

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.ДВ.01.02.03 АНАЛИЗ ДАННЫХ

Прикладной статистический анализ данных

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

01.04.02 Прикладная математика и информатика

Направленность (профиль)

01.04.02.01 Математическое моделирование

Форма обучения

очная

Год набора

2022

Красноярск 2022

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

к.ф.-м.н., Доцент, Семенова Дарья Владиславовна

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью преподавания дисциплины «Прикладной статистический анализ данных» является углубление знаний прикладной математической статистики, расширение умений и навыков практического решения задач, возникающих в гуманитарных и социально-экономических науках, связанных с обработкой многомерных данных.

1.2 Задачи изучения дисциплины

Задачами дисциплины являются

- освоение методологии разработки и реализации статистических методов обработки и анализа многомерных данных,
- разработка и анализ эффективности статистических методов анализа многомерных данных, а также отработка навыков применения этой методологии в научных исследованиях и решении практических задач обработки данных и математического моделирования, возникающих в гуманитарных и социально-экономических науках;
- формирование умения самостоятельно разбираться в имеющихся концепциях, методах и моделях прикладного статистического анализа данных и применять их для решения прикладных задач, проводить научные исследования в области разработки и применения методов математической статистики.

1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
ПК-1: Способен разрабатывать и исследовать математические модели, методы и алгоритмы по тематике проводимых исследований	
ПК-1.1: Обладает достаточными фундаментальными теоретическими и практическими знаниями математических и естественных наук, основ программирования и информационных технологий для проведения в конкретной области профессиональной деятельности	методы, направленные на решение задач аналитического характера, предполагающих выбор и многообразие актуальных способов решения задач парадигму и основные концепции развития прикладной статистики, современные подходы и методы проведения научных исследований, современные и классические математические модели сложных стохастических процессов особенности и границы применимости современных моделей, аналитических и численных методов решения стохастических задач, знает методы построения новых математических моделей для решения прикладных задач моделирования стохастических процессов. применять классические и современные методы многомерного статистического анализа данных для решения задач в гуманитарных и социально-экономических науках;

	<p>исследовать и разрабатывать математические модели, методы и алгоритмы по тематике проводимых исследований;</p> <p>самостоятельно выбирать эффективные методы решения поставленных задач и разрабатывать новые методы для получения новых научных и прикладных результатов;</p> <p>владеть методами математического и алгоритмического моделирования при решении прикладных задач</p> <p>владеть способностью передавать результат проведенных физико-математических и прикладных исследований в виде конкретных рекомендаций, выраженных в терминах предметной области изучавшегося явления</p> <p>Владеет навыками выполнения научно-исследовательской работы, самостоятельной разработки новых математических моделей стохастических систем и процессов, применения и модификации известных и самостоятельно разработанных математических моделей для получения новых научных и прикладных результатов.</p>
<p>ПК-1.2: Решает научные задачи в соответствии с поставленной целью и в соответствии с выбранной методикой</p>	<p>основные методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных;</p> <p>классические и современные методы, применяемые в многомерном статистическом анализе данных, необходимые и достаточные условия их реализации. современные тенденции и направления в научных исследованиях, проводимых в мире.</p> <p>Формулировать в рамках поставленной цели совокупность взаимосвязанных задач, обеспечивающих ее достижение. Определять ожидаемые результаты решения выделенных задач грамотно использовать программные комплексы и библиотеки при решении статистических задач</p> <p>Представлять и защищать самостоятельно разработанный проект любого типа, включая исследовательскую работу, с обоснованием ресурсов и ограничений при его разработке и реализации. инструментальными средствами для многомерного анализа данных;</p> <p>навыками выбора подходящего математического метода и программного обеспечения для решения прикладных задач;</p> <p>методами разработки и анализа концептуальных и теоретических моделей решаемых научных проблем и задач.</p>

1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

URL-адрес и название электронного обучающего курса: <https://e.sfu-kras.ru/course/view.php?id=11896>.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Сем естр	
		1	2
Контактная работа с преподавателем:	2 (72)		
занятия лекционного типа	1 (36)		
практические занятия	1 (36)		
Самостоятельная работа обучающихся:	1 (36)		
курсовое проектирование (КП)	Нет		
курсовая работа (КР)	Нет		
Промежуточная аттестация (Зачёт) (Экзамен)	1 (36)		

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

		Контактная работа, ак. час.							
№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
				Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
		Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
1. Распределение вероятностей случайных величин									
	1. Предварительные замечания вероятностного и статистического характера.	2							
	2. Дискретные распределения. Непрерывные распределения.	3							
	3. Предварительные замечания об используемом математическом аппарате.			2					
	4. Многомерное нормальное распределение. Распределение линейной комбинации нормально распределенных величин; независимость величин; частные распределения. Условные распределения и множественный коэффициент корреляции. Характеристическая функция; моменты.			2					
	5. Экспоненциальные семейства распределений							4	
2. Оценки параметров распределений вероятностей									

1. Общие замечания. Оценки типа максимального правдоподобия (М-оценки). Линейные комбинации порядковых статистик (L-оценки). Оценки, получаемые в ранговых критериях (R-оценки). Оценки метода минимального расстояния (R-оценки). Байесовское оценивание.	4							
2. Исследование свойств оценок параметров распределений вероятностей по эмпирическим данным			2					
3. Байесовское оценивание			2					
4. Робастность в статистике. Общие основы. Зачем нужны робастные процедуры. Качественная робастность. Количественная робастность. Инфинитезимальные аспекты. Общая идея вычисления робастных оценок. Робастность М-оценок. Параметрическая процедура отбраковки аномальных наблюдений	3							
5. Экспериментальное исследование робастности оценок			2					
6. Решение задач по теме "Оценки параметров распределений"							2	
7. Планирование экспериментов для оценки параметров распределений							2	
3. Методы анализа законов распределения вероятностей случайных величин								
1. Общие критерии согласия	2							
2. Специальные критерии согласия	5							
3. Критерии проверки отклонения распределения от нормального закона			2					
4. Специальные критерии согласия			4					

5. Подбор кривых распределения вероятностей по экспериментальным данным			3					
6. Разработка алгоритмов и программ							8	
7. Зачёт								
4. Проверка гипотез о значениях параметров распределения								
1. Сравнение параметров распределения.	2							
2. Непараметрические критерии однородности статистических данных. Критерии тренда и случайности. Толерантные пределы.	2							
5. Методы исследования связей между случайными величинами								
1. Дисперсионный анализ	2							
2. Корреляционный анализ	3							
3. Регрессионный анализ	4							
4. Дисперсионный анализ			2					
5. Корреляционный анализ			2					
6. Алгоритмы восстановления регрессии. МНК. Непараметрическая регрессия. Многомерная линейная регрессия. Нелинейные обобщения линейной регрессии. Логистическая регрессия			4					
7. Контрольные карты							1	
8. Математико-статистические методы планирования эксперимента							1	
6. Статистические методы анализа многомерных данных								
1. Методы снижения размерности. Методы факторного анализа. Многомерное шкалирование.	2							
2. Методы снижения размерности. Методы факторного анализа.			2					

3. Многомерное шкалирование.			2					
4. Решение задач по теме "Методы снижения размерности"							2	
7. Оценивание плотности и функции вероятности								
1. Оценивание плотности и функции вероятности. Параметрические оценки плотности. Гистограммы и ядерные оценки плотности. Свойства ядерной оценки одномерной плотности Выбор ширины окна. Референтные эвристические правила. Подстановка. Кросс-валидация на основе наименьших квадратов. Кросс-валидация на основе правдоподобия. Бутстраповские методы. Частотная и ядерная оценка функции вероятности. Ядерное оценивание плотности в случае смешанных дискретных и непрерывных данных Построение доверительных интервалов Проклятие размерности. Оценивание условной плотности. Ядерное оценивание условной плотности. Ядерное оценивание функции условного распределения. Ядерное оценивание условной квантили. Модели бинарного выбора и счетных данных.	2							
2. Оценка плотности совместного распределения			5					
3. Разработка алгоритмов построения и исследования непараметрических оценок плотностей вероятности ядерного типа. Программная реализация разработанных алгоритмов							5	
4. Решение задач по теме "Алгоритмы восстановления регрессии"							1	
5. Проект							10	

Bcero	36		36				36	
-------	----	--	----	--	--	--	----	--

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Михайлов Г. А., Войтишек А. В. Численное статистическое моделирование. Методы Монте-Карло: учебное пособие для вузов по направлению подготовки "Прикладная математика"(Москва: Академия).
2. Магнус Я. Р., Нейдеккер Х., Айвазян С. А. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике: перевод с английского(Москва: Физматлит).
3. Кендалл М., Стьюарт А., Колмогоров А. Н. Теория распределений: перевод с английского(Москва: Наука. Главная редакция физико-математической литературы [Физматлит]).
4. Дэйвисон М. Л. Многомерное шкалирование. Методы наглядного представления данных: перевод с английского(Москва: Финансы и статистика).
5. Иберла К., Дубров А. М. Факторный анализ: перевод с немецкого (Москва: Статистика).
6. Андерсон Т. У., Гнеденко Б. В. Введение в многомерный статистический анализ: перевод с английского(Москва: Государственное издательство физико-математической литературы [Физматгиз]).
7. Лапко В. А. Непараметрические методы обработки данных: учебное пособие(Красноярск: ИПЦ КГТУ).
8. Иберла К. Факторный анализ: научное издание(М.: Статистика).
9. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях: учебник для студентов экономических специальностей вузов(Москва: ЮНИТИ-ДАНА).
10. Кендалл М., Стьюарт А., Пресман Э.Л., Ротаря В.И. Многомерный статистический анализ и временные ряды: научное издание(Москва: Наука, Гл. ред. физ.-мат. лит.).
11. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников(Москва: Физматлит).
12. Лемешко Б. Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению(Москва: ООО "Научно-издательский центр ИНФРА-М").
13. Лемешко Б. Ю., Постовалов С. Н., Лемешко С. Б., Чимитова Е. В. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход(Москва: ООО "Научно-издательский центр ИНФРА-М").
14. Михайлов Г.А., Войтишек А.В. Численное статистическое моделирование. Методы Монте-Карло: учебное пособие для студентов вузов.; допущено МО РФ(М.: Академия).
15. Вучков И. Н., Бояджиева Л. Н., Солаков Е. Б. Прикладной линейный регрессионный анализ: перевод с болгарского(Москва: Финансы и статистика).
16. Горькова К. А., Абрамов Ю. Ш. Факторный анализ (метод главных

- компонент): учебное пособие(Ленинград: Ленинградский финансово-экономический институт).
17. Леман Э. Л., Прохоров А. В. Проверка статистических гипотез: перевод с английского(Москва: Наука. Главная редакция физико-математической литературы [Физматлит]).
 18. Лоули Д. Н., Максвелл А. Е. Факторный анализ как статистический метод: перевод с английского(Москва: Мир).
 19. Себер Д. А. Ф., Малютов М. Б. Линейный регрессионный анализ: перевод с английского(Москва: Мир).
 20. Уилкс С., Линник Ю. В. Математическая статистика: перевод с английского(Москва: Наука, Гл. ред. физ.-мат. лит.).
 21. Смит Г. Прикладной регрессионный анализ: Кн. 2: в 2-х кн.: пер. с англ. (Москва: Финансы и статистика).
 22. Соболев И.М. Метод Монте-Карло: научное издание(Москва: Наука).
 23. Лапко А.В., Лапко В.А. Непараметрические системы обработки неоднородной информации(Новосибирск: Наука).
 24. Носко В. П. Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов: монография(Москва: Институт экономики переходного периода (ИЭПП)).
 25. Айвазян С. А., Фантацини Д. Эконометрика - 2: продвинутый курс с приложениями в финансах: Учебник(Москва: Издательство "Магистр").
 26. Айвазян С. А. Методы эконометрики: Учебник(Москва: Издательство "Магистр").
 27. Лапко В. А. Методы и модели передачи данных: методические указания к лабораторным работам для студентов специальности 230201 "Информационные системы и технологии"(Красноярск: Красноярский университет цветных металлов и золота [ГУЦМиЗ]).
 28. Лапко В. А. Теория принятия решений. Непараметрические модели восстановления стохастических зависимостей и распознавания образов: метод. указ. по лаб. работам №1-2 для студентов укрупненной группы направления подготовки спец. 230000 (спец. 230102.65)(Красноярск: ИПЦ КГТУ).
 29. Лапко В. А. Компьютерное моделирование систем и статистический анализ данных: методические указания по лабораторным работам (Красноярск: ИПЦ КГТУ).
 30. Лапко А. В., Лапко В. А., Соколов М. И., Молоков В. В., Шерстяных А. К., Шарков М. А., Лапко О. А. Обучающиеся системы принятия решений: электрон. учеб.-метод. комплекс дисциплины(Красноярск: ИПК СФУ).
 31. Воскобойников Ю. Е. Регрессионный анализ данных в пакете Mathcad: учеб. пособие(Санкт-Петербург: Лань).
 32. Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение: нормативно-технический материал(М.: Изд-во стандартов).

33. Семенова Д.В. Многомерный статистический анализ данных: [учеб-метод. материалы к изучению дисциплины для ...01.04.02.06 Прикладная математика и информатика в гуманитарных и социально-экономических науках](Красноярск: СФУ).

4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):

1. Семинарские занятия проводятся в компьютерном классе с удаленным доступом к сети Интернет с установленным программным обеспечением:
2. Visual Studio,
3. система компьютерной вёрстки TeX,
4. Python, R (The R Project for Statistical Computing)
5. Visual Studio Code
6. MS Office,
7. Adobe Acrobat.

4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:

1. научная библиотека СФУ <http://bik.sfu-kras.ru/>;
2. поисковые системы: Google или Яндекс.

5 Фонд оценочных средств

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Используется проектор (для лекций или семинаров), слайды мультимедийных презентаций и компьютеры с предустановленным программным обеспечением и доступ в Интернет.