

МИНОБРНАУКИ РОССИИ
ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ЭКОНОМИКИ И СЕРВИСА
КАФЕДРА МАТЕМАТИКИ И МОДЕЛИРОВАНИЯ

Рабочая программа дисциплины (модуля)
КОМПЬЮТЕРНЫЙ АНАЛИЗ ДАННЫХ

Направление и направленность (профиль)

38.03.05 Бизнес-информатика. Бизнес-аналитика

Год набора на ОПОП
2020

Форма обучения
очная

Владивосток 2021

Рабочая программа дисциплины (модуля) «Компьютерный анализ данных» составлена в соответствии с требованиями ФГОС ВО по направлению(ям) подготовки 38.03.05 Бизнес-информатика (утв. приказом Минобрнауки России от 11.08.2016г. №1002) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 05.04.2017 г. N301).

Составитель(и):

Мартышенко С.Н., кандидат технических наук, профессор, Кафедра математики и моделирования, sergey.martishenko@vvsu.ru

Утверждена на заседании кафедры математики и моделирования от 14.04.2021 , протокол № 10

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Мазелис Л.С.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575656200
Номер транзакции	000000000745D5F
Владелец	Мазелис Л.С.

Заведующий кафедрой (выпускающей)

Мазелис Л.С.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575656200
Номер транзакции	000000000745D60
Владелец	Мазелис Л.С.

1. Цель и задачи освоения дисциплины (модуля)

Целью освоения дисциплины «Компьютерный анализ данных» является формирование теоретических и практических компетенций в области целостного представления, понимания места и роли, а также применения методов статистического анализа данных, а также обучение студентов современным программным средствам, в которых реализованы модули, осуществляющие решение задач анализа данных.

Задачи дисциплины:

- дать бакалаврам теоретические основы по спектру наиболее распространенных статистических методов анализа данных и условий их применения;
- дать основы количественных методов оценки адекватности и точности построенных моделей;
- привить навыки и умения практического применения компьютерных технологий при анализе и прогнозировании социально-экономических показателей (построение линейных и нелинейных моделей прогнозирования на основе регрессионного анализа, оценка их параметров, расчёт всех необходимых статистик для анализа моделей);
- изучение концепции и технологии современного анализа данных на компьютере;
- изучение принципов работы программных средств, предназначенных для статистического анализа данных;
- изучение современных визуальных методов анализа данных и использования их для статистического вывода и формулировки гипотез о структуре данных;
- выработка умения самостоятельного решения задач по выбору методов анализа в практических ситуациях;
- получение навыков применения программных систем; предназначенных для статистического анализа данных, а также тестировании программных модулей на модельных данных;
- изучение рынков программного обеспечения по анализу данных;
- обеспечить бакалаврам прочное и осознанное желание владеть системой компьютерный анализ данных, способствующей их профессиональной успешности и востребованности на рынке труда.

2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Планируемыми результатами обучения по дисциплине являются знания, умения, навыки, соотнесенные с компетенциями, которые формирует дисциплина, и обеспечивающие достижение планируемых результатов по образовательной программе в целом. Перечень компетенций, формируемых в результате изучения дисциплины (модуля), приведен в таблице 1.

Таблица 1 – Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код компетенции	Формулировка компетенции	Планируемые результаты обучения	
38.03.05 «Бизнес-информатика» (Б-БИ)	ПК-18	Способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации	Знания:	возможностей инструментальных сред анализа и извлечения знаний
			Умения:	применять методы анализа данных для исследования деятельности организаций и в других практических ситуациях

		информации по теме исследования	Навыки:	владения методами группирования, классификации данных
	ОПК-3	Способность работать с компьютером как средством управления информацией, работать с информацией из различных источников, в том числе в глобальных компьютерных сетях	Знания:	возможности наиболее распространенных онлайн-сервисов по сбору данных в глобальных компьютерных сетях
			Знания:	основные принципы работы с онлайн-сервисами по сбору данных в глобальных компьютерных сетях
			Умения:	работать с большими объемами информации
			Умения:	создавать аналитические отчеты, на основе собранных данных в глобальных компьютерных сетях
			Умения:	создавать анкетные формы в среде Google различной сложности
			Навыки:	организовывать опросы в сети, с использованием коммуникаций в социальных сетях
			Навыки:	работать с большими объемами реальных данных

3. Место дисциплины (модуля) в структуре основной образовательной программы

Дисциплина «Компьютерный анализ данных» относится к базовой части «Блока 1 Дисциплины (модули)» учебного плана направления 38.03.05 «Бизнес-информатика».

Входными требованиями, необходимыми для освоения дисциплины, является наличие у обучающихся компетенций, сформированных при изучении дисциплин и/или прохождении практик «Информатика и программирование модуль 1», «Теория вероятностей и математическая статистика». На данную дисциплину опираются «Бизнес-аналитика», «Методы финансового анализа», «Многомерные статистические методы», «Эконометрика».

4. Объем дисциплины (модуля)

Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудоемкость (З.Е.)	Объем контактной работы (час)					СРС	Форма аттестации	
					Всего	Аудиторная			Внеаудиторная			
						лек.	прак.	лаб.	ПА			КСР
38.03.05 Бизнес-информатика	ОФО	Бл1.Б	4	4	73	0	72	0	1	0	71	ДЗ

5. Структура и содержание дисциплины (модуля)

5.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Кол-во часов, отведенное на				Форма текущего контроля
		Лек	Практ	Лаб	СРС	
1	Многомерные выборки. Предварительный анализ многомерных данных	0	6	0	7	теоретический опрос, отчет по практической работе №1
2	Методы моделирования случайных величин	0	8	0	7	теоретический опрос, отчет по практической работе №2
3	Робастное статистическое оценивание	0	6	0	7	теоретический опрос, отчет по практической работе №3
4	Методы статистического оценивания и сравнения выборок	0	8	0	8	теоретический опрос, отчет по практической работе №4, индивидуальное домашнее задание
5	Непараметрические методы проверки однородности выборок	0	8	0	8	теоретический опрос, отчет по практической работе №5
6	Дисперсионный анализ	0	8	0	8	теоретический опрос, отчет по практической работе №6
7	Методы обработки ранговых данных	0	8	0	8	теоретический опрос, отчет по практической работе №7
8	Компонентный анализ	0	10	0	8	теоретический опрос, отчет по практической работе №8
9	Методы многомерной классификация данных	0	10	0	10	теоретический опрос, отчет по практической работе №9, контрольная работа
Итого по таблице		0	72	0	71	

5.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Многомерные выборки. Предварительный анализ многомерных данных.

Содержание темы: Предмет, цели и задачи дисциплины Анализ данных. Шкалы измерений. Преобразование признаков, измеренных в разных шкалах. Характеристики признаков, рассчитываемые в предварительном анализе. Графическое представление данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 2 Методы моделирования случайных величин.

Содержание темы: Метод неравномерной рулетки. Метод отбраковки. Быстрый способ моделирования одномерного нормального распределения. Моделирование многомерного нормального распределения. Способы генерации данных в EXCEL. Методы размножения выборок (бустреб-методы).

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 3 Робастное статистическое оценивание.

Содержание темы: Грубые ошибки и методы их выявления. Методы вычисления устойчивых статистических оценок: Пуанкаре, Винзора, Хубера.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 4 Методы статистического оценивания и сравнения выборок.

Содержание темы: Статистические гипотезы в анализе данных. Подгонка вероятностных распределений к реальным данным. Проверка гипотез о равенстве вектора средних значений постоянному вектору. Проверка гипотез о равенстве двух векторов средних. Проверка гипотез о равенстве ковариационных матриц.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 5 Непараметрические методы проверки однородности выборок.

Содержание темы: U-критерий Манна-Уитни (Вилкоксона).

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе, выполнение ИДЗ.

Тема 6 Дисперсионный анализ.

Содержание темы: Математическая модель. Формы представления данных. Составляющие дисперсии. Проверка гипотезы с помощью F-критерия.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 7 Методы обработки ранговых данных.

Содержание темы: Измерение связи между двумя дихотомными переменными. Коэффициент ассоциации Юла. Ранговая корреляция. Критерий знаков.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 8 Компонентный анализ.

Содержание темы: Линейная модель главных компонент. Получение матрицы весовых коэффициентов по алгоритму метода главных компонент. Квадратичные формы и главные компоненты.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе.

Тема 9 Методы многомерной классификация данных.

Содержание темы: Классификация без обучения. Расстояние между объектами и мера близости. Расстояние между кластерами. Функционалы качества разбиения. Иерархические кластер процедуры. Метод к-средних.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к теоретическому опросу, подготовка к практическому занятию, подготовка отчета по практической работе, подготовка к итоговой контрольной работе.

6. Методические указания по организации изучения дисциплины (модуля)

Для обеспечения систематической и регулярной работы по изучению дисциплины и успешного прохождения текущих и промежуточных контрольных испытаний студенту рекомендуется придерживаться следующего порядка обучения:

- самостоятельно определить объем времени, необходимого для проработки каждой темы;
- регулярно изучать каждую тему дисциплины, используя различные формы индивидуальной работы;
- согласовывать с преподавателем виды работы по изучению дисциплины.

Целью самостоятельной работы студентов является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности.

Самостоятельная работа студента включает следующие виды, выполняемые в соответствии с ФГОС ВО и рабочим учебным планом:

- аудиторная самостоятельная работа студента под руководством и контролем преподавателя на практических занятиях. При проведении практических занятиях применяется «Метод кооперативного обучения»: студенты работают в малых группах (3 – 4 чел.) над индивидуальными заданиями, в процессе выполнения которых они могут совещаться друг к другу. Преподаватель, в свою очередь, наблюдает за работой малых групп, а также поочередно разъясняет новый учебный материал малым группам, которые закончили работать над индивидуальными заданиями по предыдущему материалу;
- внеаудиторная самостоятельная работа студента под руководством и контролем преподавателя: изучение теоретического материала, подготовка к практическим занятиям.

Контроль успеваемости осуществляется в соответствии с рейтинговой системой оценки знаний студентов. Оценка по дисциплине определяется по 100-бальной шкале как сумма баллов, набранных студентом в результате работы в семестре. Распределение баллов доводится до студентов в начале семестра.

Баллами оценивается посещение занятий, теоретические опросы, отчет по практической работе, выполнение индивидуального задания, итоговая контрольная работа.

Основной формой промежуточного контроля уровня подготовки студентов является дифференцированный зачет, который проводится в форме контрольной работы.

Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов.

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

7. Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

8. Учебно-методическое и информационное обеспечение дисциплины (модуля)

8.1 Основная литература

1. Козлов Андрей Юрьевич. Статистический анализ данных в MS Excel : Учебник [Электронный ресурс] : ИНФРА-М , 2016 - 320 - Режим доступа: <http://znanium.com/go.php?id=558444>

2. Кулаичев А.П. Методы и средства комплексного статистического анализа данных : Учебное пособие [Электронный ресурс] : Инфра-М , 2018 - 484 - Режим доступа: <https://znanium.com/catalog/document?id=320726>

3. Миркин Б. Г. ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ. Учебник и практикум [Электронный ресурс] , 2020 - 174 - Режим доступа: <https://urait.ru/book/vvedenie-v-analiz-dannyh-450262>

8.2 Дополнительная литература

1. Анализ данных качественных исследований [Электронный ресурс] , 2018 - 108 - Режим доступа: <https://lib.rucont.ru/efd/705202>

2. Ниворожкина Людмила Ивановна. Многомерные статистические методы в экономике : Учебник [Электронный ресурс] : РИОР , 2017 - 203 - Режим доступа: <http://znanium.com/go.php?id=615064>

8.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Единое окно доступа к информационным ресурсам. Математика и математическое образование (<http://window.edu.ru>)

2. Математический форум Math Help Planet (<http://mathhelpplanet.com/static.php>)

3. СПС КонсультантПлюс - Режим доступа: <http://www.consultant.ru/>

4. ЭБС Консультант студента (www.studentlibrary.ru)

5. Электронная библиотечная система «РУКОНТ» - Режим доступа: <https://lib.rucont.ru/>

6. Электронная библиотечная система ZNANIUM.COM - Режим доступа: <http://znanium.com/>

7. Электронная библиотечная система ZNANIUM.COM - Режим доступа: <https://znanium.com/>

8. Электронно-библиотечная система издательства "Юрайт" - Режим доступа: <https://urait.ru/>

9. Open Academic Journals Index (ОАИ). Профессиональная база данных - Режим доступа: <http://oaji.net/>

10. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: <https://www.prlib.ru/>

9. Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

Основное оборудование:

- Ист.бесп.эл.питания Smart-UPS 3000VA
- Мультипроектор №1 Panasonic PT-LX26HE
- Облачный монитор 23" LG CAV42K
- Облачный монитор LG Electronics черный +клавиатура+мышь
- Сетевой монитор:Нулевой клиент Samsung SyncMaster NC240

Программное обеспечение:

- Microsoft Office Professional Plus 2013 Russian
- Windows Office 2007 Suites Russian

10. Словарь основных терминов

Выбросы - это нетипичные или редкие значения, которые существенно отклоняются от распределения остальных выборочных данных. Эти данные могут отражать истинные свойства изучаемого явления (переменной), а могут быть связаны с ошибками измерения или аномальными явлениями, и поэтому не должны включаться в модель.

Группировка - это процедура, позволяющая вычислить описательные статистики и корреляции для зависимых переменных, наблюдения которых разбиты на группы, определяемые одной (или более) группирующей (независимой) переменной. Эта процедура используется в проверке гипотез или в разведочных методах.

Дискриминантный анализ используется для принятия решения о том, какие переменные дискриминируют или разделяют объекты на две или более естественно возникающих групп (его используют как метод проверки гипотез или как метод разведочного анализа).

Дисперсионного анализа (ANOVA) является проверка значимости различия между средними с помощью сравнения (т.е. анализа) дисперсий. А именно, разделение общей дисперсии на несколько источников (связанных с различными эффектами в плане), позволяет сравнить дисперсию, вызванную различием между группами, с дисперсией, с дисперсией, вызванной внутригрупповой изменчивостью.

Доверительные интервалы для некоторой статистики (например, среднего значения или линии регрессии) показывают диапазон вокруг значения статистики, в котором находится истинное значение этой статистики (с определенным уровнем надежности или доверия).

Интервальная шкала - эта шкала измерений позволяет не только упорядочить наблюдения, но и количественно выразить расстояния между ними (при этом на шкале не обязательно присутствует абсолютная нулевая отметка).

Кластерный анализ (впервые ввел Тьюн, 1939) в действительности включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры, т.е. развернуть таксономии (обычно в [разведочном анализе](#)) или определить кластеры схожих объектов

Коэффициент детерминации - квадрат множественного коэффициента корреляции. Он показывает, какая доля дисперсии резульативного признака объясняется влиянием независимых переменных.

Критерий Вилкоксона является непараметрической альтернативой [t-критерию для зависимых выборок](#). Он осуществляет проверку гипотезы о параметре положения (медианы) выборочного распределения. Как правило, такой критерий применяется при сравнении совпадающих пар данных, например, "до" и "после". В этом случае критерий осуществляет сравнения разности медиан с нулем

Порядковая шкала измерений позволяет ранжировать значения переменных. Измерения в порядковой шкале содержат информацию только о порядке следования величин, но не позволяют сказать "насколько одна величина больше другой", или "насколько она меньше другой".

Расстояние Махаланобиса определяется как расстояние от наблюдаемой точки до центра тяжести в многомерном пространстве, определяемом [коррелированными](#) (неортогональными) независимыми переменными (если независимые переменные некоррелированы, расстояние Махаланобиса совпадает с обычным [евклидовым расстоянием](#)). Эта мера позволяет, в частности, определить является ли данное наблюдение [выбросом](#) по отношению к остальным значениям независимых переменных

t-критерий для зависимых выборок. t-критерий для зависимых выборок очень полезен в тех довольно часто возникающих на практике ситуациях, когда важный источник внутригрупповой вариации (или ошибки) может быть легко определен и исключен из анализа.

t-критерий для одной выборки. В t-критерии для одной выборки, наблюдаемое среднее (из одной выборки) ставняется с ожидаемым средним популяции (например, некое теоретическое среднее), а вариация в популяции подсчитывается на основе вариации в наблюдаемой выборке.