



Преподавание сквозных цифровых технологий в
рамках дисциплины

Программные средства анализа и обработки
медико-биологических данных

по специальности 32.05.01 Медико-профилактическое
дело

по специализации Медико-профилактическое дело

Шайхутдинова Асия Равильевна
доцент кафедры медицинской и биологической
физики с информатикой и медицинской
аппаратурой КГМУ

asia.shaikhutdinova@kazan-gmu.ru

Цель и задачи дисциплины



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

• ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

Цель дисциплины:

Освоение теоретических и практических знаний и приобретение практических умений и навыков в области использования среды программирования R и Python для анализа медико-биологических **данных** в рамках решения профессиональных задач с использованием **новых коммуникационных интернет-технологий и работы в интеллектуальных системах поддержки принятия решений**.

Задачи дисциплины:

1. Познакомить обучающихся с особенностями медико-биологических **данных**.
2. Научить пользоваться средой R для проведения анализа данных и их визуализации с использованием **новых коммуникационных интернет-технологий**.
3. Дать представление о базовых концепциях программирования на Python, основах проведения научных вычислений Python с использованием библиотек pandas, scipy, scikit-learn, о том как подобные вычисления могут использоваться в **интеллектуальных системах поддержки принятия решений**.



Процесс изучения дисциплины направлен на формирование следующих компетенций: ИУК-1.2., ИУК-1.5., ИОПК-7.5., ИОПК-12.2.

№ п/п	Код и наименование компетенции	Код и наименование индикатора достижения компетенции (ИДК)	Результаты обучения по дисциплине
1	2	3	4
1	УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	ИУК-1.2. Идентифицирует проблемные ситуации и выдвигает версии решения проблемы, формулирует гипотезу, предполагает конечный результат	Знать: что такое медико-биологические данные. Способы их представления и характерные отличия. Уметь: формулировать гипотезы на основе полученных данных, предлагает способы их решения. Владеть: навыками представления научных данных для подтверждения выдвинутых гипотез.
2	УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	ИУК-1.5. Проводит критическую оценку информации в цифровой среде, ее достоверность	Знать: что такое информация и способы оценки информации, ее достоверности. Понятие большие данные, использование их для работы интеллектуальных систем поддержки принятия решений. Уметь: строить логические умозаключения на основании поступающих данных, используя новые коммуникационные интернет-технологии. Владеть: навыками внесения, обработки, структурирования баз данных с помощью разных программных средств (например среды программирования R), обработки научной графики.
3	ОПК -7 Способен применять современные методики сбора и обработки информации, проводить статистический анализ и интерпретировать результаты, изучать, анализировать, оценивать тенденции, прогнозировать развитие событий и состояние популяционного здоровья населения	ИОПК-7.5. Владеет программными средствами для анализа и обработки медико-биологических данных.	Знать: основы работы в среде программирования (R, Python) для анализа данных, основные понятия, относящиеся к искусственному интеллекту, основы работы интеллектуальных систем поддержки принятия решений. Уметь: применять на практике универсальные пакеты прикладных программ для обработки медико-биологических данных, включая большие данные. Владеть: навыками сбора, анализа, очистки, статистической обработки данных, графического представления данных.
4	ОПК-12 Способен применять информационные технологии в профессиональной деятельности и соблюдать правила информационной безопасности	ИОПК-12.2. Соблюдает правила информационной безопасности в профессиональной деятельности	Знать: основы информационной безопасности при работе с медико-биологическими данными. Уметь: работать с медико-биологическими данными в глобальных сетях с учетом правил информационной безопасности. Владеть: основами защиты данных при работе в среде программирования R, Python, платформе moodle.

Структура дисциплины и объем тематических разделов по видам учебных занятий, индикаторы достижения формируемых компетенций, соотнесенные с разделами дисциплины:



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

№ раздела	Наименование раздела дисциплины	Очная форма обучения			СРС, час.	Код ИДК
		Лекции, час.	Занятия семинарского типа, час.			
			Семинары, практические занятия и др.	Лабораторные работы, лабораторные практикумы		
1.	Особенности медико-биологических данных. Способы получения. Структура.	2	3		2	ИУК-1.2., ИУК-1.5., ИОПК-12.2.
2.	Основы работы в среде программирования R	4	15		15	ИУК-1.2., ИУК-1.5., ИОПК-7.5., ИОПК-12.2.
3.	Основы проведения научных вычислений Python с использованием библиотек pandas, scipy, scikit-learn	4	12		15	ИУК-1.2., ИУК-1.5., ИОПК-7.5., ИОПК-12.2.
	Итого:	10	30		32	

Содержание дисциплины

№ раздела	Наименование раздела дисциплины	Тема занятия. Краткое содержание
1	Особенности медико-биологических данных.	Информация в медицине. Понятие. Объективность. Достоверность. Доступность. Актуальность. Меры. Большие данные.
2	Язык R. Преимущества и недостатки. Начало работы.	<p>Установка RStudio. Интерфейс RStudio. Подключение пакетов. Основы организации переменных в RStudio. Структуры данных. Векторы. Матрицы. Таблицы данных. Массивы. Датафреймы. Извлечение данных. Импорт данных из разных программ. Визуализация данных. Создание диаграмм. Графические параметры. Символы. Линии. Цвета. Заголовки. Объединение диаграмм.</p> <p>Сортировка данных. Объединение наборов данных. Математические и статистические функции. Использование статистических пакетов в R. Описательные статистики. Таблицы частот и таблицы сопряженности. Корреляция и ковариация. Проверка гипотез. Тесты Стьюдента. Непараметрические методы. Визуализация групповых различий. Линейная регрессия. Работа с базами данных. Использование новых коммуникационных интернет-технологий для поиска данных. Использование результатов обработки данных для рекомендательных систем и интеллектуальных систем поддержки принятия решений.</p>
3	Основы языка Python	<p>Основы программирования на языке Python: типы данных и методы работы с ними (переменные, листы, словари, кортежи). Особенности типов данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).</p> <p>Введение в модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным. Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных. Работа с данными Gapminder в Pandas (https://www.gapminder.org): разведывательный анализ, преобразование данных. Воссоздание визуализации Gapminder (зависимость продолжительности жизни от дохода). Использование новых коммуникационных интернет-технологий для поиска данных. Использование результатов обработки данных для рекомендательных систем и интеллектуальных систем поддержки принятия решений.</p>



Текущий контроль успеваемости обучающихся

№ п/п	Наименование раздела (темы) дисциплины	Форма текущего контроля	Оценочные средства	Код ИДК
1	2	3	4	5
1	Особенности медико-биологических данных.	Тестирование Презентация	Тесты Темы для мультимедийных презентаций	ИУК-1.2., ИУК-1.5., ИОПК-12.2.
2	Основы работы в среде программирования R	Тестирование Кейс-задачи на написание программ по обработке данных Онлайн-проект	Тесты Задания для отработки навыков программирования Темы для онлайн-проектов	ИУК-1.2., ИУК-1.5., ИОПК-7.5., ИОПК-12.2.
3	Основы проведения научных вычислений Python с использованием библиотек pandas, scipy, scikit-learn новы языка Python	Тестирование Кейс-задачи на написание программ по обработке данных Онлайн-проект	Тесты Задания для отработки навыков программирования Темы для онлайн-проектов	ИУК-1.2., ИУК-1.5., ИОПК-7.5., ИОПК-12.2.

Промежуточная аттестация осуществляется в форме зачета.

Способ проведения промежуточной аттестации:

Первый этап – тестирование

Второй этап – защита онлайн-проектов.

Темы докладов



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

- Системы кодирования и классификации медико-биологической информации.
- Системы медицинского кодирования и классификации. Основные виды классификаций.
- Системы классификации. Международная классификация болезней.
- Телемедицина, понятие и составляющие.
- Особенности сбора медико-биологической информации во время консультации с использованием телемедицины.
- Основные функции и области применения телемедицинских систем.
- Структура электронной истории болезни.
- Основные требования к разработке электронных историй болезни.
- Защита информации.
- Защита медико-биологических данных.
- Ресурсы по поиску доказательной медицинской информации.
- Критерии доказательности медико-биологической информации.



Задания для отработки навыков программирования



Казанский Государственный
Медицинский Университет

INNOPOLIS
UNIVERSITY

ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР

1. Загрузите RStudio. Загрузите исходные данные. Постройте простой график $\sin(x)$ в диапазоне от -4π до $+4\pi$. Сделайте подпись, надпись, обозначение графиков. Добавьте к существующему графику сетку с шагом $\pi/2$. Далее постройте график произвольной функции из предложенного списка. Оформите его. Создайте вектор из названий профессий (диагнозов). В векторе должно быть не меньше 6 разных наименований. Длина вектора не меньше 12. Постройте круговую диаграмму. Посчитайте процент каждого наименования. Выведите полученные данные на диаграмму.
2. Работа с данными Gapminder в Pandas (<https://www.gapminder.org>): разведывательный анализ, преобразование данных. Воссоздание визуализации Gapminder (влияние дохода на продолжительность жизни как пример или влияние курения на здоровья).





Структура проекта

Онлайн проект состоит из трех этапов:

1. Информационный (планирование и организация работы. Сбор и подготовка **больших данных** к анализу **используя новые коммуникационные интернет-технологий для поиска данных**).
2. Аналитический (анализ данных и их визуализация с помощью среды программирования R или Python).
3. Интерпретация и подготовка отчета с учетом **использования полученных результатов обработки данных для рекомендательных систем и интеллектуальных систем поддержки принятия решений**.

Реализация первого этапа осуществляется в следующей последовательности:

Первый этап.

1. Студенты самостоятельно разбиваются на группы для общекомандной работы по 3 человека. Для общей работы можно использовать Яндекс.Диск.
2. Изучаются компоненты (индикаторы) уровня жизни населения с обоснованием актуальности ее исследования. Например, потребление домашних хозяйств, материальное благосостояние, образование, здравоохранение. Далее следует определиться с количественными переменными, не менее трех.
3. Данные для анализа отбираются по выбранной компоненте по открытым источникам для всех субъектов РФ используя сайты министерств и ведомств.
4. Подготовка данных к анализу включает в себя проверку на полноту и сопоставимость, оценку выбросов, уникальности и дублирования данных.

Второй этап:

Полученный массив данных загрузить в используемую программу. Провести по количественным переменным расчет основных статистик. Рассчитать коэффициент корреляции. Построить гистограмму. Построить регрессионную модель для одной компоненты.

Третий этап.

По полученным данным сделать выводы. Представить презентацию по результатам проделанной работы и сделать предположение об использовании полученных результатов обработки данных для рекомендательных систем и интеллектуальных систем поддержки принятия решений.



1. Онлайн проект студентов в дисциплине **Программные средства анализа и обработки медико-биологических данных** состоит из:
 - 1) Одного этапа
 - 2) Двух этапов
 - 3) Трех этапов
2. Выбор программных продуктов в дисциплине **Программные средства анализа и обработки медико-биологических данных** обоснован тем, что:
 - 1) Программы свободно распространяются
 - 2) Имеют большую поддержку в виде справочников, форумов, сообществ, которые занимаются улучшением работы программ
 - 3) Имеют широкие возможности для анализа медико-биологических задач
 - 4) Все варианты
3. Укажите сквозную цифровую компетенцию, которая не имеет отношения к дисциплине **Программные средства анализа и обработки медико-биологических данных**:
 - 1) Большие данные.
 - 2) Искусственный интеллект (рекомендательные системы и интеллектуальные системы поддержки принятия решений)
 - 3) Новые коммуникационные интернет-технологии
 - 4) Интернет вещей (специализированное оборудование и медицинские изделия при решении профессиональных задач)



INNOPOLIS
UNIVERSITY

● ОПОРНЫЙ
ОБРАЗОВАТЕЛЬНЫЙ
ЦЕНТР



Казанский Государственный
Медицинский Университет

Спасибо
за внимание