

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

СОГЛАСОВАНО

Заведующий кафедрой

Кафедра высшей и прикладной
математики (ВПМ_ИМФИ)

наименование кафедры

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий ОП ВО

УТВЕРЖДАЮ

Заведующий кафедрой

Кафедра высшей и прикладной
математики (ВПМ_ИМФИ)

наименование кафедры

С.Г. Мысливец

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий дисциплину

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
АНАЛИЗ ДАННЫХ
ПРИКЛАДНОЙ
СТАТИСТИЧЕСКИЙ АНАЛИЗ
ДАННЫХ**

Дисциплина Б1.В.ДВ.01.02.03 АНАЛИЗ ДАННЫХ
Прикладной статистический анализ данных

Направление подготовки / 01.04.02 Прикладная математика и
специальность информатика Магистерская программа
01 04 02 01 Математическое моделирование

Направленность
(профиль)

Форма обучения

очная

Год набора

2021

Красноярск 2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования с учетом профессиональных стандартов по укрупненной группе

010000 «МАТЕМАТИКА И МЕХАНИКА»

Направление подготовки /специальность (профиль/специализация)

Направление 01.04.02 Прикладная математика и информатика

Магистерская программа 01.04.02.01 Математическое моделирование

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

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

ПК-1:Способен разрабатывать и исследовать математические модели, методы и алгоритмы по тематике проводимых исследований	
Уровень 1	основные методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных;
Уровень 2	классические и современные методы, применяемые в многомерном статистическом анализе данных, необходимые и достаточные условия их реализации.
Уровень 3	современные тенденции и направления в научных исследованиях, проводимых в мире.
Уровень 1	применять классические и современные методы многомерного статистического анализа данных для решения задач в гуманитарных и социально-экономических науках;
Уровень 2	исследовать и разрабатывать математические модели, методы и алгоритмы по тематике проводимых исследований;

Уровень 3	самостоятельно выбирать эффективные методы решения поставленных задач и разрабатывать новые методы для получения новых научных и прикладных результатов;
Уровень 1	инструментальными средствами для многомерного анализа данных;
Уровень 2	навыками выбора подходящего математического метода и программного обеспечения для решения прикладных задач;
Уровень 3	методами разработки и анализа концептуальных и теоретических моделей решаемых научных проблем и задач.

1.4 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Прикладной статистический анализ данных» входит в вариативную часть Блока 1 «Дисциплины (модули)» образовательной программы.

Для освоения дисциплины «Прикладной статистический анализ данных» необходимо изучение следующих дисциплин

Основы метода Монте-Карло

Дисциплина «Прикладной статистический анализ данных» предшествует изучению следующих дисциплин.

Основы метода Монте-Карло

Научно-исследовательский семинар

Выполнение и защита выпускной квалификационной работы

Некорректные задачи

Подготовка к сдаче и сдача государственного экзамена

Преддипломная практика

1.5 Особенности реализации дисциплины

Язык реализации дисциплины Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

<https://e.sfu-kras.ru/course/view.php?id=11896>

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Семестр	
		1	2
Общая трудоемкость дисциплины	4 (144)	1,5 (54)	2,5 (90)
Контактная работа с преподавателем:	2 (72)	1,06 (38)	0,94 (34)
занятия лекционного типа	1 (36)	0,53 (19)	0,47 (17)
занятия семинарского типа			
в том числе: семинары			
практические занятия	1 (36)	0,53 (19)	0,47 (17)
практикумы			
лабораторные работы			
другие виды контактной работы			
в том числе: групповые консультации			
индивидуальные консультации			
иная внеаудиторная контактная работа:			
групповые занятия			
индивидуальные занятия			
Самостоятельная работа обучающихся:	1 (36)	0,44 (16)	0,56 (20)
изучение теоретического курса (ТО)			
расчетно-графические задания, задачи (РГЗ)			
реферат, эссе (Р)			
курсовое проектирование (КП)	Нет	Нет	Нет
курсовая работа (КР)	Нет	Нет	Нет
Промежуточная аттестация (Зачёт) (Экзамен)	1 (36)		1 (36)

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа (акад. час)	Занятия семинарского типа		Самостоятельная работа, (акад. час)	Формируемые компетенции
			Семинары и/или Практические занятия (акад. час)	Лабораторные работы и/или Практикумы (акад. час)		
1	2	3	4	5	6	7
1	Распределение вероятностей случайных величин	5	4	0	2	ПК-1
2	Оценки параметров распределений вероятностей	7	6	0	4	ПК-1
3	Методы анализа законов распределения вероятностей случайных величин	7	9	0	10	ПК-1
4	Проверка гипотез значениях параметров распределения	4	0	0	0	ПК-1
5	Методы исследования связей между случайными величинами	9	8	0	6	ПК-1
6	Статистические методы анализа многомерных данных	2	4	0	2	ПК-1
7	Оценивание плотности функции вероятности	2	5	0	12	ПК-1

Всего	36	36	0	36	
-------	----	----	---	----	--

3.2 Занятия лекционного типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	Предварительные замечания вероятностного и статистического характера.	2	0	0
2	1	Дискретные распределения. Непрерывные распределения.	3	0	0
3	2	Общие замечания. Оценки типа максимального правдоподобия (M-оценки). Линейные комбинации порядковых статистик (L-оценки). Оценки, получаемые в ранговых критериях (R-оценки). Оценки метода минимального расстояния (R-оценки). Байесовское оценивание.	4	0	0
4	2	Робастность в статистике. Общие основы. Зачем нужны робастные процедуры. Качественная робастность. Количественная робастность. Инфинитезимальные аспекты. Общая идея вычисления робастных оценок. Робастность M-оценок. Параметрическая процедура отбраковки аномальных наблюдений	3	0	0

5	3	Общие критерии согласия	2	0	0
6	3	Специальные критерии согласия	5	0	0
7	4	Сравнение параметров распределения.	2	0	0
8	4	Непараметрические критерии однородности статистических данных. Критерии тренда и случайности. Толерантные пределы.	2	0	0
9	5	Дисперсионный анализ	2	0	0
10	5	Корреляционный анализ	3	0	0
11	5	Регрессионный анализ	4	0	0
12	6	Методы снижения размерности. Методы факторного анализа. Многомерное шкалирование.	2	0	0

13	7	<p>Оценивание плотности и функции вероятности. Параметрические оценки плотности. Гистограммы и ядерные оценки плотности. Свойства ядерной оценки одномерной плотности</p> <p>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</p> <p>Построение доверительных интервалов</p> <p>Проклятие размерности. Оценивание условной плотности. Ядерное оценивание условной плотности. Ядерное оценивание функции условного распределения. Ядерное оценивание условной квантили. Модели бинарного выбора и счетных данных.</p>	2	0	0
Всего			26	0	0

3.3 Занятия семинарского типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в acad. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме

1	1	Предварительные замечания об используемом математическом аппарате.	2	0	0
2	1	Многомерное нормальное распределение. Распределение линейной комбинации нормально распределенных величин; независимость величин; частные распределения. Условные распределения и множественный коэффициент корреляции. Характеристическая функция; моменты.	2	0	0
3	2	Исследование свойств оценок параметров распределений вероятностей по эмпирическим данным	2	0	0
4	2	Байесовское оценивание	2	0	0
5	2	Экспериментальное исследование робастности оценок	2	0	0
6	3	Критерии проверки отклонения распределения от нормального закона	2	0	0
7	3	Специальные критерии согласия	4	0	0
8	3	Подбор кривых распределения вероятностей по экспериментальным данным	3	0	0
9	5	Дисперсионный анализ	2	0	0
10	5	Корреляционный анализ	2	0	0
11	5	Алгоритмы восстановления регрессии. МНК. Непараметрическая регрессия. Многомерная линейная регрессия. Нелинейные обобщения линейной регрессии. Логистическая регрессия	4	0	0
12	6	Методы снижения размерности. Методы факторного анализа.	2	0	0

13	6	Многомерное шкалирование.	2	0	0
14	7	Оценка плотности совместного распределения	5	0	0
Всего			26	0	0

3.4 Лабораторные занятия

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
Всего					

4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

	Авторы, составители	Заглавие	Издательство, год
Л1.1	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], 2005
Л1.2	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, 2006
Л1.3	Лапко В. А.	Компьютерное моделирование систем и статистический анализ данных: методические указания по лабораторным работам	Красноярск: ИПЦ КГТУ, 2006
Л1.4	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, 2007

Л1.5		Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение: нормативно-технический материал	М.: Изд-во стандартов, 2004
Л1.6	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, 2017

5 Фонд оценочных средств для проведения промежуточной аттестации

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

6.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Михайлов Г. А., Войтишек А. В.	Численное статистическое моделирование. Методы Монте-Карло: учебное пособие для вузов по направлению подготовки "Прикладная математика"	Москва: Академия, 2006
Л1.2	Магнус Я. Р., Нейдеккер Х., Айвазян С. А.	Матричное дифференциальное исчисление с приложениями к статистике и эконометрике: перевод с английского	Москва: Физматлит, 2002
Л1.3	Кендалл М., Стьюарт А., Колмогоров А. Н.	Теория распределений: перевод с английского	Москва: Наука. Главная редакция физико-математической литературы [Физматлит], 1966
Л1.4	Дэйвисон М. Л.	Многомерное шкалирование. Методы наглядного представления данных: перевод с английского	Москва: Финансы и статистика, 1988
Л1.5	Иберла К., Дубров А. М.	Факторный анализ: перевод с немецкого	Москва: Статистика, 1980

Л1.6	Андерсон Т. У., Гнеденко Б. В.	Введение в многомерный статистический анализ: перевод с английского	Москва: Государственное издательство физико-математической литературы [Физматгиз], 1963
Л1.7	Лапко В. А.	Непараметрические методы обработки данных: учебное пособие	Красноярск: ИПЦ КГТУ, 2001
Л1.8	Иберла К.	Факторный анализ: научное издание	М.: Статистика, 1980
Л1.9	Айвазян С.А., Мхитарян В.С.	Прикладная статистика в задачах и упражнениях: учебник для студентов экономических специальностей вузов	Москва: ЮНИТИ-ДАНА, 2001
Л1.1 0	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., 1976
Л1.1 1	Кобзарь А. И.	Прикладная математическая статистика. Для инженеров и научных работников	Москва: Физматлит, 2006
Л1.1 2	Лемешко Б. Ю.	Критерии проверки отклонения распределения от нормального закона. Руководство по применению	Москва: ООО "Научно-издательский центр ИНФРА-М", 2015
Л1.1 3	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", 2015
Л1.1 4	Михайлов Г.А., Войтишек А.В.	Численное статистическое моделирование. Методы Монте-Карло: учебное пособие для студентов вузов.; допущено МО РФ	М.: Академия, 2006
6.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Вучков И. Н., Бояджиева Л. Н., Солаков Е. Б.	Прикладной линейный регрессионный анализ: перевод с болгарского	Москва: Финансы и статистика, 1987
Л2.2	Горькова К. А., Абрамов Ю. Ш.	Факторный анализ (метод главных компонент): учебное пособие	Ленинград: Ленинградский финансово-экономический институт, 1981

Л2.3	Леман Э. Л., Прохоров А. В.	Проверка статистических гипотез: перевод с английского	Москва: Наука. Главная редакция физико- математической литературы [Физматлит], 1979
Л2.4	Лоули Д. Н., Максвелл А. Е.	Факторный анализ как статистический метод: перевод с английского	Москва: Мир, 1967
Л2.5	Себер Д. А. Ф., Малютов М. Б.	Линейный регрессионный анализ: перевод с английского	Москва: Мир, 1980
Л2.6	Уилкс С., Линник Ю. В.	Математическая статистика: перевод с английского	Москва: Наука, Гл. ред. физ.-мат. лит., 1967
Л2.7	Смит Г.	Прикладной регрессионный анализ: Кн. 2: в 2-х кн.: пер. с англ.	Москва: Финансы и статистика, 1987
Л2.8	Соболь И.М.	Метод Монте-Карло: научное издание	Москва: Наука, 1985
Л2.9	Лапко А.В., Лапко В.А.	Непараметрические системы обработки неоднородной информации	Новосибирск: Наука, 2007
Л2.1 0	Носко В. П.	Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов: монография	Москва: Институт экономики переходного периода (ИЭПП), 2004
Л2.1 1	Айвазян С. А., Фантащини Д.	Эконометрика - 2: продвинутый курс с приложениями в финансах: Учебник	Москва: Издательство "Магистр", 2014
Л2.1 2	Айвазян С. А.	Методы эконометрики: Учебник	Москва: Издательство "Магистр", 2010
6.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	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], 2005

ЛЗ.2	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, 2006
ЛЗ.3	Лапко В. А.	Компьютерное моделирование систем и статистический анализ данных: методические указания по лабораторным работам	Красноярск: ИПЦ КГТУ, 2006
ЛЗ.4	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, 2007
ЛЗ.5	Воскобойников Ю. Е.	Регрессионный анализ данных в пакете Mathcad: учеб. пособие	Санкт- Петербург: Лань, 2011
ЛЗ.6		Статистические методы. Правила определения и методы расчета статистических характеристик по выборочным данным. Часть 1. Нормальное распределение: нормативно- технический материал	М.: Изд-во стандартов, 2004
ЛЗ.7	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, 2017

7 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

Э1	Курс: Многомерный статистический анализ данных	https://e.sfu-kras.ru/course/view.php?id=11896
Э2	Материалы и статьи по математической статистике	http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%B0%D1%82%D0%B5%D0%B3%D0%BE%D1%80%D0%B8%D1%8F:%D0%9C%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0

Э3	Электронный учебник StatSoft	http://statsoft.ru/home/textbook/default.htm
Э4	MachineLearning.ru	http://www.machinelearning.ru
Э5	Общероссийский математический портал Math-Net.Ru	http://www.mathnet.ru/
Э6	Журнал "Прикладная эконометрика"	http://www.applieconometrics.ru/

8 Методические указания для обучающихся по освоению дисциплины (модуля)

Проведение лекционных занятий по дисциплине основывается на активном методе обучения, при котором учащиеся не пассивные слушатели, а активные участники занятия, отвечающие на вопросы преподавателя. Вопросы преподавателя нацелены на активизацию процессов усвоения материала, а также на развитие математического мышления. Преподаватель заранее намечает список вопросов, стимулирующих ассоциативное мышление и установления связей с ранее освоенным материалом.

В ходе реализации учебного процесса по дисциплине проводятся практические занятия. Кратко представляются теоретические темы, затем они закрепляются решением практических заданий, по вопросам, вызывающим затруднения, проводятся консультации. В рамках курса студенты должны получить практические навыки использования современных инструментальных средств и ЭВМ при программной реализации математических моделей, рассматриваемых в рамках данного курса.

В течение каждого семестра студентам необходимо выполнить разные типы заданий:

- конкурсные задания;
- практические работы;
- теоретические задания.

Количество заданий - 6 заданий в каждом семестре. Задания, крайние сроки сдачи заданий и инструкция по сдаче заданий выкладываются на странице курса <https://e.sfu-kras.ru/course/view?id=11896>

Решения следует сдавать через систему Moodle. Для каждого задания указан крайний срок сдачи задания (соответственно, 4, 6, 8, 10, 12, 14 недели семестра)

Виды самостоятельной работы

1. Самостоятельная работа с учебными материалами, разбор изученных теоретических тем разбор решенных на занятии заданий.

2. Самостоятельная работа предполагает изучение изданий из списка основной и дополнительной литературы, а также изучение публикаций по данной предметной области в сети Интернет.

3. Самостоятельная работа предусматривает решение конкурсных заданий. Самостоятельное решение задач включает разработку, кодирование, тестирование и отладку программ реализации одной задачи (по выбору), исследование и сравнительный анализ алгоритмов ее решения.

По самостоятельной работе оформляется отчет в электронном виде. Отчет по СРС размещается в LMS в разделе «Самостоятельная работа». В установленный срок студент загружает в LMS архив, содержащий полностью оформленный отчет и программу решения контрольного домашнего задания. Оценка за задание выставляется с учетом полноты выполнения задания и оформления результатов.

9 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

9.1 Перечень необходимого программного обеспечения

9.1.1	Семинарские занятия проводятся в компьютерном классе с удаленным доступом к сети Интернет с установленным программным обеспечением:
9.1.2	Visual Studio,
9.1.3	система компьютерной вёрстки TeX,
9.1.4	Python, R (The R Project for Statistical Computing)
9.1.5	Visual Studio Code
9.1.6	MS Office,
9.1.7	Adobe Acrobat.

9.2 Перечень необходимых информационных справочных систем

9.2.1	научная библиотека СФУ http://bik.sfu-kras.ru/ ;
9.2.2	поисковые системы: Google или Яндекс.

10 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.