

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

Б1.В.ДВ.02.03 Анализ некомплектных данных

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

02.03.01 Математика и компьютерные науки

Направленность (профиль)

02.03.01.31 Математическое и компьютерное моделирование

Форма обучения

очная

Год набора

2022

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

к.ф.н., доцент, Олейников Б.В.

должность, инициалы, фамилия

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

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
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	<ul style="list-style-type: none">- 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- 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;- 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 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)	е
		1
Контактная работа с преподавателем:	1,89 (68)	
занятия лекционного типа	0,94 (34)	
практические занятия	0,94 (34)	
Самостоятельная работа обучающихся:	1,11 (40)	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	

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. Введение в некомплектные данные. Причины некомплектности данных. Качество и типы данных									
	1. Наличие пропусков в данных (причины, механизмы) и проблема анализа некомплектных данных. Пути решения проблемы анализа некомплектных данных. Требование качества данных	2							
	2. Основы теории измерений. Типы исходных величин	4							
	3. Наличие пропусков в данных (причины, механизмы) и проблема анализа некомплектных данных. Пути решения проблемы анализа некомплектных данных.			1					
	4. Основы теории измерений. Типы исходных величин			1					
2. Введение в теорию вероятностей									
	1. Основы теории вероятностей	10							
	2. Основы теории вероятностей. Решение задач на случайные события, случайные величины, числовые характеристики случайных величин (одномерных и многомерных), асимптотические теоремы			8					

3. Введение в мат статистику и работу с базовым стат пакетом								
1. Основы мат статистики и работы с базовым стат пакетом, включающим модуль по работе с данными	10							
2. 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					
4. Локальные алгоритмы восстановления пропусков								
1. Основные модели данных, содержащих пропуски (MAR, MCAR, NMAR). Количественные и качественные данные. Подходы к восстановлению, оценка погрешности и чувствительности	2							
2. 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	1							
3. Использование методов кластерного анализа и интерполяции. Байесовское оценивание, МП-и EM-алгоритмы	1							
4. Основные модели данных, содержащих пропуски (MAR, MCAR, NMAR). Количественные и качественные данные. Подходы к восстановлению, оценка погрешности и чувствительности			2					

5. 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					
6. Использование методов кластерного анализа и интерполяции. Байесовское оценивание, МП-и EM-алгоритмы			2					
5. Многомерные алгоритмы восстановления пропусков. Алгебраический подход								
1. Многомерная регрессия. Алгоритмы семейства Zet (Wanga)	2							
2. Метод максимального правдоподобия и FIML-алгоритм	1							
3. Множественное вменение - Multiple Imputation (MI-алгоритм)	1							
4. Многомерная регрессия. Алгоритмы семейства Zet (Wanga)			2					
5. Метод максимального правдоподобия и FIML-алгоритм			2					
6. Множественное вменение - Multiple Imputation (MI-алгоритм)			2					
7. Использование модулей базового пакета для восстановления пропущенных данных			4					
8. Самостоятельная работа по изучению материалов курса							40	
Всего	34		34				40	

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Пфанцагль И., Бауманн В., Хубер Г., Овчинников С. В. Теория измерений: перевод с английского(Москва: Мир).
2. Литтл Р. Д., Рубин Д. Б. Статистический анализ данных с пропусками: научное издание(М.: Финансы и статистика).
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистики: учебное пособие для бакалавров.; рекомендован МО РФ(М.: Юрайт).
4. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для бакалавров(Москва: Юрайт).
5. 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).

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

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

1. Не требуются.

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)

Материально-техническое обеспечение образовательного процесса должно включать.

Компьютерные классы, с рабочим местом преподавателя, подключенные к сети Интернет.

Средства интерактивного отображения и взаимодействия (видеопроектор, интерактивная доска).

Wi-Fi сеть для организации взаимодействия с мобильными устройствами обучаемых.

Желательна организация сетевого интерактивного взаимодействия, предусматриваемого, например, такими технологиями, как Smart class.

Кроме этого каждый студент должен иметь средства индивидуального хранения информации (flash-карта, portable disk и т.п.)

Для выполнения самостоятельных работ желательно, чтобы каждый студент имел собственный индивидуальный компьютер.