

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



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ

Проф. Талорский Д.А.



_____ 20__ г.

подписано электронно-цифровой подписью

Программа дисциплины

Прикладной многомерный статистический анализ Б1.В.ОД.5

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Методы прикладной математической статистики

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Симушкин С.В.

Рецензент(ы):

Володин И.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 981916

Казань
2016

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 прикладные 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.ОД.5 Дисциплины (модули)" основной образовательной программы 01.04.02 Прикладная математика и информатика и относится к обязательным дисциплинам. Осваивается на 2 курсе, 3 семестр.

Курс "Прикладной многомерный статистический анализ" входит в число курсов профессионального цикла подготовки магистра по направлению "Прикладная математика и информатика".

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 на 2 курсе обучения.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1 (общекультурные компетенции)	способность к абстрактному мышлению, анализу, синтезу
ОК-3 (общекультурные компетенции)	готовность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3 (профессиональные компетенции)	способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять своё научное мировоззрение
ОПК-4 (профессиональные компетенции)	способность использовать и применять углубленные знания в области прикладной математики и информатики
ПК-8 (профессиональные компетенции)	способность разрабатывать корпоративные стандарты и профили функциональной стандартизации приложений, систем, информационной инфраструктуры
ПК-9 (профессиональные компетенции)	способность к преподаванию математических дисциплин и информатики в образовательных организациях основного общего, среднего общего, среднего профессионального и высшего образования

В результате освоения дисциплины студент:

- проводить самостоятельную исследовательскую работу по темам курса;
- применять методы математического анализа, функционального анализа, теории вероятности;
- работать с распределениями вероятности случайных величин, вычислять их теоретические и выборочные характеристики;
- применять на практике основные методы многомерного статистического анализа.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) 144 часа(ов).

Форма промежуточного контроля дисциплины экзамен в 3 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 и его применения в прикладных задачах. Многомерное нормальное распределение.	3	1-3	3	0	8	отчет
2.	Тема 2. Методы сокращения размерности пространства признаков	3	3-6	2	0	9	отчет
3.	Тема 3. Статистические оценки многомерных параметров и их распределения	3	6-10	4	0	10	отчет
4.	Тема 4. Проверка гипотез о параметрах многомерных совокупностей	3	10-12	2	0	7	отчет
5.	Тема 5. Приложения многомерного статистического анализа	3	13-14	3	0	8	реферат
	Тема . Итоговая форма контроля	3		0	0	0	экзамен
	Итого			14	0	42	

4.2 Содержание дисциплины

Тема 1. Теоретико-вероятностные аспекты многомерного анализа и его применения в прикладных задачах. Многомерное нормальное распределение.

лекционное занятие (3 часа(ов)):

Введение в многомерный анализ. Многомерные случайные величины и их характеристики. Понятие функции распределения в мультивариативном случае. Многомерное нормальное распределение.

лабораторная работа (8 часа(ов)):

Подготовка данных. Удаление аутлайеров. Вычисление основных числовых характеристик -- вектора средних, матрицы ковариаций, матрицы корреляций. Множественная корреляция. Частная корреляция. Каноническая корреляция. Эллипсоид рассеяния. Графическая иллюстрация -- построение гистограмм, эмпирических функций распределения, эллипсоидов рассеяния. Вычисление частных и канонических корреляций.

Тема 2. Методы сокращения размерности пространства признаков

лекционное занятие (2 часа(ов)):

Метод главных компонент. Построение матрицы нагрузок вектора главных компонент. Выделение наиболее информативных характеристик. Интерпретация результатов. Построение факторной модели. Факторный анализ.

лабораторная работа (9 часа(ов)):

Метод главных компонент. Понижение размерности. Способы интерпретации. Связь сокращенного вектора главных компонент с матрицей исходных данных. Модель факторного анализа. Фундаментальное уравнение факторного анализа. Методы построения факторной модели -- метод максимального правдоподобия, итерационные методы. Различные способы оценки факторных нагрузок. Проблема вращения, метод варимакса.

Тема 3. Статистические оценки многомерных параметров и их распределения

лекционное занятие (4 часа(ов)):

Распределения основных выборочных характеристик. Построение доверительных границ для корреляций. Доверительный интервал коэффициентов регрессии. Поверхности линейной регрессии.

лабораторная работа (10 часа(ов)):

Построение доверительных границ для корреляций, коэффициентов регрессии, поверхности линейной регрессии по реальным данным с использованием средств пакета программ Mathematica. Оценки по методу моментов многомерных характеристик. Оценки максимального правдоподобия. Дельта-метод при асимптотическом анализе свойств статистических оценок.

Тема 4. Проверка гипотез о параметрах многомерных совокупностей

лекционное занятие (2 часа(ов)):

Критерий Фишера для статистического сравнения нескольких групп. Точность критерия Фишера. Теоретическое описание задачи анализа механической обработки дисков авиатурбин. Статистические критерии проверки гипотез о параметрах многомерного нормального распределения, основанные на точном и асимптотическом распределении выборочных статистик.

лабораторная работа (7 часа(ов)):

Реализация критерия Фишера для сравнения нескольких групп по реальным данным механической обработки дисков авиатурбин. Анализ точности критерия Фишера методом Монте-Карло.

Тема 5. Приложения многомерного статистического анализа

лекционное занятие (3 часа(ов)):

Прикладные задачи многомерного статистического анализа в медицине. Редукция сложности статистической задачи при клинических испытаниях медицинских препаратов. Задача классификации демографических данных.

лабораторная работа (8 часа(ов)):

Компьютерное моделирование прикладных задач. Вычисление и анализ характеристик полученных моделей. Применение метода Монте-Карло для анализа оптимальных свойств решений задач.

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 и его применения в прикладных задачах. Многомерное нормальное					

распределение.

3

1-3

подготовка к
отчету

9

отчет

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
2.	Тема 2. Методы сокращения размерности пространства признаков	3	3-6	подготовка к отчету	12	отчет
3.	Тема 3. Статистические оценки многомерных параметров и их распределения	3	6-10	подготовка к отчету	13	отчет
4.	Тема 4. Проверка гипотез о параметрах многомерных совокупностей	3	10-12	подготовка к отчету	10	отчет
5.	Тема 5. Приложения многомерного статистического анализа	3	13-14	подготовка к реферату	8	реферат
	Итого				52	

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 и его применения в прикладных задачах. Многомерное нормальное распределение.

отчет , примерные вопросы:

Устный опрос и проверка корректности практической части отчёта по темам: многомерное нормальное распределения, линейная регрессия, частный коэффициент корреляции и т.д.

Тема 2. Методы сокращения размерности пространства признаков

отчет , примерные вопросы:

Устный опрос и проверка корректности практической части отчёта по темам: метод главных компонент, понижение размерности, факторный анализ и т.д.

Тема 3. Статистические оценки многомерных параметров и их распределения

отчет , примерные вопросы:

Устный опрос и проверка корректности практической части отчёта по темам: метод моментов для оценивания многомерных характеристик, дельта-метод, доверительные интервалы и т.д.

Тема 4. Проверка гипотез о параметрах многомерных совокупностей

отчет , примерные вопросы:

Устный опрос и проверка корректности практической части отчёта по темам: критерий Фишера, статистика Стьюдента для коэффициента корреляции, дискриминантные признаки и т.д.

Тема 5. Приложения многомерного статистического анализа

реферат , примерные темы:

Устный опрос и доклад по результатам изучения материала на тему приложения многомерного статистического анализа в практических задачах.

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Список вопросов (вопросы теоретического характера отмечены *) при проведении экзамена:

1. Определение, формула эллипсоида рассеяния.
2. Плотность многомерного нормального закона. Двумерное нормальное распределение. Связь с эллипсом рассеяния.
3. Характеристическая функция нормального закона.
4. Линейные преобразования от нормального вектора.
5. Распределение квадратичной формы от нормального вектора.
6. Определение главных компонент.
- 7*. Способ вычисления главных компонент.
8. Свойства главных компонент.
- 9*. Определение и формула линейной регрессии.
- 10*. Определение и свойства матриц ковариации и корреляции.
11. Определение, свойства, формулы для вычисления множественных и частных коэффициентов корреляции.
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. Асимптотическое распределение наилучшего линейного прогноза.
- 18*. Доверительная полоса Уоркинга-Хотеллинга.
19. Распределение Уишарта и его связь с хи-квадрат распределением.
- 20*. Преобразование Стьюдента для коэффициентов корреляции.
- 21 Дельта-метод асимптотического анализа

- 22*. Преобразование Фишера, стабилизирующее дисперсию.
23. Критерий Фишера сравнения двух групп.
24. Асимптотическое распределение выборочного коэффициента регрессии.
25. Определение элементарных и неэлементарных дискриминантных признаков.
26. Критерий выделения наименее значимых признаков.
27. Проблема вращения, метод варимакса.
28. Оценки параметров многомерного случайного вектора по методу моментов.
29. Оценки максимального правдоподобия для нормальной модели данных.
30. Асимптотическое распределение парного коэффициента корреляции.
31. Асимптотическое распределение коэффициента парной регрессии.

7.1. Основная литература:

1. Теория вероятностей и математическая статистика: Учебник / Е.С. Кочетков, С.О. Смерчинская, В.В. Соколов. - 2-е изд., испр. и перераб. - М.: Форум: НИЦ ИНФРА-М, 2014. - 240 с.
ЭБС "ZnaniUM.COM": <http://znanium.com/bookread.php?book=447828>
2. Статистический анализ данных в MS Excel: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - М.: ИНФРА-М, 2014. - 320 с.
ЭБС "ZnaniUM.COM": <http://znanium.com/bookread.php?book=429722>
3. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. - СПб.: Лань, 2011. - 256 с.
ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=2026
4. Туганбаев А.А., Крупин В.Г. Теория вероятностей и математическая статистика. - СПб.: Лань, 2011. - 320 с.
ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=652
5. Свешников А.А. Прикладные методы теории вероятностей. - СПб.: Лань, 2012. - 480 с.
ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=3184

7.2. Дополнительная литература:

1. Бьорк Т. Теория арбитража в непрерывном времени. - М.: МЦНМО, 2010. - 560 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/9302/>
2. Коралов Л.Б., Синай Я.Г. Теория вероятностей и случайные процессы. - м.: МЦНМО, 2013. - 408 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/56404/>
3. Алон Н., Спенсер Дж. Вероятностный метод. - М.: БИНОМ. Лаборатория знаний, 2015. - 323 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/70705/>
4. Лагутин М.Б. Наглядная математическая статистика. - М.: БИНОМ. Лаборатория знаний, 2013. - 472 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/56887/>

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 "Прикладной многомерный статистический анализ" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

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.

Занятия проводятся как в компьютерном классе, так и в аудитории, снабженной доской и мелом (маркером)

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе Методы прикладной математической статистики .

Автор(ы):

Симушкин С.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Володин И.Н. _____

"__" _____ 201__ г.