

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор
по образовательной деятельности КФУ
Проф. Таюрский Д.А.

"__" _____ 20__ г.

Программа дисциплины

Многомерный статистический анализ Б1.В.ДВ.6

Направление подготовки: 01.03.02 - Прикладная математика и информатика

Профиль подготовки: Системное программирование

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Симушкин С.В.

Рецензент(ы):

Гумеров Р.Н.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Турилова Е. А.

Протокол заседания кафедры No ____ от " ____ " _____ 201__ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No ____ от " ____ " _____ 201__ г

Регистрационный No

Казань
2019

Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) доцент, к.н. (доцент) Симушкин С.В. кафедра математической статистики отделение прикладной математики и информатики ,
Sergey.Simushkin@kpfu.ru

1. Цели освоения дисциплины

изучить методы обработки статистических данных, зависящих от ряда контролируемых факторов, возникающих при анализе физических, экономических, медицинских процессов естествознания;

ознакомить с основными принципами проведения статистических экспериментов;

научить решать практические задачи с использованием методов многомерного статистического анализа;

дать представление о математических (теоретико-вероятностных) основаниях построения процедур проверки многомерных гипотез и оценивания многомерных характеристик

2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел "Б1.В.ДВ.6 Дисциплины (модули)" основной образовательной программы 01.03.02 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 3, 4 курсах, 6, 7 семестры.

Курс "Многомерный статистический анализ" входит в число курсов по выбору профессионального цикла подготовки бакалавра по направлению "Прикладная математика и информатика".

Изучению курса предшествует изучение базовых дисциплин "Линейная алгебра", "Математический анализ" и "Теория вероятностей и математическая статистика". В результате освоения предшествующих дисциплин студент должен:

знать:

основные понятия и методы линейной алгебры (решение систем линейных уравнений, обращение матриц, приведение квадратичных форм к главным осям, формирование базиса линейных пространств);

основные способы решения оптимизационных задач (метод Лагранжа), основные интегральные соотношения (интегралы Эйлера, Дирихле, Лапласа);

основные вероятностные законы (нормальный, Фишера, хи-квадрат), метод характеристических функций и дельта-метод отыскания асимптотического распределения;

принципы формирования статистических гипотез и их основных вероятностных характеристик (ошибки 1-го и 2-го рода, мощность критерия, среднеквадратическая ошибка, распределение статистик, асимптотическое распределение);

уметь:

производить алгебраические операции над матрицами и векторами;

решать линейные уравнения в матричной форме;

решать задачи на экстремум функций многих переменных;

использовать разложение в ряд Тейлора функцию многих переменных;

использовать основные алгебраические и тригонометрические тождества для преобразования алгебраических выражений;

находить распределение статистик и их основные характеристики (среднее значение, ковариация, корреляция);

владеть:

навыками использования математических справочников и таблиц;

приемами работы в основных пакетах прикладных программ ("Excel", "Mathematica").

Дисциплина читается на 3-4 курсах обучения.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОПК-4 (профессиональные компетенции)	способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности
ПК-10 (профессиональные компетенции)	способность к реализации решений, направленных на поддержку социальнозначимых проектов, на повышение информационной грамотности населения, обеспечения общедоступности информационных услуг
ПК-11 (профессиональные компетенции)	способность к организации педагогической деятельности в конкретной предметной области (математика и информатика)
ПК-12 (профессиональные компетенции)	способность к планированию и осуществлению педагогической деятельности с учетом специфики предметной области в общеобразовательных и профессиональных образовательных организациях
ПК-13 (профессиональные компетенции)	способность применять существующие и разрабатывать новые методы и средства обучения
ПК-8 (профессиональные компетенции)	способность приобретать и использовать организационно-управленческие навыки в профессиональной и социальной деятельности
ПК-9 (профессиональные компетенции)	способность составлять и контролировать план выполняемой работы, планировать необходимые для выполнения работы ресурсы, оценивать результаты собственной работы

В результате освоения дисциплины студент:

1. должен знать:

основные вероятностные законы, описывающие поведение случайных векторов (нормальный закон, мультиномиальный закон);

основные принципы построения моделей регрессионных связей между исследуемыми характеристиками (регрессия с линейной зависимостью от неизвестных параметров, полная, частная и множественная корреляция);

основные формы представления многомерных данных и особенности работы с ними; методы, применяемые для анализа и обработки многомерных совокупностей (полный корреляционный анализ, метод главных компонент, факторный анализ);

2. должен уметь:

применять методы многомерного статистического анализа для обработки реальных числовых данных, учитывая границы применимости математической модели;

применять специализированные программные продукты для проведения вычислительных процедур многомерного анализа;

реализовывать процедуры многомерного статистического анализа в рамках имеющихся средств обработки данных;

выбирать методику статистического исследования экспериментальных данных

3. должен владеть:

методами построения математической модели профессиональных задач и содержательной интерпретации полученных результатов.

навыками выбора математических методов обработки экспериментальных данных, адекватных целям исследования;

навыками реализации математических методов обработки экспериментальных данных в виде прикладных программных продуктов;

навыками составления отчетов по методикам исследования и их реализации в виде ПО, анализа результатов обработки

4. должен демонстрировать способность и готовность:

применять на практике основные методы многомерного статистического анализа

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) 144 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 6 семестре; зачет в 7 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Теоретико-вероятностные аспекты многомерного анализа Многомерное нормальное распределение	6	1-9	18	0	0	Устный опрос
2.	Тема 2. Методы сокращения размерности пространства признаков	6	10-18	18	0	0	Письменная работа

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
3.	Тема 3. Статистические оценки многомерных параметров и их распределения	7	1-10	18	0	0	Устный опрос
4.	Тема 4. Проверка гипотез о параметрах многомерных совокупностей	7	10-18	18	0	0	Устный опрос
	Тема . Итоговая форма контроля	6		0	0	0	Зачет
	Тема . Итоговая форма контроля	7		0	0	0	Зачет
	Итого			72	0	0	

4.2 Содержание дисциплины

Тема 1. Теоретико-вероятностные аспекты многомерного анализа Многомерное нормальное распределение

лекционное занятие (18 часа(ов)):

1. Многомерные случайные величины и их числовые характеристики 2. Числовые характеристики связности случайных величин - полный коэффициент корреляции, множественный коэффициент корреляции, частный коэффициент корреляции, каноническая корреляция. 3. Геометрическая характеристика - эллипсоид рассеяния. 5. Нормальное распределение - плотность, числовые характеристики, характеристическая функция, условное распределение, связь с хи-квадрат законом.

Тема 2. Методы сокращения размерности пространства признаков

лекционное занятие (18 часа(ов)):

1. Главные компоненты 2. Потеря информации при переходе к части главных компонент 3. Интерпретация результатов анализа главных компонент 4. Факторная модель выделения латентных признаков 5. Решение факторной теоремы 6. Интерпретация результатов

Тема 3. Статистические оценки многомерных параметров и их распределения

лекционное занятие (18 часа(ов)):

1. Оценки параметров методом моментов - их статистические свойства 2. Оценки параметров метода максимального правдоподобия - свойства и асимптотическое распределение. 3. Распределение оценок в нормальной модели данных. 4. Асимптотическое распределение оценок 5. Преобразования стабилизирующие дисперсию

Тема 4. Проверка гипотез о параметрах многомерных совокупностей

лекционное занятие (18 часа(ов)):

1. Критерии проверки гипотез, основанные на точном или асимптотическом распределении оценок параметров. 2. Проверка гипотез о значимости коэффициентов связности - полного, частного, множественного коэффициентов корреляции 3. Проверка гипотез о параметрах регрессии - доверительная полоса Уоркинга-Хотеллинга 4. Статистические критерии сравнения многомерных гипотез а) для двух зависимых групп б) для двух независимых групп в) для многих групп

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Теоретико-вероятностные аспекты многомерного анализа Многомерное нормальное распределение	6	1-9	подготовка к устному опросу	36	Устный опрос
2.	Тема 2. Методы сокращения размерности пространства признаков	6	10-18	подготовка к письменной работе	36	Письменная работа
	Итого				72	

5. Образовательные технологии, включая интерактивные формы обучения

Часть материала теоретического характера преподается в виде лекций у доски или в виде презентаций на мультимедийном экране. Предполагается использование диалоговой формы ведения занятий с постановкой и решением проблемных задач, обсуждением дискуссионных моментов.

Некоторые разделы курса студенты изучают самостоятельно по указанным методическим материалам или по Интернет-источникам с последующим докладом в виде презентаций и дискуссией.

При проведении практических занятий, на которых теоретический материал применяется к конкретным данным, используются готовые статистические процедуры стандартных пакетов, а также предлагается студентам разработать самостоятельные вычислительные процедуры.

Перед каждым занятием, как лекционной, так и практической направленности проводится экспресс-опрос по пройденному теоретическому материалу.

В целях выработки навыков работы в коллективе и развития коммуникативных способностей, часть вычислительных заданий, а также проработку новейших методов многомерного статистического анализа студенты выполняют, разбившись на творческие группы

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

Тема 1. Теоретико-вероятностные аспекты многомерного анализа Многомерное нормальное распределение

Устный опрос , примерные вопросы:

Ответить на следующие вопросы: 1. Что такое многомерная случайная величина 2. Какие числовые характеристики используются при анализе многомерных данных 3. Дайте определение полного коэффициента корреляции 4. Дайте определение множественного коэффициента корреляции 5. Дайте определение частного коэффициента корреляции 6. Дайте определение канонического коэффициента корреляции. 7. Дайте определение эллипсоида рассеяния. 8. Привести вид плотности нормального распределения 9. Чему равно математическое ожидание и матрица ковариаций нормально закона 10. Привести вид характеристической функции нормальной модели 11. Какое условное распределение имеет один из признаков нормального вектора 12. Как связана нормальная модель с хи-квадрат законом.

Тема 2. Методы сокращения размерности пространства признаков

Письменная работа , примерные вопросы:

1. Дать определение главных компонент 2. Сохранение дисперсии при переходе к главным компонентам 3. Потеря информации при переходе к главным компонентам 4. Способ нахождения главных компонент 5. Основная теорема факторного анализа 6. Способы решения факторной задачи 7. Методы вращения факторного решения 8. Пример построения факторного решения с интерпретацией результатов

Тема 3. Статистические оценки многомерных параметров и их распределения

Тема 4. Проверка гипотез о параметрах многомерных совокупностей

Итоговая форма контроля

зачет (в 6 семестре)

Итоговая форма контроля

зачет (в 7 семестре)

Примерные вопросы к зачету:

Проводится устный опрос по пройденному материалу

Итоговая оценка по дисциплине (из расчета 100 баллов) складывается из текущей работы (50 баллов = 36 баллов за ответы на коллоквиумах + 14 баллов за подготовку док-лада по выбранной теме) и ответа на билет во время зачёта (50 баллов).

Билет включает в себя три блока вопросов.

В первом блоке содержится 5 вопросов, для правильного ответа на которые необходимо привести формулировку определений и утверждений из списка определений и основных теорем (30% экзаменационной оценки).

Во втором блоке необходимо сформулировать теоретическое утверждение из списка основных теорем и привести подробную схему его доказательства (50% оценки).

В третьем блоке решить задачу с применением любых доступных средств обработки статистических данных (20% итоговой оценки)

Список вопросов (вопросы теоретического характера отмечены *)

- 1*. Определение и формула линейной регрессии.
- 2*. Определение и свойства матриц ковариации и корреляции.
3. Определение, свойства, формулы для вычисления множественных и частных коэффициентов корреляции.
4. Определение, формула эллипсоида рассеяния.
5. Плотность многомерного нормального закона. Двумерное нормальное распределение. Связь с эллипсом рассеяния.
6. Характеристическая функция нормального закона.
7. Линейные преобразования от нормального вектора.
8. Распределение квадратичной формы от нормального вектора.
9. Определение главных компонент.
- 10*. Способ вычисления главных компонент.
11. Свойства главных компонент.
12. Способ сокращения пространства признаков.
13. Связь сокращенного вектора главных компонент с матрицей исходных данных.
14. Интерпретация нагрузочных коэффициентов.
15. Модель факторного анализа.
16. Фундаментальное уравнение факторного анализа.
17. Проблема вращения, метод варимакса.
18. Оценки параметров многомерного случайного вектора по методу моментов.
19. Оценки максимального правдоподобия для нормальной модели данных.
20. Асимптотическое распределение парного коэффициента корреляции.

21. Асимптотическое распределение коэффициента парной регрессии.
22. Асимптотическое распределение наилучшего линейного прогноза.
- 23*. Доверительная полоса Уоркинга-Хотеллинга.
24. Распределение Уишарта и его связь с хи-квадрат распределением.
- 25*. Преобразование Стьюдента для коэффициентов корреляции.
- 26 Дельта-метод асимптотического анализа
- 27*. Преобразование Фишера, стабилизирующее дисперсию.
28. Критерий Фишера сравнения двух групп.
29. Асимптотическое распределение выборочного коэффициента регрессии.
30. Определение элементарных и неэлементарных дискриминантных признаков.
31. Критерий выделения наименее значимых признаков.

7.1. Основная литература:

1. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. - СПб.: Лань, 2011. - 256с.
ЭБС 'Лань': http://e.lanbook.com/books/element.php?pl1_id=2026
2. Королев В.Ю., Бенинг В.Е., Шоргин С.Я. Математические основы теории риска. - М.: Физматлит, 2011. - 620 с.
ЭБС 'Лань': <http://e.lanbook.com/book/2742>
3. Коралов Л.Б., Синай Я.Г. Теория вероятностей и случайные процессы. - м.: МЦНМО, 2013. - 408 с.
ЭБС 'Лань': <http://e.lanbook.com/book/71821>
4. Свешников А.А. Прикладные методы теории вероятностей.- СПб.: Лань, 2012. - 480 с.
ЭБС 'Лань': <http://e.lanbook.com/book/3184>

7.2. Дополнительная литература:

1. Свешников А.А. Прикладные методы теории марковских процессов.- СПб.: Лань, 2007. - 192 с.
ЭБС 'Лань': http://e.lanbook.com/books/element.php?pl1_id=590
2. Свешников А.А. Прикладные методы теории случайных функций.- СПб.: Лань, 2011. - 464с
ЭБС 'Лань': http://e.lanbook.com/books/element.php?pl1_id=656
3. Алон Н., Спенсер Дж. Вероятностный метод. - М.: БИНОМ. Лаборатория знаний, 2015. - 320 с.
ЭБС 'Лань': <http://e.lanbook.com/book/70705>

7.3. Интернет-ресурсы:

- StatPlus аналитическая статистика - <http://www.statplus.net.ua/ru/>
StatSoft электронный учебник по статистике - <http://www.statsoft.ru/home/textbook/>
Иллюстрированный самоучитель по SPSS - <http://www.learnspss.ru/>
Профессиональный информационно-аналитический ресурс - <http://www.machinelearning.ru/>
Федеральный образовательный портал ЭСМ - <http://ecsocman.hse.ru/>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Многомерный статистический анализ" предполагает использование следующего материально-технического обеспечения:

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен студентам. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, УМК, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего профессионального образования (ФГОС ВПО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань" , доступ к которой предоставлен студентам. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Аудитория с учебной доской и мелом

Мультимедийные средства

Компьютерный класс

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.03.02 "Прикладная математика и информатика" и профилю подготовки Системное программирование .

Автор(ы):

Симушкин С.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Гумеров Р.Н. _____

"__" _____ 201__ г.