

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

СОГЛАСОВАНО

Заведующий кафедрой

Базовая кафедра
вычислительных и
информационных технологий
(ВиИТ_ФМиИ)

наименование кафедры

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий ОП ВО

УТВЕРЖДАЮ

Заведующий кафедрой

Базовая кафедра вычислительных
и информационных технологий
(ВиИТ_ФМиИ)

наименование кафедры

д.ф.-м.н., профессор Шайдуров
В.В.

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий дисциплину

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
АНАЛИЗ НЕКОМПЛЕКТНЫХ
ДАнных

Дисциплина Б1.В.ДВ.05.03 Анализ некомплектных данных

Направление подготовки /
специальность 02.03.01 Математика и компьютерные науки
Профиль 02.03.01.31 Математическое и
компьютерное моделирование

Направленность
(профиль)

Форма обучения очная

Год набора 2020

Красноярск 2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

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

020000 «КОМПЬЮТЕРНЫЕ И ИНФОРМАЦИОННЫЕ НАУКИ»

Направление подготовки /специальность (профиль/специализация)

Направление 02.03.01 Математика и компьютерные науки Профиль

02.03.01.31 Математическое и компьютерное моделирование

Программу
составили

к.ф.н., доцент, Олейников Б.В.

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Основная цель преподавания дисциплины - изучение данных, содержащих пропуски (некомплектных данных), и способов приведения их к комплектным.

1.2 Задачи изучения дисциплины

Основными задачами изучения дисциплин являются:

- изучение требований к данным и способов представления данных;
- изучение причин появления пропусков в данных (появление некомплектных данных);
- изучение методов анализа данных с пропусками (некомплектных данных), направленных на их восстановление;
- изучение доступных пакетов программ для работы с некомплектными данными.

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

ПК-1:Способен применять в научно-исследовательской деятельности базовые знания математических и естественных наук, основ программирования и информационных технологий	
ПК-1.1:Применяет теоретические и практические знания математических и естественных наук, основ программирования и информационных технологий при проведении исследований в конкретной области профессиональной деятельности	
Уровень 1	- основные понятия и разделы научной дисциплины, её базовые идеи; - основные понятия связанные с анализом некомплектных данных; - основные приоритетные направления и критические технологии в научно- исследовательской и практической работе
Уровень 1	- анализировать информационные источники (сайты, форумы, периодические издания); - ориентироваться в круге основных проблем и использовать методы анализа и синтеза для получения новых научных знаний; - адаптировать задачи из различных областей науки и практики для представления их в терминах дисциплины;
Уровень 1	- навыками организации самообразования, технологиями приобретения, использования и обновления профессиональных знаний; - способностью к освоению новых алгоритмов и программ в рамках тематики дисциплины; - навыками систематизации и выбора необходимой информации

	согласно поставленной задаче; - навыками использования основных методов математического моделирования;
--	---

1.4 Место дисциплины (модуля) в структуре образовательной программы

Для успешного освоения дисциплины необходимо знание основных сведений, даваемых в курсе линейной алгебры и курсе теории вероятностей и мат. статистики.

Основные сведения, получаемые в рамках изучения дисциплины «Анализ некомплектных данных» могут быть использованы при изучении многих дисциплин профессиональной и вариативной частей учебного плана программы подготовки бакалавров, включая «Численные методы», «Линейная алгебра», «Математическое моделирование», «Анализ данных» и др., а также при выполнении соответствующих лабораторных работ и научных исследований в рамках подготовки выпускных работ.

1.5 Особенности реализации дисциплины

Язык реализации дисциплины Русский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	Семестр
		4
Общая трудоемкость дисциплины	3 (108)	3 (108)
Контактная работа с преподавателем:	1,89 (68)	1,89 (68)
занятия лекционного типа	0,94 (34)	0,94 (34)
занятия семинарского типа		
в том числе: семинары		
практические занятия	0,94 (34)	0,94 (34)
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:	1,11 (40)	1,11 (40)
изучение теоретического курса (ТО)		
расчетно-графические задания, задачи (РГЗ)		
реферат, эссе (Р)		
курсовое проектирование (КП)	Нет	Нет
курсовая работа (КР)	Нет	Нет
Промежуточная аттестация (Зачёт)		

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	Введение в некомплектные данные. Причины некомплектности данных. Качество и типы данных	6	2	0	0	
2	Введение в теорию вероятностей	10	8	0	0	
3	Введение в мат статистику и работу с базовым стат пакетом	10	8	0	0	
4	Локальные алгоритмы восстановления пропусков	4	6	0	0	
5	Многомерные алгоритмы восстановления пропусков. Алгебраический подход	4	10	0	40	
Всего		34	34	0	40	

3.2 Занятия лекционного типа

№	№ раздела	Наименование занятий	Объем в акад. часах
---	-----------	----------------------	---------------------

п/п	дисциплины		Всего	в том числе, в инновационной форме	в том числе, в электронной форме
1	1	Наличие пропусков в данных (причины, механизмы) и проблема анализа некомплектных данных. Пути решения проблемы анализа некомплектных данных. Требование качества данных	2	0	0
2	1	Основы теории измерений. Типы исходных величин	4	0	0
3	2	Основы теории вероятностей	10	0	0
4	3	Основы мат статистики и работы с базовым стат пакетом, включающим модуль по работе с данными	10	0	0
5	4	Основные модели данных, содержащих пропуски (MAR, MCAR, NMAR). Количественные и качественные данные. Подходы к восстановлению, оценка погрешности и чувствительности	2	0	0
6	4	Простые методы анализа некомплектных данных. Удаление пробельных данных и проблема уменьшения выборки. Заполнения средними. Заполнение подбором. Параметрические (регрессия) и непараметрические (бутстреппинг, hot-deck) методы замены пропусков	1	0	0

7	4	Использование методов кластерного анализа и интерполяции. Байесовское оценивание, МП-и EM-алгоритмы	1	0	0
8	5	Многомерная регрессия. Алгоритмы семейства Zet (Wanga)	2	0	0
9	5	Метод максимального правдоподобия и FIML-алгоритм	1	0	0
10	5	Множественное вменение - Multiple Imputation (MI-алгоритм)	1	0	0
Всего			24	0	0

3.3 Занятия семинарского 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	Наличие пропусков в данных (причины, механизмы) и проблема анализа некомплектных данных. Пути решения проблемы анализа некомплектных данных.	1	0	0
2	1	Основы теории измерений. Типы исходных величин	1	0	0
3	2	Основы теории вероятностей. Решение задач на случайные события, случайные величины, числовые характеристики случайных величин (одномерных и многомерных), асимптотические теоремы	8	0	0

4	3	Основы мат статистики и работы с базовым стат пакетом, включающим модуль по работе с данными, содержащими пропуски. Работа с выборкой, оценка функций распределений, оценка параметров, построение доверительных интервалов, проверка гипотез. Проведение анализа данных с использованием модулей пакета по основным направлениям	8	0	0
5	4	Основные модели данных, содержащих пропуски (MAR, MCAR, NMAR). Количественные и качественные данные. Подходы к восстановлению, оценка погрешности и чувствительности	2	0	0
6	4	Простые методы анализа некомплектных данных. Удаление пробельных данных и проблема уменьшения выборки. Заполнения средними. Заполнение подбором. Параметрические (регрессия) и непараметрические (бутстрепинг, hot-deck) методы замены пропусков	2	0	0
7	4	Использование методов кластерного анализа и интерполяции. Байесовское оценивание, МП-и EM-алгоритмы	2	0	0
8	5	Многомерная регрессия. Алгоритмы семейства Zet (Wanga)	2	0	0
9	5	Метод максимального правдоподобия и FIML-алгоритм	2	0	0

10	5	Множественное вменение - Multiple Imputation (MI-алгоритм)	2	0	0
11	5	Использование модулей базового пакета для восстановления пропущенных данных	4	0	0
Всего			24	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	Кустицкая Т.А.	Теория вероятностей, математическая статистика и теория случайных процессов: [учеб-метод. материалы к изучению дисциплины для ...01.03.04 - Прикладная математика]	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	Пфанцагль И., Бауманн В., Хубер Г., Овчинников С. В.	Теория измерений: перевод с английского	Москва: Мир, 1976
Л1.2	Литтл Р. Д., Рубин Д. Б.	Статистический анализ данных с пропусками: научное издание	М.: Финансы и статистика, 1990

Л1.3	Гмурман В.Е.	Руководство к решению задач по теории вероятностей и математической статистики: учебное пособие для бакалавров.; рекомендован МО РФ	М.: Юрайт, 2013
6.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Гмурман В.Е.	Теория вероятностей и математическая статистика: учеб. пособие для бакалавров	Москва: Юрайт, 2013
6.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	Кустицкая Т.А.	Теория вероятностей, математическая статистика и теория случайных процессов: [учеб-метод. материалы к изучению дисциплины для ...01.03.04 - Прикладная математика]	Красноярск: СФУ, 2017

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

Э1	An introduction to modern missing data analyses	http://scs.math.yorku.ca/images/6/6d/Enders_jofschoolpsyc.pdf
Э2	Using modern methods for missing data analysis with the social relations model: A bridge to social network analysis	Using modern methods for missing data analysis with the social relations model: A bridge to social network analysis
Э3	IBM SPSS Missing Values 22	http://www.sussex.ac.uk/its/pdfs/SPSS_Missing_Values_22.pdf

8 Методические указания для обучающихся по освоению дисциплины (модуля)

Организация процесса работы по дисциплине (модулю) «Анализ некомплектных данных» помимо аудиторной работы (лекции, практические занятия) предусматривает и самостоятельную работу студентов, которая в основном направлена для

1) самостоятельного изучения теоретического материала, необходимого для усвоения дисциплины и решения индивидуальных задач

2) выполнения индивидуальных заданий с использованием базового стат пакета в рамках проведения практических занятий

Литература для самостоятельного изучения теоретического материала приводится в данной рабочей программе. Организация самостоятельной работы производится в соответствии с графиком учебного процесса и самостоятельной работы, который доводится преподавателем до студентов.

Формой промежуточного контроля по дисциплине является экзамен. Экзамен проводится в устной форме. Студенту предлагается билет, состоящий из двух теоретических вопросов. В фонде оценочных средств приведен список вопросов и примеры экзаменационных билетов. При недостаточно полном ответе студенту могут быть заданы дополнительные вопросы.

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

9.1 Перечень необходимого программного обеспечения

9.1.1	MS Power Point 2007 или позднее
9.1.2	Embarcadero Delphi 2009 или более поздних версий Возможно (особенно в рамках самостоятельной работы студентов) использование PascalABC.Net, Microsoft Excel 2007
9.1.3	Пакет программ статистического анализа данных: SPSS.

9.2 Перечень необходимых информационных справочных систем

9.2.1	Не требуются.
-------	---------------

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

Материально-техническое обеспечение образовательного процесса должно включать.

- 1) Компьютерные классы, с рабочим местом преподавателя, подключенные к сети Интернет.
- 2) Средства интерактивного отображения и взаимодействия (видеопроектор, интерактивная доска).
- 3) Wi-Fi сеть для организации взаимодействия с мобильными устройствами обучающихся.
- 4) Желательна организация сетевого интерактивного взаимодействия, предусматриваемого, например, такими технологиями, как Smart class.
- 5) Кроме этого каждый студент должен иметь средства индивидуального хранения информации (flash-карта, portable disk и т.п.)
- 6) Для выполнения самостоятельных работ желательно, чтобы каждый студент имел собственный индивидуальный компьютер.