

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет имени В. Н. Татищева»
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО
Руководитель ОПОП
_____ И.М.
Ажмухамедов

«2» июня 2022 г.

УТВЕРЖДАЮ
И.о. заведующего кафедрой ИБ
_____ Р.Ю. Демина
протокол заседания кафедры № 2
от «2» июня 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ МЕТОДЫ В АНАЛИЗЕ
ДАННЫХ
(наименование)

Составитель(-и)	Кошкарлова Т.А., к.т.н., доцент кафедры ЦТ;
Направление подготовки	Кузнецова В.Ю., к.т.н., доцент кафедры ИБ 09.03.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
Направленность (профиль) ОПОП	Безопасность информационных систем
Квалификация (степень)	бакалавр
Форма обучения	очно-заочная
Год приема	2021
Курс	2
Семестр	3

Астрахань, 2022

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1. Целью освоения дисциплины «Вероятностно-статистические методы в анализе данных» является формирование у студентов современных теоретических знаний о вероятностных и статистических закономерностях, практических навыков в решении и исследовании прикладных задач теоретико-вероятностного и статистического характера, выработка у студентов теоретико-вероятностной интуиции, необходимой при решении разнообразных задач, возникающих в профессиональной деятельности.

Изучение дисциплины «Вероятностно-статистические методы в анализе данных» способствует развитию у студентов склонности и способности к творческому мышлению, выработке системного подхода к исследуемым явлениям.

В ходе изучения дисциплины предусматривается проведение лекционных и лабораторных занятий, выполнение практических работ.

В лекциях излагается содержание тем программы с учётом требований, установленных для специалиста в квалификационной характеристике. Основное внимание уделяется наиболее сложным вопросам курса.

Лабораторные занятия проводятся в учебных группах с целью закрепления теоретических основ, излагаемых в лекционном курсе, и получения практических навыков в применении методов теории вероятностей и математической статистики в практической инженерной деятельности.

Бакалавр, изучив дисциплину «Вероятностно-статистические методы в анализе

1.2. Задачи освоения дисциплины:

1) приобретение студентами необходимых знаний основ теории вероятностей и математической статистики;

2) овладение навыками исследования случайных величин, вычисления их основных характеристик, статистического анализа выборок;

3) приобретение знаний и навыков моделирования случайных событий, обработки статистических данных, точечного и интервального оценивания параметров распределений, проверки статистических гипотез;

4) формирование умения интерпретировать результаты вероятностных и статистических исследований и применять их при решении практических задач;

5) применение полученных знаний, навыков и умений в последующей профессиональной деятельности.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

2.1. Учебная дисциплина «Вероятностно-статистические методы в анализе данных» относится к обязательной части учебного плана направления подготовки 09.03.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ, профиль «Безопасность информационных систем» 2021 года набора.

Изучение курса «Вероятностно-статистические методы в анализе данных» рассчитано на один семестр (3 семестр) и предусматривает сдачу студентами зачета на основе рейтинго-балльной системы оценивания. Общая трудоемкость дисциплины – 3 кредита (ЗЕТ) 108 часов в 3-ем семестре.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения и навыки, формируемые предшествующими учебными дисциплинами:

- Математические основы информационных технологий и вычислительной техники.
- Информатика.

В результате освоения этих дисциплин, студент должен:

знать:

- основные понятия и методы математического анализа;
- основные понятия и методы аналитической геометрии;
- основные понятия и методы линейной алгебры и теории алгебраических систем;
- основные понятия информатики,

уметь:

- использовать математические методы и модели для решения прикладных задач;
- использовать программные и аппаратные средства персонального компьютера,

владеть:

- владеть методами количественного анализа процессов обработки, поиска и передачи информации;
- навыками поиска информации в глобальной информационной сети Интернет и работы с офисными приложениями (текстовыми процессорами, электронными таблицами, средствами подготовки презентационных материалов, СУБД и т.п.)..

2.3. Последующие учебные дисциплины (модули) и (или) практики, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной (модулем):

- Теория принятия решений и методы оптимизации.
- Методы искусственного интеллекта.
- Интеллектуальные системы и технологии.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Процесс освоения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки (специальности):

а) универсальных (УК):

- УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

б) общепрофессиональных (ОПК):

- ОПК-7. Способен осуществлять выбор платформ и инструментальных программно- аппаратных средств для реализации информационных систем.

Таблица 1 – Декомпозиция результатов обучения

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.	ИУК-1.1. Знать: методики поиска, сбора и обработки информации; актуальные российские и зарубежные источники информации в сфере профессиональной деятельности; метод системного анализа.	ИУК-1.2. Уметь: применять методики поиска, сбора и обработки информации; осуществлять критический анализ и синтез информации, полученной из разных источников;	ИУК-1.3. Владеть: методами поиска, сбора и обработки, критического анализа и синтеза информации; методикой системного подхода для решения поставленных задач.
ОПК-7. Способен осуществлять выбор	ИОПК-7.1. Знать: основные	ИОПК-7.2. Уметь: осуществлять	ИОПК-7.3. Иметь навыки: владения

платформ и инструментальных программно-аппаратных средств для реализации информационных систем	платформы, технологии и инструментальные программно-аппаратные средства для реализации информационных систем и обеспечения их безопасности.	выбор платформ и инструментальных программно-аппаратных средств для реализации информационных систем, применять современные технологии реализации информационных систем и обеспечения их безопасности.	технологиями и инструментальными программно-аппаратными средствами для реализации информационных систем и обеспечения их безопасности.
--	---	--	--

4. Структура и содержание дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах **3 зачетные единицы**. Всего 108 часов: 36 часов выделено на контактную работу обучающихся с преподавателем (лекции – 18 часов, лабораторные работы – 18 часов), 72 часа – на самостоятельную работу обучающихся.

Таблица 2 – Структура и содержание дисциплины (модуля)

№ п/п	Наименование раздела (темы)	Семестр	Неделя семестра	Контактная работа(в часах)			Самостоят. работа		Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				Л	П	ЛР	КР	СР	
1.	1 Раздел № 1. Основы Python и анализа данных.	3	1-4	4		4		14	Лабораторная работа 1
2.	2 Раздел № 2. Предобработка данных	3	5-7	3		3		14	Лабораторная работа 2
3.	3 Раздел № 3. Исследовательских анализ данных	3	8-11	4		4		14	Контрольная работа 1
4.	4 Раздел № 4. Статистический анализ данных	3	12-14	3		3		14	Лабораторная работа 3
5.	5 Раздел № 5. Теория вероятностей	3	15-18	4		4		16	Лабораторная работа 4 Тест
10	Итого	108		18		18		72	зачет

Примечание: Л – лекция; ПЗ – практическое занятие, семинар; ЛР – лабораторная работа; КР – курсовая работа; СР – самостоятельная работа.

Таблица 3 – Матрица соотнесения тем/разделов учебной дисциплины/модуля и формируемых компетенций

Темы, разделы дисциплины	Кол-во часов	Код компетенции (указываются компетенции перечисленные в п.3)	общее количество компетенций
		УК 1	
Раздел № 1. Основы Python и анализа данных.	22	+	1
Раздел № 2. Предобработка данных	20	+	1
Раздел № 3. Исследовательских анализ данных	22	+	1
Раздел № 4. Статистический анализ данных	20	+	1
Раздел № 5. Теория вероятностей	24	+	1
Итого	108		1

Содержание дисциплины

Раздел № 1. Основы Python и анализа данных.

Введение и основы синтаксиса. Списки и циклы. Операции с таблицами. Условия и функции. Библиотека Pandas. Предобработка данных. Анализ данных и оформление результатов. Веб-оболочка Jupyter Notebook. Процесс анализа данных

Раздел № 2. Предобработка данных

Основные принципы предобработки. Работа с пропусками. Изменение типов. Поиск дубликатов. Категоризация данных.

Раздел № 3. Исследовательских анализ данных

Введение в исследовательский анализ. Графики и выводы. Изучение срезов данных. Работа с несколькими источниками данных. Взаимосвязь данных. Валидация данных

Раздел № 4. Статистический анализ данных

Введение в статистический анализ. Описательная статистика. Математическая статистика. Выборочные методы математической статистики. Основы теории оценивания. Основы проверки статистических гипотез. Основы регрессионного анализа.

Раздел № 5. Теория вероятностей

Случайные события. Пространство элементарных событий. Статистическая вероятность. Условные вероятности. Последовательности независимых испытаний. Случайные величины. Функции и числовые характеристики случайных величин. Основные законы распределения случайных величин. Экспоненциальное распределение и его числовые характеристики. Предельные теоремы теории вероятностей.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)

При подготовке к лекционным и практическим занятиям необходимо воспользоваться учебно-методической литературой из п.8. Лекции необходимо проводить

с использованием презентаций, созданных в прикладном пакете Microsoft Office PowerPoint.

При подготовке к лабораторным занятиям необходимо воспользоваться учебно-методической литературой (основной и дополнительной) и информационно-справочными ресурсами из п.8.

Тестирование допускается проводить в бумажном или электронном виде на специализированных образовательных площадках (Moodle).

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Таблица 4 – Содержание самостоятельной работы обучающихся

<i>Номер раздела (темы)</i>	<i>Темы/вопросы, выносимые на самостоятельное изучение</i>	<i>Кол-во часов</i>	<i>Формы работы</i>
1	Раздел № 1. Основы Python и анализа данных.	14	Внеаудиторная, изучение учебных пособий
2	Раздел № 2. Предобработка данных	14	Внеаудиторная, изучение учебных пособий
3	Раздел № 3. Исследовательских анализ данных	14	Внеаудиторная, изучение учебных пособий
4	Раздел № 4. Статистический анализ данных	14	Внеаудиторная, изучение учебных пособий
5	Раздел № 5. Теория вероятностей	16	Внеаудиторная, изучение учебных пособий

5. 3. Виды и формы письменных работ, предусмотренных при освоении дисциплины, выполняемые обучающимися самостоятельно.

В процессе изучения дисциплины предусмотрены следующие виды и формы письменных работ для самостоятельного выполнения:

- 1) аудиторная контрольная работа;
- 2) типовой расчет - внеаудиторная работа;
- 3) домашнее задание, как теоретического, так и практического характера;
- 4) лабораторная работа, полностью или частично выполняемая в математических пакетах;
- 5) зачетная работа.

Контрольные работы и зачетная работа выполняется студентом в аудитории. Типовой расчет выполняется вне аудитории за определенный промежуток времени, установленный преподавателем, оформляется в отдельной тетради. В установленный срок студент сдает типовой расчет и устно отчитывается преподавателю по выполненной работе. Зачетную работу студент сдает в том случае, если по результатам семестрового контроля он набрал рейтинг менее 60 баллов. Лабораторные работы выполняются в аудитории и оцениваются как зачет/незачет. Для получения зачета по дисциплине по результатам текущего семестрового контроля студенту необходимо иметь рейтинг по

контрольным работам не менее 60 баллов и зачет по всем лабораторным работам.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

6.1. Образовательные технологии

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Раздел № 1. Основы Python и анализа данных.	Обзорная лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы
Раздел № 2. Предобработка данных	Лекция-диалог	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы
Раздел № 3. Исследовательских анализ данных	Лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы,
Раздел № 4. Статистический анализ данных	Лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы
Раздел № 5. Теория вероятностей	Обзорная лекция	Не предусмотрено	выполнение контрольной работы, выполнение лабораторной работы

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line в формах: видеолекций, лекций-презентаций, видеоконференции, собеседования в режиме чат, форума, чата, выполнения виртуальных практических и/или лабораторных работ и др.

Максимальный объем занятий обучающегося с применением электронных образовательных технологий не должен превышать 25%.

6.2. Информационные технологии

При реализации различных видов учебной и внеучебной работы используются

следующие информационные технологии:

- использование возможностей Интернета в учебном процессе (использование информационного сайта преподавателя (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление учащихся с оценками и т.д.));
- __использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источников информации;
- использование возможностей электронной почты преподавателя;
- __использование средств представления учебной информации (электронных учебных пособий и практикумов, применение новых технологий для проведения очных (традиционных) лекций и семинаров с использованием презентаций и т.д.);
- __использование интегрированных образовательных сред, где главной составляющей являются не только применяемые технологии, но и содержательная часть, т.е. информационные ресурсы (доступ к мировым информационным ресурсам, на базе которых строится учебный процесс);
- __использование виртуальной обучающей среды (или системы управления обучением LMS Moodle «Электронное образование») или иных информационных систем, сервисов и мессенджеров.

6.3. Перечень программного обеспечения и информационных справочных систем

6.3.1. Программное обеспечение:

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда
Google Chrome	Браузер
Microsoft Office 2013, Microsoft Office Project 2013 , Microsoft Office Visio 2013	Офисная программа
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты

6.3.2. Современные профессиональные базы данных и информационные справочные системы:

1. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем»: <https://library.asu.edu.ru>.
2. Электронный каталог «Научные журналы АГУ»: <http://journal.asu.edu.ru/>.
3. Универсальная справочно-информационная полнотекстовая база данных периодических изданий ООО «ИВИС»: <http://dlib.eastview.com/>
4. Электронно-библиотечная система elibrary. <http://elibrary.ru>
5. Справочная правовая система КонсультантПлюс: <http://www.consultant.ru>
6. Информационно-правовое обеспечение «Система ГАРАНТ»: <http://garant-astrakhan.ru>

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине (модулю) «Вероятностно-статистические методы в анализе данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 6 – Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

№ п/п	Контролируемые разделы (темы) дисциплины*	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Раздел № 1. Основы Python и анализа данных.	УК-1, ОПК-7	Лабораторная работа 1
2.	Раздел № 2. Предобработка данных	УК-1, ОПК-7	Лабораторная работа 2
3.	Раздел № 3. Исследовательских анализ данных	УК-1, ОПК-7	Контрольная работа 1
4.	Раздел № 4. Статистический анализ данных	УК-1, ОПК-7	Лабораторная работа 3
5.	Раздел № 5. Теория вероятностей	УК-1, ОПК-7	Лабораторная работа 4 Тест, вопросы к зачету

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Для оценки результатов обучения применяются следующие критерии:

Таблица 7 – Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2	демонстрирует существенные пробелы в знании теоретического

«неудовлетворительно»	материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры
-----------------------	--

Таблица 8 – Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, не способен применить знание теоретического материала при выполнении заданий, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание при подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задание

7.3. Контрольные задания и иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

Раздел № 1. Основы Python и анализа данных.

Лабораторная работа 1

Заказчик — кредитный отдел банка. Нужно разобраться, влияет ли семейное положение и количество детей клиента на факт погашения кредита в срок. Входные данные от банка — статистика о платёжеспособности клиентов.

Результаты исследования будут учтены при построении модели *кредитного скоринга* — специальной системы, которая оценивает способность потенциального заёмщика вернуть кредит банку.

Инструкция по выполнению

Шаг 1. Откройте таблицу и изучите общую информацию о данных

Шаг 2. Предобработка данных

1. В двух столбцах есть пропущенные значения. Один из них — `days_employed`.

Пропуски в этом столбце вы обработаете на следующем этапе. Найдите другой столбец и заполните пропущенные значения в нём медианным значением:

- опишите, какие пропущенные значения вы обнаружили;
- проверьте, какую долю составляют пропущенные значения в каждом из столбцов с пропусками;
- приведите возможные причины появления пропусков в данных;
- объясните, почему заполнить пропуски медианным значением — лучшее решение для количественных переменных.

2. В данных могут встречаться артефакты (аномалии) — значения, которые не отражают действительность и появились по какой-то ошибке. Например, отрицательное количество дней трудового стажа в столбце `days_employed`. Для реальных данных это нормально. Обработайте значения в столбцах с аномалиями и опишите возможные причины появления таких данных. После обработки аномалий заполните пропуски в `days_employed` медианными значениями по этому столбцу.
3. Замените вещественный тип данных в столбце `total_income` на целочисленный, например, с помощью метода `astype()`.
4. Если в данных присутствуют строки-дубликаты, удалите их. Также обработайте неявные дубликаты. Например, в столбце `education` есть одни и те же значения, но записанные по-разному: с использованием заглавных и строчных букв. Приведите такие значения к одному регистру. После удаления дубликатов сделайте следующее:
 - поясните, как выбирали метод для поиска и удаления дубликатов в данных;
 - приведите возможные причины появления дубликатов.
5. Создайте два новых датафрейма, в которых:
 - каждому уникальному значению из `education` соответствует уникальное значение `education_id` — в первом;
 - каждому уникальному значению из `family_status` соответствует уникальное значение `family_status_id` — во втором.

Удалите из исходного датафрейма столбцы `education` и `family_status`, оставив только их идентификаторы: `education_id` и `family_status_id`. Новые датафреймы — это те самые «словари» (не путайте с одноимённой структурой данных в Python), к которым вы сможете обращаться по идентификатору.
6. На основании диапазонов, указанных ниже, создайте столбец `total_income_category` с категориями:
 - 0–30000 — 'E';
 - 30001–50000 — 'D';
 - 50001–200000 — 'C';
 - 200001–1000000 — 'B';
 - 1000001 и выше — 'A'.

Например, кредитополучателю с доходом 25000 нужно назначить категорию 'E', а клиенту, получающему 235000, — 'B'.
7. Создайте функцию, которая на основании данных из столбца `purpose` сформирует новый столбец `purpose_category`, в который войдут следующие категории:
 - 'операции с автомобилем',
 - 'операции с недвижимостью',
 - 'проведение свадьбы',
 - 'получение образования'.

Например, если в столбце `purpose` находится подстрока 'на покупку автомобиля', то в столбце `purpose_category` должна появиться строка 'операции с автомобилем'.

Вы можете использовать собственную функцию и метод `apply()`. Изучите данные в столбце `purpose` и определите, какие подстроки помогут вам правильно определить категорию.

Шаг 3. Ответьте на вопросы

Ответы на вопросы можно разместить в ячейках тетрадок Jupyter Notebook с типом `markdown`.

- Есть ли зависимость между количеством детей и возвратом кредита в срок?
- Есть ли зависимость между семейным положением и возвратом кредита в срок?

- Есть ли зависимость между уровнем дохода и возвратом кредита в срок?
- Как разные цели кредита влияют на его возврат в срок?

Ответы сопроводите интерпретацией — поясните, о чём именно говорит полученный вами результат.

Шаг 4. Напишите общий вывод

Оформление: Задание выполните в Jupyter Notebook. Программный код заполните в ячейках типа code, текстовые пояснения — в ячейках типа markdown. Примените форматирование и заголовки.

Описание данных

- children — количество детей в семье
- days_employed — общий трудовой стаж в днях
- dob_years — возраст клиента в годах
- education — уровень образования клиента
- education_id — идентификатор уровня образования
- family_status — семейное положение
- family_status_id — идентификатор семейного положения
- gender — пол клиента
- income_type — тип занятости
- debt — имел ли задолженность по возврату кредитов
- total_income — ежемесячный доход
- purpose — цель получения кредита

Раздел № 2. Предобработка данных

Лабораторная работа 2

Описание проекта

В вашем распоряжении данные сервиса Яндекс Недвижимость — архив объявлений о продаже квартир в Санкт-Петербурге и соседних населённых пунктах за несколько лет. Нужно научиться определять рыночную стоимость объектов недвижимости. Ваша задача — установить параметры. Это позволит построить автоматизированную систему: она отследит аномалии и мошенническую деятельность.

По каждой квартире на продажу доступны два вида данных. Первые вписаны пользователем, вторые — получены автоматически на основе картографических данных. Например, расстояние до центра, аэропорта, ближайшего парка и водоёма.

Инструкция по выполнению проекта

Шаг 1. Откройте файл с данными и изучите общую информацию

Путь к файлу: /datasets/real_estate_data.csv.

Шаг 2. Предобработка данных

1. Определите и изучите пропущенные значения:
 - Для некоторых пропущенных значений можно предположить логичную замену. Например, если человек не указал число балконов — скорее всего, их нет. Такие пропуски правильно заменить на 0. Для других типов данных нет подходящего значения на замену. В этом случае правильно оставить эти значения пустыми. Отсутствие значения — тоже важный сигнал, который не нужно прятать.
 - Заполните пропуски, где это уместно. Опишите, почему вы решили заполнить пропуски именно в этих столбцах и как выбрали значения.
 - Укажите причины, которые могли привести к пропускам в данных.
2. Приведите данные к нужным типам:
 - Поясните, в каких столбцах нужно изменить тип данных и почему.

Шаг 3. Посчитайте и добавьте в таблицу

- цену квадратного метра;
- день недели, месяц и год публикации объявления;
- этаж квартиры; варианты — первый, последний, другой;
- соотношение жилой и общей площади, а также отношение площади кухни к общей.

Шаг 4. Проведите исследовательский анализ данных и выполните инструкции

- Изучите следующие параметры: площадь, цена, число комнат, высота потолков. Постройте гистограммы для каждого параметра.
- Изучите время продажи квартиры. Постройте гистограмму. Посчитайте среднее и медиану. Опишите, сколько обычно занимает продажа. Когда можно считать, что продажи прошли очень быстро, а когда необычно долго?
- Уберите редкие и выбивающиеся значения. Опишите, какие особенности обнаружили.
- Какие факторы больше всего влияют на стоимость квартиры? Изучите, зависит ли цена от площади, числа комнат, удалённости от центра. Изучите зависимость цены от того, на каком этаже расположена квартира: первом, последнем или другом. Также изучите зависимость от даты размещения: дня недели, месяца и года.
- Выберите 10 населённых пунктов с наибольшим числом объявлений. Посчитайте среднюю цену квадратного метра в этих населённых пунктах. Выделите населённые пункты с самой высокой и низкой стоимостью жилья. Эти данные можно найти по имени в столбце `locality_name`.
- Изучите предложения квартир: для каждой квартиры есть информация о расстоянии до центра. Выделите квартиры в Санкт-Петербурге (`locality_name`). Ваша задача — выяснить, какая область входит в центр. Создайте столбец с расстоянием до центра в километрах: округлите до целых значений. После этого посчитайте среднюю цену для каждого километра. Постройте график: он должен показывать, как цена зависит от удалённости от центра. Определите границу, где график сильно меняется, — это и будет центральная зона.
- Выделите сегмент квартир в центре. Проанализируйте эту территорию и изучите следующие параметры: площадь, цена, число комнат, высота потолков. Также выделите факторы, которые влияют на стоимость квартиры (число комнат, этаж, удалённость от центра, дата размещения объявления). Сделайте выводы. Отличаются ли они от общих выводов по всей базе?

Шаг 5. Напишите общий вывод

Оформление: Выполните задание в Jupyter Notebook. Заполните программный код в ячейках типа `code`, текстовые пояснения — в ячейках типа `markdown`. Примените форматирование и заголовки.

Описание данных

- `airports_nearest` — расстояние до ближайшего аэропорта в метрах (м)
- `balcony` — число балконов
- `ceiling_height` — высота потолков (м)
- `cityCenters_nearest` — расстояние до центра города (м)
- `days_exposition` — сколько дней было размещено объявление (от публикации до снятия)
- `first_day_exposition` — дата публикации
- `floor` — этаж
- `floors_total` — всего этажей в доме
- `is_apartment` — апартаменты (булев тип)
- `kitchen_area` — площадь кухни в квадратных метрах (м²)
- `last_price` — цена на момент снятия с публикации
- `living_area` — жилая площадь в квадратных метрах (м²)
- `locality_name` — название населённого пункта

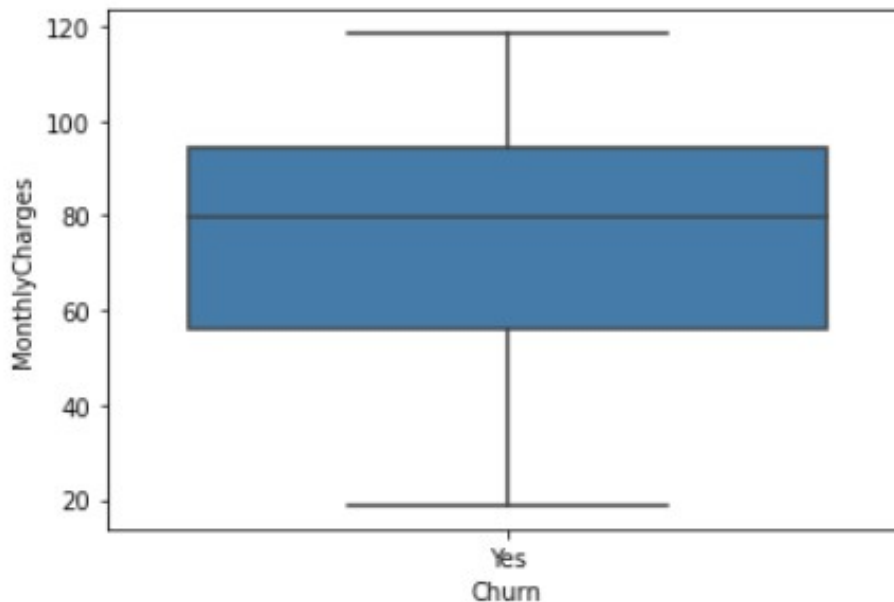
- open_plan — свободная планировка (булев тип)
- parks_around3000 — число парков в радиусе 3 км
- parks_nearest — расстояние до ближайшего парка (м)
- ponds_around3000 — число водоёмов в радиусе 3 км
- ponds_nearest — расстояние до ближайшего водоёма (м)
- rooms — число комнат
- studio — квартира-студия (булев тип)
- total_area — площадь квартиры в квадратных метрах (м²)
- total_images — число фотографий квартиры в объявлении

Раздел № 3. Исследовательских анализ данных

Контрольная работа 1

1 задание

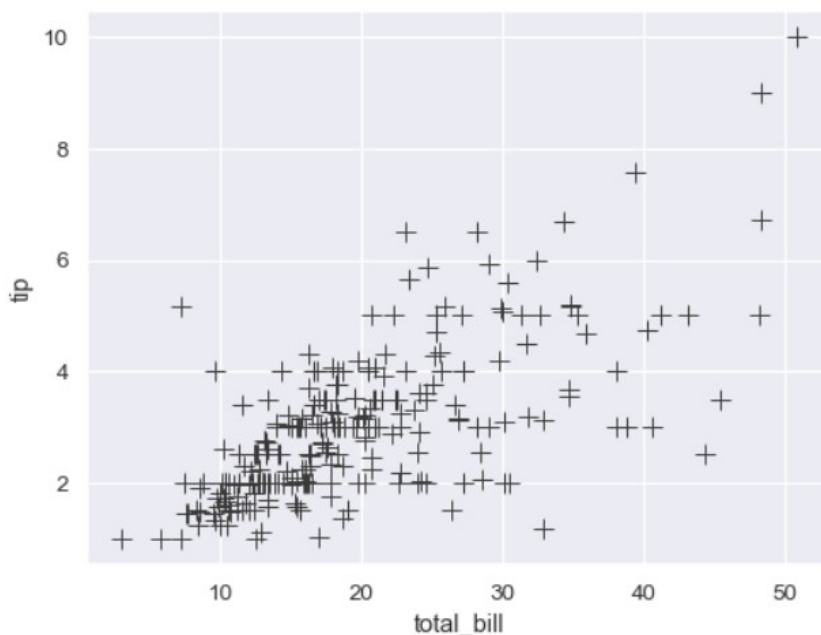
На графике нарисован boxplot или ящик с усами. Определите по нему межквартильный диапазон



(размах).

2 задание

Как называется данный тип диаграммы?



3 задание

Перед вами график изменения возраста посетителей сайта во времени. Какой вывод можно сделать исходя из данной визуализации?



Раздел № 4. Статистический анализ данных

Лабораторная работа № 3

Дискретная случайная величина.

Дискретной случайной величиной (X) называется случайная величина, которая в результате испытания принимает отдельные значения (x_1, x_2, \dots) с определёнными вероятностями (p_1, p_2, \dots). Число возможных значений дискретной случайной величины может быть конечным и бесконечным.

Соотношение, устанавливающее связь между отдельными возможными значениями

случайной величины и соответствующими им вероятностями, называется законом распределения дискретной случайной величины:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

$$\sum_{i=1}^n p_i = 1$$

Закон (ряд) распределения можно изобразить графически, в виде точек с координатами (x_i, p_i) , соединённых отрезками. Получим многоугольник распределения вероятностей (полигон распределения).

Дискретная случайная величина может быть задана функцией распределения.

Функцией распределения случайной величины X называется функция $F(x)$, выражающая вероятность того, что X примет значение, меньшее чем x :

$$F(x) = P(X < x)$$

Пример. Закон распределения случайной величины X :

X	0	1	2	3
P	0,198	0,457	0,293	0,052

$$\sum_{i=1}^4 p_i = 0,198 + 0,457 + 0,293 + 0,052 = 1$$

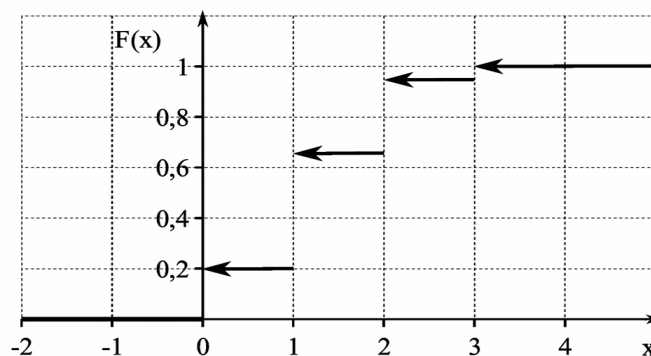
Проверка:

Функция распределения вероятностей $F(x)$ случайной величины X :

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0,198, & 0 < x \leq 1 \\ 0,198 + 0,457 = 0,655, & 1 < x \leq 2 \\ 0,198 + 0,457 + 0,293 = 0,948, & 2 < x \leq 3 \\ 0,198 + 0,457 + 0,293 + 0,052 = 1, & x > 3 \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 0,198, & 0 < x \leq 1 \\ 0,655, & 1 < x \leq 2 \\ 0,948, & 2 < x \leq 3 \\ 1, & x > 3 \end{cases}$$

График функции распределения:



Числовые характеристики дискретной случайной величины.

Математическое ожидание случайной величины X .

$$M(X) = \sum_{i=1}^n p_i \cdot x_i$$

Дисперсия случайной величины X .

$$D(X) = M\left(\left(X - M(X)\right)^2\right) = M(X^2) - (M(X))^2 = \sum_{i=1}^n p_i \cdot x_i^2 - (M(X))^2$$

Среднее квадратическое отклонение.

$$\text{sigma}(X) = \sqrt{D(X)}$$

Обобщёнными числовыми характеристиками для случайных величин в теории вероятностей, а также математической статистике являются начальные и центральные моменты.

Начальным моментом k -го порядка случайной величины X называют математическое ожидание от величины в k -ой степени:

$$\text{inu}_k = M(X^k) = \sum_{i=1}^n p_i \cdot x_i^k, \quad k = 1, 2, 3, \dots$$

Начальный момент первого порядка:

$$\text{inu}_1 = M(X) = \sum_{i=1}^n p_i \cdot x_i$$

Начальный момент второго порядка:

$$\text{inu}_2 = M(X^2) = \sum_{i=1}^n p_i \cdot x_i^2$$

Начальный момент третьего порядка:

$$\text{inu}_3 = M(X^3) = \sum_{i=1}^n p_i \cdot x_i^3$$

Центральным моментом k -го порядка случайной величины X называют математическое ожидание от величины $(X - M(X))^k$:

$$\text{itmu}_k = M\left(\left(X - M(X)\right)^k\right) = \sum_{i=1}^n p_i \cdot (x_i - M(X))^k, \quad k = 1, 2, 3, \dots$$

Центральный момент первого порядка:

$$\text{itmu}_1 = M(X - M(X)) = M(X) - M(X) = 0$$

Центральный момент второго порядка:

$$\text{itmu}_2 = M\left(\left(X - M(X)\right)^2\right) = \sum_{i=1}^n p_i \cdot (x_i - M(X))^2 = D(X) = M(X^2) - (M(X))^2 = \text{inu}_2 - \text{inu}_1^2$$

Центральный момент третьего порядка:

$$\text{itmu}_3 = M\left(\left(X - M(X)\right)^3\right) = \sum_{i=1}^n p_i \cdot (x_i - M(X))^3$$

Задание.

1. Построить многоугольник распределения.
2. Составить функцию распределения и построить её график.
3. Найти начальные и центральные моменты первого, второго и третьего порядка.
4. Найти числовые характеристики случайной величины (математическое ожидание, дисперсию, среднее квадратическое отклонение).

Вариант № 1

X	23	28	34	45	47	52	56	67	69	73
P	0,01	0,03	0,04	0,13	0,15	0,28	0,16	0,08	0,06	0,06

Вариант № 2

X	35	40	46	57	59	64	68	79	81	85
P	0,05	0,07	0,14	0,31	0,18	0,11	0,05	0,04	0,03	0,02

Вариант № 3

X	65	115	175	285	305	355	395	505	525	565
P	0,02	0,03	0,04	0,11	0,13	0,15	0,16	0,24	0,09	0,03

Вариант № 4

X	64	79	97	130	136	151	163	196	202	214
P	0,01	0,04	0,08	0,13	0,34	0,18	0,12	0,07	0,02	0,01

Вариант № 5

X	61	71	83	105	109	119	127	149	153	161
P	0,01	0,02	0,04	0,25	0,19	0,18	0,16	0,08	0,04	0,03

Раздел № 5. Теория вероятностей

Лабораторная работа 4

«Первичная обработка эмпирических данных»

Имеющийся набор эмпирических данных является выборкой из генеральной совокупности.

Набор данных расположенный в порядке возрастания, называется вариационным рядом. Если набор данных достаточно большой, то удобнее всего представить его в виде интервального вариационного ряда.

Для построения интервального вариационного ряда необходимо выполнить следующие действия:

1. Имеющиеся данные располагают в порядке возрастания.

2. В выборке определяют самое большое x_{max} и самое маленькое x_{min} значение изучаемого признака.

3. Определяют размах варьирования $R = x_{max} - x_{min}$.

4. Определяют ширину частичных интервалов $h = \frac{R}{k}$, где k — число частичных интервалов (целое число). Число интервалов приблизительно можно определить с помощью формулы Стержеса: $k = 1 + 3,32 \cdot \lg n$ (n — объем выборки).

5. Нижняя граница первого интервала x_0 выбирается так, чтобы минимальная варианта выборки x_{min} попадала примерно в середину этого интервала: $x_0 = x_{min} - \frac{1}{2} \cdot h$.

Промежуточные интервалы получают прибавляя к концу предыдущего интервала длину частичного интервала h : $x_i = x_{i-1} + h$, $i = 1, 2, \dots$. Построение шкалы интервалов на основе вычисления границ интервалов продолжается до тех пор, пока величина x_i

удовлетворяет соотношению: $x_i \leq x_{max} + \frac{1}{2} \cdot h$.

6. Подсчитывается количество значений признака, попадающих в каждый частичный интервал (частоты n_i).

7. Результаты формируются в виде таблицы, которая и является интервальным

вариационным рядом:

$[x_0; x_1)$	$[x_1; x_2)$	$[x_2; x_3)$...	$[x_{i-1}; x_i)$...	$[x_{k-1}; x_k]$
n_1	n_2	n_3	...	n_i	...	n_k

$$\sum_{i=1}^k n_i = n$$

(Проверка: сумма частот должна совпадать с объёмом выборки).

Задание.

Построить интервальный вариационный ряд.

Вариант № 1.

Пораженность отливок точечными поверхностными дефектами (ТПД).

№ п/п	ТПД, %	№ п/п	Пораженность ТПД, %
1	10,56	41	20,71
2	10,98	42	20,89
3	12,45	43	21,24
4	12,47	44	21,67
5	13,37	45	21,72
6	13,38	46	21,96
7	13,81	47	21,96
8	13,91	48	22,31
9	14,3	49	22,5
10	14,79	50	22,68
11	14,84	51	22,79
12	15,24	52	23,03
13	15,29	53	23,06
14	16,31	54	23,26
15	16,39	55	23,54
16	16,41	56	23,6
17	16,52	57	23,92
18	17,19	58	24,19
19	17,21	59	24,25
20	17,27	60	24,71
21	17,3	61	24,71
22	17,39	62	24,73
23	17,6	63	25,1
24	17,62	64	25,4
25	17,75	65	25,56
26	17,92	66	25,61
27	18,58	67	25,64
28	18,62	68	26,81
29	18,69	69	26,86
30	18,72	70	27,07
31	19,4	71	27,58
32	19,4	72	27,58
33	19,57	73	27,69
34	19,89	74	28,6
35	20,06	75	29,08
36	20,17	76	30,26
37	20,2	77	30,74
38	20,41	78	30,86
39	20,48	79	30,86
40	20,62	80	33,76

Перечень вопросов к зачету

1. Введение и основы синтаксиса. Списки и циклы.
2. Операции с таблицами. Условия и функции.
3. Библиотека Pandas. Анализ данных и оформление результатов.
4. Веб-оболочка Jupyter Notebook.

5. Процесс анализа данных.
6. Предобработка данных.
7. Основные принципы предобработки. Работа с пропусками.
8. Изменение типов. Поиск дубликатов. Категоризация данных.
9. Введение в исследовательский анализ. Графики и выводы.
10. Изучение срезов данных.
11. Работа с несколькими источниками данных.
12. Взаимосвязь данных.
13. Валидация данных.
14. Введение в статистический анализ.
15. Описательная статистика.
16. Математическая статистика.
17. Выборочные методы математической статистики.
18. Основы теории оценивания.
19. Основы проверки статистических гипотез.
20. Основы регрессионного анализа.
21. Случайные события.
22. Пространство элементарных событий.
23. Статистическая вероятность.
24. Условные вероятности.
25. Последовательности независимых испытаний.
26. Случайные величины.
27. Функции и числовые характеристики случайных величин.
28. Основные законы распределения случайных величин.
29. Экспоненциальное распределение и его числовые характеристики.
30. Предельные теоремы теории вероятностей.

Таблица 9 – Примеры оценочных средств с ключами правильных ответов

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.				
1.	Задание закрытого типа	На плоскости нарисованы две концентрические окружности, радиусы которых 6 и 12 см соответственно. Какова вероятность того, что точка брошенная наудачу в большой круг, попадет в кольцо, образованное указанными окружностями? 1) 0,5; 2) 0,65; 3) 0,12; 4) 0,75; 5) 0,60.	4	2
2.		На сборку попадают детали с двух автоматов: 80 % из первого и 20 % из второго. Первый автомат дает 10 % брака, второй – 5 % брака. Найти вероятность попадания на сборку доброкачественной детали. 1) 0,90; 2) 0,09;	3	3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		3) 0,91; 4) 0,85; 5) 0,15		
3.		Случайная величина распределена по нормальному закону, причем $M(X) = 15$. Найти $P(10 < X < 15)$, если известно, что $P(15 < X < 20) = 0,25$. 1) 0,10; 2) 0,15; 3) 0,20; 4) 0,25; 5) 0,30.	4	3
4.		По выборке объема $n = 51$ найдена смещенная оценка $Dv=3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности. 1) 3,05; 2) 3,06; 3) 3,51; 4) 3,60; 5) 0.	2	3
5.		Как называется число m_0 (наступления события в n независимых испытаниях, в каждом из которых вероятность появления события равна p), определяемое из неравенства: $np - q \leq m_0 \leq np + p$ 1) наибольшее 2) оптимальное 3) наимвероятнейшее 4) невозможное 5) минимальное	3	3
6.	Задание открытого типа	На пяти одинаковых карточках написаны буквы И, Л, О, С, Ч. Если перемешать их, и разложить наудачу в ряд две карточки, то вероятность p получить слово ИЛ равна В ответе запишите число $1/p$.	20	5
7.		Данное предприятие в среднем выпускает 20 % продукции высшего сорта и 70 % продукции первого сорта. Найти вероятность P того, что случайно взятое изделие этого предприятия будет высшего или первого сорта. В ответ записать число $30P$.	27	6
8.		Случайная величина X распределена равномерно на интервале $(2; 6)$ и $p(x)$ – ее плотность вероятности. Найти $p(3)$. В ответ записать число $40p(3)$.	10	8
9.		Студентам нужно сдать 4 экзамена за 6 дней. Сколькими способами можно составить расписание сдачи экзаменов?	360	8
10.		Вероятность появления события A в каждом из 100 независимых испытаний	64	8

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		равна 0,4. Найдите математическое ожидание и дисперсию случайной величины X – числа появлений события A. В ответ запишите их сумму.		
ОПК-7. Способен осуществлять выбор платформ и инструментальных программно- аппаратных средств для реализации информационных систем				
11.	Задание закрытого типа	Когда применяется классический способ задания вероятности: 1) пространство элементарных событий бесконечно, все события равновозможные и независимые; 2) пространство элементарных событий замкнуто, все события независимы; 3) пространство элементарных событий конечно, все события равновозможные; 4) пространство элементарных событий конечно, все элементарные события независимы.	3	2
12.		Когда применяется геометрический способ задания вероятности: 1) пространство элементарных событий бесконечно, все события равновозможные и независимые; 2) пространство элементарных событий замкнуто, все события независимы; 3) пространство элементарных событий конечно, все события равновозможные; 4) пространство элементарных событий конечно, все элементарные события независимы.	1	2
13.		Какое из утверждений относительно генеральной и выборочной совокупностей является верным? 1) выборочная совокупность – часть генеральной 2) генеральная совокупность – часть выборочной 3) выборочная и генеральная совокупности равны по численности 4) правильный ответ отсутствует	1	2
14.		Сумма частот признака равна: 1) объему выборки n 2) среднему арифметическому значений признака 3) нулю 4) единице	1	2
15.		Уточненная выборочная дисперсия S^2 случайной величины X обладает следующими свойствами: 1) является смещенной оценкой дисперсии случайной величины X 2) является несмещенной оценкой дисперсии случайной величины X 3) является смещенной оценкой среднеквадратического отклонения	2	2

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)										
		случайной величины X 4) является несмещенной оценкой среднеквадратического отклонения случайной величины X												
16.	Задание открытого типа	Случайная величина X задана функцией распределения $F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ x^2 & \text{при } 0 < x \leq 1, \\ 1 & \text{при } x > 1. \end{cases}$ <p>Найти вероятность P того, что в результате испытания случайная величина X примет значение, принадлежащее интервалу $(0,1; 0,6)$. В ответ записать число $20P$.</p>	7	8										
17.		Задан статистический ряд распределения <table border="1" style="margin-left: 20px;"> <tr> <td>Варианта x_i</td> <td>1</td> <td>2</td> <td>5</td> <td>7</td> </tr> <tr> <td>Частота n_i</td> <td>10</td> <td>50</td> <td>25</td> <td>15</td> </tr> </table> <p>Найти выборочную среднюю \bar{X}. В ответ записать число $5\bar{X}$.</p>	Варианта x_i	1	2	5	7	Частота n_i	10	50	25	15	17	8
Варианта x_i	1	2	5	7										
Частота n_i	10	50	25	15										
18.		Средний расход электроэнергии в некотором регионе составляет 40000 квт/ч. Пользуясь неравенством Маркова, оценить вероятность того, что расход электроэнергии не превысит 50000 квт/ч. В ответ запишите 10 р.	2	8										
19.		В группе из 20 студентов 4 отличника и 16 хорошистов. Вероятности успешной сдачи сессии для них соответственно равны 0,9 и 0,65. Найдите вероятность того, что наугад выбранный студент успешно сдаст сессию. В ответ запишите 10 р.	7	8										
20.		Время ремонта автомобиля есть случайная величина X , имеющая показательное распределение с параметром $\lambda = 0,1$. Найдите среднее время ремонта автомобиля.	10	8										

Полный комплект оценочных материалов по дисциплине (модулю) (фонд оценочных средств) хранится в электронном виде на кафедре, утверждающей рабочую программу дисциплины (модуля), и в Центре мониторинга и аудита качества обучения.

7.4. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю)

По дисциплине «Вероятностно-статистические методы в анализе данных» итоговой формой отчетности в четвертом семестре является зачет. Согласно действующей в АГУ системе оценивания БАРС для дисциплин, итоговой формой отчетности для которых является зачет, отводится 100 баллов (90 баллов на текущие формы контроля и до 10 баллов отводится на бонусы), которые накапливаются студентом в течение всего семестра изучения дисциплины и распределяются по возможности равномерно по всему семестру. Система накопления баллов, а также система штрафов, представлена в технологической

карте дисциплины «Вероятностно-статистические методы в анализе данных».

Отчет по лабораторной работе

Отчет по лабораторной работе представляется в электронном виде. Защита отчета проходит в форме доклада студента по выполненной работе и ответов на вопросы преподавателя. В случае, если оформление отчета и поведение студента во время защиты соответствуют указанным требованиям, студент получает максимальное количество баллов.

Основаниями для снижения количества баллов в диапазоне от max до min являются:

- небрежное выполнение,
- отсутствие выводов,
- нарушение сроков предоставления отчета.

Отчет не может быть принят и подлежит доработке в случае:

- отсутствия необходимых разделов,
- отсутствия необходимого графического материала,
- неверных результатов расчета.

Контрольные работы

Контрольная работа состоит из 2-х заданий.

Основаниями для снижения оценки за задание являются:

- ошибки в объяснениях и комментариях при верно выполненном задании;
- неполный ответ для теоретических заданий;
- небрежное выполнение;
- многократное переписывание контрольной работы.

Задание не может быть засчитано, если:

- даны два неверных ответа на теоретические вопросы.

В соответствии с балльно-рейтинговой системой БАРС по дисциплине отводится 100 баллов (90 баллов на текущие формы контроля и до 10 баллов отводится на бонусы), которые накапливаются студентом в течение всего семестра изучения дисциплины.

Оценивание студентов на зачете осуществляется в соответствии с требованиями и критериями 100-балльной шкалы. Учитываются как результаты текущего контроля, так и знания, навыки и умения, непосредственно показанные студентами в ходе зачета.

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения самостоятельных и тематических контрольных работ. Он предусматривает проверку готовности студентов к плановым занятиям, оценку качества и самостоятельности выполнения заданий на практических занятиях, проверку правильности решения задач, выданных на самостоятельную проработку.

На зачете осуществляется комплексная проверка знаний, навыков и умений студентов по всему теоретическому материалу дисциплины и с проверкой практических навыков и умений по разработке документов различных видов. Теоретические знания оцениваются путем компьютерного тестирования или на основании письменных ответов студентов по нескольким теоретическим вопросам.

Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине (модулю)

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
Основной блок				
1.	<i>Выполнение лабораторной работы</i>	4/15	60	По расписанию

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
2.	<i>Выполнение контрольной работы</i>	1/15	15	нию
3.	<i>Тест</i>	1/15	15	
Всего			90	-
Блок бонусов				
4.	<i>Посещение занятий без пропусков</i>	1	3	
5.	<i>Своевременное выполнение всех заданий</i>	1	3	
6.	<i>Активность студента на занятии</i>	1	4	
Всего			10	-
ИТОГО			100	-

Таблица 11 – Система штрафов (для одного занятия)

Показатель	Балл
<i>Опоздание на занятие</i>	- 1
<i>Нарушение учебной дисциплины</i>	- 1
<i>Неготовность к занятию</i>	- 2
<i>Пропуск занятия без уважительной причины</i>	- 2

Таблица 12 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине (модулю)

Сумма баллов	Оценка по 4-балльной шкале	
90–100	5 (отлично)	Зачтено
85–89	4 (хорошо)	
75–84		
70–74		
65–69	3 (удовлетворительно)	
60–64	2 (неудовлетворительно)	Не зачтено
Ниже 60		

При реализации дисциплины (модуля) в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

8.1. Основная литература:

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г.

Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262>

2. Теория вероятностей и математическая статистика. Практикум [Электронный ресурс]: учеб. пособие / И.Ю. Мацкевич, Н.П. Петрова, Л.И. Тарусина - Минск : РИПО, 2017. - URL: <http://www.studentlibrary.ru/book/ISBN9789855037119.html> (ЭБС «Консультант студента»).

3. Ракитский, А. А. Методы машинного обучения : учебно-методическое пособие / А. А. Ракитский. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/90591> .

4. Рашка С., Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - М. : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970604090.html>

5. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166>

8.2. Дополнительная литература:

1. Теория вероятностей. Примеры и задачи [Электронный ресурс]: учеб. пособие / Васильчик М.Ю. - Новосибирск : Изд-во НГТУ, 2014. - URL: <http://www.studentlibrary.ru/book/ISBN9785778224872.html> (ЭБС «Консультант студента»).

2. Теория вероятностей и математическая статистика [Электронный ресурс] / Балдин К. В. - М. : Дашков и К, 2014. - URL: <http://www.studentlibrary.ru/book/ISBN9785394021084.html> (ЭБС «Консультант студента»).

3. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/26444.html>

4. Билл, Фрэнкс Укрощение больших данных : как извлекать знания из массивов информации с помощью глубокой аналитики / Фрэнкс Билл ; перевод А. Баранов. — Москва : Манн, Иванов и Фербер, 2014. — 340 с. — ISBN 978-5-00057-146-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/39433.html>

5. Флах П., Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Флах П. - М. : ДМК Пресс, 2015. - 400 с. - ISBN 978-5-97060-273-7 - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970602737.html>

8.3. Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1 Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». www.studentlibrary.ru.

2 Электронно-библиотечная система IPR BOOKS <http://www.iprbookshop.ru/>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения лабораторных занятий необходима компьютерная аудитория, в которой организован доступ к сети Интернет и установлено программное обеспечение. Для проведения публичной защиты проектов, необходима мультимедийная аудитория с проектором.

Во время лекций используется ноутбук и проектор для презентаций. Семинары проходят в компьютерном классе. Слушателям курса рекомендуется использовать свои ноутбуки с установленной последней версией Anaconda для Python 3.7.

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).