

Учреждение образования
«Международный государственный экологический институт
имени А.Д. Сахарова» Белорусского государственного университета

УТВЕРЖДАЮ

Директор

М.Э.И. им. А.Д. Сахарова БГУ

О.И. Родькин

2024

Регистрационный № УД-1514-24/уч.



АНАЛИЗ МЕДИКО-БИОЛОГИЧЕСКИХ ДАННЫХ

Учебная программа учреждения образования
по учебной дисциплине для направления специальности:
1-40 05 01-07 Информационные системы и технологии (в здравоохранении)

2024 г.

Учебная программа дисциплины составлена на основе образовательного стандарта высшего образования ОСВО 1-40 05 01-2021 и учебного плана учреждения высшего образования «Международный государственный экологический институт имени А.Д. Сахарова» БГУ для направления специальности 1-40 05 01-07 Информационные системы и технологии (в здравоохранении) Рег.№130-21/уч. от 14.05.2021

СОСТАВИТЕЛЬ:

В.А. Иванюкович, доцент кафедры информационных технологий в экологии и медицине учреждения образования «Международный государственный экологический институт имени А.Д. Сахарова» БГУ, кандидат физико-математических наук, доцент

РЕЦЕНЗЕНТЫ:

Б.В. Новыш, заведующий кафедрой управления информационными ресурсами Академии управления при Президенте Республики Беларусь, кандидат физико-математических наук, доцент;

Н.А. Савастенко, заведующий кафедрой общей и медицинской физики МГЭИ им. А.Д. Сахарова БГУ, кандидат физико-математических наук, доцент

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой информационных технологий в экологии и медицине учреждения образования «Международный государственный экологический институт имени А.Д. Сахарова» БГУ
(протокол № 9 от 18 апреля 2024 г.);

Научно-методическим советом учреждения образования «Международный государственный экологический институт имени А.Д. Сахарова» БГУ
(протокол № 9 от 21.05 2024 г.)

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Предметом изучения дисциплины являются дизайн медицинских исследований и методы и алгоритмы статистического анализа медико-биологических данных.

Цель изучения дисциплины – сформировать систематизированные знания по методам обработки медицинских и биологических данных, анализа и интерпретации моделей, использующихся в исследованиях медико-биологического характера и познакомить с основными библиотеками языка программирования R для анализа медико-биологических данных.

Задачи учебной дисциплины:

- сформировать у студентов цельное представление об использовании статистических методов и моделей в анализе медико-биологических данных;
- овладеть навыками подготовки данных и проведения статистического анализа.

Учебная дисциплина «Анализ медико-биологических данных» предполагает предварительное изучение дисциплин «Математический анализ», «Линейная алгебра и аналитическая геометрия», «Дискретная математика», «Теория вероятностей и математическая статистика», «Основы алгоритмизации и программирования», «Скриптовые языки программирования», «Статистические методы обработки данных», «Общественное здоровье и здравоохранение».

В результате усвоения этой дисциплины студент должен:

знать:

- особенности и основные виды статистических исследований в медицине;
- основы дизайна исследований;
- основные характеристики данных в предварительном анализе;
- основы унивариантного и бивариантного анализа;

уметь:

- проводить дизайн исследований;
- проводить анализ рисков при множественных исходах и нескольких уровнях фактора, оценивать мощность исследования и проводить расчет объемов выборок;
- представлять данные в графическом виде;
- рассчитывать статистические коэффициенты, проводить анализ выживаемости, оценивать уровень значимости и доверительные интервалы;

владеть:

- современными методами дизайна исследований и статистической обработки медицинских данных;
- специализированными программными продуктами;
- навыками организации статистических вычислений.

Изучение дисциплины «Анализ медико-биологических данных» направлено на приобретение студентами специальной **компетенции**: Разрабатывать дизайн медицинских исследований, владеть статистическими методами доказательной медицины (СК-17).

Учебная программа дисциплины «Анализ медико-биологических данных» относится к модулю «Дисциплины направления», рассчитана на 176 учебных часов, из них 80 аудиторных учебных часов. Примерное распределение аудиторных учебных часов по видам занятий: 32 часа – лекции, 48 часов – практические занятия, самостоятельная работа составляет 96 часов.

Изучение дисциплины предусматривает систематическую самостоятельную работу студентов с литературными источниками. Эффективность самостоятельной работы студентов целесообразно проверять в ходе промежуточного и итогового контроля знаний.

В качестве промежуточного контроля в седьмом семестре предусмотрен экзамен. Текущий контроль осуществляется при проведении практических работ.

Форма получения высшего образования – очная (дневная).

Трудоемкость дисциплины составляет 5 зачетных единиц.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Основы языка программирования R

Инсталляция пакетов и подключение. Переменные. Базовые функции. Векторы и операции с ними. Факторы. Матрицы. Индексирование. Массивы. Датафреймы. Списки. Экспорт и импорт данных. Арифметические функции. Пользовательские функции. Создание функций. Операторы разветвления if, switch. Операторы создания циклов for, while, repeat. Функции семейства apply(). Анализ совокупностей данных. Групповые операции. Базовые графические пакеты graphics и lattice. Специализированные пакеты языка программирования R.

Тема 2. Основные типы дизайна исследований в медицине

Когортное исследование. Исследование случай-контроль. Одномоментное исследование. Экспериментальные или рандомизированные клинические исследования.

Тема 3. Переменные исследования и типы данных. Гипотеза исследования

Переменные исследования. Типы данных в исследованиях. Типы ошибок. Понятие уровня значимости. Мощность исследования и расчет объемов выборок. Этапы проверки статистических гипотез.

Тема 4. Графическое представление данных

Количественные данные: диаграмма частот, ящик с усами. Качественные данные: столбиковые диаграммы, круговые диаграммы, гистограммы.

Тема 5. Унивариантный анализ

Дескриптивные статистики. Сравнение с пороговым значением. Доверительные интервалы. Биноминальная переменная. Мультиномиальная переменная. Количественная переменная.

Тема 6. Бивариантный анализ: взаимосвязь двух переменных

Диаграмма рассеяния. Меры ассоциации. Характеристическая кривая (ROC-кривая).

Тема 7. Таблицы 2×2, их анализ и интерпретация

Чувствительность и специфичность. Оценка мер риска при изучении фактора риска и исхода (отношение рисков и отношение шансов). Оценка эффективности лечения. Обнаружение различий в пропорциях двух групп на основе доверительных интервалов.

Тема 8. Однофакторный дисперсионный анализ, ANOVA

Однофакторная ANOVA. Непараметрическая ANOVA. ANOVA с повторяющимися измерениями.

Тема 9. Простая линейная регрессионная модель

Линейная регрессия с бинарным предиктором. Линейная регрессия с мультиномиальным предиктором. Понятие конфаундера, ковариационный анализ. Взаимодействие предикторов в линейной модели. Преобразование переменных. Интерпретация параметров модели. Анализ остатков модели.

Тема 10. Логистическая регрессия

Логистическая регрессия с бинарным предиктором. Логистическая регрессия с мультиномиальным предиктором. Взаимодействие предикторов в логистической регрессии. Стратифицированный анализ серии таблиц 2×2 . Интерпретация параметров модели. Диагностика логистической регрессии.

Тема 11. Анализ выживаемости

Понятие цензурированных данных, событий и времен наблюдения. Функция выживаемости. Понятие функции риска. Сравнение групп по выживаемости. Оценка выживаемости по Каплан-Майеру. Оценка выживаемости по актуарным таблицам.

Тема 12. Регрессионная модель пропорциональных рисков Кокса

Регрессия Кокса с бинарным предиктором. Регрессия Кокса с количественным предиктором. Взаимодействие предикторов в модели Кокса. Интерпретация параметров модели. Проверка предположений модели регрессии Кокса.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

(очная (дневная) форма получения высшего образования)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов						Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Управляемая самостоятельная работа	Иное	
1	Основы языка программирования R	4	12					1–4
2	Основные типы дизайна исследований в медицине	2						1–4
3	Переменные исследования и типы данных. Гипотеза исследования	2						1–4
4	Графическое представление данных	2						1–4
5	Унивариантный анализ	2	4					1–4
6	Бивариантный анализ: взаимосвязь двух переменных	2	4					1–4
7	Таблицы 2×2, их анализ и интерпретация	2	4					1–4
8	Однофакторный дисперсионный анализ, ANOVA	4	4					1–4
9	Простая линейная регрессионная модель	4	8					1–4
10	Логистическая регрессия	2	4					1–4
11	Анализ выживаемости	2	4					1–4
12	Регрессионная модель пропорциональных рисков Кокса	4	4					1–4
	ИТОГО	32	48					

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Инновационные подходы и методы к преподаванию учебной дисциплины

При изучении дисциплины «Информационные технологии» рекомендуется активно использовать практико-ориентированный подход, метод группового обучения, принципы модульного обучения.

Практико-ориентированный подход предполагает: освоение содержание образования через решения практических задач; приобретение навыков эффективного выполнения разных видов профессиональной деятельности; ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры; использованию процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

Метод группового обучения представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Принцип модульного обучения формулируется из основной идеи модульного обучения – использования модулей как основного средства усвоения учащимися дозы учебной информации. Принцип модульности является основой индивидуализации обучения, поскольку обеспечивает вариативность содержания и способов его усвоения в зависимости от уровня базовой подготовленности учащихся, что в свою очередь развивает в обучающихся умение работать самостоятельно.

Рекомендуемая литература

Основная

1. James G. An Introduction to Statistical Learning with Applications in R. Second Edition / G. James, D. Witten, T. Hastie, R. Tibshirani; – Springer, 2023. – 604 p.
2. Лонг, Дж.Д. Р. Книга рецептов. Проверенные рецепты для статистики, анализа и визуализации. / Дж.Д. Лонг, П.Р. Титор; пер. с англ. Д.А. Беликова. – М.: ДМК Пресс, 2020. – 510 с.: илл.
3. Кабаков, Р. И. R в действии. Анализ и визуализация данных в программе R. Изд. 3 / Роберт И. Кабаков; пер. с англ. П. А. Волковой. – М.: ДМК Пресс, 2023. – 769 с.: илл.
4. James G. An Introduction to Statistical Learning with Applications in R. Second Edition / G. James, D. Witten, T. Hastie, R. Tibshirani; – Springer, 2023. – 604 p.
5. Уэилд Р. Аналитика в Power BI с помощью и Python. / Райан Уэилд. – М.: ДМК Пресс, 2021. – 338 с.: илл.
6. Уикем Х. Р. К вершинам мастерства. / Х. Уикем. – М.: ДМК Пресс, 2024. – 752 с.: илл.

Дополнительная

7. Красько, О.В. Статистический анализ данных в медицинских исследованиях / О.В. Красько. – Электронная версия, – 2018. – 254 с.
8. Джеймс, Г. Введение в статистическое обучение с примерами на языке R. / Г. Джеймс, Д. Уиттон, Т. Хастис, Рю Тибширани. – М.: ДМК Пресс, 2017. – 456 с.: ил.
9. Chain, B.K.C. Biostatistics for epidemiology and public health. / Bertram K.C. Chain. – Springer, 2016. – 446 p.
10. Shahbaba B. Biostatistics with R. / Babak Shahbaba. – Springer, 2012. – 352 p.
11. Флетчер, Р. Клиническая эпидемиология. Основы доказательной медицины / Р. Флетчер, С. Флетчер, Э. Вагнер. – М.: Медиа Сфера, –1998. – 352 с.
12. Шитиков, В.К. Классификация, регрессия и другие алгоритмы Data Mining с использованием R / В.К. Шитиков, С.Э. Мастицкий. – 2017. – 351 с. – Электронный ресурс, адрес доступа: <https://github.com/ranalytics/data-mining>.
13. Гланц, С. Медико-биологическая статистика / С. Гланц. – М.: Практика, – 1998. – 459 с.
14. Agresti, A. Categorical Data Analysis / A. Agresti, – NY: John Wiley & Sons, – 2002. – 734 p.
15. Hosmer D.W. Applied logistic regression / D.W. Hosmer, S. Lemeshow. 2nd ed., – NY: John Wiley & Sons, – 2000. – 376 p.
16. Hosmer, D.W. Applied Survival Analysis / D.W. Hosmer, S. Lemeshow. – NY: John Wiley & Sons, – 1999. – 386 p.

17. Harrell, F.E. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis / F.E. Harrell. – NY: Springer. – 2001. – 568 p.
18. Steyerberg, E.W. Clinical Prediction Models. A Practical Approach to Development, Validation and Updating / Ewout W. Steyerberg, – NY: Springer, – 2009. – 498 p.

Наименования и виды методических средств

№ п/п	Наименование	Вид
1	Лекционные занятия	Аудитория, компьютерная мультимедийная система
2	Практические занятия	Учебная аудитория для групповых занятий

Формы контроля знаний

№ п/п	Форма
1	Выборочный контроль на лекциях
2	Проверка конспектов лекций студентов
3	Собеседование при защите отчетов по практическим занятиям
4	Проведение экзамена по курсу

Примерный перечень тем практических занятий

1. Язык программирования R. Основные операции. Функции.
2. Циклы и ветвления в R. Графика в R.
3. Статистический анализ экспериментальных данных. Пакет Rcmdr.
4. Унивариантный анализ количественных данных.
5. Бивариантный анализ количественных и качественных данных. ROC-анализ.
6. Анализ таблиц 2×2. Таблицы сопряженности.
7. Дисперсионный анализ двух или нескольких групп данных.
8. Линейная модель. Подгонка модели. Сравнение моделей.
9. Линейная модель в дисперсионном анализе.
10. Логистическая регрессия.
11. Анализ выживаемости.
12. Регрессионная модель Кокса.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ ПО ИЗУЧАЕМОЙ УЧЕБНОЙ
ДИСЦИПЛИНЕ С ДРУГИМИ ДИСЦИПЛИНАМИ СПЕЦИАЛЬНОСТИ

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения кафедры об изменениях в содержании учебной программы учреждения высшего образования по изучаемой дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Согласование с другими дисциплинами не требуется			