

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
Прикладная статистика Б3.ДВ.2

Направление подготовки: 080500.62 - Бизнес-информатика

Профиль подготовки:

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Автор(ы):

Лернер Э.Ю.

Рецензент(ы):

Володин И.Н., Миссаров М.Д.

СОГЛАСОВАНО:

Заведующий(ая) кафедрой: Миссаров М. Д.

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) Лернер Э.Ю. кафедра анализа данных и исследования операций отделение фундаментальной информатики и информационных технологий , Eduard.Lerner@gmail.com

1. Цели освоения дисциплины

Прикладная статистика - это наука о методах обработки статистических данных. Знание её основ необходимо каждому прикладному специалисту.

Для изучения курса прикладной статистики студентам необходимо знание основ: теории вероятностей и математической статистики, линейной алгебры, математического анализа.

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

Данная учебная дисциплина включена в раздел " БЗ.ДВ.2 Профессиональный" основной образовательной программы 080500.62 Бизнес-информатика и относится к дисциплинам по выбору. Осваивается на 4 курсе, 7 семестр.

Курс изучается на 4 курсе в 8 семестре. Для изучения требуются знания по курсам "Теория вероятностей и математическая статистика" и "Специальные главы математической статистики", полезно изучение курса по выбору "Анализ временных рядов".

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
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:

1. должен знать:

- методы многомерного статистического анализа, в частности, основы факторного и кластерного анализа

2. должен уметь:

- использовать в профессиональной деятельности основные методы обработки и анализа данных наблюдений и эксперимента

- ориентироваться в сложных взаимодействиях экономических факторов;

3. должен владеть:

- навыками построения и анализа многомерных статистических моделей в экономике;

- представлением о принципах изучения взаимосвязей массовых общественных явлений, об изменениях в социально-экономических процессах.

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. Классификация методов, применяемых в прикладной статистике.	7	1	2	6	0	домашнее задание
2.	Тема 2. Параметрическая проверка гипотез.	7	2	2	6	0	домашнее задание
3.	Тема 3. Непараметрическая проверка гипотез.	7	3	2	6	0	домашнее задание
4.	Тема 4. Анализ таблиц сопряженности.	7	4	2	6	0	контрольная работа домашнее задание
5.	Тема 5. Снижение размерности.	7	5	2	6	0	домашнее задание
6.	Тема 6. Изучение многомерной корреляции данных, классификация с обучением.	7	6	4	8	0	домашнее задание
7.	Тема 7. Анализ структуры связей между объектами, классификация без обучения.	7	7-8	4	12	0	контрольная работа домашнее задание

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
8.	Тема 8. Подготовка к экзамену.	7	9	0	0	0	
.	Тема .	7		0	0	0	
	Итого			18	50	0	

4.2 Содержание дисциплины

Тема 1. Классификация методов, применяемых в прикладной статистике.

лекционное занятие (2 часа(ов)):

Классификация методов, применяемых в прикладной статистике, соотношение с методами эконометрики и анализа временных рядов.

практическое занятие (6 часа(ов)):

Рассмотрение на примерах методов, применяемых в прикладной статистике. Расстояние соотношения с методами эконометрики и анализа временных рядов.

Тема 2. Параметрическая проверка гипотез.

лекционное занятие (2 часа(ов)):

Начальные параметрические критерии для проверки гипотез. Примеры прикладных задач из бизнеса, маркетинга, медицины. Гипотеза о равенстве средних: критерий Стьюдента для одной и двух выборок. Пример: задача формирования ценовых коридоров. Гипотеза о равенстве дисперсий: критерий Фишера, критерий Бартлета. Критерии нормальности: критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий хи-квадрат (Пирсона). Исторический пример: проверка закона Менделя А.Н. Колмогоровым. Упрощенные проверки нормальности по асимметрии и эксцессу. Эмпирические подтверждения ненормальности реальных измерений.

практическое занятие (6 часа(ов)):

Изучение начальных параметрических критериев для проверки гипотез на примерах прикладных задач из бизнеса, маркетинга, медицины. Рассмотрение критерия Стьюдента для одной и двух выборок на примере задачи формирования ценовых коридоров. Разбор на примерах гипотезы о равенстве дисперсий: критерия Фишера, критерия Бартлета, критерия нормальности: критерия Шапиро-Уилка, критерия Колмогорова-Смирнова, критерия хи-квадрат (Пирсона).

Тема 3. Непараметрическая проверка гипотез.

лекционное занятие (2 часа(ов)):

Непараметрические ранговые критерии для проверки гипотез: критерий Уилкоксона, Манна-Уитни, критерий знаков, номинальные, порядковые и количественные переменные. Разновидности средних. Пример: маркетинговое исследование привлекательности продуктов (образовательных услуг).

практическое занятие (6 часа(ов)):

Практическое изучение на примерах критерия Уилкоксона, Манна-Уитни, критерия знаков, номинальных, порядковых и количественных переменных. На примере маркетингового исследования привлекательности продуктов (образовательных услуг) разбор разновидностей средних.

Тема 4. Анализ таблиц сопряженности.

лекционное занятие (2 часа(ов)):

Критерий согласия Пирсона: простая гипотеза, сложная гипотеза. Таблицы сопряженности малой и большой размерности. Парадокс хи-квадрат. Примеры: посещаемость сайтов пользователями Интернет, анализ результатов голосования, маркетинговые исследования.

практическое занятие (6 часа(ов)):

На примерах посещаемости сайтов пользователями Интернет, анализа результатов голосования, маркетинговых исследований изучение критерия согласия Пирсона, простой гипотезы, сложной гипотезы. Разбор на примерах таблицы сопряженности малой и большой размерности, парадокса хи-квадрат.

Тема 5. Снижение размерности.

лекционное занятие (2 часа(ов)):

Метод главных компонент (PCA). Диагонализация ковариационной матрицы. Идея сингулярного разложения. Простой итерационный алгоритм сингулярного разложения. Матрица преобразования к главным компонентам. Остаточная дисперсия. Нормировка после приведения к главным компонентам и до вычисления главных компонент. Факторный анализ и эвристические методы снижения размерности.

практическое занятие (6 часа(ов)):

Разбор на примерах методов снижения размерности матрицы, рассмотренных на лекциях.

Тема 6. Изучение многомерной корреляции данных, классификация с обучением.

лекционное занятие (4 часа(ов)):

Наивный байесовский подход, решающие правила, классификация и метрики. Соотношение с методом максимального правдоподобия, применяемым к смесям распределений. Метод опорных векторов (SVM). Деревья решений, построение классификационных деревьев.

практическое занятие (8 часа(ов)):

Рассмотрение на примерах методов многомерной корреляции данных: наивного байесовского подхода, метода максимального правдоподобия, применяемым к смесям распределений, метода опорных векторов (SVM).

Тема 7. Анализ структуры связей между объектами, классификация без обучения.

лекционное занятие (4 часа(ов)):

Иерархическая группировка и ультраметрика. Быстрые алгоритмы группировки. Задачи выявления структуры. Графы о-сходства, графы максимальной длины, цепные графы. Алгоритмы выделения подмножеств. Выявление структуры основных связей.

практическое занятие (12 часа(ов)):

Проведение на примерах анализа структуры связей между объектами.

Тема 8. Подготовка к экзамену.

Тема .

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. Классификация методов, применяемых в прикладной статистике.	7	1	подготовка домашнего задания	6	домашнее задание
2.	Тема 2. Параметрическая проверка гипотез.	7	2	подготовка домашнего задания	6	домашнее задание
3.	Тема 3. Непараметрическая проверка гипотез.	7	3	подготовка домашнего задания	6	домашнее задание

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
4.	Тема 4. Анализ таблиц сопряженности.	7	4	подготовка домашнего задания	6	домашнее задание
				подготовка к контрольной работе	8	контрольная работа
5.	Тема 5. Снижение размерности.	7	5	подготовка домашнего задания	6	домашнее задание
6.	Тема 6. Изучение многомерной корреляции данных, классификация с обучением.	7	6	подготовка домашнего задания	6	домашнее задание
7.	Тема 7. Анализ структуры связей между объектами, классификация без обучения.	7	7-8	подготовка домашнего задания	6	домашнее задание
				подготовка к контрольной работе	8	контрольная работа
	Итого				58	

5. Образовательные технологии, включая интерактивные формы обучения

Обучение происходит в форме лекционных и лабораторных занятий, а также самостоятельной работы студентов.

Теоретический материал излагается на лекциях. Причем конспект лекций, который остается у студента в результате прослушивания лекции не может заменить учебник. Его цель - формулировка основных утверждений и определений. Прослушав лекцию, полезно ознакомиться с более подробным изложением материала в учебнике. Список литературы разделен на две категории: необходимый для сдачи экзамена минимум и дополнительная литература.

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. Она также включает в себя подготовку к зачету.

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

Тема 1. Классификация методов, применяемых в прикладной статистике.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 2. Параметрическая проверка гипотез.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 3. Непараметрическая проверка гипотез.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 4. Анализ таблиц сопряженности.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

контрольная работа , примерные вопросы:

Выполнение индивидуальных заданий.

Тема 5. Снижение размерности.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 6. Изучение многомерной корреляции данных, классификация с обучением.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

Тема 7. Анализ структуры связей между объектами, классификация без обучения.

домашнее задание , примерные вопросы:

Углубленное изучение литературы по теме. Доработка заданий, выполняемых на лабораторных занятиях.

контрольная работа , примерные вопросы:

Выполнение индивидуальных заданий.

Тема 8. Подготовка к экзамену.

Тема .

Примерные вопросы к экзамену:

По данному курсу предусмотрено проведение экзамена.

Упражнения для самостоятельной работы находятся, в частности, в тексте учебного пособия автора программы http://kek.ksu.ru/EOS/Model/uchebnik/1_4.htm
http://kek.ksu.ru/EOS/Model/uchebnik/2_2.htm (учебное пособие расположено на кафедральном сайте в удобном для чтения и скачивания данных виде, для его просмотра достаточно скопировать вышеприведенные ссылки в браузер).

Примерные вопросы экзамена:

1. Классификация методов применяемых в прикладной статистике, соотношение с методами эконометрики и анализа временных рядов
2. Параметрическая проверка гипотез
3. Нормальные параметрические критерии для проверки гипотез, Примеры прикладных задач из бизнеса, маркетинга, медицины. Гипотеза о равенстве средних: критерий Стьюдента для одной и двух выборок. Пример: задача формирования ценовых коридоров. Гипотеза о равенстве дисперсий: критерий Фишера, критерий Бартлета. Критерии нормальности: Критерий Шапиро-Уилка, критерий Колмогорова-Смирнова, критерий хи-квадрат (Пирсона). Исторический пример: проверка закона Менделя А.Н.Колмогоровым. Упрощенные проверки нормальности по асимметрии и эксцессу. Эмпирические подтверждения ненормальности реальных измерений.
4. Непараметрическая проверка гипотез

5. Непараметрические ранговые критерии для проверки гипотез: Критерий Уилкоксона-Манна-Уитни, критерий знаков. Номинальные, порядковые и количественные переменные. Разновидности средних. Пример: маркетинговое исследование привлекательности продуктов (образовательных услуг).
6. Анализ таблиц сопряженности. Критерий согласия Пирсона: простая гипотеза, сложная гипотеза. Таблицы сопряженности малой и большой размерности. Парадокс хи-квадрат. Примеры: посещаемость сайтов пользователями Интернет, анализ результатов голосования, маркетинговые исследования.
7. Снижение размерности. Метод главных компонент (PCA). Диагонализация ковариационной матрицы. Идея сингулярного разложения. Простой итерационный алгоритм сингулярного разложения. Матрица преобразования к главным компонентам. Остаточная дисперсия. Нормировка после приведения к главным компонентам и
8. до вычисления главных компонент.
9. Факторный анализ и эвристические методы снижения размерности.
10. Изучение многомерной корреляции данных, классификация с обучением. Наивный байесовский подход, решающие правила, классификация и метрики. Соотношение с методом максимального правдоподобия, применяемым к смесям распределений. Метод опорных векторов (SVM). Деревья решений, построение классификационных деревьев.
11. Анализ структуры связей между объектами, классификация без обучения. Иерархическая группировка и ультраметрика. Быстрые алгоритмы группировки. Задачи выявления структуры. Графы α -сходства, графы максимальной длины, цепные графы. Алгоритмы выделения подмножеств. Выявление структуры основных связей.

7.1. Основная литература:

1. Общая и прикладная статистика: Учебник для студентов высшего проф. образов. / Р.Н.Пахунова, П.Ф.Аскеров, А.В.Пахунов; Под общ. ред. Р.Н.Пахуновой - М.: НИЦ ИНФРА-М, 2013. - 272 с.: 60x90 1/16. - (ВО: Бакалавр.). (п) ISBN 978-5-16-006669-1, 100 экз.
<http://www.znaniyum.com/bookread.php?book=404310>
2. Кашина О.А., Миссаров М.Д. Электронный курс "Анализ данных в среде R", 2013
<http://zilant.kpfu.ru/course/view.php?id=17341>
3. Кашина О.А., Миссаров М.Д. Электронный курс "Статистический анализ данных", 2013
<http://zilant.kpfu.ru/course/view.php?id=17260>

7.2. Дополнительная литература:

1. Прикладная математическая статистика : для инженеров и науч. работников / А.И. Кобзарь .? Москва : Физматлит, 2006 .? 813 с. ; 24 .? (Современные методы в математике) .? На 4-й с. обл. авт.: Кобзарь А.И., д.т.н., проф., лауреат премии СМ СССР .? Предм. указ.: с. 806-813 .? Библиогр.: с. 737-759 (638 назв.) .? ISBN 5-9221-0707-0, 2000.
2. Математическая статистика / А. А. Боровков .? [3-е изд., испр.] .? Москва : Физматлит, 2007 .? 703 с. ; 25 .? Библиогр.: с. 692-697 .? Предм. указ.: с. 701-703 .? ISBN 978-5-94052-141-X (В пер.) , 3000.
3. Статистика : социол. и маркетинг. исслед. : [пер. с англ.] / Дж. Хили ; под общ. ред. А.А. Руденко .? 6-е изд. ? Москва [и др.] ; Санкт-Петербург [и др.] : DiaSoft : Питер, 2005 .? 637 с. : ил. ; 24 .? Загл. и авт. ориг.: Statistics/Joseph F. Healey .? Предм. указ.: с. 635-637 .? ISBN 5-93772-122-5 .? ISBN 5-469-00813-4, 3000.

7.3. Интернет-ресурсы:

- Библиофонд - <http://www.bibliofond.ru/>
Бизнес-портал - aur.ru

Госкомстат - - gks.ru

РОЦИТ - www.rocit.ru/

Энциклопедия - <http://ru.wikipedia.org/>

8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Прикладная статистика" предполагает использование следующего материально-технического обеспечения:

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.

Лекции и практические занятия по дисциплине проводятся в аудитории, оснащенной доской и мелом(маркером).

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

Автор(ы):

Лернер Э.Ю. _____

"__" _____ 201__ г.

Рецензент(ы):

Володин И.Н. _____

Миссаров М.Д. _____

"__" _____ 201__ г.