



КГУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГУ»)

УТВЕРЖДАЮ

Директор института цифровых
технологий и экономики

Э.И. Беляев

29 ноября 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.08 Методы статистики для разработчиков

Направление подготовки

09.03.03 Прикладная информатика

Направленность (профиль)

Прикладной искусственный интеллект

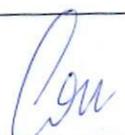
Квалификация

Бакалавр

г. Казань, 2023

Программу разработал(и):

Наименование кафедры	Должность, уч. степень, уч. звание	ФИО разработчика
УрФУ Департамент ИТиА	Доцент, к.ф.-м.н.	Белоусова В.И.
КГЭУ ИТИС	Доцент, к.т.н., доцент	Гадельшина Г.А.

Согласование	Наименование подразделения	Дата	№ протокола	Подпись
Одобрена	ИТИС	27.11.23	11	 И.о. зав. кафедрой, к.ф.-м.н., доц., Соловьев С.А.
Согласована	Учебно-методический совет ИЦТЭ	27.11.23	3	 Директор, к.т.н., Беляев Э.И.
Одобрена	Ученый совет ИЦТЭ	28.11.23	3	 Директор, к.т.н., Беляев Э.И.

1. Цель, задачи и планируемые результаты обучения по дисциплине

(Цель и задачи освоения дисциплины, соответствующие цели ОП)

Целью освоения дисциплины «Методы статистики для разработчиков» является формирование универсальных компетенций, обоснование значимости и функций статистики в анализе больших данных, компьютерном зрении, нейросетях и искусственном интеллекте; прогнозировании и разработке сценариев развития; в овладении вопросов теории и практики статистики и применение статистических методов анализа в целом, и в частности, в области разработки компьютерных приложений.

Задачами дисциплины являются:

познакомить обучающихся со статистическими методами, используемыми в анализе данных;

изучить современные информационные технологии и программное обеспечение для получения статистических оценок;

сформировать практические навыки и умения интерпретации статистических оценок, возникающих в ходе анализа данных и построения моделей в различных прикладных областях.

обучить применять статистические методы, необходимых для профессиональной деятельности.

Компетенции и индикаторы, формируемые у обучающихся:

Код и наименование компетенции	Код и наименование индикатора
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Способен осуществлять поиск, критический анализ и синтез информации из различных источников
	УК-1.2. Использует системный подход для решения поставленных задач, владеет логическими методами обработки информации, отличает факты от мнений, гипотез и интерпретаций

2. Место дисциплины в структуре ОП

Предшествующие дисциплины: «Теория вероятностей и математическая статистика», «Математические модели и методы», «Программирование».

Последующие дисциплины: «Производственная практика (преддипломная)».

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			7
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	3	108	108
КОНТАКТНАЯ РАБОТА*		54	54
АУДИТОРНАЯ РАБОТА	1,3	48	48
Лекции	0,4	16	16
Практические (семинарские) занятия	0,9	32	32
Лабораторные работы	0	0	0
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	1,7	60	60
Проработка учебного материала	1,7	60	60
Курсовой проект	0	0	0
Курсовая работа	0	0	0
Подготовка к промежуточной аттестации	0	0	0
Промежуточная аттестация:			3
			-

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Разделы дисциплины	Всего часов	Распределение трудоемкости по видам учебной работы				Формы и вид контроля	Индексы индикаторов формируемых компетенций
		лекции	лаб. раб.	пр. зан.	сам. раб.		
Раздел 1. Введение в статистический анализ и первичная статистическая обработка	14	2		4	8	ТК1	УК-1.1, УК-1.2
Раздел 2. Критерии сравнения групп	14	2		4	8	ТК1	УК-1.1, УК-1.2
Раздел 3. Корреляционный анализ	14	2		4	8	ТК2	УК-1.1, УК-1.2
Раздел 4. Регрессионный анализ	26	4		8	14	ТК2	УК-1.1, УК-1.2
Раздел 5. Дисперсионный анализ	14	2		4	8	ТК3	УК-1.1, УК-1.2
Раздел 6. Задачи классификации и кластеризации	26	4		8	14	ТК3	УК-1.1, УК-1.2
Зачет	0				0	ОМ	УК-1.1, УК-1.2
ИТОГО	108	16		32	60		

3.3. Содержание дисциплины

Раздел 1. Введение в статистический анализ и первичная статистическая обработка

Тема 1.1. Задачи и этапы статистического анализа. Типы и структуры данных.

Статистическая оценка. Точечное и интервальное оценивание. Свойства оценок. Точность. Надежность. Доверительный интервал. Понятие статистической гипотезы. Критическая область. Критические точки. Ошибки первого и второго рода. Уровень значимости. Мощность критерия.

Тема 1.2. Предварительная обработка данных

Использование различных источников данных. Технология получения данных из баз данных с помощью SQL запросов. Предобработка данных. Очистка данных.

Раздел 2. Критерии сравнения групп

Тема 2. Параметрические критерии: сравнение средних значений в двух группах (критерий Стьюдента). Непараметрические критерии: критерий Уилкоксона (Манна-Уитни) для двух групп, критерий Краскела-Уоллиса.

Раздел 3. Корреляционный анализ

Тема 3. Линейный коэффициент корреляции. Оценка значимости связи. Шкала Чеддока. Определение доверительного интервала для коэффициента корреляции.

Раздел 4. Регрессионный анализ

Тема 4.1. Линейная регрессия

Спецификация модели. Построение линейной регрессии методом наименьших квадратов. Оценка качества модели. Свойства, экономическая интерпретация и оценка параметров линейного уравнения регрессии. Проверка гипотез о значимости регрессионной модели и проверка значимости ее параметров. Критерии Стьюдента и Фишера.

Интервалы прогноза по линейному уравнению регрессии. Построение доверительных интервалов для прогнозируемых значений. Стандартные ошибки коэффициентов регрессии. Средняя ошибка аппроксимации.

Метод градиентного спуска.

Тема 4.2. Нелинейная регрессия.

Нелинейная регрессия. Основные модели. Выбор зависимости. Линеаризация нелинейных зависимостей. Схема применения метода наименьших квадратов в нелинейных моделях.

Корреляция для нелинейной регрессии. Средняя ошибка аппроксимации.

Раздел 5. Дисперсионный анализ

Тема 5. Сравнение средних нескольких генеральных совокупностей. Дисперсионный анализ. Испытуемый фактор. Уровни фактора. Внутригрупповая дисперсия, межгрупповая дисперсия. F-критерий. Однофакторный дисперсионный анализ. Двухфакторный дисперсионный анализ.

Раздел 6. Задачи классификации и кластеризации

Тема 6.1. Задачи классификации

Классификация методами бинарной и множественной регрессии.
Метрики качества классификации.

Тема 6.2. Задачи кластеризации

Кластерный анализ. Задачи кластерного анализа. Меры расстояний.
Алгоритмы кластеризации. Формы кластеров. Иерархическая кластеризация.
Кластеризация k-средним. Оценка качества кластеризации.

3.4. Тематический план практических занятий

Тема 1.1. Точечное и интервальное оценивание характеристик случайных величин. Проверка статистических гипотез.

Тема 1.2. Подготовка данных

Тема 3. Корреляционный анализ

Тема 4.1. Линейная регрессия. Отбор факторов, оценка значимости

Тема 4.2. Нелинейная регрессия.

Тема 5. Однофакторный дисперсионный анализ. Двухфакторный дисперсионный анализ.

Тема 6.1. Задачи классификации. Модели бинарного выбора. Модель множественного выбора

Тема 6.2. Задачи кластеризации. Иерархическая кластеризация.
Кластеризация k-средним.

3.5. Тематический план лабораторных работ

Данный вид работы не предусмотрен учебным планом

3.6. Курсовой проект /курсовая работа

Данный вид работы не предусмотрен учебным планом

4. Оценивание результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Способен осуществлять поиск, критический анализ и синтез информации из различных источников	знать:				
		статистические методы и подходы, используемые в анализе данных	Отличный уровень знаний в объеме, соответствующем программе подготовки, и, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок
		уметь:				
	оценивать качество и достоверность полученных результатов	Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки	

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)				
			Высокий	Средний	Ниже среднего	Низкий	
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54	
			Шкала оценивания				
			отлично	хорошо	удовлетворительно	неудовлетворительно	
		зачтено			не зачтено		
		владеть:					
		навыками интерпретации результатов статистической обработки данных и моделирования	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки	
		УК-1.2. Использует системный подход для решения поставленных задач, владеет логическими методами обработки информации, отличает факты от мнений,	знать:				
			статистические подходы для решения поставленных задач	Отличный уровень знаний в объеме, соответствующем программе подготовки, и, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок
уметь:							

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
		применять статистические методы, необходимых для профессиональной деятельности	Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки
		владеть:				
		современными информационными технологиями и программным обеспечением для получения статистических оценок	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки

Оценочные материалы для проведения текущего контроля и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины.

Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика.

5. Учебно-методическое и информационное обеспечение дисциплины

5.1. Учебно-методическое обеспечение

5.1.1. Основная литература

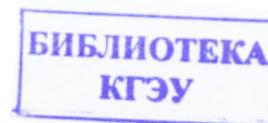
1. Буре, В. М. Методы прикладной статистики в R и Excel : учебное пособие / В. М. Буре, Е. М. Парилина, А. А. Седаков. - 4-е изд., стер. - Санкт-Петербург : Лань, 2023. - 152 с. - URL: <https://e.lanbook.com/book/319424>.
2. Чураков, Е. П. Введение в многомерные статистические методы / Е. П. Чураков. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 148 с. — ISBN 978-5-507-47141-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/330530>.
3. Ганичева, А. В. Прикладная статистика : учебное пособие / А. В. Ганичева. — 4-е изд., стер. — Санкт-Петербург : Лань, 2023. — 172 с. — ISBN 978-5-507-47980-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/336800>.

5.1.2. Дополнительная литература

1. Как устроен Python. Гид для разработчиков, программистов и интересующихся : справочное издание / М. Харрисон. - СПб. : Питер, 2019. - 272 с. - URL: <https://ibooks.ru/bookshelf/359217>.
2. Никитина, Т. П. Программирование. Основы Python для инженеров : учебное пособие / Т. П. Никитина, Л. В. Королев. - Санкт-Петербург : Лань, 2023.
3. Хливненко, Л. В. Практика нейросетевого моделирования : учебное пособие для вузов / Л. В. Хливненко, Ф. А. Пятакович. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 200 с. — ISBN 978-5-8114-8264-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/173811>.

5.2. Информационное обеспечение

5.2.1. Электронные и интернет-ресурсы



№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	Электронно-библиотечная система «Лань»	https://e.lanbook.com/
2	Электронно-библиотечная система «ibooks.ru»	https://ibooks.ru/
3	Электронно-библиотечная система «book.ru»	https://www.book.ru/
4	Портал «Открытое образование»	https://npoed.ru

8	Национальная электронная библиотека (НЭБ)	https://rusneb.ru/
9	Онлайн-курс “Прикладное программирование на языке Python”.	https://openedu.ru/course/urfu/PYAP
10	Real Python Tutorials	https://realpython.com

5.2.2. Профессиональные базы данных / Информационно-справочные системы

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
1	Официальный интернет-портал правовой информации	http://pravo.gov.ru	http://pravo.gov.ru
2	Справочная правовая система «Консультант Плюс»	http://consultant.ru	http://consultant.ru
3	Справочно-правовая система по законодательству РФ	http://garant.ru	http://garant.ru
4	Applied Science & Technology Source. EBSCO publishing	http://search.ebscohost.com	http://search.ebscohost.com
5	Wiley Online Library	http://onlinelibrary.wiley.com/	http://onlinelibrary.wiley.com/
6	Научная электронная библиотека	http://elibrary.ru	http://elibrary.ru
7	Российская государственная библиотека	http://www.rsl.ru	http://www.rsl.ru
8	Международная реферативная база данных научных изданий zbMATH	http://www.zbmath.org	http://www.zbmath.org
9	Международная реферативная база данных научных изданий Springerlink	http://link.springer.com	http://link.springer.com
10	Образовательный портал	http://www.ucheba.com	http://www.ucheba.com

5.2.3. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Способ распространения (лицензионное/свободно)	Реквизиты подтверждающих документов
1	Операционная система Microsoft Windows 10	Пользовательская операционная система	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
3	Microsoft Office 2019	Пакет офисных приложений	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл.

			право, срок действия лицензии - бессрочно
4	LMS Moodle	ПО для эффективного онлайн-взаимодействия преподавателя и студента	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
5	Visual Studio Express	Инструмент создания приложений	https://visualstudio.microsoft.com/ru/vs/express/

6. Материально-техническое обеспечение дисциплины

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
Лекции	Учебная аудитория для проведения занятий лекционного типа	Специализированная учебная мебель, технические средства обучения, служащие для представления учебной информации большой аудитории (мультимедийный проектор, компьютер (ноутбук), экран), демонстрационное оборудование, учебно-наглядные пособия
Лабораторные работы	Учебная лаборатория программной инженерии, ауд. В-608	Специализированное лабораторное оборудование по профилю лаборатории программной инженерии, специализированная учебная мебель на 50 посадочных мест, 24 компьютера с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-610	Специализированная учебная мебель на 42 посадочных места, 17 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория информационной безопасности, ауд. В-615	Специализированное лабораторное оборудование по профилю лаборатории информационной безопасности, специализированная учебная мебель на 35 посадочных мест, 15 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
		(мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-617	Специализированная учебная мебель на 24 посадочных места, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-619	Специализированная учебная мебель на 26 посадочных мест, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-621	Специализированная учебная мебель на 35 посадочных мест, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория реинжиниринга и управления бизнес-процессами, ауд. В-623	Специализированное лабораторное оборудование по профилю лаборатории реинжиниринга и управления бизнес-процессами, специализированная учебная мебель на 34 посадочных места, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, ноутбук, экран), видеокамеры, необходимое лицензионное программное обеспечение
Самостоятельная работа	Компьютерный класс с возможностью выхода в Интернет и обеспечением	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультимедийный проектор, компьютер (ноутбук),

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
	доступа в ЭИОС В-600	экран), видеокамеры, программное обеспечение
	Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, экран, мультимедийный проектор, программное обеспечение

7. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета www/kgeu.ru. Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);
- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);
- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;
- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;
- обеспечения возможности для обучающегося получить адресную консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению

подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;

- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;

- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;

- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;

- обеспечивается необходимый уровень освещенности помещений;

- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ, инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

8. Методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.

Методическое обеспечение процесса воспитания обучающихся выступает одним из определяющих факторов высокого качества образования. Преподаватель вуза, демонстрируя высокий профессионализм, эрудицию, четкую гражданскую позицию, самодисциплину, творческий подход в решении профессиональных задач, в ходе образовательного процесса способствует формированию гармоничной личности.

При реализации дисциплины преподаватель может использовать следующие методы воспитательной работы:

- методы формирования сознания личности (беседа, диспут, внушение, инструктаж, контроль, объяснение, пример, самоконтроль, рассказ, совет, убеждение и др.);

- методы организации деятельности и формирования опыта поведения (задание, общественное мнение, педагогическое требование, поручение, приучение, создание воспитывающих ситуаций, тренинг, упражнение, и др.);

- методы мотивации деятельности и поведения (одобрение, поощрение социальной активности, порицание, создание ситуаций успеха, создание ситуаций для эмоционально-нравственных переживаний, соревнование и др.)

При реализации дисциплины преподаватель должен учитывать следующие направления воспитательной деятельности:

Гражданское и патриотическое воспитание:

- формирование у обучающихся целостного мировоззрения, российской

идентичности, уважения к своей семье, обществу, государству, принятым в семье и обществе духовно-нравственным и социокультурным ценностям, к национальному, культурному и историческому наследию, формирование стремления к его сохранению и развитию;

- формирование у обучающихся активной гражданской позиции, основанной на традиционных культурных, духовных и нравственных ценностях российского общества, для повышения способности ответственно реализовывать свои конституционные права и обязанности;

- развитие правовой и политической культуры обучающихся, расширение конструктивного участия в принятии решений, затрагивающих их права и интересы, в том числе в различных формах самоорганизации, самоуправления, общественно-значимой деятельности;

- формирование мотивов, нравственных и смысловых установок личности, позволяющих противостоять экстремизму, ксенофобии, дискриминации по социальным, религиозным, расовым, национальным признакам, межэтнической и межконфессиональной нетерпимости, другим негативным социальным явлениям.

Духовно-нравственное воспитание:

- воспитание чувства достоинства, чести и честности, совестливости, уважения к родителям, учителям, людям старшего поколения;

- формирование принципов коллективизма и солидарности, духа милосердия и сострадания, привычки заботиться о людях, находящихся в трудной жизненной ситуации;

- формирование солидарности и чувства социальной ответственности по отношению к людям с ограниченными возможностями здоровья, преодоление психологических барьеров по отношению к людям с ограниченными возможностями;

- формирование эмоционально насыщенного и духовно возвышенного отношения к миру, способности и умения передавать другим свой эстетический опыт.

Культурно-просветительское воспитание:

- формирование эстетической картины мира;

- формирование уважения к культурным ценностям родного города, края, страны;

- повышение познавательной активности обучающихся.

Научно-образовательное воспитание:

- формирование у обучающихся научного мировоззрения;

- формирование умения получать знания;

- формирование навыков анализа и синтеза информации, в том числе в профессиональной области.

Вносимые изменения и утверждения на новый учебный год

№ п/п	№ раздела внесения изменений	Дата внесения изменений	Содержание изменений	«Согласовано» Зав. каф. реализующей дисциплину	«Согласовано» председатель УМК института (факультета), в состав которого входит выпускающая кафедра)
1	2	3	4	5	6
1					
2					
3					



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГУУ»)

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
по дисциплине

Методы статистики для разработчиков

(Наименование дисциплины в соответствии с РУП)

Направление подготовки

09.03.03 Прикладная информатика

Квалификация

бакалавр

г. Казань, 2023

Оценочные материалы по дисциплине *Методы статистики для разработчиков* предназначены для оценивания результатов обучения на соответствие индикаторам достижения компетенций.

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля (ТК) и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

1. Технологическая карта

Семестр 7

Наименование раздела	Формы и вид контроля	Рейтинговые показатели							
		I текущий контроль	Дополнительные баллы к ТК1	II текущий контроль	Дополнительные баллы к ТК2	III текущий контроль	Дополнительные баллы к ТК3	Итого	Промежуточная аттестация
Раздел 1. Введение в статистический анализ и первичная статистическая обработка Раздел 2. Критерии сравнения групп	ТК1	15	0-15					15-30	15-30
Защита лабораторной работы		8							
Отчет по лабораторной работе		7							
Раздел 3. Корреляционный анализ Раздел 4. Регрессионный анализ	ТК2			15	0-15			15-30	15-30
Защита лабораторной работы				8					
Отчет по лабораторной работе				7					
Раздел 5. Дисперсионный анализ Раздел 6. Задачи классификации и кластеризации	ТК3					25	0-15	25-40	25-40
Защита лабораторной работы						15			
Отчет по лабораторной работе						10			
Промежуточная аттестация (зачет)	ОМ								0-45
Тестирование									0-45

2. Оценочные материалы текущего контроля и промежуточной аттестации

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Способен осуществлять поиск, критический анализ и синтез информации из различных источников	знать:				
		статистические методы и подходы, используемые в анализе данных	Отличный уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок
		уметь:				
		оценивать качество и достоверность полученных результатов	Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки
		владеть:				

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
		навыками интерпретации результатов статистической обработки данных и моделирования	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки
	УК-1.2. Использует системный подход для решения поставленных задач, владеет логическими методами обработки информации, отличает факты от мнений, гипотез и интерпретаций	знать:				
статистические подходы для решения поставленных задач		Отличный уровень знаний в объеме, соответствующем программе подготовки, и, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок	
уметь:						
		применять статистические методы, необходимых для профессиональной деятельности	Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами, выполнены все	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено			не зачтено
			задания в полном объеме			
владеть:						
современными информационными технологиями и программным обеспечением для получения статистических оценок	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки		

Оценка **«отлично»** выставляется за выполнение *расчетных работ в семестре; тестовых заданий; глубокое понимание методов расчета, полные и содержательные ответы на вопросы билета.*

Оценка **«хорошо»** выставляется за выполнение *расчетных работ в семестре; тестовых заданий; понимание методов расчета, ответы на вопросы билета.*

Оценка **«удовлетворительно»** выставляется за выполнение *расчетных работ в семестре и тестовых заданий; понимание базовых методов расчета, ответы на вопросы билета с частичными ошибками.*

Оценка **«неудовлетворительно»** выставляется за *частичное выполнение расчетных работ в семестре и тестовых заданий; ответы на вопросы билета с ошибками.*

3. Перечень оценочных средств

Краткая характеристика оценочных средств, используемых при текущем контроле успеваемости и промежуточной аттестации обучающегося по дисциплине:

Наименование оценочного средства	Краткая характеристика оценочного средства	Описание оценочного средства
Отчет по лабораторной работе (ОЛР)	Выполнение лабораторной работы, обработка результатов испытаний, измерений, эксперимента. Оформление отчета, защита результатов лабораторной работы по отчету	Перечень заданий и вопросов для защиты лабораторной работы, перечень требований к отчету
Тест (Тест)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося	Комплект тестовых заданий

4. Перечень контрольных заданий или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций в процессе освоения дисциплины

Для текущего контроля ТК1:

Проверяемая компетенция: УК-1.1, УК-1.2

Комплект лабораторных работ

Для оценки сформированности компетенций используются следующие задания для выполнения лабораторных работ. Данные необходимо взять из файлов, размещенных на сетевом ресурсе.

Работа 1. Точечное и интервальное оценивание характеристик случайных величин

Цель работы: освоить методы расчета основных статистических характеристик.

Задание. Дана месячная заработная плата рабочих (тыс.руб.)

Данные взять из файла *rab1.xlsx* в соответствии с вариантом

MS Excel

1.1. По исходным данным:

Вычислить следующие характеристики:

- *среднюю заработную плату;*
- *оценку дисперсии;*
- *оценку среднеквадратического отклонения.*

Проверить. Для проверки использовать функции Excel (СРЗНАЧ, ДИСП.В, СТАНДОТКЛОН.В)

Сделать вывод о средней заработной плате рабочих предприятия.

1.2. На новом листе из исходных данных получить сгруппированные данные.

По сгруппированным данным найти Среднюю, Дисперсию и Среднеквадратическое отклонение. Сравнить с результатами предыдущих расчетов.

1.3. Получить данные с интервальной группировкой. Для этого:

- Найти интервал разбиения (интервал группировки) $h = \frac{x_{\max} - x_{\min}}{n^*}$, где n^* определяется по формуле Стерджесса $n^* = [1 + 3.322 \lg n]$, n – объем выборки, а квадратные скобки обозначают округление до целого;
- определить границы интервалов $x_i' = x_{i-1}' + h$
- найти для каждого интервала сумму частот, попавшие в i –ый интервал.

По данным с интервальной группировкой найти Среднюю, Дисперсию и Среднеквадратическое отклонение. Сравнить с результатами предыдущих расчетов.

1.4. Определить значение t_n для нормального распределения с заданной надежностью, пользуясь функцией НОРМ.СТ.ОБР (Мастер функций, категория Статистические).

Рассчитать точность $\delta = t \cdot \frac{s}{\sqrt{n}}$.

1.5. Проверить найденное значение точности с помощью функции ДОВЕРИТ.НОРМ, которая также использует для определения t нормальное распределение. (Мастер функций, категория Статистические).

1.6. Найти границы доверительного интервала $\bar{x} - t \cdot \frac{s}{\sqrt{n}} < a < \bar{x} + t \cdot \frac{s}{\sqrt{n}}$.

Сделать выводы.

1.7. Определить значение t_s для распределения Стьюдента, пользуясь функцией СТЬЮДЕНТ.ОБР.2Х (Мастер функций, категория Статистические). Сравнить с величиной t_n . Повторить пункты 4-6. В п.5 для проверки использовать функцию ДОВЕРИТ.СТЬЮДЕНТ. Сделать выводы.

ППП Statistica

2.1. Запустить Statistica. Создать новую таблицу. Задать имя (X) и свойство поля для данных о зарплате.

2.2. Скопировать данные из файла Excel в таблицу Statistica.

2.3. Вывести данные описательной статистики и сравнить с п.1.1.

2.4. Получить сгруппированные данные. Сравнить с п.1.2.

2.5. Получить данные с интервальной группировкой. Сравнить с п.1.3.

Python

3.1. Запустить Python. Подключить библиотеку Pandas.

3.2. Получить данные из файла Excel с листа со своим вариантом с присвоением имени столбцу с данными (справка: https://pandas.pydata.org/docs/reference/api/pandas.read_excel.html#)

3.3. Используя функции, рассчитать среднее, дисперсию и стандартное отклонение и сравнить с п.1.1.

(справка:

<https://pandas.pydata.org/docs/reference/api/pandas.Series.html?highlight=series#pandas.Series>)

3.4. Используя групповые операции, получить сгруппированные данные. Сравнить с п.1.2.

(справка:

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html?highlight=groupby#pandas.DataFrame.groupby>)

3.5 По исходным данным получить интервальную группировку (параметр *bins* метода *value_counts*).

3.6 Рассчитать доверительный интервал

Работа 2. Проверка статистических гипотез

Цель работы: освоить методы проверки статистических гипотез.

Для каждой из задач, представленных ниже:

- сформулировать гипотеза H_0 и H_1
- записать критерий
- обосновать тип критической области
- найти расчетное и критическое значение критерия
- сформулировать вывод

Решение получить в Excel, ППП Statistica и Python

Задача 1.

Двумя методами проведены измерения температуры жидкости. Получены следующие данные:

x	9,6	10	9,8	10,2	10,6
y	10,4	9,7	10	10,3	

Определить, дают ли оба этих метода одинаковую точность (сравнить дисперсии). Уровень значимости α принять равным 0,1. В качестве альтернативной гипотезы принять $D(X) \neq D(Y)$

Задача 2. В результат длительного хронометража времени сборки узла различными сборщиками установлено, что дисперсия этого времени $\sigma_0^2 = 2$ мин². Результаты работы новичка таковы:

x_i	56	58	60	62	64
n_i	1	4	10	3	2

Можно ли при уровне значимости 0,05 считать, что новичок работает ритмично (разброс его времени существенно не отличается от остальных сборщиков)?

Работа 3. Подготовка данных

Цель работы: освоить методы подготовки данных для дальнейшего анализа данных.

Расчеты провести в Excel и Python

1. Открыть файл `log.csv`. Ознакомиться с данными.
2. Посмотрите все уникальные значения в `sample.csv`
3. Посмотрите, сколько непустых значений в колонке `City`
4. Файл `log.csv`, в колонке `user_id` есть записи, которые содержат технические ошибки. Укажите, что записано в поле `user_id` в строчках с ошибкой
5. Создайте новый датафрейм `sample2`, в который будут входить только записи о людях в возрасте меньше 30 лет
6. Создайте новый датафрейм `log_win`, в который будут входить только записи, где пользователь выиграл. Посчитайте, сколько таких записей, и сохраните в переменной `win_count`
7. Создайте новый датафрейм `sample2`, в который будут входить только записи о рабочих младше 30 лет
8. Найдите записи, где в городах нет буквы "о", и сохраните в переменную `sample4`. Не забудьте про `NaN` и параметр `na`
9. Сохраните в переменную `new_log` датафрейм, из которого удалены записи с ошибкой в поле `user_id`
10. С помощью `apply` и лямбда-функции увеличьте возраст во всех записях на 1 год и сохраните в `sample2`. В переменной `sample2` должен содержаться весь датафрейм `sample`
11. С помощью `apply` и `lambda`-функции замените все буквы в поле `City` на маленькие и сохраните в `sample2`. Вам может понадобиться функция `s.lower()`
12. Напишите функцию `profession_code`, которая на вход получает строку, а на выход возвращает:
 - 0 — если на вход поступила строка "Рабочий"
 - 1 — если на вход поступила строка "Менеджер"
 - 2 — в любом другом случае
13. Напишите функцию `age_category`, которая на вход получает число, а на выход отдаёт:
 - "молодой" — если возраст меньше 23
 - "средний" — если возраст от 23 до 35
 - "зрелый" — если возраст больше 35

Для текущего контроля ТК2:

Проверяемая компетенция: УК-1.1, УК-1.2

Комплект лабораторных работ

Работа 4. Линейная регрессия

Цель работы: освоить методы расчета коэффициентов линейной регрессии .

MS Excel

1. Рассчитать параметры уравнения регрессии и записать полученное уравнение.
2. Рассчитать по построенной модели y_p .

3. Построить график зависимости y и y_p (должна получиться прямая!!).
4. Запустить Регрессионный анализ и сравнить полученные значения.
5. Оценить значимость уравнения, сделать вывод.
6. Оценить значимость параметров уравнения, сделать вывод.
7. Оценить качество уравнения по R^2 . Сделать вывод.
8. Рассчитать доверительные интервалы с надежностью 90% и 95% для параметров уравнения регрессии. Попадает ли в интервалы ноль? Сделать выводы.
9. Рассчитать доверительные интервалы для \hat{y} с надежностью 95%. Добавить на график п.3 границы доверительных интервалов. Убедиться, что не менее 95% значений y_i находятся внутри оцененных границ.

Statistica

Получиться все оценки через ППП Statistica

Python

Повторить все пункты, используя библиотеку scikit-learn

В отчет

1. Записать уравнения регрессии.
2. Записать, что показывает коэффициент b_1 .
3. Перерисовать (распечатать) график.
4. Оценить значимость уравнения (2 способа!), записать вывод.
5. Оценить значимость параметров уравнения (2 способа!), записать вывод.
6. Оценить качество уравнения по R^2 . Записать вывод.
7. Записать доверительные интервалы для надежности 90% и 95% для параметров уравнения регрессии. Записать выводы.
8. Зарисовать график с доверительными интервалами. Записать сколько процентов значений y_i оказалось внутри оцененных границ.

Контрольные вопросы

1. Парная регрессия
2. Множественная регрессия
3. Парная линейная регрессия
4. МНК
5. Формулы и смысл $SS_{\text{общ}}$, $SS_{\text{ост}}$, $SS_{\text{факт}}$, $S^2_{\text{общ}}$, $S^2_{\text{ост}}$, $S^2_{\text{факт}}$
6. Оценка значимости уравнения (включая формулировку гипотез)
7. Формулы m_{b_0} , m_{b_1} , m_y , t_b
8. Оценка значимости параметров уравнения (включая формулировку гипотез)
9. Коэффициент детерминации.
10. Доверительные интервалы для b_j и \hat{y} .

Работа 5. Отбор факторов, оценка значимости

Цель работы: освоить методы отбора факторов при построении модели линейной регрессии .

Задание 1.

1. По исходным данным из файла текущей работы сформировать матрицу X и вектор Y .
2. Найти оценку вектора B коэффициентов уравнения множественной линейной регрессии: $B = (X'X)^{-1} X'Y$.
3. Найти вектор E и $SS_{\text{ост}}$: $E = Y - XB$, $SS_{\text{ост}} = E^T \cdot E$
4. Найти стандартные ошибки коэффициентов регрессии: $m_b = \sqrt{S_b^2}$, где $S_b^2 = S_{\text{ост}}^2 (X'X)^{-1}$.
5. Рассчитать t-статистики для коэффициентов регрессии. Рассчитать р-значение для t-статистик.
6. Выполнить Анализ данных – Регрессия, сравнить результаты.

В отчет

9. Записать уравнение регрессии и t-статистики.

Задание 2.

1. Перейти к стандартизованным переменным.
2. Найти оценку вектора β коэффициентов уравнения множественной линейной регрессии в стандартизованных переменных, используя матричный метод.
3. Используя полученные β коэффициенты, найти коэффициенты b_j исходного уравнения. Сравнить с результатами из Задания 1.

В отчет

1. Записать уравнение регрессии в стандартизованных переменных и t-статистики.

Задание 3.

1. Построить частные уравнения регрессии
2. Рассчитать частные коэффициенты эластичности и их средние. Сделать выводы.
3. Рассчитать частные коэффициенты корреляции. Сделать выводы.
4. Рассчитать частные F-критерии.
5. Рассчитать t-критерии, используя частные F-критерии. Сравнить с результатами из Задания 1.

В отчет

1. Записать частные уравнение регрессии.
2. Записать средние частные коэффициенты эластичности и выводы.
3. Записать частные коэффициенты корреляции и выводы.
4. Записать частные F-критерии.
5. Записать t-критерии.

Задание 4.

Провести корреляционный анализ. Сделать выводы.

В отчет

Записать таблицу корреляционного анализа. записать выводы.

Задание 5.

Метод пошагового исключения

1. Определить параметры уравнения множественной линейной регрессии с включением всех факторов.
2. Оценить значимость параметров уравнения регрессии
3. Исключить наиболее незначимый фактор
4. Перестроить уравнение регрессии по оставшимся факторам.
5. Повторять пп 2-6 до тех пор, пока в уравнении не останутся только значимые факторы. Построить уравнение регрессии по значимым факторам (с учетом результатов корреляционного анализа).
6. Провести оценку значимости уравнения и параметров уравнения регрессии.

Метод пошагового включения

7. Построить уравнения линейной регрессии с одним фактором.
8. Оценить значимость уравнения регрессии
9. Добавить следующий фактор. Перестроить регрессию с дополнительным фактором.
10. Провести оценку значимости. Если включенные факторы значимы, то добавить следующий фактор. Если есть незначимый фактор, то его следует исключить. Повторять пп. 14-15 пока в уравнении не останутся только значимые факторы.
11. Построить уравнение регрессии по значимым факторам (с учетом результатов корреляционного анализа).
12. Провести оценку значимости уравнения и параметров уравнения регрессии.

Регрессия в стандартизованных переменных

13. Перейти к стандартизованным переменным: $t = \frac{(y - \bar{y})}{\sigma_y}$ $t_{xy} = \frac{(x - \bar{x}_j)}{\sigma_{x_j}}$
14. Построить уравнения линейной регрессии в стандартизованных переменных.
15. Оценить значимость уравнения регрессии и параметров
16. Сделать выводы.
17. Получить расчетные значения $u_{\text{расч}}$.
18. Рассчитать $SS_{\text{общ}}$, $SS_{\text{ост}}$, $SS_{\text{факт}}$, $S^2_{\text{общ}}$, $S^2_{\text{ост}}$, $S^2_{\text{факт}}$
19. Рассчитать коэффициент множественной детерминации, коэффициент множественной корреляции, нормированный коэффициент множественной детерминации, F и $F_{\text{кр}}$. Сделать выводы.
20. Построить графики y_i (точками) и $u_{\text{расч}}$ (линией).

В отчет

1. Записать полученную последовательность шагов по методу пошагового исключения. На каждом шаге указать R^2 , нормированный R^2 , t критерий и его значимость для каждого коэффициента регрессии.
2. Записать полученную последовательность шагов по методу пошагового включения. На каждом шаге указать R^2 , нормированный R^2 , t критерий и его значимость для каждого коэффициента регрессии.

3. Записать итоговое уравнение регрессии, t критерии для каждого коэффициента, P значение, выводы о значимости.
4. Записать уравнение регрессии в стандартизованных переменных и сделать выводы.
5. Записать полученные $SS_{\text{общ}}$, $SS_{\text{ост}}$, $SS_{\text{факт}}$, $S^2_{\text{общ}}$, $S^2_{\text{ост}}$, $S^2_{\text{факт}}$.
6. Записать полученные: коэффициент множественной детерминации, коэффициент множественной корреляции, нормированный коэффициент множественной детерминации, F и $F_{\text{кр}}$. Сделать выводы.
7. Зарисовать график y_i (точками) и $y_{\text{расч}}$ (линией)..

1. По результатам текущей работы рассчитать ошибки в MS Excel
2. По результатам текущей работы рассчитать ошибки в Python
3. Сравнить

В отчет

8. Записать полученные значения ошибок.

Контрольные вопросы

1. Абсолютные и относительные ошибки.
2. Формулы расчета:
 - Средняя абсолютная ошибка (Mean Absolute Error)
 - Средняя квадратичная ошибка (Mean Squared Error)
 - Среднеквадратическая ошибка (Root Mean Squared Error)
 - Средняя абсолютная процентная ошибка, средняя ошибка аппроксимации (Mean Absolute Percentage Error)
 - Коэффициент детерминации

Контрольные вопросы

3. Множественная линейная регрессия.
4. Отбор факторов.
5. Корреляционный анализ
6. Коллинеарность и мультиколлинеарность факторов
7. Оценка значимости в множественной линейной регрессии.
8. Метод исключения
9. Метод включения
10. Коэффициент множественной детерминации, коэффициент множественной корреляции, нормированный коэффициент множественной детерминации.
11. Уравнение регрессии в стандартизованных переменных.
12. Формулы расчета вектора B , $SS_{\text{ост}}$, стандартных ошибок, t -статистик.
13. Формулы стандартизованных переменных, их связь с b коэффициентами.
14. Формулы частных коэффициентов корреляции и частных F -критериев.

Работа 6. Нелинейная регрессия

Цель работы: освоить методы линеаризации и оценки коэффициентов нелинейной регрессии .

Задание

10. Построить график (точечный) по исходным данным.
 11. Выбрать подходящие для моделирования зависимости.
- Для каждой из выбранных зависимостей*
12. Сделать линеаризацию, предварительно ввести новые переменные.
 13. По новым переменным найти параметры преобразованной модели.
 14. Найти параметры исходной нелинейной зависимости.
 15. Рассчитать $Y_{\text{расч}}$ (по нелинейной! модели).
 16. Найти индекс детерминации, индекс корреляции, среднюю ошибку аппроксимации.
 17. Построить графики исходных и расчетных данных (исходные – точками, расчетные – линия без маркеров).
 18. Сделать вывод.

Метод средних точек

19. Для каждой из стандартных зависимостей рассчитать координаты средних точек.
20. Добавить средние точки на график п.1. Сделать вывод, какая из функций лучше будет описывать данные. Совпадает ли этот вывод с выводом п.9?

В отчет

10. Зарисовать графики исходных и расчетных данных.
11. Рядом с каждым графиком записать:
 - уравнения регрессии,
 - процесс линеаризации (сведения нелинейной зависимости к линейной),
 - индекс детерминации,
 - индекс корреляции,
 - среднюю ошибку аппроксимации.
12. Зарисовать график со средними точками. Сделать вывод.
13. Записать общий вывод по работе с учетом п.2 и п.3.

Контрольные вопросы

15. Нелинейная регрессия
16. Классификация моделей нелинейной регрессии
17. В чем заключается линеаризация?
18. Индекс детерминации.
19. Индекс корреляции.
20. Средняя ошибка аппроксимации.
21. Метод средних точек.

Работа 7. Метод градиентного спуска

Цель работы: освоить метод градиентного спуска .

Задание 1. Разбиение выборки

1. Открыть файл с датасетом Iris:

```
from sklearn.datasets import load_iris
```

```
np=load_iris()
```

```
X=pd.DataFrame(np.data, columns=['sepal length (cm)',  
    'sepal width (cm)',  
    'petal length (cm)',  
    'petal width (cm)'])
```

```
y=pd.DataFrame(np.target,columns=['sort'])
```

где sort – это

0 - 'setosa', 1 - 'versicolor', 2 - 'virginica'

2. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую без параметров. Вывести `y_train`.

3. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую с долей `train 0,6`. Вывести `y_train`.

4. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую с числом `test 20`. Вывести `y_test`.

5. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую с долей `train 0,1`. Обеспечить равное вхождение всех сортов ирисов. Вывести `y_train`.

Задание 2. Построение модели

1. Открыть файл с датасетом boston: `from sklearn.datasets import load_boston`

2. Изучить признаки: какие имеет смысл включать в модель?

3. Провести корреляционный анализ. Выявить коллинеарные факторы.

4. Используя библиотеку `statsmodels` и метод пошагового исключения, отобрать факторы в модель (не забудьте про свободный член!!). Учесть результаты корреляционного анализа.

5. Разделить выборку на тренировочную и тестовую случайным(!) образом.

6. Обучить модель на тренировочной выборке.

7. Рассчитать значение ошибок MSE и RMSE и R^2 для тренировочных данных:

8. Рассчитать значение ошибок MSE и RMSE и R^2 для тестовых данных:

Повторить шаги 5-7 несколько раз (отдельный код писать не надо – просто несколько раз запустить). Что можно сказать по поводу построенной модели?

9. Какие факторы вошли в модель (словами, а не именами переменных)?

10. Записать модель

11. Дать экономическую интерпретацию коэффициентам регрессии.

пп. 9-11 записать словами как комментарии (Markdown) в файл блокнота.

Задание 3. Градиентный спуск

12. Продолжить работу с датасетом boston
13. Подключить библиотеку SGDRegressor
14. Построить модель по исходным данным (без нормализации). Вывести коэффициенты. Сравнить результаты с Заданием 2.
15. Нормализовать факторы X.
16. Построить модель по нормализованным переменным.
17. Вывести коэффициенты. Вывести R2. Сравнить результаты с Заданием 2.
18. Рассчитать время обучения моделей SGDRegressor, LinearRegression и statsmodels. Какая в данном случае работает быстрее?

Для текущего контроля ТКЗ:

Проверяемая компетенция: УК-1.1, УК-1.2

Комплект лабораторных работ

Работа 8. Дисперсионный анализ

Цель работы: *освоение методики проведения одно- и двухфакторного дисперсионного анализа.*

1. Однофакторный анализ

Задача 1. Заданы значения объемов продаж некоторого товара по пяти регионам.

Провести дисперсионный анализ для установления факта: зависит ли объем продаж товара от уровней фактора, т.е. от региона?

Порядок выполнения работы:

1. В ячейку A1 ввести текст Однофакторный дисперсионный анализ, A2 - Объем продаж по регионам, в ячейки A3:E11 - данные из таблицы.

2. Вызвать команду Сервис, Анализ данных, Однофакторный дисперсионный анализ. Ввести, используя мышь, в поле Входной интервал адреса ячеек, содержащих имена столбцов R1- R5 и значения элементов

R1	R2	R3	R4	R5
79,8	87,3	42,45	76	70,7
86,3	69,6	64,3	83,5	64,65
86,5	81,75	78,9	72,8	38,5
92,3	77,95	61	89	77
76,5	83,65	31,3	76,5	91,5
87,05	64,8	72,85	87,45	68
82,5	67,3	58,65	74,5	38,05
90	75,45	52,5	93,15	79,95

выборки. Нажать кнопку Группирование по столбцам, установить флажок Метки, указать в поле Выходной интервал адрес A12 левой верхней ячейки блока ячеек, в который будут помещены результаты анализа, ОК.

3. Проанализировать полученные результаты и сделать вывод о том, имеются ли статистически значимые отличия между объемами продаж в различных регионах.

4. Выявить так называемые *гомогенные* группы, т.е. группы, которые по значениям испытуемого признака незначимо отличаются друг от друга. Для этого:

- в ячейки F15, G15, H15 ввести названия столбцов Дов.инт. (доверительный интервал), Н.гран. (нижняя граница), В.гран. (верхняя граница), соответственно;

- в ячейке F16 вычислить значение доверительного интервала для среднего значения испытуемого признака в первой группе, используя функцию Мастера функций ДОВЕРИТ (доверительные интервалы) и введя в соответствующие поля ввода значения $\text{Альфа} = 0,05$, $\text{Размер} = 8$ (размер выборки). Величину *Стандартное_откл* (стандартное отклонение) вычислить непосредственно в поле ввода функции ДОВЕРИТ как положительное значение квадратного корня из величины дисперсии (ячейка E16), с помощью *вложенной* функции КОРЕНЬ (она вызывается щелчком по полю *Имени* в левом верхнем углу окна рабочей книги). Скопировать формулу для средних значений в других группах;

- в столбцах С и Н вычислить нижние и верхние границы доверительного интервала, вычитая и прибавляя к среднему групповому значению испытуемого признака величину доверительного интервала, соответственно;

- построить график, включив в исходный диапазон адреса ячеек D15:D20; G15:H20 и выбрав *Тип диаграммы - Биржевая*, а *Вид графика - набор из трех значений (самый высокий курс, самый низкий курс, курс закрытия)*', установить размер графика на всю высоту окна рабочей книги;

- проанализировать представленную на диаграмме информацию с целью установления перекрывающихся (общих) областей доверительных интервалов для разных групп (уровней фактора).

Задача 2. Считается, что кофеин при приеме его внутрь оказывает возбуждающее действие, результат и разброс которого зависят от величины принятой дозы. Чтобы получить сведения о действии кофеина при выполнении физической деятельности, был проведен следующий простой эксперимент.

Использовались три уровня обработки (три дозы): 0, 100 и 200 мг кофеина. Было отобрано и обучено быстро ударять пальцем по клавише 30 студентов мужского пола одинакового возраста с примерно одинаковыми физическими данными. После обучения из них были скомплектованы случайным образом три группы по 10 человек. Студенты получали одну из доз, однако никто из испытуемых и наблюдающих физиологов заранее не знал, какую дозу будет принимать тот или иной человек. Это знал только статистик, который занимался обработкой данных. Через два часа после приема доз каждый из испытуемых должен был выполнить предусмотренное упражнение. Результаты (число ударов в минуту) представлены в таблице.

Обработки	Наблюдения
0 мг кофеина	242, 245, 244, 248, 247, 248, 242, 244, 246, 242
1 00 мг кофеина	248, 246, 245, 247, 248, 250, 247, 246, 243, 244
200 мг кофеина	246, 248, 250, 252, 248, 250, 246, 248, 245, 250

Необходимо выяснить, имеются ли значимые отличия в результатах у студентов трех групп.

2. Двухфакторный анализ

Задача 3. Установить, зависит ли объем продаж товара от времени продажи и от региона. Данные представлены в таблице. *Порядок выполнения работы:*

1. Перейти на новый лист, сделать заголовки, аналогичные заголовкам Листа 1, оформить исходные данные согласно таблице.

2. Вызвать команду Сервис, Анализ данных, Д в у х ф а к т о р н ы й дисперсионный анализ с повторениями. Ввести, используя мышь, в поле Входной интервал адреса ячеек, содержащих имена столбцов и строк, а также значения

	<i>k1</i>	<i>k2</i>	<i>k3</i>	<i>k4</i>
<i>r1</i>	13,2 13,9	18,9 21	7,3 8,5	20 20,8
<i>r2</i>	4,7 5,8	19,8 17,9	38,2 37,7	60,1 60,9
<i>r3</i>	53,4 48,3	14 13,2	5,1 5,9	19,6 18,5
<i>r4</i>	13,6 13,2	9,5 8,6	54,4 55,2	58,2 59,7

элементов выборки. Задать число строк для выборки, равное двум, так как выборочные данные, ориентированные по строкам (регионы), расположены на двух строках. Указать в поле Выходной интервал адрес A14 левой верхней ячейки блока ячеек, в который будут помещены результаты анализа, ОК.

3. Проанализировать полученные результаты и сделать вывод о том, имеются ли статистически значимые отличия между объемами продаж в различных регионах и по разным кварталам. Обратить внимание на то, что результаты представлены в виде расчетов отдельно по регионам и итоговые расчеты. Выводы о наличии влияния уровней факторов на испытываемый признак сделать по данным, представленным в таблице *ANOVA*, попарно сравнивая между собой оценки дисперсий *MS*: *Внутри* и *Выборка*, *Внутри* и *Столбцы*, а также соответствующие значения *P* и *P-критическое*, *P-значение* и $\alpha = 0,05$.

Задача 4. Изучается влияние на содержание витамина С в яблоках двух факторов: сорта яблок (фактор *A*) и способа опыления (фактор *B*). Были выбраны четыре сорта яблок: джонатан, антоновка, коричное, ренет Симиренко и два способа опыления: пылью антоновки и пылью ренета.

Данные по содержанию витамина С в яблоках представлены в таблице.

Оценить влияние факторов на содержание витамина С в яблоках разных сортов.

Способ опыления	Сорт			
	Джонатан	Антоновк а	Коричное	Ренет Симиренк
Пылью антоновки	18,7	19,5	16,5	24,5
	20,1	22,1	18,7	21,2
	17,4	21,7	15,8	20
	16,7	26,2	0	22,1
Пылью ренета	18,3	20,5	13,5	24,5

	19,1	22,1	15,7	20,2
	17,4	18,7	12,8	19
	16,8	25,2	14,2	20,2

Работа 9. Модели бинарного выбора

Цель работы: освоить методику расчета с использованием модели бинарного выбора .

1. Загрузить исходные данные в специализированный ППП.
2. Получить оценки параметров модели бинарного выбора с помощью логит-модели.
3. Получить оценки параметров модели бинарного выбора с помощью пробит-модели.
4. По полученным результатам выбрать подходящую модель.
5. В MS Excel получить расчетные значения. Оценить точность модели.
6. По построенной модели получить прогнозные значения для заданных условий.
7. Построить модель в Python и сравнить

В отчет

1. Записать полученную логит-модель.
2. Записать полученную пробит-модель.
3. Сделать вывод.
4. Записать результаты прогнозирования.

Контрольные вопросы

1. Модели бинарного выбора.
2. Свойства функции моделирования вероятности.
3. Логит-модель.
4. Пробит-модель.
5. Оценка качества модели.
6. Оценка значимости коэффициентов модели.

Работа 10. Модель множественного выбора

Цель работы: освоить методику расчета с использованием модели бинарного выбора .

1. Загрузить исходные данные в специализированный ППП.
2. Получить оценки параметров модели множественного выбора с помощью логит-модели.
3. В MS Excel получить расчетные значения. Оценить точность модели.
4. По построенной модели получить прогнозные значения для заданных условий.

В отчет

1. Записать полученную логит-модель.
2. Оценить качество и значимость параметров модели. Сделать вывод.
3. Записать результаты прогнозирования.

Контрольные вопросы

1. Модели множественного выбора.
2. Логит-модель множественного выбора.
3. Пробит-модель множественного выбора.
4. Оценка качества модели.
5. Оценка значимости коэффициентов модели.
6. Предельные эффекты.

Работа 11. Задачи классификации

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Проведите первичный анализ данных соревнования
2. Выполните предварительную обработку данных
3. Построить модель логистической регрессии. Учтите, что в модели должны остаться только значимые факторы.
4. Какие факторы вошли в модель (словами, а не именами переменных)?
5. Записать модель
6. Получить расчетные значения
7. Проанализировать метрики качества модели
8. Дать экономическую интерпретацию коэффициентам регрессии.
9. Проанализировать результаты

Критериями оценки выполнения задания, согласно достигнутого уровня, являются: Высокий уровень. Знания теоретического материала, правильно выполнены все задания в соответствии с требованиями, полные ответы на вопросы, правильно выполнены домашние задания, своевременно предоставлен отчет о выполнении работы - 4 балла. Средний уровень Теоретический материал знает, правильно выполнены все задания, ответы на вопросы не полные, домашние задания выполнены не в полном объеме, предоставлен отчет о выполнении работы, либо в случае несвоевременного предоставления отчета или с наличием несущественных ошибок в выполнении лабораторных заданиях - 3 балла. Ниже среднего уровень. Выполнено не все, но более 50% заданий лабораторной работы, домашнее задание не выполнены, несвоевременно предоставлен отчет о выполнении работы - 2 балла. Низкий уровень. Выполнено менее 50% лабораторной работы, не выполнено домашнее задание, отчет о выполнении работы не предоставлен – 1 балл
Количество баллов за Отчёт по лабораторной работе: минимум – 1 б.
Количество баллов за Отчёт по лабораторной работе (с учетом коэффициента сложности): максимум – 4 б.

Для контроля ОМ:

Проверяемая компетенция: УК-1.1, УК-1.2

Тест

Примерный вариант

Линейный коэффициент корреляции предназначен для измерения...

Выберите один ответ:

- a. гомоскедастичности
- b. значимости уравнения
- c. тесноты связи между факторами ✓
- d. значимости факторов
- e. гетероскедастичности

Дан вариационный ряд, где x_i – варианты, n_i – частоты, w_i – относительные частоты, n – объем совокупности. \bar{x} – среднее

Отметить формулу (формулы) для расчета дисперсии

Выберите один или несколько ответов:

- a. $\frac{\sum x_i n_i}{n}$
- b. $\frac{\sum (x_i - \bar{x})^2 \cdot w_i}{n}$
- c. $\frac{\sum (x_i - \bar{x})^2}{n}$ ✓
- d. $\sum (x_i - \bar{x})^2 \cdot w_i$
- e. $\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

В таблице представлена выборка. Отметить вариант с исходными данными

x_i	1	2	3
n_i	2	3	5

Выберите один ответ:

- a. 3, 1, 3, 1, 2, 3, 2, 2, 3, 1 ✗
- b. 3, 3, 2, 3, 3, 2, 1, 1, 2, 3
- c. 2, 3, 2, 2, 1, 2, 2, 3, 1, 1
- d. 2, 1, 2, 2, 1, 1, 3, 3, 1, 2

Коэффициент эластичности - это...

Выберите один или несколько ответов:

- a. $\frac{1}{n} \cdot \sum \left| \frac{y - y_p}{y} \right| \cdot 100$
- b. $f'(x) \frac{x}{y}$
- c. $\sqrt{1 - \frac{SS_{ост}}{SS_{общ}}} \times$
- d. $\frac{100}{\bar{y}} \cdot \sqrt{\frac{\sum (y - y_p)^2}{n}}$
- e. $\frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1}$
- f. $\frac{100}{\bar{y}} \cdot \frac{\sum (y - y_p)^2}{n}$

Общая дисперсия рассчитывается по формуле

n - объем выборки

m - число параметров уравнения регрессии

Выберите один ответ:

- a. $\sum (y_i - \hat{y}_i)^2$
- b. $\sum (y_i - \bar{y})^2$
- c. $\frac{\sum (y_i - \hat{y}_i)^2}{n - m}$
- d. $\sum (\hat{y}_i - \bar{y})^2$
- e. $\frac{\sum (y_i - \bar{y})^2}{n - 1}$ ✓
- f. $\frac{\sum (\hat{y}_i - \bar{y})^2}{m - 1}$

По результатам регрессионного анализа отметить верные утверждения

Регрессионная статистика					
Множественный R	0,940				
R-квадрат	0,884				
Нормированный R-квадрат	0,826				
Стандартная ошибка	43,227				
Наблюдения	16				
Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	5	142693	28539	15,27315	0,000210
Остаток	10	18686	1869		
Итого	15	161379			
	Коэффициенты	Станд. ошибка	t-статистика	P-Значение	
Y-пересечение	-3008,975	1136,175	-2,648	0,024	
x1	-6,674	5,383	-1,240	0,243	
x2	6,678	3,153	2,118	0,0001	
x3	30,664	11,975	2,561	0,028	
x4	13,949	15,157	0,920	0,379	
x5	-10,216	17,045	-0,599	0,562	

Выберите один или несколько ответов:

- a. факторная дисперсия 142693
- b. объем выборки 14 X
- c. факторная дисперсия 15,27 X
- d. факторная дисперсия 28539
- e. фактор x3 значим
- f. фактор x4 значим X

Для чего используется метод наименьших квадратов?

Выберите один ответ:

- a. для оценки значимости уравнения регрессии X
- b. для оценки качества модели
- c. для нахождения y расчетного
- d. для оценки значимости параметров уравнения регрессии
- e. для нахождения числа параметров уравнения регрессии
- f. для нахождения параметров уравнения регрессии

Каков порядок частного коэффициента корреляции?

$$r_{yx1,x2,x3,x6}$$

Ответ ввести цифрой.

Свойства стандартизованных переменных:

Выберите один или несколько ответов:

- a. $\sigma[t]=1$ ✓
- b. $M[t]=1$
- c. $D[t]=0$
- d. $\sigma[t_y]=\sigma[t_{y0}]+\sum\sigma[t_{y_j}]$ X
- e. $M[t]=0$

Получено уравнение регрессии:

$$y = 60 - 3 \cdot x$$

где x - число автобусов на маршруте

y - время ожидания автобуса, мин

Сколько автобусов можно снять с маршрута, чтобы время ожидания увеличилось не более, чем на 3 мин?

Ответ ввести числом.

Отметить подходящее линеаризирующее преобразование

$$y = b_0 \cdot e^{b_1 x}$$

Выберите один ответ:

- a. $\ln y = b_0 + b_1 x$ ✗
- b. $\ln y = \ln b_0 + b_1 \ln x$
- c. $y = b_0 + b_1 x_1$
- d. $\ln y = \ln b_0 + b_1 x$
- e. $y = b_0 + b_1 \frac{1}{x}$
- f. $y = b_0 + b_1 x_1 + b_2 x_2$

Ранговый коэффициент корреляции Спирмена может изменяться в интервале

Выберите один ответ:

- a. должен быть больше t критического
- b. должен быть меньше t критического
- c. от 0 до 1
- d. от 0 до бесконечности
- e. от -1 до 1 ✓

В множественной линейной регрессии каким должно быть соотношение между объемом выборки n и числом факторов m ?

Выберите один ответ:

- a. $n > m$
- b. $n > m + 1$ ✗
- c. $n < m$
- d. $n > 2 \cdot m$
- e. $n > 5 \cdot m$

В модели бинарного выбора можно оценить...

Выберите один ответ:

- a. математическое ожидание $M(Y)$ где $Y=\{-1;0;+1\}$
- b. вероятность $P(Y=1)$ где $Y=\{-1;+1\}$ ✓
- c. вероятность $P(Y=0)$ где $Y=\{-1;0;+1\}$
- d. математическое ожидание $M(Y)$ где $Y=\{-1;+1\}$
- e. вероятность $P(Y=1)$ где $Y=\{-1;0;+1\}$

В результате оценки параметров множественной **logit-модели** в ППП Statistica была получена следующая таблица

y - Parameter estimates (множеств. sta)									
Distribution : MULTINOMIAL									
Link function: LOGIT									
Effect	Level of Effect	Level of Response	Column	Estimate	Standard Error	Wald Stat.	Lower CL 95, %	Upper CL 95, %	p
Intercept 1		0	1	5,35856	3,473264	2,380242	-1,44891	12,16604	0,122878
x2		0	2	-0,08802	0,942748	0,008717	-1,93577	1,75973	0,925612
x3		0	3	-2,37098	1,095301	4,685864	-4,51773	-0,22423	0,030412
Intercept 2		1	4	5,84634	3,381311	2,989492	-0,78091	12,47358	0,083806
x2		1	5	-0,63935	0,903627	0,500603	-2,41042	1,13173	0,479235
x3		1	6	-1,93432	1,016099	3,623960	-3,92583	0,05720	0,056953
Scale				1,00000	0,000000		1,00000	1,00000	

Рассчитать значение линейной формы $u(y_i=0)$ для $x_2=0$ и $x_3=0$.

Ответ округлить до сотых, ввести через точку.

Пример: 0.24

Пример: 0.40

Значение линейного коэффициента корреляции $r(x,y)=0,095$

Выберите один ответ:

- a. связь логарифмическая
- b. связь обратная
- c. существует линейная зависимость от времени
- d. связь статистически значима
- e. связь статистически не значима
- f. связь полиномиальная ✗
- g. возможна нелинейная зависимость от времени

Дана выборка. Рассчитать значение линейного коэффициента корреляции. Среднеквадратические отклонения принять

$S_x=1,6$

$S_y=2,6$

Ответ округлить до двух знаков после запятой.

x	0	-2	2	0
y	-2	0	2	4

В результате построения модели множественного выбора была получена следующая матрица ошибок

Observed	Predicted 0	Predicted 1	Predicted 2
0	6	1	1
1	3	4	1
2	1	1	3

Чему равняется точность в категории 2?

Ответ округлить до сотых, ввести через точку, ответ вводить НЕ в процентах!

Пример: 0.24

Пример: 0.40

Пусть выборка задана рядом распределения относительных частот признака X.

Вычислите среднее арифметическое выборки

x_i	1	2	4	7
w_i	0,4	0,4	0,1	0,1

К методам снижения размерности относятся...

Выберите один или несколько ответов:

- a. метод главных компонент ✓
- b. регрессионный анализ
- c. многофакторный дисперсионный анализ
- d. кластерный анализ
- e. дискриминантный анализ ✓

Имеется следующий ряд данных.

0
1
0
-1
-1
0
2
2

Определить ранг для числа "0"