

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. Таюрский Д.А.

"__" _____ 20__ г.

Программа дисциплины
Статистический анализ данных Б2.ДВ.1

Направление подготовки: 080500.62 - Бизнес-информатика

Профиль подготовки:

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Симушкин С.В.

Рецензент(ы):

Миссаров М.Д.

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
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

Курс посвящён изучению основ статистического анализа данных и его применению.

Освоение курса статистического анализа данных должно:

- научить основным методам дескриптивного статистического анализа, доверительного оценивания;
- дать представление об основных методах регрессионного анализа;
- изучить способы применения статистических пакетов при прикладном статистическом анализе;
- ознакомить с классическими концепциями гарантийного статистического вывода.

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

Данная учебная дисциплина включена в раздел " Б2.ДВ.1 Общепрофессиональный" основной образовательной программы 080500.62 Бизнес-информатика и относится к дисциплинам по выбору. Осваивается на 3 курсе, 5 семестр.

Данная дисциплина осваивается на 3 курсе (5 семестр).

Изучению курса предшествует изучение дисциплин "Линейная алгебра", "Математический анализ" и "Теория вероятностей и математическая статистика".

Методы, изучаемые в курсе, важны во всех дальнейших курсах, связанных со статистическими исследованиями, анализом данных и т.д.

В результате освоения предшествующих дисциплин студент должен:

- знать основные методы и определения теории вероятностей и математического анализа;
- уметь преобразовывать алгебраические выражения, вероятностные соотношения, случайные величины;
- проявлять готовность изучать сложный математический материал, производить численные исследования.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-16 (общекультурные компетенции)	способен работать с информацией из различных источников
ПК-19 (профессиональные компетенции)	использовать основные методы естественнонаучных дисциплин в профессиональной деятельности для теоретического и экспериментального исследования
ПК-20 (профессиональные компетенции)	использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования

В результате освоения дисциплины студент:

- применять пакеты прикладных программ для первичного статистического анализа выборочных данных;
- применять теоретические методы исследования и преобразования при статистическом анализе;
- применять на практике методы гарантийного статистического вывода.

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 3 зачетных(ые) единиц(ы) 108 часа(ов).

Форма промежуточного контроля дисциплины экзамен в 5 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. Дескриптивный анализ выборочных данных.	5	1-4	4	0	4	отчет
2.	Тема 2. Проверка статистических гипотез относительно выборочных данных.	5	5-9	5	0	5	отчет
3.	Тема 3. Доверительное оценивание и доверительные интервалы.	5	9-13	5	0	4	отчет
4.	Тема 4. Независимость случайных величин и регрессионный анализ.	5	14-18	4	0	5	отчет
	Тема . Итоговая форма контроля	5		0	0	0	экзамен
	Итого			18	0	18	

4.2 Содержание дисциплины

Тема 1. Дескриптивный анализ выборочных данных.

лекционное занятие (4 часа(ов)):

Определение понятий случайных величин, выборки, математического ожидания. Основные теоретические моментные характеристики случайных величин. Эмпирическое распределение, гистограмма.

лабораторная работа (4 часа(ов)):

Произвести вычисление основных выборочных характеристик заданной выборки. Построить гистограмму с графиком подогнанной теоретической функции плотности. Построить эмпирическую функцию распределения.

Тема 2. Проверка статистических гипотез относительно выборочных данных.

лекционное занятие (5 часа(ов)):

Общие принципы формулировки статистических гипотез и построения статистических критериев для их проверки. Основные критерии проверки однородности.

лабораторная работа (5 часа(ов)):

Произвести проверку гипотезы однородности: одновыборочным критерием Стьюдента, критерием знаков, двухвыборочным критерием Стьюдента, критерием Вилкоксона.

Тема 3. Доверительное оценивание и доверительные интервалы.

лекционное занятие (5 часа(ов)):

Определение понятия доверительного множества. Постановка статистической задачи. Методы построения доверительных границ для основных задач.

лабораторная работа (4 часа(ов)):

Построение выборочных доверительных границ для среднего и дисперсии нормального закона; для вероятности успеха биномиального закона.

Тема 4. Независимость случайных величин и регрессионный анализ.

лекционное занятие (4 часа(ов)):

Основные свойства независимых случайных величин. Коэффициент корреляции. Проверка гипотез независимости случайных величин. Линейная регрессия.

лабораторная работа (5 часа(ов)):

Проверить гипотезу независимости критерием хи-квадрат и критерием Стьюдента. Построить линию регрессии по заданным выборочным данным.

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. Дескриптивный анализ выборочных данных.	5	1-4	подготовка к отчету	7	отчет
2.	Тема 2. Проверка статистических гипотез относительно выборочных данных.	5	5-9	подготовка к отчету	10	отчет
3.	Тема 3. Доверительное оценивание и доверительные интервалы.	5	9-13	подготовка к отчету	10	отчет
4.	Тема 4. Независимость случайных величин и регрессионный анализ.	5	14-18	подготовка к отчету	9	отчет

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
	Итого				36	

5. Образовательные технологии, включая интерактивные формы обучения

Обучение производится только в форме лекционных и лабораторных 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 подготовки к экзамену; 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. Дескриптивный анализ выборочных данных.

отчет , примерные вопросы:

Отчёт производится по результатам выполнения дескриптивного анализа выборочных данных. При сдаче отчёта студенту задаются теоретические вопросы по теме.

Тема 2. Проверка статистических гипотез относительно выборочных данных.

отчет , примерные вопросы:

Отчёт производится по результатам выполнения заданий по проверке статистических гипотез однородности. При сдаче отчёта студенту задаются теоретические вопросы по теме.

Тема 3. Доверительное оценивание и доверительные интервалы.

отчет , примерные вопросы:

Отчёт производится по результатам построения доверительных границ для среднего, дисперсии, вероятности успеха. При сдаче отчёта студенту задаются теоретические вопросы по теме.

Тема 4. Независимость случайных величин и регрессионный анализ.

отчет , примерные вопросы:

Отчёт производится по результатам выполнения задач по проверки гипотезы независимости и построению линии регрессии. При сдаче отчёта студенту задаются теоретические вопросы по теме.

Тема . Итоговая форма контроля

Примерные вопросы к экзамену:

Примерные вопросы к экзамену:

1. Определение понятия случайной величины, распределения случайной величины.
2. Определение понятия выборки. Основные выборочные характеристики.

3. Построение гистограммы. Связь между гистограммой и функцией плотности случайной величины.
4. Построение эмпирической функции распределения. Связь эмпирической функции распределения с теоретической функцией распределения.
5. Формулировка статистических гипотез. Основные принципы построения критериев проверки гипотез.
6. Одновыборочный и двухвыборочный критерии Стьюдента проверки гипотезы однородности.
7. Критерий знаков проверки гипотезы однородности.
8. Критерий Вилкоксона проверки гипотезы однородности.
9. Критерий хи-квадрат проверки гипотезы согласия.
10. Определение понятия доверительного множества. Построение доверительных интервалов. Связь доверительных интервалов и критериев проверки гипотез.
11. Построение доверительного интервала для среднего нормального закона.
12. Построение доверительного интервала для дисперсии нормального закона.
13. Построение доверительного интервала для вероятности успеха биномиального закона.
14. Критерий хи-квадрат проверки гипотезы независимости.
15. Критерий Стьюдента проверки гипотезы независимости.
16. Линейная регрессия.

7.1. Основная литература:

1. Володин И.Н. Лекции по теории вероятностей и математической статистики / И.Н. Володин. - Казань: Изд-во КГУ, 2006. - 271 с.
URL: http://libweb.ksu.ru/ebooks/09_66%20_ds006.pdf
2. 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
3. Спирина М.С. Теория вероятностей и математическая статистика: учебник для студ. вузов / М. С. Спирина, П. А. Спирин. - Москва: Академия, 2007. - 352 с.
4. Симушкин С.В. Теоретические аспекты заданий курсового проекта по математической статистике / Казань: Казан. гос. ун-т, 2004. - 67 с.
5. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / В. Е. Гмурман. - 12-е изд., перераб.. - Москва: Высш. образование: Юрайт-Издат, 2009. - 478 с.

7.2. Дополнительная литература:

1. Боровков А.А. Математическая статистика. - СПб.: Лань, 2010. - 704 с.
http://e.lanbook.com/books/element.php?pl1_id=3810
2. Свешников А.А. Прикладные методы теории вероятностей. - СПб.: Лань, 2012. - 480 с.
http://e.lanbook.com/books/element.php?pl1_id=3184
3. Булинский А.В., Ширяев А.Н. Теория случайных процессов. - М.:Физматлит, 2005. - 400 с.
ЭБС "Лань": <http://e.lanbook.com/view/book/59319/>

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 данных" 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.

Компьютерный класс с доской и с установленным пакетом прикладных программ, пригодным для проведения статистического анализа

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 080500.62 "Бизнес-информатика" .

Автор(ы):

Симушкин С.В. _____

"__" _____ 201__ г.

Рецензент(ы):

Миссаров М.Д. _____

"__" _____ 201__ г.