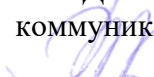




**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»

Кафедра прикладной информатики и документоведения

УТВЕРЖДАЮ

Декан факультета бизнес-коммуникаций и информатики

В.К. Карнаухова

«25» марта 2022 г.

Рабочая программа дисциплины (модуля)

Наименование дисциплины (модуля)

Б1.В.02 Анализ данных

(индекс дисциплины по учебному плану, наименование дисциплины (модуля)).

Направление подготовки:

09.03.03 Прикладная информатика

(код, наименование направления подготовки)

Направленность (профиль) подготовки:

Прикладная информатика в управлении

Квалификация выпускника – **бакалавр**

Форма обучения: **очно-заочная** *(с использованием электронного обучения и дистанционных образовательных технологий в полном объеме)*

(очная, заочная (с использованием электронного обучения и дистанционных образовательных технологий), очно-заочная (с использованием электронного обучения и дистанционных образовательных технологий*))*

Согласовано с УМК факультета
бизнес-коммуникаций и информатики

Рекомендовано кафедрой прикладной
информатики и документоведения

Протокол № 7 от «16» марта 2022 г.

Протокол № 8 от «04» марта 2022 г.

Председатель  В.К. Карнаухова

и.о.зав. кафедрой  А.В. Рохин

Иркутск – 2022

СОДЕРЖАНИЕ

	<i>стр.</i>
I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ (МОДУЛЯ)	3
II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО	3
III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	3
IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ	4
4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов	4
4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине.	4
4.3 Содержание учебного материала	5
4.3.1. Перечень семинарских, практических занятий и лабораторных работ	6
4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение самостоятельной работы студентов	6
4.4. Методические указания по организации самостоятельной работы студентов	7
4.5. Примерная тематика курсовых работ (проектов)	12
V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)	13
а) основная литература	13
б) дополнительная литература.....	13
в) периодическая литература	13
г) базы данных, информационно-справочные и поисковые системы.....	13
VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	14
6.1. Учебно-лабораторное оборудование:	14
6.2. Программное обеспечение:	14
6.3. Технические и электронные средства:	15
VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ	15
VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ	16
8.1. Оценочные средства текущего контроля.....	17
8.2. Оценочные средства для промежуточной аттестации	23

I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ (МОДУЛЯ)

Цели:

сформировать у студентов компетентностный подход к анализу больших данных и создать целостную, логически замкнутую систему знаний, идей и методов анализа данных, а также уметь применять полученные знания на практике.

Задачи:

- формирование представлений о науке о данных;
- развитие логического мышления, научного кругозора;
- выработка навыков компьютерной обработки больших данных для решения математических и профессиональных задач;
- ознакомление с математическими методами изучения закономерностей однородных случайных массовых явлений, с общими идеями создания математических моделей, выполнения работ и проведения исследований в профессиональной деятельности.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

2.1. Учебная дисциплина относится к обязательной части программы направления 09.03.03 «Прикладная информатика».

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами: Математика, Прикладная математика, Теория вероятностей и математическая статистика, Дискретная математика, Программирование

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной: Машинное обучение и большие данные, Теория систем и системный анализ, Прикладной системный анализ.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

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

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
ПК-1 Способность выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы	ПК-1.1	Знать выполнение работ по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы
	ПК-1.2	Уметь выполнять работы по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы
	ПК-1.3	Владеть навыками проведения работ по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов, в том числе 8 часов на контроль.

Форма промежуточной аттестации: зачет с оценкой.

4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов очно-заочная форма обучения

№ п/п	Раздел дисциплины/темы	Семестр	Всего часов	Из них практическая подготовка обучающихся	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Формы текущего контроля успеваемости; Форма промежуточной аттестации (по семестрам)
					Контактная работа преподавателя с обучающимися			Самостоятельная работа	
					Лекции	Семинарские (практические занятия)	Консультации, контроль		
1.	Раздел 1: Введение в науку о данных.	5				12		22	УО
2.	Раздел 2. Статистические методы анализа данных.	5				12		22	УО
3.	Раздел 3. Оптимизация данных.	5				12		20	УО
4.	Промежуточная аттестация	5					8		Зачет с оценкой
Итого часов			108			36	8	64	

4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Затраты времени (час.)		
5	Раздел 1: Введение в науку о данных.	Визуализация данных	2 нед	20	СР-1	презентация
5	Раздел 2. Статистические методы анализа данных.	Парная регрессия	7 нед	20	СР-4	презентация
5	Раздел 3. Оптимизация данных.	Градиентный спуск	14 нед	24	СР-5	презентация
Общая трудоемкость самостоятельной работы по дисциплине (час)				64		

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Затраты времени (час.)		
	Из них объем использования дистанционных образовательных технологий (час)	самостоятельной работы с использованием электронного обучения	с и	64		
	Бюджет времени самостоятельной работы, предусмотренный учебным планом для данной дисциплины (час)			64		

4.3 Содержание учебного материала

Трудоемкость дисциплины (з.е.)	3
Наименование основных разделов (модулей)	<p>Раздел 1: Введение в науку о данных.</p> <p>Тема 1. Визуализация данных Библиотека matplotlib языка Python. Столбчатые диаграммы. Линейные графики. Точечные диаграммы.</p> <p>Тема 2. Обработка данных и работа с векторами и матрицами Исследование одномерных данных. Двумерные данные. Многомерные данные. Шкалирование. Снижение размерности. Сложение, вычитание векторов, умножение вектора на скаляр. Скалярное, векторное произведения векторов. Евклидово расстояние между векторами. Операции над матрицами и решение линейных систем уравнений.</p> <p>Тема 3. Теория вероятностей. Классическое определение вероятности. Условная и безусловная вероятность, проверка на PYTHON. Теорема Байеса. Нормальное распределение. Интегральная и дифференциальная функции нормального распределения. Распределение Бернулли. Центральная предельная теорема. Проверка статистических гипотез. A/B тестирование.</p> <p>Раздел 2. Статистические методы анализа данных.</p> <p>Тема 4. Описательная статистика Нахождение среднего, дисперсии, стандартного отклонения, вариации. Корреляция и ковариация.</p> <p>Тема 5. Регрессионный анализ. Простая линейная регрессия. Парная регрессия. Множественная регрессия. Стандартные ошибки коэффициентов регрессии. Логистическая регрессия. Прогнозирование по регрессионным моделям.</p> <p>Раздел 3. Оптимизация данных.</p> <p>Тема 6. Численные методы оптимизации. Градиентный спуск. Вычисление градиента. Использование градиента. Выбор оптимального размера шага. Стохастический градиентный спуск.</p> <p>Тема 7. Кластеризация данных. Кластеры. Метод K-средних. Выбор числа кластеров. Иерархическая кластеризация. Кластеризация цвета.</p>
Формы текущего контроля	тесты, контрольные работы, практические занятия

Форма промежуточной аттестации	Зачет с оценкой
--------------------------------	-----------------

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

п/п	№ раздела и темы дисциплины (модуля)	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)		Оценочные средства	Формируемые компетенции
			Всего часов	Из них практическая подготовка		
	2	3	4		5	6
1	Раздел 1. Тема 1.	Визуализация данных	2		СР-1	ПК-1
2	Раздел 1. Тема 2.	Обработка данных и работа с векторами и матрицами	4		СР-2	ПК-1
3	Раздел 2. Тема 4.	Описательная статистика.	2		СР-3	ПК-1
4	Раздел 2. Тема 5.	Парная регрессия.	2		СР-4	ПК-1
5	Раздел 3. Тема 6.	Градиентный спуск.	4		СР-5	ПК-1
6	Раздел 1. Тема 3.	Теория вероятности.	4		СР-7	ПК-1
7	Раздел 3. Тема 6.	Метод деления отрезка пополам.	4		СР-8	ПК-1
8	Раздел 1. Тема 3.	Проверка статистических гипотез.	2		СР-9	ПК-1
9	Раздел 2. Тема 5.	Стохастический градиентный спуск для нахождения коэффициентов регрессии.	2		СР-6	ПК-1
10	Раздел 2. Тема 5.	Множественная регрессия.	2		СР-10	ПК-1
11	Раздел 2. Тема 4.	Оценка коэффициентов регрессии.	2		СР-10	ПК-1
12	Раздел 2. Тема 5.	Логистическая регрессия.	2		СР-11	ПК-1
13	Раздел 3. Тема 7.	Кластеризация данных.	4		СР-12	ПК-1
			36			

4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение самостоятельной работы студентов

№ п/п	Тема	Задание	Формируемая компетенция	ИДК
	Раздел 1. Тема 2.	Матрицы. Библиотека NumPy http://numpy.org для работы с матрицами.	ПК-1	ПК-1.1 – ПК.1.3
	Раздел 2. Тема 4.	Библиотека scipy.stats http://docs.scipy.org/doc/scipy/reference/stats.html	ПК-1	ПК-1.1 – ПК.1.3
	Раздел 2.	Модуль SGDClassifier стохастического	ПК-1	ПК-1.1 –

	Тема 5.	градиентного спуска библиотеки scikit-learn http://scikit-learn.org/stable/modules/sgd.html		ПК.1.3
	Раздел 2. Тема 5.	Гребневая регрессия и Лассо-регрессия	ПК-1	ПК-1.1 – ПК.1.3
	Раздел 2. Тема 5.	Модуль linear_model линейной регрессии библиотеки scikit-learn http://scikit-learn.org/stable/modules/linear_model.html	ПК-1	ПК-1.1 – ПК.1.3
	Раздел 2. Тема 5.	Библиотека Statsmodels (http://statsmodels.sourceforge.net) (содержит модели линейной регрессии, ДФР и ИФР для популярных распределений вероятностей)	ПК-1	ПК-1.1 – ПК.1.3
	Раздел 2. Тема 5.	Метод опорных векторов	ПК-1	ПК-1.1 – ПК.1.3
	Раздел 3. Тема 7.	Восходящий метод иерархической кластеризации. Библиотека scikit-learn содержит модуль sklearn.cluster. Библиотека SciPy (http://www.scipy.org) располагает двумя моделями кластеризации: scipy.cluster.vq и scipy.cluster.hierarchy.	ПК-1	ПК-1.1 – ПК.1.3

4.4. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов всех форм и видов обучения является одним из обязательных видов образовательной деятельности, обеспечивающей реализацию требований Федеральных государственных стандартов высшего профессионального образования. Согласно требованиям нормативных документов самостоятельная работа студентов является обязательным компонентом образовательного процесса, так как она обеспечивает закрепление получаемых на лекционных занятиях знаний путем приобретения навыков осмысления и расширения их содержания, навыков решения актуальных проблем формирования общекультурных и профессиональных компетенций, научно-исследовательской деятельности, подготовки к семинарам, лабораторным работам, сдаче зачетов и экзаменов. Самостоятельная работа студентов представляет собой совокупность аудиторных и внеаудиторных занятий и работ. Самостоятельная работа в рамках образовательного процесса в вузе решает следующие задачи:

- закрепление и расширение знаний, умений, полученных студентами во время аудиторных и внеаудиторных занятий, превращение их в стереотипы умственной и физической деятельности;
- приобретение дополнительных знаний и навыков по дисциплинам учебного плана;
- формирование и развитие знаний и навыков, связанных с научно-исследовательской деятельностью;
- развитие ориентации и установки на качественное освоение образовательной программы;
- развитие навыков самоорганизации;
- формирование самостоятельности мышления, способности к саморазвитию, самосовершенствованию и самореализации;
- выработка навыков эффективной самостоятельной профессиональной

теоретической, практической и учебно-исследовательской деятельности.

Подготовка к лекции. Качество освоения содержания конкретной дисциплины прямо зависит от того, насколько студент сам, без внешнего принуждения формирует у себя установку на получение на лекциях новых знаний, дополняющих уже имеющиеся по данной дисциплине. Время на подготовку студентов к двухчасовой лекции по нормативам составляет не менее 0,2 часа.

Подготовка к практическому занятию. Подготовка к практическому занятию включает следующие элементы самостоятельной деятельности: четкое представление цели и задач его проведения; выделение навыков умственной, аналитической, научной деятельности, которые станут результатом предстоящей работы. Выработка навыков осуществляется с помощью получения новой информации об изучаемых процессах и с помощью знания о том, в какой степени в данное время студент владеет методами исследовательской деятельности, которыми он станет пользоваться на практическом занятии. Подготовка к практическому занятию нередко требует подбора материала, данных и специальных источников, с которыми предстоит учебная работа. Студенты должны дома подготовить к занятию 3–4 примера формулировки темы исследования, представленного в монографиях, научных статьях, отчетах. Затем они самостоятельно осуществляют поиск соответствующих источников, определяют актуальность конкретного исследования процессов и явлений, выделяют основные способы доказательства авторами научных работ ценности того, чем они занимаются. В ходе самого практического занятия студенты сначала представляют найденные ими варианты формулировки актуальности исследования, обсуждают их и обосновывают свое мнение о наилучшем варианте. Время на подготовку к практическому занятию по нормативам составляет не менее 0,2 часа.

Подготовка к семинарскому занятию. Самостоятельная подготовка к семинару направлена: на развитие способности к чтению научной и иной литературы; на поиск дополнительной информации, позволяющей глубже разобраться в некоторых вопросах; на выделение при работе с разными источниками необходимой информации, которая требуется для полного ответа на вопросы плана семинарского занятия; на выработку умения правильно выписывать высказывания авторов из имеющихся источников информации, оформлять их по библиографическим нормам; на развитие умения осуществлять анализ выбранных источников информации; на подготовку собственного выступления по обсуждаемым вопросам; на формирование навыка оперативного реагирования на разные мнения, которые могут возникать при обсуждении тех или иных научных проблем. Время на подготовку к семинару по нормативам составляет не менее 0,2 часа.

Подготовка к коллоквиуму. Коллоквиум представляет собой коллективное обсуждение раздела дисциплины на основе самостоятельного изучения этого раздела студентами. Подготовка к данному виду учебных занятий осуществляется в следующем порядке. Преподаватель дает список вопросов, ответы на которые следует получить при изучении определенного перечня научных источников. Студентам во внеаудиторное время необходимо прочитать специальную литературу, выписать из нее ответы на вопросы, которые будут обсуждаться на коллоквиуме, мысленно сформулировать свое мнение по каждому из вопросов, которое они выскажут на занятии. Время на подготовку к коллоквиуму по нормативам составляет не менее 0,2 часа.

Подготовка к контрольной работе. Контрольная работа назначается после изучения определенного раздела (разделов) дисциплины и представляет собой совокупность развернутых письменных ответов студентов на вопросы, которые они заранее получают от преподавателя. Самостоятельная подготовка к контрольной работе включает в себя: — изучение конспектов лекций, раскрывающих материал, знание которого проверяется контрольной работой; повторение учебного материала, полученного при подготовке к семинарским, практическим занятиям и во время их проведения;

изучение дополнительной литературы, в которой конкретизируется содержание проверяемых знаний; составление в мысленной форме ответов на поставленные в контрольной работе вопросы; формирование психологической установки на успешное выполнение всех заданий. Время на подготовку к контрольной работе по нормативам составляет 2 часа.

Подготовка к зачету (в том числе к дифференцированному при отсутствии экзамена по дисциплине). Самостоятельная подготовка к зачету должна осуществляться в течение всего семестра. Подготовка включает следующие действия: перечитать все лекции, а также материалы, которые готовились к семинарским и практическим занятиям в течение семестра, соотнести эту информацию с вопросами, которые даны к зачету, если информации недостаточно, ответы находят в предложенной преподавателем литературе. Рекомендуется делать краткие записи. Время на подготовку к зачету по нормативам составляет не менее 4 часов.

Формы внеаудиторной самостоятельной работы

Подготовка доклада Цель самостоятельной работы: расширение научного кругозора, овладение методами теоретического исследования, развитие самостоятельности мышления студента. Доклад — публичное сообщение или документ, которые содержат информацию и отражают суть вопроса или исследования применительно к данной ситуации. Виды докладов: 1. Устный доклад — читается по итогам проделанной работы и является эффективным средством разъяснения ее результатов. 2. Письменный доклад: — краткий (до 20 страниц) — резюмирует наиболее важную информацию, полученную в ходе исследования; — подробный (до 60 страниц) — включает не только текстовую структуру с заголовками, но и диаграммы, таблицы, рисунки, фотографии, приложения, сноски, ссылки, гиперссылки. Выполнение задания: 1) четко сформулировать тему (например, письменного доклад); 2) изучить и подобрать литературу, рекомендуемую по теме, выделив три источника библиографической информации: — первичные (статьи, диссертации, монографии и т. д.); — вторичные (библиография, реферативные журналы, сигнальная информация, планы, граф-схемы, предметные указатели и т. д.); — третичные (обзоры, компилятивные работы, справочные книги и т. д.); 3) написать план, который полностью согласуется с выбранной темой и логично раскрывает ее; 4) написать доклад, соблюдая следующие требования: к структуре доклада — она должна включать: краткое введение, обосновывающее актуальность проблемы; основной текст; заключение с краткими выводами по исследуемой проблеме; список использованной литературы; к содержанию доклада — общие положения надо подкрепить и пояснить конкретными примерами; не пересказывать отдельные главы учебника или учебного пособия, а изложить собственные соображения по существу рассматриваемых вопросов, внести свои предложения; 5) оформить работу в соответствии с требованиями. Планируемые результаты самостоятельной работы: способность студентов анализировать результаты научных исследований и применять их при решении конкретных образовательных и исследовательских задач; готовность использовать индивидуальные креативные способности для оригинального решения исследовательских задач; способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

Составление глоссария Цель самостоятельной работы: повысить уровень информационный культуры; приобрести новые знания; отработать необходимые навыки в предметной области учебного курса. Глоссарий — словарь специализированных терминов и их определений. Статья глоссария — определение термина. Содержание задания: сбор и систематизация понятий или терминов, объединенных общей специфической тематикой, по одному либо нескольким источникам. Выполнение задания: 1) внимательно прочитать работу; 2) определить наиболее часто встречающиеся термины; 3) составить список

терминов, объединенных общей тематикой; 4) расположить термины в алфавитном порядке; 5) составить статьи глоссария: — дать точную формулировку термина в именительном падеже; — объемно раскрыть смысл данного термина. Планируемые результаты самостоятельной работы: способность студентов решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

Выполнение кейс-задания Цель самостоятельной работы: формирование умения анализировать в короткие сроки большой объем неупорядоченной информации, принятие решений в условиях недостаточной информации. Кейс-задание (англ. case — случай, ситуация) — метод обучения, основанный на разборе практических проблемных ситуаций — кейсов, связанных с конкретным событием или последовательностью событий. Виды кейсов: иллюстративные, аналитические, связанные с принятием решений. Выполнение задания: 1) подготовить основной текст с вопросами для обсуждения: — титульный лист с кратким запоминающимся названием кейса; — введение, где упоминается герой (герои) кейса, рассказывается об истории вопроса, указывается время начала действия; — основная часть, где содержится главный массив информации, внутренняя интрига, проблема; — заключение (в нем решение проблемы, рассматриваемой в кейсе, иногда может быть не завершено); 2) подобрать приложения с подборкой различной информации, передающей общий контекст кейса (документы, публикации, фото, видео и др.); 3) предложить возможное решение проблемы. Планируемые результаты самостоятельной работы: — способность студентов анализировать результаты научных исследований и применять их при решении конкретных исследовательских задач; — готовность использовать индивидуальные креативные способности для оригинального решения исследовательских задач; — способность решать нестандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий.

Составление тематического портфолио работ Цель самостоятельной работы: развитие способности к систематизации и анализу информации по выбранной теме, работе с эмпирическими данными, со способами и технологиями решения проблем. Тематическое портфолио работ — материалы, отражающие цели, процесс и результат решения какой-либо конкретной проблемы в рамках той или иной темы курса (модуля). Портфолио работ состоит из нескольких разделов (согласуются с преподавателем). Структура тематического портфолио работ: — сопроводительный текст автора портфолио с описанием цели, предназначения и краткого описания документа; — содержание или оглавление; органайзер (схемы, рисунки, таблицы, графики, диаграммы, гистограммы); лист наблюдений за процессами, которые произошли за время работы; письменные работы; видеофрагменты, компьютерные программы; рефлексивный журнал (личные соображения и вопросы студента, которые позволяют обнаружить связь между полученными и получаемыми знаниями). Выполнение задания: 1) обосновать выбор темы портфолио и дать название своей работе; 2) выбрать рубрики и дать им названия; 3) найти соответствующий материал и систематизировать его, представив в виде конспекта, схемы, кластера, интеллект-карты, таблицы; 4) составить словарь терминов и понятий на основе справочной литературы; 5) подобрать необходимые источники информации (в том числе интернет-ресурсы) по теме и написать тезисы; 6) подобрать статистический материал, представив его в графическом виде; сделать выводы; 7) подобрать иллюстративный материал (рисунки, фото, видео); 8) составить план исследования; 9) провести исследование, обработать результаты; 10) проверить наличие ссылок на источники информации. Планируемые результаты самостоятельной работы: — готовность студентов использовать индивидуальные креативные способности для оригинального решения исследовательских задач; — повышение информационной культуры студентов и

обеспечение их готовности к интеграции в современное информационное пространство; — способность использовать современные способы и технологии решения проблем.

Информационный поиск Цель самостоятельной работы: развитие способности к проектированию и преобразованию учебных действий на основе различных видов информационного поиска. Информационный поиск — поиск неструктурированной документальной информации. Список современных задач информационного поиска: решение вопросов моделирования; классификация документов; фильтрация, классификация документов; проектирование архитектур поисковых систем и пользовательских интерфейсов; извлечение информации (аннотирование и реферирование документов); выбор информационно-поискового языка запроса в поисковых системах. Содержание задания по видам поиска: поиск библиографический — поиск необходимых сведений об источнике и установление его наличия в системе других источников. Ведется путем разыскания библиографической информации и библиографических пособий (информационных изданий); поиск самих информационных источников (документов и изданий), в которых есть или может содержаться нужная информация; — поиск фактических сведений, содержащихся в литературе, книге (например, об исторических фактах и событиях, о биографических данных из жизни и деятельности писателя, ученого и т. п.). Выполнение задания: 1) определение области знаний; 2) выбор типа и источников данных; 3) сбор материалов, необходимых для наполнения информационной модели; 4) отбор наиболее полезной информации; 5) выбор метода обработки информации (классификация, кластеризация, регрессионный анализ и т.д.); 6) выбор алгоритма поиска закономерностей; 7) поиск закономерностей, формальных правил и структурных связей в собранной информации; 8) творческая интерпретация полученных результатов. Планируемые результаты самостоятельной работы: — способность студентов решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности; готовность использовать знание современных проблем науки и образования при решении образовательных и профессиональных задач.

Использование инфографики Цель самостоятельной работы: усвоение отношений между понятиями или отдельными разделами темы с помощью инфографики. Инфографика — «область коммуникативного дизайна, в основе которой лежит графическое представление информации, связей, числовых данных и знаний» (В. В. Лаптев). Вариант задания: представить информацию по заданной теме с помощью зрительных форм — знаков, графического дизайна, рисунков, иллюстраций. Выполнение задания: 1) выбор темы; 2) сбор информации (документальной и визуальной); 3) систематизация собранной информации; 4) создание плана презентации: — классификация информации по типу; — выбор тематики действия (инструктивная, исследовательская, имитационная); — выбор коммуникативной тактики (дискуссии и дебаты для точной передачи идеи); — выбор творческой тактики (создание новых форм и подходов к изучению и представлению информации); — систематизация информации по какому-либо принципу (по алфавиту, по времени, по категориям, по иерархии); 5) создание эскиза (для печатной инфографики) и раскадровка (для интернет-инфографики); 6) планирование и работа над графикой (создание основного и второстепенных объектов). Планируемые результаты самостоятельной работы: — готовность студентов использовать индивидуальные креативные способности для оригинального решения исследовательских задач; — усвоение отношений между понятиями или отдельными разделами темы.

Разработка мультимедийной презентации Цели самостоятельной работы (варианты): — освоение (закрепление, обобщение, систематизация) учебного материала; — обеспечение контроля качества знаний; — формирование специальных компетенций, обеспечивающих возможность работы с информационными технологиями; — становление общекультурных компетенций. Мультимедийная презентация —

представление содержания учебного материала, учебной задачи с использованием мультимедийных технологий. Выполнение задания: 1. Этап проектирования: — определение целей использования презентации; — сбор необходимого материала (тексты, рисунки, схемы и др.); — формирование структуры и логики подачи материала; — создание папки, в которую помещен собранный материал. 2. Этап конструирования: — выбор программы MS PowerPoint в меню компьютера; — определение дизайна слайдов; — наполнение слайдов собранной текстовой и наглядной информацией; — включение эффектов анимации и музыкального сопровождения (при необходимости); — установка режима показа слайдов (титольный слайд, включающий наименование кафедры, где выполнена работа, название презентации, город и год; содержательный — список слайдов презентации, сгруппированных по темам сообщения; заключительный слайд содержит выводы, пожелания, список литературы и пр.). 3. Этап моделирования — проверка и коррекция подготовленного материала, определение продолжительности его демонстрации. Планируемые результаты самостоятельной работы: — повышение информационной культуры студентов и обеспечение их готовности к интеграции в современное информационное пространство; — способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности; — способность к критическому восприятию, обобщению, анализу профессиональной информации, постановке цели и выбору путей ее достижения; — способность применять современные методики и технологии организации и реализации образовательного процесса на различных образовательных ступенях в различных образовательных учреждениях; — готовность использовать индивидуальные креативные способности для оригинального решения исследовательских задач.

Построение сводной (обобщающей) таблицы Цель самостоятельной работы: усвоение отношений между понятиями или отдельными разделами темы с помощью построения таблицы. Сводная (обобщающая) таблица — концентрированное представление отношений между изучаемыми феноменами, выраженными в форме переменных. Варианты задания: — представить функциональные отношения между элементами какой-либо системы, выраженными в тексте в форме понятий или категорий; — представить междисциплинарные связи изучаемой темы (дисциплины). Правила составления таблицы: 1) таблица должна быть выразительной и компактной, лучше делать несколько небольших по объему, но наглядных таблиц, отвечающих задаче исследования; 2) название таблицы, заглавия граф и строк следует формулировать точно и лаконично; 3) в таблице обязательно должны быть указаны изучаемый объект и единицы измерения; 4) при отсутствии каких-либо данных в таблице ставят многоточие либо пишут «Нет сведений», если какое-либо явление не имело места, то ставят тире; 5) значения одних и тех же показателей приводятся в таблице с одинаковой степенью точности; 6) таблица должна иметь итоги по группам, подгруппам и в целом; 7) если суммирование данных невозможно, то в этой графе ставят знак умножения; 8) в больших таблицах после каждых пяти строк делается промежуток для удобства чтения и анализа. Планируемые результаты самостоятельной работы: — готовность студентов использовать индивидуальные креативные способности для оригинального решения исследовательских задач; — усвоение отношений между понятиями или отдельными разделами темы.

В ФБГОУ ВО «ИГУ» организация самостоятельной работы студентов регламентируется Положением о самостоятельной работе студентов, принятым Ученым советом ИГУ 22 июня 2012 г.

4.5. Примерная тематика курсовых работ (проектов)

По данной дисциплине выполнение курсовых проектов (работ) не

предусматривается.

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

а) основная литература

1. Тюрин, Ю. Н. Анализ данных на компьютере [Текст] : учеб. пособие / Ю. Н. Тюрин. - 4-е изд., перераб. - М. : Форум, 2010. - 368 с. : ил. - (Высшее образование). - ISBN 978-5-8199-0356-8 – 5 экз.

2. Анализ данных в пакете StatGraphics [Текст] : учеб. пособие для студ. вузов, обуч. по спец.: математика, компьютер. науки, прикл. информатика (в экономике): Лабораторные работы / Л.И. Бродская, Ю.И. Бродский, М.И. Логинов, Г.С. Шелементьев ; Урал. гос. ун-т им. А.М. Горького. - Екатеринбург : Изд-во Урал. гос. ун-та, 2004. - 131 с. ; 21 см. - Библиогр.: с. 130-131. - ISBN 5-86037-100-4 – 1 экз.

3. Ключников, М. В. Применение Microsoft Word и Excel в финансовых расчетах [Текст] : учеб. пособие / М. В. Ключников. - М. : Маркет ДС Корпорейшн, 2006. - 211 с. : ил. ; 21 см. - Библиогр.: с. 211. - ISBN 5-7958-0111-9—1 экз.

б) дополнительная литература

1. Грасс Дж. Data Science. Наука о данных с нуля: Пер. с англ.—СПб.: БХВ-Петербург, 2017. – Точка доступа <http://i.uran.ru/webcab/system/files/bookspdf/data-science-nauka-o-dannyyh-s-nulya/dana.pdf>

в) периодическая литература

г) базы данных, информационно-справочные и поисковые системы

1. Открытая электронная база ресурсов и исследований «Университетская информационная система РОССИЯ» [Электронный ресурс] : сайт. – Режим доступа: <http://uisrussia.msu.ru> бессрочный

2. Государственная информационная система «Национальная электронная библиотека» [Электронный ресурс] : сайт. – Режим доступа: <http://нэб.рф>. бессрочный

3. Научная электронная библиотека «ELIBRARY.RU» [Электронный ресурс] : сайт. – Режим доступа: <http://elibrary.ru/defaultx.asp>. - Контракт № 148 от 23.12.2020 г. Акт от 24.12.2020 г. срок действия по 31.12. 2021 г. доступ: <http://elibrary.ru/>

4. ЭБС «Издательство Лань». Контракт № 100 от 13.11.2020 г. Акт № 671 от 14.11.2020 г.; Срок действия по 13.11.2021 г. доступ: www.e.lanbook.com, Контракт № 100 от 13.11.2020 г. Акт № Э 656 от 14.11.2020 г. ; Срок действия по 13.11.2021 г. доступ: www.e.lanbook.com

5. ЭБС ЭЧЗ «Библиотех». Государственный контракт № 019 от 22.02.2011 г. ООО «Библиотех». Лицензионное соглашение к Государственному контракту № 019 от 22.02.2011. Адрес доступа: <https://isu.bibliotech.ru/> Срок действия: бессрочный.

6. ЭБС «Рукопт» Контракт № 98 от 13.11.2020 г.; Акт № 6К-5415 от 14.11.20 г. Срок действия по 13.11.2021г. доступ: <http://rucont.ru/>

7. ЭБС «Айбукс.ру/ibooks.ru» Контракт № 99 от 13.11.2020г.; Акт № 99А от 13.11.2020 г. Срок действия по 13.11.2021 г. доступа: <http://ibooks.ru>

8. ООО «Электронное издательство Юрайт». Контракт № 60 от 23.09.2020г. Акт приема-передачи № 3263 от 18.10.2020; Срок действия по 17.10. 2021 г. доступ: <https://urait.ru/>

Лицензионный контракт № 04-Е-0258 от 20.09.2021г. Акт приема-передачи № 5684 от 18.10.2021; Срок действия по 17.10. 2022 г. доступ: <https://urait.ru/>

9. ООО «ИВИС», контракт № 157 от 25. 12.2020 г.; Акт от 25.12.2020 г. Срок действия с 01.01.2021 по 31.12.2021 г. доступ: <http://dlib.eastview.com>

10. ООО «ИД «Гребенников», контракт № 147 от 23. 11.2020 г.; Акт от 25.12.2020 г. Срок действия с 01.01.2021 по 31.12.2021 г. доступ: <http://grebennikon.ru>

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование:

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
<p>Специальные помещения: Учебная аудитория для проведения занятий лекционного и семинарского типа, текущего контроля, промежуточной аттестации.</p>	<p>Аудитория оборудована специализированной учебной мебелью, техническими средствами обучения, служащими для представления информации большой аудитории: Ноутбук (AserAspirev3-5516 (AMDA10-4600M 2300 МГц)) (1 штука) с неограниченным доступом к сети Интернет, с неограниченным доступом к сети Интернет; Проектор Vivitek, экран ScreenVtdiaEcot- 3200*200MW 1:1, колонки, наборы демонстрационного оборудования и учебно-наглядных пособий, обеспечивающие тематические иллюстрации, соответствующие рабочей программе дисциплины.</p> <p>Учебная лаборатория: компьютеры для проведения практических работ (Системный блок AMDAthlon-64 X3 445 3100 МГц), Монитор LG F1742S (2 штуки), Монитор ViewSonic VA703b(24 штуки) с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации; проектор Sony XGA VPLSX535, экран ScreenVtdiaEcot- 3200*200MW 1:1</p>	<p>OC Windows: DreamSpark Premium, Договор № 03-016-14 от 30.10.2014</p> <p>Microsoft Office: 0365ProPiusOpenStudents ShrdSvr ALNG subs VL NL I MthAcdmsStdnt w/Faculty (15000 лицензий)</p> <p>Kaspersky Endpoint Security для бизнеса- стандартный Russian Edition. 1500-2499 Node 1 year Educational License № 1B08170221054045730177</p>
<p>Специальные помещения: компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), организации самостоятельной работы, в том числе, научно-исследовательской</p>	<p>Аудитория оборудована специализированной учебной мебелью, техническими средствами обучения: компьютеры (системный блок AMD Athlon 64 X2 DualCore 3600+ 1900 МГц (15 штук), Монитор LGFlatron L1742SE (14 штук), Монитор ViewSonic VG720) с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.</p>	<p>OC Windows: DreamSpark Premium, Договор № 03-016-14 от 30.10.2014</p> <p>Microsoft Office: 0365ProPiusOpenStudents ShrdSvr ALNG subs VL NL I MthAcdmsStdnt w/Faculty (15000 лицензий)</p> <p>Kaspersky Endpoint Security для бизнеса- стандартный Russian Edition. 1500-2499 Node 1 year Educational License № 1B08-170221054045730177</p>

6.2. Программное обеспечение:

№	Наименование Программного продукта	Кол-во	Обоснование для пользования ПО	Дата выдачи лицензии	Срок действия права пользования
1.	AdobeAcrobat XI Лицензия АЕ для акад.организаций Русская	12	11447921 Государственный	19.06.2013	бессрочно

	версия MultipleLicense RU (65195558)Platforms		контракт № 03-019-13		
2.	Microsoft Office Professional Plus 2007 Russian Academic OPEN No Level	25	Номер Лицензии Microsoft 46211164 Гос.контракт № 03-162-09 от 01.12.2009	01.12.2009	бессрочно
3.	Microsoft Office Professional Plus 2007 Russian Academic OPEN No Level	10	Номер Лицензии Microsoft 42095516	27.04.2007	бессрочно
4.	Microsoft SQL Server 2012	1	Номер Лицензии Microsoft 65343111		бессрочно
5.	Microsoft Windows Server 2008 r2 Enterprise	1	Номер Лицензии Microsoft 49413875		бессрочно
6.	Microsoft® Windows® Professional 7 Russian Upgrade Academic OPEN No Level Promo	12	Номер Лицензии Microsoft 46211164 Гос.контракт № 03-162-09 от 01.12.2009	01.12.2009	бессрочно
7.	Microsoft®WinSL 8.1 Russian Academic OLP 1License NoLevel Legalization GetGenuine	130	Microsoft Invoice Number: 9564547610 ООО 'ИЦ 'Сиброн'	22.12.2014	бессрочно
8.	OpenOffice 4.1.3	Условия правообладателя	Условия использования по ссылке: https://www.openoffice.org/licenses/PDL.html	Условия правообладателя	бессрочно
9.	Protege	100	Условия использования по ссылке: http://protege.stanford.edu/support.php	Условия правообладателя	бессрочно
10.	Python 3	Условия правообладателя	Условия использования по ссылке: https://docs.python.org/3/license.html	Условия правообладателя	бессрочно
11.	UbuntuLinux 16.04.1	Условия правообладателя	Условия использования по ссылке: https://www.ubuntu.com/legal/terms-and-policies/terms	Условия правообладателя	бессрочно
12.	VirtualBox 5.1	Условия правообладателя	Условия использования по ссылке: https://www.virtualbox.org/wiki/VirtualBox_PUEL	Условия правообладателя	бессрочно

6.3. Технические и электронные средства:

Методической концепцией преподавания предусмотрено использование технических и электронных средств обучения и контроля знаний студентов: мультимедийные презентации, фрагменты фильмов.

VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При реализации программы данной дисциплины используются различные образовательные технологии.

1.	Разноуровневое обучение	У преподавателя появляется возможность помогать слабому, уделять внимание сильному, реализуется желание сильных учащихся быстрее и глубже продвигаться в образовании. Сильные учащиеся утверждаются в своих способностях, слабые получают возможность испытывать учебный успех, повышается уровень мотивации ученья.
2.	Проектные методы обучения	Работа по данной методике дает возможность развивать индивидуальные творческие способности учащихся, более осознанно подходить к профессиональному и социальному самоопределению
3.	Исследовательские методы в обучении	Дает возможность учащимся самостоятельно пополнять свои знания, глубоко вникать в изучаемую проблему и предполагать пути ее решения, что важно при формировании мировоззрения. Это важно для определения индивидуальной траектории развития каждого обучающегося
4.	Лекционно-семинарско-зачетная система	Данная система дает возможность сконцентрировать материал в блоки и преподнести его как единое целое, а контроль проводить по предварительной подготовке обучающихся
5.	Информационно-коммуникационные технологии	Изменение и неограниченное обогащение содержания образования, использование интегрированных курсов, доступ в ИНТЕРНЕТ.
6.	Систему инновационной оценки «портфолио»	Формирование персонифицированного учета достижений обучающегося как инструмента педагогической поддержки социального самоопределения, определения траектории индивидуального развития личности

Наименование тем занятий с использованием активных форм обучения:

№	Тема занятия	Вид занятия	Форма / Методы интерактивного обучения	Кол-во часов
1	Все темы	Практика	Разбор типичных методов и решение конкретных задач, примеров и тестов. Рассматриваются также задачи с недостаточным количеством данных, необходимых для решения или наоборот с «избытком», из которого нужно отобрать самые важные для нахождения решения.	4
Итого часов				4

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

8.1. Оценочные средства текущего контроля

8.1.1. Оценочные средства для входного контроля (могут быть в виде тестов с закрытыми или открытыми вопросами).

Входной тест по анализу данных.

1 Математическое ожидание непрерывной, случайной величины можно вычислить по формуле:

а)
$$MX = \sum_{i=1}^n x_i p_i;$$

б)
$$MX = \sum_{i=1}^n x_i p_i^2;$$

в)
$$MX = \int_{-\infty}^{\infty} x f(x) dx;$$

г)
$$MX = \int_{-\infty}^{\infty} x^3 f(x) dx$$

2. Дисперсия константы C равна:

а) самой константе C ;

б) C^2 ;

в) нулю;

г) нельзя ответить определенно.

3. Математическое ожидание произведения случайной величины X на константу C равно:

а) $MC + MX$;

б) $C \cdot MX$;

в) $C^2 \cdot MX$;

г) $C \cdot [MX]^2$.

4. Математическое ожидание случайной величины, распределенной по равномерному закону равно:

а) $1/(b-a)$;

б) $(b+a)/2$;

в) $(a-b)^2/12$;

г) a .

5. Дисперсия случайной величины, распределенной по равномерному закону равна:

а) $1/(b-a)$;

б) $(b+a)/2$;

в) $(b-a)^2/12$;

г) $(a+b)^2/12$;

6. Выборочная средняя для выборки, заданной статистическим распределением

x_i	3	5	x_3	10
w_i	0,2	0,1	w_3	0,2

равна $\bar{x}_B = 7,1$. Значение x_3 равно _____

7. Выборочная дисперсия выборки, заданной статистическим распределением

x_i	1	2	5	6
n_i	2	1	6	1

, равна _____

8. Решить систему уравнений тремя способами (матричным, Крамера, Гаусса)

$$x - y + 2z = 1$$

$$\begin{aligned}x-2y-z &= 2 \\ 3x-y+5z &= 3\end{aligned}$$

8.1.2. Оценочные средства текущего контроля формируются в соответствии с ЛНА университета (могут быть в виде тестов, ситуационных задач, деловых и ролевых игр, диспутов, тренингов и др. Назначение оценочных средств ТК – выявить сформированность компетенций или их составляющих частей– указать каких конкретно).

Самостоятельная работа-1. Визуализация данных

Численность населения городов-миллионников России на 2002г.

город	Население, тыс. чел.
Волгоград	1013
Екатеринбург	1293
Казань	1105
Москва	10358
Нижний Новгород	1311
Новосибирск	1426
Омск	1134
Пермь	1000
Ростов-на-Дону	1070
Самара	1158
Санкт-Петербург	4669
Уфа	1042
Челябинск	1078

В ОТК были измерены диаметры 300 валиков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала даны в таблице.

Границы отклонений	Число валиков
-30 ... -25	3
-25 ... -20	8
-20 ... -15	15
-15 ... -10	35
-10 ... -5	40
-5 ... 0	60
0 ... 5	55
5 ... 10	30
10 ... 15	25
15 ... 20	14
20 ... 25	8
25 ... 30	7

Ориентировочное число посетителей 2-х сайтов в зависимости от времени.

Время суток	% посетителей	% посетителей
2-00	4	5
4-00	2	8

6-00	2	10
8-00	3	12
10-00	9	18
12-00	13	33
14-00	14	45
16-00	15	61
18-00	12	73
20-00	7	82
22-00	8	93
24-00	7	100

Самостоятельная работа-2. Векторы

- Построить графики в оболочке Python для иллюстрации скалярного и векторного произведения, подобные тем, что есть в Интернете.
- Привести коды.

Самостоятельная работа-3. Описательная статистика

Выполнить задание согласно Вашему варианту в MSExcel и на Python. В таблице представлены данные о годовых доходах и расходах на личное потребление (долл. США) для 30 североамериканских семей. Для выполнения работы берете данные из таблицы для 10 семей начиная с Вашего номера в группе. Если до конца таблицы данных не хватает, продолжаете брать их с начала таблицы. Например, 25-ый вариант берет данные с номерами 25-30 и 1-4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
годовой доход	2508	2572	2408	2522	2700	2531	2390	2595	2524	2685	2435	2354	2404	2381	2581
расходы на личное потребление	2406	2464	2336	2281	2641	2385	2297	2416	2001	2448	2311	2278	2240	2181	2408

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
годовой доход	2529	2562	2624	2407	2448	2690	2735	2222	2985	2634	2512	2408	2396	2864	2750
расходы на личное потребление	2379	2378	2554	2019	2356	2602	2652	2159	2905	2511	2379	2357	2309	2455	2698

1. Построить график
2. Найти средние доходы и расходы
3. Найти медианные расходы и 25% и 75% -квантили
4. Найти размах варьирования расходов, дисперсию и среднеквадратическое отклонение доходов
5. Найти ковариацию доходов и расходов и коэффициент их корреляции, сделайте выводы.
6. Найдите выбросы. Уберите выбросы и переделайте п. 1-5.

Самостоятельная работа-4. Линейная регрессия

месяц	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
-------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

январь	164	147	139	151	138	132	134	136	138	140	142	150	152	154	152	132	160
февраль	148	133	143	134	136	133	134	135	136	137	138	139	140	141	138	143	144
март	152	163	150	164	152	160	151	161	152	162	153	163	154	164	153	165	166
апрель	144	150	154	126	127	140	129	126	148	149	151	152	130	136	146	145	155
май	155	129	137	131	151	152	136	138	146	148	150	137	135	146	155	159	161
июнь	125	131	129	125	130	128	127	126	127	128	134	130	126	133	134	135	136
июль	153	145	128	127	119	120	118	122	129	131	144	142	143	150	151	153	155
август	146	137	140	143	153	138	139	141	142	144	150	147	148	149	150	155	160
сентябрь	138	138	143	143	140	137	137	139	139	140	145	141	141	142	142	149	151
октябрь	190	168	151	160	167	150	155	160	165	170	180	180	185	190	195	199	200
ноябрь	192	176	177	190	185	175	178	183	186	189	193	177	180	191	192	197	199
декабрь	192	188	184	182	185	181	185	183	189	187	193	182	186	188	193	199	202

Месячная продажа пива в Австралии

Для данных по вашему месяцу за 17 лет

построить модель линейной регрессии в MS Excel и на Python.

- Выписать вид модели
- Подсчитать ошибки аппроксимации
- Вычислить MSE, R², MAD, MAPE и F
- Дайте интерпретацию MAD, MAPE
- Построить графики
- Сделать прогноз на следующий год по линейной модели
- Сделать выводы

Самостоятельная работа-5. Метод градиентного спуска

- Методом градиентного спуска найдите минимум функции согласно Вашему варианту. Предоставьте скрипт.
- Что выдаст Ваш скрипт в случае, если у функции нет минимума?
- Что выдаст Ваш скрипт в случае, если у функции несколько локальных минимумов?

вариант	Функция
1	$f=x^2-2x+y^2-4y+z^2-6z$
2	$f=x^2+z^2+y^2-xy+x-2z$
3	$f=x^2+4y^2-2xy-x$
4	$f=x^2-x\sqrt{y+y}-6x+10$
5	$f=x^2-2y^2-2xy+x$
6	$f=(x^3-1)^4+(y-x)^2-6$
7	$f=x^2-3x+y^2-4y+z^2-5z$
8	$f=2x^2+2x+y^2+4y+z^2-6z$
9	$f=x^4-(x+y)^2+y^4$
10	$f=x^2-2xy+4y^3$

вариант	Функция
1	$f=\exp(-x^2-y^2)$
2	$f=(x-1)^3+(y^3-x)^2$
3	$f=x^2+4y^2-2xy-x$
4	$f=x^2-x\sqrt{y+y-6x+10}$
5	$f=x^2-2y^2-2xy+x$
6	$f=(x^3-1)^4+(y-x)^2-6$
7	$f=x^2-3x+y^2-4y+z^2-5z$
8	$f=2x^2+2x+y^2+4y+z^2-6z$
9	$f=x^4-(x+y)^2+y^4$
10	$f=x^2-2xy+4y^3$

Самостоятельная работа-6. Нахождение коэффициентов простой линейной регрессии методом градиентного спуска

за 17 лет

- Для данных по вашему месяцу найти коэффициенты линейной регрессии с помощью стохастического градиентного спуска на Python. Сравнить с найденными в самостоятельной работе-3.
- найти коэффициенты линейной регрессии с помощью spss. Сравнить с найденными в п.1.

Самостоятельная работа-7а*. Теория вероятностей

- У строгого проверяющего на проверку работы уходит t_1 мин, где t_1 распределено по нормальному распределению с параметрами $\mu_1 = 30$, $\sigma_1 = 10$. Добрый проверяющий тратит на проверку работы t_2 мин, где t_2 распределено по нормальному распределению с параметрами $\mu_2 = 20$, $\sigma_2 = 5$. Для каждой работы проверяющий выбирается случайным образом с вероятностью 0,5. Если известно, что на проверку работы ушло 10 мин, найти условную вероятность, что работа проверялась добрым проверяющим.
- Решить задачу аналитически и провести компьютерный эксперимент для проверки этого результата.

Самостоятельная работа-7б*. Теория вероятностей

Случайная величина ξ распределена по распределению Пуассона с параметром λ . Если $\xi = k$, проводятся k испытаний по схеме Бернулли с вероятностью успеха в одном испытании p . Пусть случайная величина η определяется как число успехов в схеме Бернулли. Докажите аналитически и с помощью компьютерной симуляции, что η имеет распределение Пуассона с параметром $p\lambda$.

Самостоятельная работа-8. Метод деления отрезка пополам

- Методом деления отрезка пополам найти нули функции согласно вашему варианту.
- Нарисуйте функцию

вариант	Функция	нули
1	$f=x^3+10x^2-27x-36$	-12, 3, -1
2	$f=x^3-16x^2+35x+52$	13, -1, 4
3	$f=\sin(x)$	0 и тд
4	$f=x^3+14x^2+35x+22$	-11, -1, -2
5	$f=\cos(x)$	
6	$f=\text{tg}(x)$	М.б. беск
7	$f=x^3-12.5x^2-31x+140$	14 2,5 -4
8	$f=x^3+16x^2-2.25x-36$	-3/2, 3/2, -16
9	$f=x^3-9x^2-82x-120$	15, -2, -4
10	$f=x/\sin(x)$	Беск в π , $-\pi$

Самостоятельная работа-9. Проверка статистических гипотез

Независимому статистику поручено проверить информацию маркетинговой службы некоторого туристического бюро о том, что $a\%$ клиентов выбирают в качестве

формы обслуживания полупансион. Статистик проводит опрос b случайно выбранных туристов. При каком количестве опрошенных, предпочитающих полупансион, он должен отклонить первоначальную гипотезу на уровне значимости критерия в $c\%$? Смоделируйте опрос b человек на Python, и ответьте на этот вопрос для односторонней и двусторонней критической области. Найдите уровень значимости для d человек, предпочитающих пансион и прокомментируйте его. Начертите данное биномиальное распределение и его аппроксимацию нормальным распределением на Python.

вариант	1	2	3	4	5	6	7	8	9	10
a	70	75	75	80	90	85	70	65	55	60
b	150	200	500	400	1000	300	600	700	800	550
c	5	5	10	10	5	10	10	2	3	15
d	84	85	86	189	950	240	402	420	488	402

Самостоятельная работа-10. Множественная регрессия

- Пусть к данным о продаже пива в Австралии добавились данные о затратах на рекламу этого пива в год.
- Построить модель множественной регрессии на Python и в SPSS для Ваших данных за 17 лет плюс затраты на рекламу текущего года и затраты на рекламу предыдущего года. Предоставить код.
- Записать уравнение регрессии, сделать прогноз на следующий год, нарисовать график предсказанных и истинных значений, выписать суммарную квадратическую ошибку и проанализировать другие статистики метода.
- Сделать выводы

год	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
затраты на рекламу в год	10	9	10	9	9,1	9,2	9,3	9,4	9,5	10	10	10	10	10	9	10	10

Самостоятельная работа-10а. Анализ коэффициентов множественной регрессии

- К работе 10 проверить статистические гипотезы о равенстве коэффициентов модели нулю, подсчитать доверительные интервалы для коэффициентов регрессии, сделать выводы.
- Убрать одну из независимых переменных и посмотреть, которая из двух моделей окажется лучше.
- Построить прогноз на следующий год.
- Выводы

Самостоятельная работа-11. Логистическая регрессия

На платежную дисциплину компаний по отношению к уплате налогов влияют два фактора: Отношение быстрых активов к текущим пассивам и Отношение размера прибыли к процентной ставке заимствования. Оба они измеряются по шкале от 300 до 900. Данные согласно Вашему варианту берете так: 8 компаний, начиная с Вашего номера варианта, из первой группы (из тех, что не платят налоги) и 10 компаний из второй группы (которые платят налоги). Если не хватает, начинаете брать по кругу с начала. Например, 7-й вариант берет из первой группы №№7,8,9,10,1,2,3,4 и из второй группы №№7,8,9,10,11,12,13,1,2,3.

	Компании, которые не платят налоги		Компании, которые платят налоги	
	Отношение активов и пассивов	Отношение прибыли к ставке заимствований	Отношение активов и пассивов	Отношение прибыли к ставке заимствований
1	740	680	750	590
2	670	600	360	600
3	560	550	720	750
4	540	520	540	710
5	590	540	570	700
6	590	700	520	670
7	470	600	590	790
8	560	540	670	700
9	540	630	620	730
10	500	600	690	840
11			610	680
12			550	730
13			590	750

- Создайте функцию, которая бы предсказывала, будет ли компания платить налоги на основе логистической регрессии.
- Определите качество подбора модели (точность и полноту прогнозов).
- Нарисуйте разделяющую гиперплоскость.
- Проверьте точность прогнозов для данных, которые вы не использовали для построения модели (два значения из первой группы и три из второй).

Самостоятельная работа-12. Кластеризация данных

Определить основной цвет картинки (Задача Алексея Быкова).

- Для этого: сгруппировать цвета картинки по трем кластерам. Определить самый многочисленный кластер и выдать средний цвет этого кластера.
- Картинка должна быть без фона.
- Определить основной цвет банана и зебры.

8.2. Оценочные средства для промежуточной аттестации

Вопросы для подготовки к зачету:

1. Математические методы визуализации данных.
2. Первичная статистическая обработка данных
3. Методы подгонки.
4. Критерии оценки модели.
5. Линейная регрессия.
6. Парная регрессия.
7. Логистическая регрессия.
8. Множественная регрессия.
9. Прогнозы..
10. Выбросы.
11. Фиктивные регрессоры.
12. Понятие градиента. Нахождение градиента.
13. Метод скорейшего спуска.
14. Метод деления отрезка пополам.

15. Проверка статистических гипотез.
16. Операции над векторами, матрицами.
17. Кластеризация. Метод K-средних. Выбор числа кластеров.

Итоговый тест по Анализу данных. Вариант 1.

1. Визуализация применяется для:
 - a. исследования данных,
 - b. передачи данных,
 - c. обнаружения связи между двумя парными выборками.
2. Для чего лучше подходят столбчатые диаграммы:
 - a. для исследования варьированности величины,
 - b. для исследования тенденций изменения величины,
 - c. для иллюстрации связи между данными.
3. Что такое скалярное произведение векторов?
4. Для чего аналитик данных может использовать векторы:
 - a. для представления данных по нескольким категориям для одного человека,
 - b. для изображения положения героя компьютерной игры,
 - c. для получения скалярного произведения списков.
5. Что такое среднее? Укажите достоинства и недостатки этой характеристики набора данных.
6. Что такое дисперсия? Для чего нужно знать дисперсию данных?
7. Почему нужно удалять выбросы?
8. Если X и Y сильно коррелированы, это может означать:
 - a. X влечет за собой Y,
 - b. Y влечет за собой X,
 - c. Они взаимообусловлены,
 - d. Они оба зависят от какого-то третьего фактора,
 - e. Ничего не значит.
9. Поясните фразу: Королем распределений является нормальное распределение.
10. Что определяет кумулятивная (интегральная) функция распределения?
11. В чем состоит метод подгонки (регрессии)?
12. Укажите критерии определения лучшей кривой подгонки, и какие из значений этих критериев лучше.
13. Когда при анализе данных лучше применять логарифмическую кривую?
14. Что такое фиктивные регрессоры?
15. Если модель плохо описывает исходные данные, от нее стоит ожидать плохих результатов при прогнозировании
 - a. Всегда
 - b. Только вне диапазона данных
 - c. Иногда
 - d. Почти никогда
 - e. Никогда
16. Поясните фразу: Начиная добавлять переменные (в модель множественной регрессии), следует позаботиться о том, чтобы они имели «смысл».
17. За что в модели отвечает случайная ошибка?
18. Какую величину, как и почему, используют вместо коэффициента детерминации при решении вопроса, нужно ли вводить дополнительные регрессоры в модель?
19. Исходя из приведенных 95%-ных интервалов и соответствующих уровней значимости, установить, какие регрессоры следует оставить в модели и почему.

$$\theta_0 \in [1094,20;1098,86], \theta_1 \in [1,08;5,69], \theta_2 \in [-2,01;2,61], \theta_3 \in [-2,44;2,17]$$

$$Sig_0=0,0001; Sig_1=0,015; Sig_2=0,61; Sig_3=0,68$$

Что изменится, если взять уровень значимости для коэффициентов регрессии в 90%?

20. Метод скорейшего спуска применяется для

- a. нахождения корней уравнения
- b. нахождения экстремумов функции
- c. нахождения значения аргумента по значению функции,
- d. нахождения значений коэффициентов регрессии.

21. Если вектор-градиент функции в точке меньше нуля, то он показывает направление убывания функции?

- a. Да.
- b. Нет.

Разработчики:



(подпись)

профессор

(занимаемая
должность)

О.А. Николайчук

(инициалы, фамилия)

Документ составлен в соответствии с требованиями ФГОС ВО по направлению 09.03.03 «Прикладная информатика», утвержденного приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. № 922, с учетом требований профессиональных стандартов 06.013 «Специалист по информационным ресурсам», 06.015 «Специалист по информационным системам» и 06.024 «Специалист по технической поддержке информационно-коммуникационных систем»

Программа рассмотрена на заседании кафедры прикладной информатики и документоведения «04» марта 2022 г.

Протокол № 8. Зав. кафедрой



А.В. Рохин

Настоящая программа, не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.