

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Алтайский государственный университет»

УТВЕРЖДАЮ
Декан (директор)

_____ г.

**Информационные технологии сравнительного анализа
статистических данных
рабочая программа дисциплины (модуля)**

Закреплена за кафедрой **Кафедра математического анализа**

Учебный план 09_03_03_ПИ-4-2017.plx
09.03.03 Прикладная информатика

Форма обучения **очная**

Общая трудоемкость **4 ЗЕТ**

Часов по учебному плану 144

в том числе:

аудиторные занятия 54

самостоятельная работа 90

Виды контроля в семестрах:
зачеты 7

Распределение часов дисциплины по семестрам

Семестр (<Курс>. <Семестр на курсе>)	7 (4.1)		Итого	
	уп	рпд	уп	рпд
Неделя	19			
Вид занятий	уп	рпд	уп	рпд
Лекции	6	6	6	6
Практические	48	48	48	48
Итого ауд.	54	54	54	54
Контактная работа	54	54	54	54
Сам. работа	90	90	90	90
Итого	144	144	144	144

Программу составил(и):

Дронов С.В., к.ф. - м.н., доцент _____

Рецензент(ы):

Баянова Н.В., к.ф. - м.н., доцент _____

Рабочая программа дисциплины

Информационные технологии сравнительного анализа статистических данных

разработана в соответствии с ФГОС:

Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА (уровень бакалавриата) (приказ Минобрнауки России от 12.03.2015г. №207)

составлена на основании учебного плана:

09.03.03 Прикладная информатика

утвержденного учёным советом вуза от _____ протокол № ____.

Рабочая программа одобрена на заседании кафедры

Кафедра математического анализа

Протокол от _____ г. № ____

Срок действия программы: 2017-2018 уч.г.

Зав. кафедрой Саженов А.Н, к.ф. - м.н., доцент _____

Визирование РПД для исполнения в очередном учебном году

Утверждаю: Декан (директор)

_____ г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в _____ - _____ учебном году на заседании кафедры**Кафедра математического анализа**Протокол от _____ г. № ____
Зав. кафедрой Саженок А.Н, к.ф. - м.н., доцент

Визирование РПД для исполнения в очередном учебном году

Утверждаю: Декан (директор)

_____ г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в _____ - _____ учебном году на заседании кафедры**Кафедра математического анализа**Протокол от _____ г. № ____
Зав. кафедрой Саженок А.Н, к.ф. - м.н., доцент

Визирование РПД для исполнения в очередном учебном году

Утверждаю: Декан (директор)

_____ г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в _____ - _____ учебном году на заседании кафедры**Кафедра математического анализа**Протокол от _____ г. № ____
Зав. кафедрой Саженок А.Н, к.ф. - м.н., доцент

Визирование РПД для исполнения в очередном учебном году

Утверждаю: Председатель НМСС

_____ г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в _____ - _____ учебном году на заседании кафедры**Кафедра математического анализа**Протокол от _____ г. № ____
Зав. кафедрой Саженок А.Н, к.ф. - м.н., доцент

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Курс "Информационные технологии сравнительного анализа статистических данных" имеет в основном практическую направленность. Целью его изучения является систематизация и активация знаний, полученных студентами в рамках изучения различных предметов и дисциплин, связанных с обработкой данных практических наблюдений.
-----	--

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:	Б1.В.ДВ.10
2.1	Требования к предварительной подготовке обучающегося:
2.1.1	Студент должен владеть математическим аппаратом дифференцирования и интегрирования, умением работать в многомерных евклидовых и векторных пространствах, быть знакомым с понятиями линейной алгебры (включая теорию собственных чисел и собственных пространств матриц) и аналитической геометрии. Необходима хорошая ориентация в комбинаторных и вероятностных задачах, владение основными комбинаторными и матричными алгоритмами, хороший уровень самостоятельного написания несложных компьютерных программ.
2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:
2.2.1	1. Линейная алгебра и аналитическая геометрия.
2.2.2	2. Математический анализ.
2.2.3	3. Теория вероятностей.
2.2.4	4. Математическая статистика.
2.2.5	5. Прикладная статистика с элементами эконометрики.
2.2.6	6. Офисные приложения.
2.2.7	7. Программирование и практикум на ЭВМ.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ПК-1: способностью проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе	
Знать:	
Уровень 1	
Уровень 2	
Уровень 3	
Уметь:	
Уровень 1	
Уровень 2	
Уровень 3	
Владеть:	
Уровень 1	
Уровень 2	
Уровень 3	

В результате освоения дисциплины обучающийся должен

3.1	Знать:
3.1.1	Разнообразные методы и приемы первоначальной обработки "сырых" статистических данных, включая их нечисловые и бинарные варианты;
3.1.2	Способы исключения грубых ошибок наблюдения;
3.1.3	Методы группировки многомерных данных и их визуализации;
3.1.4	Приемы предварительной, а затем и количественной оценки степени связи показателей различных типов;
3.1.5	Подходы к организации классификации многомерных данных и построения дискриминирующих функций;
3.1.6	Понятие информационной и прогностической силы показателя и способы их сортировки и/или исключения;
3.1.7	Виды и формы статистически обоснованных заключений о характере представленных данных и возможных методах их обработки.
3.2	Уметь:
3.2.1	Производить быстрый прикидочный анализ и осмотр структуры имеющихся данных;
3.2.2	Цензурировать данные, включая несложные многомерные случаи;

3.2.3	Производить группировку визуализацию данных, производя параллельно сокращение их размерности, если это необходимо;
3.2.4	Строить статистически обоснованные классификации данных различными методами и предлагать дискриминационные процедуры вплоть до их существенно нелинейной формы;
3.2.5	Обосновывать предложенные подходы и писать несложные компьютерные программы для их реализации.
3.3	Владеть:
3.3.1	Работы с объемными многомерными статистическими данными.
3.3.2	Решения задач цензурирования и сокращения размерности;
3.3.3	Перевода нечисловых категоризованных данных в числовую форму (включая бинарный случай);
3.3.4	Изучения силы и вида связей как числовых, так и категоризованных показателей. Изучения и обработки четырехпольных таблиц и общих таблиц сопряженности;
3.3.5	Проверки разного рода предположений о структуре изучаемых статистических данных.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
Раздел 1. Общие принципы работы с данными					
1.1	Данные и их виды. Размерность. Способы перевода одного из видов данных в другой. /Лек/	7	2		Л1.1 Л2.1
1.2	Приемы сокращения размерностей и визуализации данных /Лек/	7	2		Л1.1
1.3	Квантификация и задачи информационной ценности классов /Ср/	7	10		Л1.1 Л3.1
1.4	иды наиболее часто встречающихся практических задач обработки данных и подходы к их решению /Лек/	7	2		Л1.1 Л2.1
1.5	Алгоритмы некоторых статистических методик, комбинаторные алгоритмы /Ср/	7	10		Л2.1
Раздел 2. Работа с "сырыми" данными					
2.1	Цензурирование данных /Пр/	7	4		Л1.1 Л3.1
2.2	Построение латентных классов и обработка нечетких категорий. Алгоритм Ростовцева /Ср/	7	10		Л1.1 Л2.1
2.3	Визуализация. Поля корреляции, ЭЛР, коэффициент Пирсона /Пр/	7	4		Л1.1 Л3.1
2.4	Деревья зависимостей и корреляционные плеяды /Ср/	7	10		Л1.1 Л2.1
2.5	Четырехпольные и шестипольные таблицы. Методы их обработки /Пр/	7	4		Л1.1 Л3.1
2.6	Применение таблиц сопряженности для обработки генетической информации. Критерий Ханделя - Манзеля. /Ср/	7	10		Л1.1 Л2.1
2.7	Нечисловые данные. Анализ соответствий /Пр/	7	8		Л1.1 Л3.1
2.8	Применение множественного анализа соответствий к задачам квантификации /Ср/	7	10		Л1.1
2.9	Классические задачи кластерного анализа /Пр/	7	8		Л1.1 Л3.1
2.10	Коэффициенты корреляции для нечисловых данных /Ср/	7	10		Л1.1
Раздел 3. Проверка предположений о структуре данных					
3.1	Полная или частичная однородность данных /Пр/	7	4		Л1.1 Л3.1
3.2	Гипотеза о виде распределения /Пр/	7	4		Л1.1 Л2.1
3.3	Множественная регрессия с ограничениями /Пр/	7	4		Л1.1 Л3.1
3.4	Задачи нелинейной регрессии /Ср/	7	10		Л2.1
3.5	Дисперсионный анализ и post-hoc задача кластерного анализа /Пр/	7	4		Л1.1
3.6	Предположения ДА. Дисперсионный анализ, свободный от вида распределения (ранговый анализ) /Ср/	7	10		Л1.1 Л2.1
3.7	Метод главных компонент и дискриминантный анализ /Пр/	7	4		Л1.1 Л3.1
3.8	/Зачёт/	7	3		

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Контрольные вопросы и задания

Поскольку курс является обобщающим, практическим, то контроль теоретических знаний ограничен одним тестированием.

1. Задача проверки статистических гипотез может быть поставлена с целью:
 - a. Верификации некоторого предположения о явлении, наделенном чертами случайности;
 - b. Подтверждения утверждения синоптика о том, что 10 мая будет хорошая погода;
 - c. Выяснения последствий принятого управленческого решения;
 - d. Обеспечения работой сотрудников института статистических исследований.(правильный ответ а)
2. Статистический критерий – это
 - a. Коллектив критически настроенных экспертов;
 - b. Эталонный образец правильности высказываемого предположения;
 - c. Способ представления отчетности по результатам статистического исследования;
 - d. Правило, указывающее, какую из высказанных гипотез следует принять по результатам наблюдений.(правильный ответ d)
3. Главное при построении наилучшего критерия
 - a. Простота исполнения требуемых им действий;
 - b. Интуитивная прозрачность результатов его применения;
 - c. Минимизация вероятностей все возможных ошибок одновременно;
 - d. Его максимальная мощность.(правильный ответ d)
4. Критическое множество статистического критерия – это
 - a. Совокупность всех аргументов против использования этого критерия;
 - b. Набор всех выборок, на которых он принимает альтернативную гипотезу;
 - c. Множество тех границ, превышение которых приводит к отвержению основной гипотезы;
 - d. Ресурсный запас, показывающий, какое количество раз еще допустимо применение этого критерия.(правильный ответ b)
5. Критерием согласия называют
 - a. Способ выработки единого мнения экспертов;
 - b. Проверку значимости отличий наблюдаемой картины от идеальной;
 - c. Алгоритм проверки гипотезы об отсутствии существенных различий наборов данных;
 - d. Универсальную методику проверки гипотез, принятую на Всемирном конгрессе математиков в 1924 году.(правильный ответ b)
6. Критерий хи-квадрат был предложен
 - a. Р. Фишером;
 - b. Дж. Нейманом;
 - c. К. Пирсоном;
 - d. Э. Пирсоном.(правильный ответ c)
7. Классический вариант критерия хи-квадрат был ориентирован на проверку
 - a. Гипотезы нормальности выборки;
 - b. Гипотезы о виде распределения;
 - c. Гипотезы независимости связанных выборок;
 - d. Гипотезы однородности двух или нескольких рядов числовых данных.(правильный ответ b)
8. Простой гипотезой является
 - a. Та, которую несложно проверить;
 - b. Наиболее понятная из высказанных гипотез;
 - c. Та, которой удовлетворяет лишь одно распределение;
 - d. Гипотеза, формулировка которой укладывается не более, чем в десять слов.(правильный ответ c)
9. Рандомизация критерия заключается в
 - a. Принятии решения случайным образом на границе критического множества;
 - b. Выборе решающего правила из списка возможных путем использования таблицы случайных чисел;
 - c. Добавлении недостающих для принятия решения данных путем их случайного выбора из генеральной совокупности;
 - d. Отказе от использования сомнительных данных и исключении их с использованием далее функции RANDOM в электронных таблицах для их замещения.(правильный ответ а)
10. Статистическая связь обязательно характеризуется
 - a. Наличием точных и понятных закономерностей;
 - b. Немедленным изменением одной из связанных величин при изменении другой;
 - c. Изменением распределения одной из величин при изменении другой;
 - d. Отсутствием каких-либо функциональных связей между изучаемыми величинами.(правильный ответ c)
11. Правильное присвоение числовых меток значениям нечислового показателя позволяет
 - a. Сократить объемы хранимой информации;
 - b. Оценить различия между категориями показателя;
 - c. Применить к нечисловым данным методы непараметрической статистики;

d. Удачно отчитаться перед вышестоящими органами по сбору статистических данных.
(правильный ответ b)

12. Четырехпольные таблицы

- a. Содержат информацию о четырех видах сельскохозяйственных культур;
- b. Применяются для задания совместного распределения двух бинарных показателей;
- c. Дают наглядное представление о взаимодействии четырех изучаемых показателей;
- d. Позволяют в числовом виде представить информацию о поле корреляции изучаемых показателей.

(правильный ответ b)

13. Маркером наличия статистической связи между нечисловыми категоризованными показателями НЕ является

- a. Возможность задать совместное распределение этих показателей с помощью таблицы сопряженности;
- b. Возможность с достаточной точностью предсказать категорию одного из них по заданной категории другого;
- c. Близкий к диагональному вид таблицы сопряженности показателей;
- d. Наличие в каждой строке или каждом столбце таблицы сопряженности доминирующей группы близких категорий.

(правильный ответ a)

14. Черно-белые алгоритмы обработки связанных выборок позволяют

- a. Изучать нечисловые показатели и оцифровывать их;
- b. Изображать данные эксперимента с помощью черно-белого рисунка;
- c. Выявлять наличие и характер связи между показателями;
- d. Задать границы категорий, наилучшим образом проявляющие связь показателей.

(правильный ответ d).

5.2. Темы письменных работ

Не предусмотрены

5.3. Фонд оценочных средств

Представлен файлом с текстами задач по темам курса для самостоятельного решения. Файл расположен в локальной сети университета по адресу P:\courses\DronovSV

Курс представлен также в виде ЭУКД в системе Moodle, где уже предусмотрены контрольные вопросы по каждой из тем

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л1.1	Дронов С.В.	Методы и задачи многомерной статистики: учебник	АлтГУ, 2015	

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л2.1	Дронов С.В.	Математическая статистика	АлтГУ, 2016	http://elibrary.asu.ru/handle/asu/2845

6.1.3. Методические разработки

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л3.1	Дронов С.В.	Задачник по статистике и дополнительным главам вероятности: методическое пособие	Изд-во АлтГУ, 2013	

6.3.1 Перечень программного обеспечения

7.3.1.1	Офисные программы: MS Word, MS Excel или подобные им.
7.3.1.2	Пакет статистической обработки данных класса IBM SPSS - для самоконтроля
7.3.1.3	Программы для анализа соответствий и поиска главных компонент - собственной разработки,
7.3.1.4	Свидетельство о регистрации программы для ЭВМ № 2015612979 от 27 февраля 2015 г (авторы - Дронов С.В. и Сазонова А.С.)

6.3.2 Перечень информационных справочных систем

7.3.2.1	Не нужны
---------	----------

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Аудитория
7.2	Специализированный компьютерный класс с необходимым программным обеспечением

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Данный курс представляет собой систематизацию, обобщение знаний, полученных в рамках предыдущих курсов по

обработке статистических данных. Поэтому лекционный материал содержит исключительно напоминание некоторых моментов, а от обучающегося ожидается прежде всего самостоятельная практическая работа.

В файле 445 - информационные технологии.doc, содержащемся в компьютерной сети университета на диске Р в папке courses\DronovSV, содержатся задачи для решения. Сформулированные задачи решаются в компьютерном классе в течение всего периода освоения курса (учебный семестр).

В этом же файле содержатся небольшие подсказки теоретических моментов, нужных для решения задач по соответствующей теме. Полную теорию можно посмотреть в основном учебнике курса, который в электронном виде содержится в той же папке.

Этот же курс имеется полностью в подробном виде в системе Moodle.

После окончания основной части курса студент получает одну из решавшихся на практике задач. Для получения зачета он должен представить компьютерную программу, написанную на любом из языков программирования, решающую эту задачу. При этом программа должна давать ответ на все вопросы задачи в автоматическом режиме при вводе в нее любых исходных данных в том же формате, в каком были предложены данные в исходной задаче. Объем массивов информации при этом, по возможности, должен быть произвольным, не обязательно совпадающим с данными задачи.

Впрочем, все детали конкретного зачетного задания обсуждаются после его получения с преподавателем в индивидуальном порядке.

Удачи и успехов в освоении курса!