

МИНОБРНАУКИ РОССИИ
ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ЭКОНОМИКИ И СЕРВИСА

КАФЕДРА МАТЕМАТИКИ И МОДЕЛИРОВАНИЯ

Рабочая программа дисциплины (модуля)
МНОГОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ

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) «Многомерные статистические методы» 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	0000000000745D62
Владелец	Мазелис Л.С.

Заведующий кафедрой (выпускающей)

Мазелис Л.С.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575656200
Номер транзакции	0000000000745D63
Владелец	Мазелис Л.С.

1. Цель и задачи освоения дисциплины (модуля)

Целью изучения дисциплины «Многомерные статистические методы» является формирование у студентов фундаментальных теоретических знаний по вопросам методики и практики применения методов многомерного статистического анализа данных, а также обучение студентов современным программным средствам в которых реализованы модули, осуществляющие решение задач многомерного анализа.

В ходе изучения курса у студента должно формироваться представление о конкретных практических ситуациях, в которых необходимо использование методов многомерного статистического анализа.

Задачи дисциплины:

- изучение теоретических основ по спектру наиболее распространенных статистических методов анализа данных и условий их применения;
- изучение концепции и технологии современного анализа данных на компьютере;
- изучение принципов работы программных средств, предназначенных для многомерного 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 многомерного статистического анализа данных, а также тестировании программных модулей на модельных данных;
- изучение рынков программного обеспечения по анализу данных.

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-17	Способность использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования	Знания:	1) основных принципов, методов и результатов современных многомерных статистических методов; 2) общности понятий и представлений многомерных статистических методов с другими математическими и экономическими дисциплинами
			Умения:	проводить факторный анализ, используя метод главных компонент и метод максимального правдоподобия, оценивать значимость построенной факторной модели

			Навыки:	1) владения основными аналитическими приемами многомерного и статистического анализа; 2) владения методиками проведения вероятностных расчетов, навыками расчета основных характеристик, возникающих при проведении вероятностного анализа в практических задачах;
	ПК-18	Способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования	Знания:	основных свойств и характеристик многомерных случайных величин, идеологии методов многомерного анализа, назначение и структуру соответствующих модулей ППП STATISTICA.
			Умения:	ставить задачи в области прикладного многомерного анализа экономических данных и применять многомерные статистические методы для их решения.
			Навыки:	владения технологией работы с соответствующими модулями в ППП STATISTICA
	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 «Многомерные статистические методы» относится к базовой части «Блока 1 Дисциплины (модули)» учебного плана направления «Бизнес-информатика».

Входными требованиями, необходимыми для освоения дисциплины, является наличие у обучающихся компетенций, сформированных при изучении дисциплин и/или прохождении практик «Компьютерный анализ данных», «Математический анализ модуль 1», «Математический анализ модуль 2», «Теория вероятностей и математическая статистика». На данную дисциплину опираются «Методы и модели стратегического управления», «Моделирование процессов и систем».

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 обуче- ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо- емкость	Объем контактной работы (час)						СРС	Форма аттес- тации
				(З.Е.)	Всего	Аудиторная			Внеауди- торная			
						лек.	прак.	лаб.	ПА	КСР		
38.03.05 Бизнес- информатика	ОФО	Бл1.Б	6	3	37	0	36	0	1	0	71	3

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	Прак	Лаб	СРС	
1	Методы моделирование данных	0	4	0	8	теоретический опрос, отчет по практической работе №1, индивидуальное домашнее задание
2	Преобразование данных, представленных в различных шкалах измерения. Предварительный анализ многомерных данных	0	4	0	8	теоретический опрос, отчет по практической работе №2, индивидуальное домашнее задание
3	Методы анализа достоверности данных и исправление ошибок	0	6	0	12	теоретический опрос, отчет по практической работе №3
4	Методы проверки статистических гипотез	0	4	0	8	теоретический опрос, отчет по практической работе №4, индивидуальное домашнее задание
5	Кластерный анализ непрерывных признаков	0	6	0	12	теоретический опрос, отчет по практической работе №5
6	Кластерный анализ бинарных признаков	0	6	0	12	теоретический опрос, отчет по практической работе №6
7	Дискриминантный анализ	0	6	0	11	теоретический опрос, отчет по практической работе №7
Итого по таблице		0	36	0	71	

5.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Методы моделирование данных.

Содержание темы: Генерация случайных чисел. Свойства датчика случайных чисел Excel. Преобразование случайных чисел к различным диапазонам. Генерация выборки нормального закона путем преобразования случайных чисел. Генерация данных различных законов распределений средствами Excel. Генерация случайных выборок методом неравномерной рулетки. Генерация данных методом отбраковки. Генерация многомерного нормального распределения. Средства визуализации данных при генерации данных случайных выборок (трехмерные графики).

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.

Тема 2 Преобразование данных, представленных в различных шкалах измерения. Предварительный анализ многомерных данных.

Содержание темы: Преобразование данных, представленных в различных шкалах измерения. Определение диапазона значений признаков. Расчет характеристик выборки в предварительном анализе многомерных данных. Средства визуального анализа в предварительном анализе.

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 Методы анализа достоверности данных и исправление ошибок.

Содержание темы: Предварительный анализ данных анкетных опросов. Оценка интервьюеров. Анализ многомерной выборки на отсутствие данных. Выделение выбросов по многомерной выборке непрерывных признаков. Выделение выбросов по многомерной выборке бинарных признаков. Обобщенный способ выделения выбросов по многомерным данным. Средства визуального анализа достоверности данных и повышения достоверности 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 4 Методы проверки статистических гипотез.

Содержание темы: Проверка согласование данных выборки конкретному закону распределения по критерию. Оценка зависимости признаков с помощью критерия. Средства визуального анализа, используемые в задачах. Дисперсионный анализ. Нелинейная регрессия. Оценка составляющих дисперсии. Использование критерия Фишера для оценки качества регрессии. Средства визуального анализа 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 5 Кластерный анализ непрерывных признаков.

Содержание темы: Расчет матриц близости объектов и классов по непрерывным признакам. Оценка функции качества классификации. Построение дендрограмм и другие средства визуального анализа данных.

Формы и методы проведения занятий по теме, применяемые образовательные

технологии: стандартная, 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 Кластерный анализ бинарных признаков.

Содержание темы: Расчет матриц близости объектов и классов по бинарным признакам. Оценка функции качества классификации. Построение дендрограмм и другие средства визуального анализа 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 7 Дискриминантный анализ.

Содержание темы: Классификация многомерных выборок с обучением. Оценка качества классификации с использованием метода скользящего экзамена. Средства визуального анализа данных в задачах классификации с обучением.

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.

Основным видом работы по изучению дисциплины является выполнение практических заданий на практических занятиях. Специфика дисциплины состоит в изучении сложных наукоемких алгоритмов. Большинство известных алгоритмов реализованы в пакетах прикладных программ. Однако, для эффективного использования программных модулей необходимо хорошо представлять все этапы работы алгоритмов. Для освоения алгоритмов разработаны специальные шаблоны в среде EXCEL. При выполнении практических заданий студенты используют разработанные преподавателем шаблоны. Такой методический прием позволяет сосредоточить внимание студента на содержании, изучаемого алгоритма, исключая непроизводительные затраты времени на оформление промежуточных результатов.

Продвинутые методы многомерного статистического анализа реализованы в виде программных модулей в специализированных пакетах прикладных программ. В процессе изучения дисциплины отрабатываются навыки работы с одним из самых распространенных пакетов статистической обработки данных ППП STATISTICA.

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 занятий 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 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. Многомерный статистический анализ [Электронный ресурс] , 2019 - 146 - Режим доступа: <https://lib.rucont.ru/efd/706335>

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

8.2 Дополнительная литература

1. Кулаичев А.П. Методы и средства комплексного статистического анализа данных : Учебное пособие [Электронный ресурс] : ИНФРА-М , 2017 - 484 - Режим доступа: <https://znanium.com/catalog/document?id=264006>

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

9. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: <https://www.prilib.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
- Усилитель-распределитель VGA/XGA Kramer VP-200

Программное обеспечение:

- Microsoft Office Professional Plus 2013 Russian
- Statsoft Statistica Base for Windows v.6 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 Tryon, 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 результативного 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) сравнивается с ожидаемым средним популяции (например, некое

теоретическое среднее), а вариация в популяции подсчитывается на основе вариации в наблюдаемой выборке.