

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



подписано электронно-цифровой подписью

Программа дисциплины
Многомерный статистический анализ Б3.В.6

Направление подготовки: 010400.62 - Прикладная математика и информатика

Профиль подготовки: Теория вероятностей и математическая статистика

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 9170014

Казань
2014

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 " БЗ.В.6 Профессиональный" основной образовательной программы 010400.62 Прикладная математика и информатика и относится к вариативной части. Осваивается на 4 курсе, 7 семестр.

Курс "Многомерный статистический анализ" входит в число курсов по выбору профессионального цикла подготовки бакалавра по направлению "Прикладная математика и информатика" (профиль "Теория вероятностей и математическая статистика"). Входит в раздел В.6 ООП.

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 на 4 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)	владение культурой мышления, способность к обобщению, анализу, восприятию информации, постановке цели и выбору путей ее достижения
ОК-14 (общекультурные компетенции)	способность использовать в научной и познавательной деятельности, а также в со-циальной сфере профессиональные навыки работы с информационными и компьютерны-ми технологиями
ОК-15 (общекультурные компетенции)	способность работы с информацией из различных источников, включая ресурсы сети Интернет, для решения профессиональных и социальных задач
ПК-20 (профессиональные компетенции)	использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования
ПК-5 (профессиональные компетенции)	владеет основными приемами обработки и представления экспериментальных дан-ных

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 5 зачетных(ые) единиц(ы) 180 часа(ов).

Форма промежуточного контроля дисциплины экзамен 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	7	1-5	0	5	15	
2.	Тема 2. Методы сокращения размерности пространства признаков	7	6-9	0	4	12	
3.	Тема 3. Статистические оценки многомерных параметров и их распределения	7	10-13	0	4	12	
4.	Тема 4. Проверка гипотез о параметрах многомерных совокупностей	7	14-18	0	5	15	
	Тема . Итоговая форма контроля	7		0	0	0	экзамен

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
Итого				0	18	54	

4.2 Содержание дисциплины

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

практическое занятие (5 часа(ов)):

Подготовка данных. Удаление аутлайеров. Вычисление основных числовых характеристик -- вектора средних, матрицы ковариаций, матрицы корреляций. Построение линейного прогноза для каждой характеристики. Графическая иллюстрация -- построение гистограмм, эмпирических функций распределения, эллипсоидов рассеяния. Вычисление частных и канонических корреляций.

лабораторная работа (15 часа(ов)):

Линейная регрессия. Множественная корреляция. Частная корреляция. Каноническая корреляция. Эллипсоид рассеяния. Способы задания многомерной нормальной модели -- характеристическая функция, линейные преобразования, квадратичные формы.

Тема 2. Методы сокращения размерности пространства признаков

практическое занятие (4 часа(ов)):

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

лабораторная работа (12 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

практическое занятие (4 часа(ов)):

Построение доверительных границ для корреляций, коэффициентов регрессии, поверхности линейной регрессии по реальным данным с использованием средств пакета программ Mathematica.

лабораторная работа (12 часа(ов)):

Оценки по методу моментов многомерных характеристик. Оценки максимального правдоподобия. Дельта-метод при асимптотическом анализе свойств статистических оценок. Асимптотическое распределение основных статистических характеристик -- коэффициенты корреляции, коэффициенты регрессии, поверхность линейной регрессии. Доверительные интервалы.

Тема 4. Проверка гипотез о параметрах многомерных совокупностей

практическое занятие (5 часа(ов)):

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

лабораторная работа (15 часа(ов)):

Статистические критерии проверки гипотез о параметрах многомерного нормального распределения, основанные на точном и асимптотическом распределении выборочных статистик. Статистика Стьюдента для коэффициента корреляции. Сравнение векторов средних значений. Критерий Фишера и его асимптотические аналоги. Элементарные и неэлементарные дискриминантные признаки. Отсев наименее значимых признаков.

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	7	1-5	Изучение литературных первоисточников и интернет-ресурсов	20	коллоквиум
2.	Тема 2. Методы сокращения размерности пространства признаков	7	6-9	Изучение литературных первоисточников и интернет-ресурсов	16	коллоквиум
3.	Тема 3. Статистические оценки многомерных параметров и их распределения	7	10-13	Изучение литературных первоисточников и интернет-ресурсов	16	коллоквиум
4.	Тема 4. Проверка гипотез о параметрах многомерных совокупностей	7	14-18	Изучение литературных первоисточников и интернет-ресурсов	20	коллоквиум
	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

коллоквиум, примерные вопросы:

Ответить на вопросы теоретического плана: 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 нормального закона. 6. Характеристическая функция нормального закона. 7. Линейные преобразования, квадратичные формы от нормального вектора.

Тема 2. Методы сокращения размерности пространства признаков

коллоквиум, примерные вопросы:

Ответить на вопросы теоретического плана: 1. Определение и способ вычисления главных компонент. 2. Свойства главных компонент. 3. Связь сокращенного вектора главных компонент с матрицей исходных данных. 4. Интерпретация нагрузочных коэффициентов. 5. Модель факторного анализа. 6. Фундаментальное уравнение факторного анализа. 7. Проблема вращения, метод варимакса.

Тема 3. Статистические оценки многомерных параметров и их распределения

коллоквиум, примерные вопросы:

Ответить на вопросы теоретического плана: 1. Оценки по методу моментов. 2. Оценки максимального правдоподобия для нормальной модели данных. 3. Асимптотическое распределение парного коэффициента корреляции, коэффициента парной регрессии, линейной регрессии. 4. Доверительная полоса Уоркинга-Хотеллинга. 5. Распределение Уишарта и его связь с хи-квадрат распределением.

Тема 4. Проверка гипотез о параметрах многомерных совокупностей

коллоквиум, примерные вопросы:

Ответить на вопросы теоретического плана: 1. Преобразование Стьюдента для коэффициентов корреляции. 2. Преобразование Фишера, стабилизирующее дисперсию. 3. Критерий Фишера сравнения двух групп. 4. Асимптотическое распределение выборочного коэффициента регрессии. 5. Определение элементарных и неэлементарных дискриминантных признаков. 6. Критерий выделения наименее значимых признаков.

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

По первым трем разделам дисциплины проводится коллоквиум в виде опроса по пройденному материалу

Экзаменационная 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 экзаменационный билет (50 баллов).

Экзаменационный 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% экзаменационной 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 Стюдента 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.Боровков А.А. Математическая статистика.- СПб.: Лань, 2010. - 704 с. ЭБС "Лань":
http://e.lanbook.com/books/element.php?pl1_id=381
- 3.Свешников А.А. Прикладные методы теории вероятностей.- СПб.: Лань, 2012. - 480 с. ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=318
- 4.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
- 5.Свешников А.А. Прикладные методы теории случайных функций.- СПб.: Лань, 2011. - 464с ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=656

7.2. Дополнительная литература:

1. Вероятность и прикладная статистика, Орлов, Александр Иванович, 2010г.
- 2.Прикладная математическая статистика, Кобзарь, Александр Иванович, 2006г.
- 3.Математическая и прикладная статистика, Харин, Юрий Семенович;Жук, Евгений Евгеньевич, 2005г.

4. Андерсон Т. Введение в многомерный статистический анализ. - М.: Физматгиз, 1963. - 500 с.
5. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. - М.: Финансы и стат., 2003. - 350 с.
6. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976.-736 с.
7. Справочник по прикладной статистике. Под редакцией Э.Ллойда, У.Ледермана. Т.1 - М.: Финансы и статистика, 1989. - 511 с.
8. Справочник по прикладной статистике. Под редакцией Э.Ллойда, У.Ледермана. Т.2 - М.: Финансы и статистика, 1990. - 528 с.
9. Симушкин С.В. Многомерный статистический анализ. Методические разработки по специальному курсу. Ч.1. - Изд-во КГУ, 2006. - 99 с.

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:

Аудитория с учебной доской и мелом

Мультимедийные средства

Компьютерный класс

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 010400.62 "Прикладная математика и информатика" и профилю подготовки Теория вероятностей и математическая статистика .

Автор(ы):

Симушкин С.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Гумеров Р.Н. _____

"__" _____ 201__ г.