясь на размах в три сигмы. Предположите, что меньшее время чем нижняя граница практически невозможно.

Сколько целых минут составит самый быстрый путь от момента заказа до прибытия в точку В?

## **Примерные задания для контрольной работы**

### **Задание 1.**

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

- а) порядок выбора не важен;
  - б) порядок выбора важен;
  - в) можно выбирать одного и того же человека несколько раз?
- Объясните, к каким видам комбинаторики относятся эти задачи.

### **Задание 2.**

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

### **Задание 3.**

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

### **Задание 4.**

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

### **Задание 5.**

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

### **Задание 6.**

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

### **Задание 7.**

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

### **Задание 8.**

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

### **Задание 9.**

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

#### **Задание 10.**

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

### **Примерный перечень вопросов к экзамену**

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