Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ КемГУ
Дата и время: 2025-04-23 00:00:00
471086fad29a3b30e244c728abc3661ab35e9d50310dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Кемеровский государственный университет»
Кузбасский гуманитарно-педагогический институт
федерального государственного бюджетного образовательного учреждения
высшего образования
«Кемеровский государственный университет»
Факультет информатики, математики и экономики

УТВЕРЖДАЮ
Декан А.В. Фомина
10 февраля 2022 г.

Рабочая программа дисциплины

Б1.В.ДВ.01.02 Прикладной анализ данных
Код, название дисциплины

Направление подготовки
09.03.03 Прикладная информатика
Код, название направления

Направленность (профиль) подготовки
Прикладная информатика в экономике

Программа бакалавриата

Квалификация выпускника
бакалавр

Форма обучения
Очная

Год набора 2021

Новокузнецк 2022

Оглавление

1 Цель дисциплины. 3
1.1 Формируемые компетенции 3
1.2 Индикаторы достижения компетенций 3
1.3 Знания, умения, навыки (ЗУВ) по дисциплине 4
2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации. 4
3. Учебно-тематический план и содержание дисциплины. 5
3.1 Учебно-тематический план 5
3.2. Содержание занятий по видам учебной работы 6
4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации. 8
5 Учебно-методическое обеспечение дисциплины. 9
5.1 Учебная литература 9
5.2 Программное и информационное обеспечение освоения дисциплины. ... 10
5.2.1 Программное обеспечение 10
5.3.2 Современные профессиональные базы данных и информационные справочные системы. 10
6 Иные сведения и (или) материалы. 11
6.1.Примерные темы письменных учебных работ..... 11
6.2. Примерные вопросы и задания / задачи для промежуточной аттестации 14

1 Цель дисциплины.

В результате освоения данной дисциплины у обучающегося должны быть сформированы компетенции основной профессиональной образовательной программы бакалавриата (далее - ОПОП):

ПК-1 Способен разрабатывать прототипы ИС на базе типовой ИС

Содержание компетенций как планируемых результатов обучения по дисциплине см. таблицы 1 и 2.

1.1 Формируемые компетенции

Таблица 1 - Формируемые дисциплиной компетенции

Наименование вида компетенции *(универсальная, общепрофессиональная, профессиональная)*	Наименование категории (группы) компетенций	Код и название компетенции
профессиональные		ПК-1 Способен разрабатывать прототипы ИС на базе типовой ИС

1.2 Индикаторы достижения компетенций

Таблица 2 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Дисциплины и практики, формирующие компетенцию ОПОП
ПК-1 Способен разрабатывать прототипы ИС на базе типовой ИС	ПК 1.1 Собирает и анализирует данные о запросах и потребностях заказчика применительно к типовой ИС ПК 1.2 Разрабатывает прототип ИС на базе типовой ИС ПК 1.3 Тестирует прототип ИС на корректность архитектурных решений ПК 1.4 Согласовывает пользователь-ский интерфейс и предполагаемые из-менения с заказчиком ПК 1.5 Осуществляет разработку, инсталляцию, модификацию и эксплуатацию алгоритмов и программного обеспечения СИИ	Б1.В.01 Информационные системы в экономике Б1.В.02 Экономика предприятия Б1.В.04 Корпоративные информационные системы Б1.В.05 Проектирование информационных систем Б1.В.08 Разработка программного обеспечения информационных систем Б1.В.ДВ.01.01 Исследование операций **Б1.В.ДВ.01.02 Прикладной анализ данных** Б1.В.ДВ.03.01 Интеллектуальные информационные системы Б1.В.ДВ.03.02 Системы искусственного интеллекта Б2.В.01(П) Производственная практика. Технологическая (проектно-технологическая) практика Б2.В.02(П) Производственная практика. Профильная практика Б3.01(Д) Выполнение и защита выпускной квалификационной работы

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Дисциплины и практики, формирующие компетенцию ОПОП
		ФТД.02 Разработка эконометрических моделей

1.3 Знания, умения, навыки (ЗУВ) по дисциплине

Таблица 3 – Знания, умения, навыки, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
ПК-1 Способен разрабатывать прототипы ИС на базе типовой ИС	ПК 1.1 Собирает и анализирует данные о запросах и потребностях заказчика применительно к типовой ИС	Знать: статистические методы анализа данных; классификацию статистических данных; методы анализа временных рядом; Уметь: проводить анализ статистических данных; использовать прикладные методы статистики, средства моделирования и программные средства для анализа данных и выявлении закономерностей экономических и социальных процессов; Владеть: навыками применения методов статистического анализа данных; навыками анализа временных рядов.

2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.

Таблица 4 – Объем и трудоемкость дисциплины по видам учебных занятий

Общая трудоемкость и виды учебной работы по дисциплине, проводимые в разных формах	Объём часовпо формам обучения
	ОФО
1 Общая трудоемкость дисциплины	108
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	
Аудиторная работа (всего):	36
в том числе:	
лекции	18
практические занятия, семинары	18
практикумы	
лабораторные работы	
в интерактивной форме	
в электронной форме	
Внеаудиторная работа (всего):	72
в том числе, индивидуальная работа обучающихся с преподавателем	
подготовка курсовой работы/контактная работа	
групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие групповую	

или индивидуальную работу обучающихся с преподавателем)	
творческая работа (эссе)	
3 Самостоятельная работа обучающихся (всего)	72
4 Промежуточная аттестация обучающегося и объём часов, выделенный на промежуточную аттестацию:	
4 семестр – зачет	

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 5 - Учебно-тематический план очнойформы обучения

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость *(всего час.)*	Трудоемкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости
			ОФО			
			Аудиторн. занятия		СРС	
			лекц.	практ.		
Семестр 4						
	1. Предварительный анализ данных					
	1.1 Классификация статистических данных	6	2	-	4	
	1.2 Анализ одномерных категориальных и количественных данных	6	-	2	4	Индивидуальное задание №1
	2. Генеральная и выборочная совокупности					
	2.1 Характеристики генеральной совокупности	3	1	-	2	
	2.2 Статистическое оценивание и проверка гипотез о параметрах генеральной совокупности	9	1	2	6	Тест №1
	2.3 Робастное оценивание параметров и непараметрические модели генеральной совокупности	12	2	2	8	Индивидуальное задание №2
	3. Анализ данных					
	3.1 Корреляционный анализ данных	6	1	1	4	
	3.2 Регрессионный анализ данных	6	1	1	4	Индивидуальное задание №3
	3.3 Снижение размерности признакового пространства	12	2	2	8	Тест №2
	4. Многомерные наблюдения					
	4.1 Кластерный анализ, непараметрическая классификация без обучения	12	2	2	8	
	4.2 Классификация с обучением.	12	2	2	8	

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость *(всего час.)*	Трудоемкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости
			ОФО			
			Аудиторн. занятия		СРС	
			лекц.	практ.		
	Дискриминантный анализ					
	4.3 Параметрическая классификация без обучения	12	2	2	8	Тест №3
	5. Анализ временных данных					
	5.1 Введение в анализ временных данных	6	1	-	2	
	5.2 Статистический анализ и применение адаптивных моделей	12	2	2	8	
	5.3 Модели авторегрессии	9	1	2	6	Тест №4
19	Промежуточная аттестация - зачет					
	Всего:	108	18	18	72	

3.2. Содержание занятий по видам учебной работы

Таблица 6 – Содержание дисциплины

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия
	Семестр 4	
Содержание лекционного курса		
1	*Предварительный анализ данных*	
1.1	Классификация статистических данных	*Критерии классификации данных. Классификация данных по числу переменных. Классификация данных по наличию или отсутствию упорядочения. Классификация данных по тпу шкалы измерения признака. Классификация данных по способу их получения*
2	*Генеральная и выборочная совокупности*	
2.1	Характеристики генеральной совокупности	*Распределение генеральной совокупности. Характеристики одномерной и многомерной генеральной совокупностей. Выборка из генеральной совокупности*
2.2	Статистическое оценивание и проверка гипотез о параметрах генеральной совокупности	*Статистическое оценивание параметров одномерных совокупностей. Оценки параметров многомерной генеральной совокупности. Статистическая проверка гипотез для одномерной совокупности. Статистическая проверка гипотез для многомерных генеральных совокупностей*
2.3	Робастное оценивание параметров и непараметрические модели генеральной совокупности	*Аномальные значения. Методы обнаружения засорения выборки Устойчивые параметрические методы оценивания. Оценки на основе порядковых статистик. Непараметрические модели распределений. Оценки методами бутстреп-анализа*
3	*Анализ данных*	
3.1	Корреляционный анализ	*Основные понятия корреляционного анализа. Корреляционный*

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия	
	данных	*анализ взаимосвязи количественных признаков. Корреляционный анализ взаимосвязи качественных признаков. Канонические корреляции и канонические величины генеральной совокупности. Оценка канонических корреляций и канонических величин*	
3.2	Регрессионный анализ данных	*Двумерная линейная модель регрессии. Множественная линейная модель регрессии. Нелинейные модели регрессии и их линеаризация*	
3.3	Снижение размерности признакового пространства	*Основные понятия и задачи снижения размерности. Компонентный анализ. Факторный анализ. Эвристические методы снижения размерности. Многомерное шкалирование*	
4	*Многомерные наблюдения*		
4.1	Кластерный анализ, непараметрическая классификация без обучения	*Основные понятия и определения кластерного анализа. Расстояние между объектами (кластерами) и меры близости групп объектов. Иерархические кластер-процедуры. Функционалы качества разбиения. Итерационные алгоритмы классификации. Метод k-средних. Иерархические алгоритмы, использующие понятие порога*	
4.2	Классификация с обучением. Дискриминантный анализ	*Функции потерь и вероятности неправильной классификации. Построение оптимальных (байесовских) процедур классификации.Параметрический дискриминантый анализ в случае нормальных классов*	
4.3	Параметрическая классификация без обучения	*Общая постановка задачи расщепления смеси вероятностных распределений и алгоритм ее выполнения. Пример параметрической модели классификации*	
5.	*Анализ временных данных*		
5.1	Введение в анализ временных данных	*Показатели динамики временны́х рядов. Прогнозирование с помощью показателей динамики. Методы сглаживания временны́х данных и моделирования тенденции развития*	
5.2	Статистический анализ и применение адаптивных моделей	*Статистический анализ и прогнозирование сезонных колебаний во временны́х данных. Применение адаптивных моделей, основанных на экспоненциальном сглаживании, для краткосрочного прогнозирования*	
5.3	Модели авторегрессии	*Модели стационарных временны́х рядов. Методология применения моделей ARIMA*	
Содержание практических занятий			5.3 М
1	*Предварительный анализ данных*		
1.2	Анализ одномерных категориальных и количественных данных	*Построение интервального вариационного ряда для непрерывных количественных данных. Нормирование (стандартизация) и унификация данных*	
2	*Генеральная и выборочная совокупности*		
2.1	Расчет характеристик генеральной совокупности	*Характеристики одномерной генеральной совокупности* *Характеристики многомерной генеральной совокупности* *Многомерная нормально распределенная генеральная совокупность*	

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия
2.2	Проверка гипотез о параметрах генеральной совокупности	*Статистическая проверка гипотез для одномерной совокупности. Статистическая проверка гипотез для многомерных генеральных совокупностей*
2.3	Устойчивые параметрические методы оценивания	*Проведение оценок генеральной совокупности*
3.	*Анализ данных*	
3.1	Проведение корреляционного анализа данных	*Корреляционный анализ взаимосвязи количественных признаков. Корреляционный анализ взаимосвязи качественных признаков*
3.2	Проведение регрессионного анализа данных	*Двумерная линейная модель регрессии. Множественная линейная модель регрессии.*
3.3	Компонентный и факторный анализ	*Компонентный анализ. Факторный анализ*
4.	*Многомерные наблюдения*	
4.1	Проведение кластерного анализа	*Иерархические кластер-процедуры*
4.2	Проведение дискриминантного анализа	*Параметрический дискриминантый анализ в случае нормальных классов*
4.3	Декомпозиция смесей вероятностных распределений	*Параметрическая модель классификации*
5.	*Анализ временных данных*	
5.2	Применение адаптивных моделей	*Применение адаптивных моделей, основанных на экспоненциальном сглаживании, для краткосрочного прогнозирования*
5.3	Использование моделей авторегрессии	*Модели стационарных временны́х рядов*
	Промежуточная аттестация - *зачет*	

4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.

Для положительной оценки по результатам освоения дисциплины обучающемусянеобходимо выполнить все установленные виды учебной работы. Оценка результатов работы обучающегося в баллах (по видам) приведена в таблице 7.

Пример заполнения таблицы

Таблица 7 - Балльно-рейтинговая оценка результатов учебной работы обучающихся по видам (БРС) 1 семестр

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации	Баллы (17 недель)
Текущая учебная работа в семестре (Посещение занятий по расписанию и	**60**	Лекционные занятия (конспект) (9 занятий)	**0,5 балла** посещение 1 лекционного занятия	5
		Лабораторные работы (отчет о выполнении	**2 балл** - посещение 1 практического занятия и выполнение работы на 51-65%	18 - 27

выполнение заданий)		лабораторной работы) (9 работ).	**3 балла** – посещение 1 занятия и существенный вклад на занятии в работу всей группы, самостоятельность и выполнение работы на 85,1-100%	
		Индивидуальные задания (отчет о выполнении) (3 работы)	**За одну ИД :** **1 балла** (выполнено 66 - 85% заданий) **5 баллов** (выполнено 86 - 100% заданий)	3 - 15
		Тесты (4 теста)	**За один:** **1 балла** (пороговое значение) **3 баллов** (максимальное значение)	4 - 12
Итого по текущей работе в семестре				31 - 60
Промежуточная аттестация (экзамен)	40 (100% /баллов приведенной шкалы)	Теоретический вопрос 1.	**3 балла** (пороговое значение) **10 баллов** (максимальное значение)	5 - 10
		Теоретический вопрос 2.	**3 балла** (пороговое значение) **5 баллов** (максимальное значение)	7 - 14
		Решение задачи 1.	**3 балла** (пороговое значение) **5 баллов** (максимальное значение)	8 - 16
Итого по промежуточной аттестации (экзамену)				(51 – 100% по приведенной шкале) 20 – 40 б.
Суммарная оценка по дисциплине:		Сумма баллов текущей и промежуточной аттестации		51 – 100 б.

5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.

5.1 Учебная литература

Основная учебная литература

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

Дополнительная учебная литература

Черткова, Е. А. Статистика. Автоматизация обработки информации : учебное пособие для вузов / Е. А. Черткова ; под общей редакцией Е. А. Чертковой. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 195 с. — (Высшее образование). — ISBN 978-5-534-01429-7.— URL: https://urait.ru/bcode/452447 . — Текст : электронный // ЭБС Юрайт [сайт].

Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0.— URL: https://urait.ru/bcode/450262. — Текст : электронный // ЭБС Юрайт [сайт].

5.2 Материально-техническое и программное обеспечение дисциплины.

Учебные занятия по дисциплине проводятся в учебных аудиториях НФИ КемГУ:

Таблица8 – Информационные технологии и программное обеспечение аудиторных занятий и самостоятельной работы

Наименование помещений для проведения всех видов учебной деятельности, предусмотренной учебным планом, в том числе помещения для самостоятельной работы	Перечень основного оборудования, учебно-наглядных пособий и используемого программного обеспечения	Адрес (местоположение) помещений для проведения всех видов учебной деятельности, предусмотренной учебным планом
508 Компьютерный класс. Учебная аудитория (мультимедийная) для проведения: - занятий лекционного типа; - занятий семинарского (практического) типа; - групповых и индивидуальных консультаций; - самостоятельной работы; - текущего контроля и промежуточной аттестации.	Специализированная (учебная) мебель: доска меловая, кафедра, столы, стулья. Оборудование для презентации учебного материала:стационарное - компьютер преподавателя, проектор, экран. Оборудование: стационарное – компьютеры для обучающихся (18 шт.). Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по сублицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО), R (свободно распространяемое ПО), MicrosoftVisualStudio (MicrosoftImaginePremium 3 year по сублицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), FoxitReader (свободно распространяемое ПО), Java (бесплатная версия), Eclipse(свободно распространяемое ПО). Интернет с обеспечением доступа в ЭИОС.	654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19

5.3 Современные профессиональные базы данных и информационные справочные системы.

Перечень СПБД и ИСС по дисциплине

1. CITForum.ru - on-line библиотека свободно доступных материалов по информационным технологиям на русском языке - http://citforum.ru

2. Научная электронная библиотека eLIBRARY.RU – крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты - www.elibrary.ru

3. Единое окно доступа к образовательным ресурсам - http://window.edu.ru/

6 Иные сведения и (или) материалы.

6.1.Примерные темы письменных учебных работ

Темы индивидуальных заданий

Индивидуальное задание №1

1. Определите в представленной таблице тип каждой переменой и данных в целом.

Число умерших человек в Российской Федерации за 2011—2013 гг.

Год	2011	2012	2013
Женщины	928 226	929 081	917 419
Мужчины	997 494	977 254	954 390

2. Определите тип каждой переменной — количественная (если да — дискретная или непрерывная) или категориальная (если да — номинальная или порядковая) и тип данных по упорядоченности во времени — пространственные, временные, пространственно-временные. Укажите шкалу измерений:

а) число смс-сообщений, посылаемых вами ежедневно;

б) динамика вашего веса и роста за всю жизнь;

в) вес и рост студентов вашей группы, измеренный на последней диспансеризации;

г) ежемесячный счет за пользование вашим мобильным телефоном;

д) номера студенческих групп в вашем вузе;

е) оценки за один из последних экзаменов студентов вашей группы по шкале «отлично — удовлетворительно»;

ж) ответы на вопросы анкеты о качестве продукта (понравился; скорее понравился; не могу определиться; скорее не понравился; не понравился);

з) число пар в расписании каждый день;

и) длина вашего прыжка с места на всех соответствующих занятиях физкультурой.

3. По данным обследования группы крови студентов некой студенческой группы, представленным в таблице, определите шкалу измерения признака, постройте частотную таблицу и дайте графическое представление в виде круговой диаграммы.

частотную таблицу и дайте графическое представление в виде круговой диаграммы.

№ п/п	Группа крови	№ п/п	Группа крови	№ п/п	Группа крови	№ п/п	Группа крови	№ п/п	Группа крови
1	I	6	II	11	II	16	I	21	II
2	II	7	I	12	II	17	I	22	III
3	I	8	I	13	IV	18	III	23	II
4	III	9	II	14	I	19	II	24	III
5	IV	10	III	15	II	20	I	25	I

4. По данным представленной таблицы наблюдений над непрерывным признаком, принимающим только неотрицательные значения, постройте с помощью формулы Стерджеса интервальный вариационный ряд распределения. Изобразите его графически в виде гистограммы, полигона и кумуляты относительных частот интервального ряда распределения.

8	37	79	45	17	3
15	17	33	62	41	97
89	25	77	39	74	99
61	56	38	87	79	72
43	91	87	23	41	41

5. По данным представленной таблицы 100 наблюдений выполните следующее. 1. Постройте с помощью формулы Стерджеса интервальный вариационный ряд распределения.

2. Вычислите основные числовые характеристики по исходным несгруппированным данным и по вариационному ряду:

а) среднюю арифметическую;

б) начальные и центральные моменты первого, второго, третьего и четвертого порядка;

в) дисперсию и среднее квадратическое отклонение;

г) коэффициенты асимметрии и эксцесса;

д) медиану и моду;

е) коэффициент вариации.

Сравните полученные по исходным несгруппированным данным и по вариационному ряду оценки, объясните различия.

3. Изобразите графически исходные данные в виде столбиковой диаграммы, постройте гистограмму, полигон и кумуляту частот вариационного ряда.

Объемы реализации продукции 100 текстильных предприятий за год

6,90	12,60	10,20	5,10	4,30	3,30	4,90	10,20	1,87	10,20
6,90	6,90	4,90	8,90	7,20	6,90	4,20	9,10	1,84	12,60
6,90	10,20	4,90	2,20	3,30	8,90	2,20	4,90	8,90	4,90
9,10	6,90	7,20	8,90	4,90	5,10	6,10	2,20	4,90	11,60
4,90	6,90	1,30	5,10	1,30	6,00	5,80	6,20	6,90	11,60
5,10	6,90	3,90	5,90	2,20	6,90	4,40	5,80	7,40	5,10
5,10	5,10	4,80	3,80	4,90	5,30	6,30	4,90	6,10	7,40
5,10	6,90	4,30	8,80	8,90	11,60	5,70	5,80	11,60	4,90
7,10	6,90	4,80	4,90	5,10	11,60	10,20	5,90	7,30	5,10
12,60	3,30	5,10	8,90	3,60	3,30	5,10	2,20	10,20	2,80

6. На основе представленного интервального вариационного ряда вычислите для изучаемого признака: а) среднюю арифметическую, дисперсию и среднее квадратическое отклонение; б) центральные моменты третьего и четвертого порядка; в) коэффициенты асимметрии и вариации; г) моду и медиану интервального ряда.

Средняя дневная выручка киоска, тыс. руб.	128—132	132—136	136—140	140—144	144—148
Число киосков	4	13	16	8	3

Индивидуальное задание №2

1. В процессе испытания четырех приборов не было ни одного отказа. Определите с доверительной вероятностью 0,95 границы интервальной оценки вероятности отказа прибора, если число отказов имеет биномиальное распределение.

2. При испытании зерна на всхожесть получено, что из 12 зерен первой партии взошло шесть, а из 100 зерен второй партии взошло 50. Определите с доверительной вероятностью 0,99 интервальную оценку вероятности прорастания одного зерна каждой партии, если число проросших зерен имеет биномиальное распределение.

3. При проверке двух предприятий розничной торговли установлено, что в одном магазине по данным 25 счетов среднее сальдо счета равно 90 тыс. руб., а среднее квадратическое отклонение равно 4 тыс. руб. В другом магазине по данным 17 счетов среднее сальдо счета равно 108 тыс. руб., а среднее квадратическое отклонение — 6 тыс. руб. Определите с доверительной вероятностью 0,95 границы интервальной оценки разности средних сальдо счетов двух магазинов.

4. Из 1500 семей города взята выборка объемом 300 семей и выяснено, что 20% семей намерены в предстоящем году приобрести автомобиль. Требуется с доверительной вероятностью 0,91 определить: а) интервальную оценку доли семей в городе, которые предполагают купить автомобиль; б) наибольшее число автомобилей, которые могут быть приобретены в городе.

5.Оценка Хьюбера для среднего значения предполагает, что наблюдения, превышающие пороговое значение k, принимаются равными k, а наблюдения, меньшие – k, берутся равными – k. Покажите, что это эквивалентно экспоненциальному убыванию веса наблюдения за пределами интервала (-k;k) от ближайшей его границы.

6. Представлены данные о средней производительности труда (выработка на

одного работающего) 10 предприятий подотрасли. Проверьте наличие грубых ошибок.

Номер предприятия	1	2	3	4	5	6	7	8	9	10
Выработка	12	11	13	12	14	12	18	15	14	13

Индивидуальное задание №3

1.В таблице приведены данные по 14 машиностроительным предприятиям. По приведенным данным проведите регрессионный анализ зависимости индекса снижения себестоимости продукции от трудоемкости и удельного веса покупных изделий.

№ п/п	y	x_1	x_2	№ п/п	y	x_1	x_2
1	12,2	4795	69	7	12,7	4888	55
2	7,6	6962	82	8	10,5	6237	81
3	10,	6571	87	9	15,1	2997	65
4	49,9	4249	92	10	10,6	2990	98
5	15,7	9540	23	11	15,2	1748	100
6	14,0	3488	31	12	17,2	2128	69

2. Докажите, что вектор b, полученный методом наименьших квадратов, является методом наименьших квадратов вектора генеральных коэффициентов регрессии β.

3. Докажите, что выражение $b=(X^TX)^{-1}X^TY$ есть оценка методом наименьших квадратов вектора генеральных коэффициентов регрессии β.

6.2. Примерные вопросы и задания / задачи для промежуточной аттестации

Семестр 4

Таблица 9- Примерные теоретические вопросы и практические задания к экзамену

Разделы и темы	Примерные теоретические вопросы	Примерные практические задания
1. *Предварительный анализ данных*		
1.1 Классификация статистических данных	1. Критерии классификации данных. 2. Классификация данных по числу переменных. 3. Классификация данных по наличию или отсутствию упорядочения. 4. Классификация данных по типу шкалы измерения признака. 5. Классификация данных по способу их получения	Определите тип каждой переменной — количественная (если да — дискретная или непрерывная) или категориальная (если да — номинальная или порядковая) и тип данных по упорядоченности во времени — пространственные, временные, пространственно-временные.

1.2 Анализ одномерных категориальных и количественных данных	6. Построение интервального вариационного ряда для непрерывных количественных данных. 7. Нормирование (стандартизация) и унификация данных	На основе представленного интервального вариационного ряда вычислите для изучаемого признака: а) среднюю арифметическую, дисперсию и среднее квадратическое отклонение; б) центральные моменты третьего и четвертого порядка; в) коэффициенты асимметрии и вариации; г) моду и медиану интервального ряда.
2. Классификация прикладного программного обеспечения и информационных систем		
2.1 Характеристики генеральной совокупности	8. Распределение генеральной совокупности. 9. Характеристики одномерной и многомерной генеральной совокупностей. 10. Выборка из генеральной совокупности	
2.2 Статистическое оценивание и проверка гипотез о параметрах генеральной совокупности	11. Статистическое оценивание параметров одномерных совокупностей. 12. Оценки параметров многомерной генеральной совокупности. 13. Статистическая проверка гипотез для одномерной совокупности. 14. Статистическая проверка гипотез для многомерных генеральных совокупностей	При испытании зерна на всхожесть получено, что из 12 зерен первой партии взошло шесть, а из 100 зерен второй партии взошло 50. Определите с доверительной вероятностью 0,99 интервальную оценку вероятности прорастания одного зерна каждой партии, если число проросших зерен имеет биномиальное распределение.
2.3 Робастное оценивание параметров и непараметрические модели генеральной совокупности	15. Методы обнаружения засорения выборки 16. Устойчивые параметрические методы оценивания. 17. Оценки на основе порядковых статистик.	
3. Анализ данных		
3.1 Корреляционный анализ данных	18. Основные понятия корреляционного анализа. 19. Корреляционный анализ взаимосвязи количественных 20. признаков. Корреляционный анализ взаимосвязи	

	качественных признаков. 21. Канонические корреляции и канонические величины генеральной совокупности. 22. Оценка канонических корреляций и канонических величин	

3.2 Регрессионный анализ данных	23. Двумерная линейная модель регрессии. 24. Множественная линейная модель регрессии. 25. Нелинейные модели регрессии и их линеаризация	Докажите, что выражение $b=(X^TX)^{-1}X^TY$ есть оценка методом наименьших квадратов вектора генеральных коэффициентов регрессии β.
3.3 Снижение размерности признакового пространства	26. Основные понятия и задачи снижения размерности. 27. Компонентный анализ. 28. Факторный анализ. 29. Эвристические методы снижения размерности. 30. Многомерное шкалирование	По данным 15 предприятий отрасли рассчитан коэффициент корреляции между себестоимостью продукции x(1) и производительностью труда x(2), который оказался равен -0.7. Построить матрицу собственных значений и матрицу факторных нагрузок. Определить дисперсию и относительный вклад первой главной компоненты в обобщенную дисперсию исходных признаков.
4. Многомерные наблюдения		
4.1 Кластерный анализ, непараметрическая классификация без обучения	31. Основные понятия и определения кластерного анализа. 32. Расстояние между объектами (кластерами) и меры близости групп объектов. 33. Иерархические кластер-процедуры.	
4.2 Классификация с обучением. Дискриминантный анализ	34. Функции потерь и вероятности неправильной классификации 35. Параметрический дискриминантый анализ в случае нормальных классов	
4.3 Параметрическая классификация без обучения	36. Общая постановка задачи расщепления смеси вероятностных распределений и алгоритм ее выполнения.	
5. Анализ временных данных		
Применение адаптивных моделей	37. Показатели динамики временны́х рядов. 38. Прогнозирование с помощью показателей динамики. 39. Методы сглаживания временны́х данных и моделирования тенденции развития	
Использование моделей авторегрессии	40. Статистический анализ и прогнозирование	

	сезонных колебаний во временны́х данных. 41. Применение адаптивных моделей, основанных на экспоненциальном сглаживании, для краткосрочного прогнозирования	
Модели авторегрессии	42. Модели стационарных временны́х рядов. 43. Методология применения моделей ARIMA	

Составитель (и): Штейнбрехер О.А., канд. техн. наук, доцент кафедры информатики и вычислительной техники им. В.К. Буторина