

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

Статистический анализ выборочных данных Б1.В.ДВ.13

Направление подготовки: 01.03.02 - Прикладная математика и информатика

Профиль подготовки: Системное программирование

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Володин И.Н. , Симушкин С.В.

Рецензент(ы):

Халиуллин С.Г.

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 944317

Казань
2017

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) профессор, д.н. (профессор) Володин И.Н. кафедра математической статистики отделение прикладной математики и информатики , igornvolodin@gmail.com ; 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 статистического моделирования оценок и тестовых статистик на конкретных прикладных задачах;

изучить способы применения статистических пакетов при прикладном статистическом анализе;

ознакомить с классическими и современными концепциями гарантийного статистического вывода;

научить способам статистического моделирования точностных свойств оценок;

научить способам статистического моделирования мощностных свойств статистических критериев.

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

Данная учебная дисциплина включена в раздел " Б1.В.ДВ.13 Дисциплины (модули)" основной образовательной программы 01.03.02 Прикладная математика и информатика и относится к дисциплинам по выбору. Осваивается на 4 курсе, 8 семестр.

Лабораторное задание по теме "Статистический анализ выборочных данных" входит в вариативную часть подготовки бакалавра по направлению "Прикладная математика".

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;

способы вычисления числовых характеристик случайных величин и случайных векторов

уметь:

производить алгебраические операции над матрицами и векторами;

решать линейные уравнения;

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 4 курсе обучения.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-1 (профессиональные компетенции)	способность использовать стандартные пакеты прикладных программ для решения практических задач на ЭВМ, отлаживать, тестировать прикладное программное обеспечение
ПК-10 (профессиональные компетенции)	готовностью применять математический аппарат для решения поставленных задач, способностью применить соответствующую процессу математическую модель и проверить ее адекватность, провести анализ результатов моделирования, принять решение на основе полученных результатов
ПК-12 (профессиональные компетенции)	способностью самостоятельно изучать новые разделы фундаментальных наук

В результате освоения дисциплины студент:

1. должен знать:

- основные принципы описания статистических структур;
- основные способы построения оптимальных процедур статистического вывода (переход к достаточным статистикам, информационные неравенства, байесовские и минимаксные правила, процедуры с минимальной функцией риска);
- основные принципы планирования экспериментов;
- методы построения гарантийных статистических правил;
- методы построения оптимальных планов для полиномиальной и тригонометрической регрессий.

2. должен уметь:

- применять методы построения оптимальных правил для обработки реальных числовых данных, учитывая границы применимости математической модели;
- применять специализированные программные продукты для проведения вычислительных процедур дисперсионного анализа;
- реализовывать процедуры дисперсионного анализа в рамках имеющихся средств обработки данных;
- выбрать методику статистического исследования экспериментальных данных.

3. должен владеть:

- методами построения оптимальных процедур статистического вывода.

4. должен демонстрировать способность и готовность:

- способность и готовность применять на практике методы оптимального и гарантийного статистического вывода.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины экзамен в 8 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. Статистический анализ данных экономических показателей малых предприятий	8	1-4	0	0	8	Отчет
2.	Тема 2. Статистический анализ данных на надежность	8	5-8	0	0	7	Отчет
3.	Тема 3. Статистический анализ данных уровня бъефа Чебоксарской ГЭС	8	9-11	0	0	6	Реферат
4.	Тема 4. Статистический анализ данных неврологических испытаний	8	12-14	0	0	7	Отчет

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
5.	Тема 5. Статистический анализ данных биологических испытаний	8	15-17	0	0	8	Реферат
	Тема . Итоговая форма контроля	8		0	0	0	Экзамен
	Итого			0	0	36	

4.2 Содержание дисциплины

Тема 1. Статистический анализ данных экономических показателей малых предприятий лабораторная работа (8 часа(ов)):

Оценка популяционной суммы по данным дохода малых предприятий. Расслоенный выбор. Оценка дисперсии.

Тема 2. Статистический анализ данных на надежность лабораторная работа (7 часа(ов)):

Тестирование вероятностных моделей развития трещины в метрологических испытаниях. Анализ данных по содержанию примесей в нефтепродуктах. Проверка модельных предположений.

Тема 3. Статистический анализ данных уровня бьефа Чебоксарской ГЭС лабораторная работа (6 часа(ов)):

Построение вероятностной модели гидрологических испытаний. Наилучший прогноз уровня бьефа. Применение нейронных сетей.

Тема 4. Статистический анализ данных неврологических испытаний лабораторная работа (7 часа(ов)):

Анализ данных с помощью таблиц сопряженности. Полный корреляционный анализ показателей, снятых с кривой вызванного потенциала.

Тема 5. Статистический анализ данных биологических испытаний лабораторная работа (8 часа(ов)):

Выявление генов, ответственных за патологические изменения в организме. Полный корреляционный и регрессионный анализ данных анестезиологии.

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. Статистический анализ данных экономических показателей малых предприятий	8	1-4	подготовка к отчету	6	отчет

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
2.	Тема 2. Статистический анализ данных на надежность	8	5-8	подготовка к отчету	6	отчет
3.	Тема 3. Статистический анализ данных уровня бьефа Чебоксарской ГЭС	8	9-11	подготовка к реферату	6	реферат
4.	Тема 4. Статистический анализ данных неврологических испытаний	8	12-14	подготовка к отчету	8	отчет
5.	Тема 5. Статистический анализ данных биологических испытаний	8	15-17	подготовка к реферату	6	реферат
	Итого				32	

5. Образовательные технологии, включая интерактивные формы обучения

Обучение производится в форме практических занятий.

Теоретический материал излагается перед каждым занятием с последующим углублённым самостоятельным изучением студентом. Список литературы разделен на две категории: необходимый для успешного выполнения практических заданий и дополнительная литература.

Выполнение практических заданий подразумевает овладение теоретическим материалом и навыков компьютерного моделирования с использованием современных статистических пакетов.

Самостоятельная работа предполагает выполнение домашних работ. Практические задания, выполненные в аудитории, предназначены для указания общих методов решения задач определенного типа. Закрепить навыки можно лишь в результате самостоятельной работы.

Кроме того, самостоятельная работа включает подготовку к экзамену. При подготовке к сдаче экзамена весь объем работы рекомендуется распределять равномерно по дням, отведенным для подготовки к экзамену, контролировать каждый день выполнения работы.

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

Тема 1. Статистический анализ данных экономических показателей малых предприятий

отчет , примерные вопросы:

Отчет должен содержать: 1) описание структуры выборочных данных; 2) описание статистических методов анализа данных; 3) полученные оценки популяционных сумм и дисперсии.

Тема 2. Статистический анализ данных на надежность

отчет , примерные вопросы:

Отчет должен содержать: 1) описание структуры выборочных данных; 2) описание статистических методов анализа данных; 3) полученные р-значения при тестировании моделей.

Тема 3. Статистический анализ данных уровня бьефа Чебоксарской ГЭС

реферат , примерные темы:

Реферат должен содержать: 1) описание структуры выборочных данных; 2) описание статистических методов наилучшего прогноза; 3) аналитический обзор имеющейся литературы; 4) программный продукт.

Тема 4. Статистический анализ данных неврологических испытаний

отчет , примерные вопросы:

Отчет должен содержать: 1) описание структуры выборочных данных; 2) описание статистических методов анализа данных; 3) полученные результаты полного корреляционного анализа.

Тема 5. Статистический анализ данных биологических испытаний

реферат , примерные темы:

Реферат должен содержать: 1) описание структуры выборочных данных; 2) описание статистических методов наилучшего прогноза; 3) аналитический обзор имеющейся литературы; 4) результаты тестирования множественных гипотез.

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).

БИЛЕТЫ К ЭКЗАМЕНУ

Билет ♦ 1

1. Вероятностная модель биномиального распределения.
2. Обратный биномиальный выбор.

Билет ♦ 2

1. Вероятностная модель пуассоновского распределения.
2. Наиболее точные нижние доверительные границы для параметра распределения Пуассона.

Билет ♦ 3

1. Модель "отсутствие последствие".
2. Функциональное уравнение модели и его решение.

Билет ♦ 4

1. Вероятностная модель нормального распределения.
2. Наиболее мощный критерий проверки простой гипотезы при простой альтернативе.

Билет ♦ 5

1. Вероятностная модель старения и износа (гамма-распределение).
2. Принцип двойственности в построении доверительных множеств и проверке гипотез.

Билет ♦ 6

1. Вероятностная модель пуассоновского распределения.
2. Проверка гипотезы о вероятности успеха в испытаниях Бернулли.

Билет ♦ 7

1. Как найти необходимый объем выборки?

2. Двухступенчатая процедура Стейна для оценки среднего значения нормального распределения при неизвестном значении дисперсии.

Билет ♦ 8

1. Вероятностная модель пуассоновского распределения.
2. Байесовская оценка при квадратичной функции потерь.

Билет ♦ 9

1. Вероятностная модель старения и износа (гамма-распределение).
2. Равномерно наиболее мощный критерий для распределений с монотонным отношением правдоподобия.

Билет ♦ 10

1. Вероятностная модель пуассоновского распределения.
2. Несмещенность наиболее мощного критерия.

Билет ♦ 11

1. Несмещенная оценка с минимальным риском для надежности (показательное распределение долговечности).
2. Дать определение критерия, критической области, функции мощности, размера критерия, уровня значимости и критического уровня значимости.

Билет ♦ 12

1. Вероятностные модели роста (логарифмически нормальное распределение, распределение Бирнбайма-Саундерса).
2. Минимаксная оценка вероятности успеха в испытаниях Бернулли при нормированной квадратичной функции потерь.

Билет ♦ 13

1. Вероятностная модель старения и износа (гамма-распределение).
2. Равномерно наиболее точная нижняя доверительная граница для вероятности успеха в испытаниях Бернулли.

Билет ♦ 14

1. Вероятностная модель нормального распределения.
2. Полная достаточная статистика для семейства биномиальных распределений.

Билет ♦ 15

1. Критерий, критическая функция, уровень значимости, функция мощности.
2. Метод построения несмещенных оценок с равномерно минимальным риском.

Билет ♦ 16

1. Вероятностные модели роста (логарифмически нормальное распределение, распределение Бирнбайма-Саундерса).
2. Максимальные инварианты и их роль при построении инвариантных критериев.

Билет ♦ 17

1. Вероятностная модель старения и износа (гамма-распределение).
2. Двухступенчатая процедура Стейна для построения доверительного интервала фиксированной ширины.

Билет ♦ 18

1. Распределение Бирнбаума-Саундерса.
2. Построение равномерно наиболее точных доверительных областей на основе PNM критериев.

Билет ♦ 19

1. Вероятностные модели роста (логарифмически нормальное распределение, распределение Бирнбайма-Саундерса).
2. Лемма Неймана-Пирсона (необходимое условие наибольшей мощности критерия отношения правдоподобия).

Билет ♦ 20

1. Вероятностная модель нормального распределения.
2. Гарантийная оценка среднего.

Билет ♦ 21

1. Обратный биномиальный выбор.
2. Определение границ продолжения наблюдений в последовательном критерии отношения правдоподобия.

Билет ♦ 22

1. Вероятностная модель старения и износа (гамма-распределение).
2. Асимптотика необходимого объема выборки при различении сближающихся параметрических гипотез.

Билет ♦ 23

1. Вероятностная модель показательного распределения.
2. Равномерно наиболее мощный критерий для распределений с монотонным отношением правдоподобия.

Билет ♦ 24

1. Вероятностная модель нормального распределения.
2. Методы построения несмещенных оценок с равномерно минимальным риском.

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. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. - М.:ФОРУМ: инфра-м, 2013. - 464 с.

ЭБС "Знаниум": <http://znanium.com/bookread.php?book=369689>

3. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистический анализ данных в MS Excel. - М.: ИНФРА-М, 2014. - 320 с.

ЭБС "Знаниум": <http://znanium.com/bookread.php?book=429722>

4. Свешников А.А. Прикладные методы теории вероятностей. - СПб.: Лань, 2012. - 480 с.

ЭБС "Лань": http://e.lanbook.com/books/element.php?pl1_id=3184

5. Ширяев А.Н., Эрлих И.Г., Яськов П.А. Вероятность в теоремах и задачах (с доказательствами и решениями). Книга 1. - М.: МЦНМО, 2013. - 648 с.

ЭБС "Лань": <http://e.lanbook.com/view/book/56417/>

7.2. Дополнительная литература:

1. Лагутин, М.Б. Наглядная математическая статистика. [Электронный ресурс] : учеб. пособие ? Электрон. дан. ? М. : Издательство 'Лаборатория знаний', 2015. ? 475 с. ? Режим доступа: <http://e.lanbook.com/book/70706> ? Загл. с экрана.

7.3. Интернет-ресурсы:

курс лекций - http://chaos.phys.msu.ru/loskutov/PDF/Lectures_time_series_analysis.pdf

Портал математических интернет ресурсов - <http://www.math.ru>

Портал математических интернет ресурсов - <http://www.allmath.com>

Примеры вычисление тестовых статистик - <http://exponenta.ru>

Сайт с материалами по естественно-научным дисциплинам - <http://en.endu.ru>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Статистический анализ выборочных данных" 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 в компьютерном классе

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):

Володин И.Н. _____

Симушкин С.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Халиуллин С.Г. _____

"__" _____ 201__ г.