Федеральное государственное бюджетное образовательное учреждение

высшего образования «УФИМСКИЙ ГОСУДАРСТВЕННЫЙ

АВИАЦИОННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Центр дополнительного образования

УТВЕРЖДАЮ

И.о. проректора по учебной работе

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Р.Д. Еникеев

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

**ПАСПОРТ**

Дополнительной профессиональной программы повышения

квалификации для системы предоставления персональных

цифровых сертификатов от государства на развитие

у трудоспособного населения компетенций цифровой экономики

**«Машинное обучение для обработки медико-биологических данных»**

**72 академических часа**

**Уфа 2020**

|  |  |
| --- | --- |
| **Версия программы** | **1** |
| **Дата Версии** | **13.10.2020** |

1. **Сведения о Провайдере**

|  |  |  |
| --- | --- | --- |
| 1.1 | Провайдер | Федеральное государственное бюджетное образовательное учреждение высшего образования Уфимский государственный авиационный технический университет (ФГБОУ ВО УГАТУ) |
| 1.2 | Логотип образовательной организации |  |
| 1.3 | Провайдер ИНН | 0274023747 |
| 1.4 | Ответственный за программу ФИО | Лакман Ирина Александровна |
| 1.5 | Ответственный должность | Доцент кафедры БМИ УГАТУ |
| 1.6 | Ответственный Телефон | +7-927-9655655 |
| 1.7 | Ответственный Е-mail | Lackmania@mail.ru |

1. **Основные Данные**

|  |  |  |
| --- | --- | --- |
| **№** | **Название** | **Описание** |
| 2.1 | Название программы | Машинное обучение для обработки медико-биологических данных |
| 2.2 | Ссылка на страницу программы | https://sdo.ugatu.su/course/view.php?id=5615 |
| 2.3 | Формат обучения | Онлайн |
|  | Подтверждение от ОО наличия возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа | ФГБОУ ВО «УГАТУ» подтверждает наличие возможности реализации образовательной программы с применением электронного обучения и (или) дистанционных образовательных технологий с возможностью передачи данных в форме элементов цифрового следа. |
| 2.4 | Уровень сложности | Базовый |
| 2.5 | Количество академических часов | 72 |
|  | Практикоориентированный характер образовательной программы: не менее 50 % трудоёмкости учебной деятельности отведено практическим занятиям и (или) выполнению практических заданий в режиме самостоятельной работы | Более 50 % трудоёмкости учебной деятельности отведено практическим занятиям и (или) выполнению практических заданий в режиме самостоятельной работы |
| 2.6 | Стоимость обучения одного обучающегося по образовательной программе, а также предоставление ссылок на 3 (три) аналогичные образовательные программы иных организаций, осуществляющих обучение, для оценки объективности стоимости или обоснование уникальности представленной образовательной программы в случае отсутствия аналогичных образовательных программ на рынке образовательных услуг | 24000 руб.  Проведённый обзор показал, что курсов на 72 часа не существует, есть курс, затрагивающий часть вопросов, на 24 часа – 15000 рублей <https://nmfo.ru/index.php/obuchenie/meropriyatiya/373-onlajn-kurs-osnovy-statisticheskoj-obrabotki-meditsinskikh-dannykh>  Курс, затрагивающий часть вопросов на 32 часа – 20000 рублей <https://bioinf.me/education/workshops/stat/2020>  Похожий курс – 80000 руб. (180 ч.) <https://new.geekbrains.ru/data-science-medicine>  **Исходя из анализа предложений, стоимость курса составляет 24000 рублей.** |
| 2.7 | Минимальное количество человек на курсе | 10 |
| 2.8 | Максимальное количество человек на курсе | 100 |
| 2.9 | Данные о количестве слушателей, ранее успешно прошедших обучение по образовательной программе | 35 (по похожему курсу, реализуемому в УГАТУ) |
| 2.10 | Формы аттестации | зачет |
|  | Указание на область реализации компетенций цифровой экономики, к которой в большей степени относится образовательная программа, в соответствии с Перечнем областей | Искусственный интеллект |

1. **Аннотация программы**

На сегодняшний день разработаны современные инструменты машинного обучения, применяемые в основном в медицине – Survival models, Difference-in-Difference и др. Целью реализации программы является формирование новых компетенций в области интеллектуального анализа медицинских и биологических данных, необходимых для проведения специалистами клинических и эпидемиологических исследований. Конечной целью освоения программы слушателями является формирование у будущих специалистов теоретических знаний и практических навыков для решения прикладных задач интеллектуального анализа доказательной медицины, освоение навыков использования инструментов машинного обучения конкретно к медицинским задачам, использование инструментов, применяемых для разработки систем поддержки принятия в медицине и медицинских информационных систем. В итоге после прохождения курсов у слушателя должна быть сформирована компетенция цифровой экономики на базовом уровне: способность управлять информацией и данными медико-биологического характера.

В результате изучения дисциплины слушатель должен:

**знать:** методы дизайна эксперимента в медицине, формирование контрольных групп и групп воздействия; методы проверки параметрических и непараметрических критериев для медицинских исследований; методы построения бинарных регрессий, предпосылки для получения достоверных оценок регрессии различными методами; методы построения таблиц выживаемости, нахождения оценок Каплана Майера, критерии на разделении выживаемости в подгруппах; регрессионные модели выживаемости (Кокса, логнормальную, экспоненциальную, нормальную); методы оценки воздействия в медицине посредством DID-анализа (разности разностей).

**уметь:** проверять гипотезы о различиях выборок в медицине, используя параметрические и непараметрические критерии; строить адекватные модели бинарной регрессии и проводить интерпретацию результатов моделирования на основе маржинальных эффектов влияния факторов на результат; проводить ROC-анализ, рассчитывать показатели специфичности и чувствительности; строить таблицы времен жизни, интерпретировать результаты анализа таблицы времен жизни, анализировать графики функции выживаемости, и функции мгновенного риска, определять оценки выживаемости методом Каплана-Майера, а также разделять и проводить анализ выживаемости по подгруппам; строить различные регрессионные модели выживаемости, проверять их адекватность реальному моделируемому процессу, интерпретировать результаты сточки зрения оценки мультипликативного эффекта для оценки риска недожития; уметь оценить эффективность воздействия в медицине, используя DiD-модель.

**владеть:** интеллектуальной обработки медицинских данных, используя RStudio; построения адекватных моделей бинарной регрессии и качественной интерпретации результатов моделирования; проведения анализа выживаемости, DID-анализа; применения RStudio для проведения анализа медико-биологических данных.

Для успешного прохождения курса слушатели должны на продвинутом уровне пользоваться компьютером, иметь базовые знания о проведении медицинского эксперимента, основные принципы доказательной медицины, знать основы теории вероятности и математической статистики. Для слушателей курсов предусмотрены входные контрольные задания по теории вероятности и математической статистике (описательные дескриптивные статистики и проверка гипотез – ошибки первого и второго рода).

Компетенция, сформированная в рамках прохождения курса, позволит развиться в профессиональной деятельности ИТ-специалистам, медицинским работникам.

Федеральное государственное бюджетное образовательное учреждение

высшего образования «УФИМСКИЙ ГОСУДАРСТВЕННЫЙ

АВИАЦИОННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Центр дополнительного образования

УТВЕРЖДАЮ

И.о. проректора по учебной работе

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Р.Д. Еникеев

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20\_\_\_ г.

Дополнительная профессиональная программа повышения

квалификации для системы предоставления персональных

цифровых сертификатов от государства на развитие

у трудоспособного населения компетенций цифровой экономики

**«Машинное обучение для обработки медико-биологических данных»**

**72 академических часа**

**Уфа 2020**

1. **Цель программы**

дать систематизированное представление о современных подходах к интеллектуального анализа медицинских и биологических данных, необходимых для проведения специалистами клинических и эпидемиологических исследований в современных информационных средах (RStudio). Конечной целью освоения программы слушателями является формирование у будущих специалистов теоретических знаний и практических навыков для решения прикладных задач интеллектуального анализа доказательной медицины, освоение навыков использования инструментов машинного обучения конкретно к медицинским задачам, использование инструментов, применяемых для разработки систем поддержки принятия в медицине и медицинских информационных систем. В итоге после прохождения курсов у слушателя должна быть сформирована компетенция цифровой экономики на базовом уровне: **способность управлять информацией и данными медико-биологического характера.**

1. **Планируемые результаты обучения:**
   1. Знание (осведомленность в областях)

2.1.1. методы дизайна эксперимента в медицине, формирование контрольных групп и групп воздействия;

2.1.2. методы проверки параметрических и непараметрических критериев для медицинских исследований;

2.1.3. методы построения бинарных регрессий, предпосылки для получения достоверных оценок регрессии различными методами;

2.1.4. методы построения таблиц выживаемости, нахождения оценок Каплана Майера, критерии на разделении выживаемости в подгруппах;

2.1.5. регрессионные модели выживаемости (Кокса, логнормальную, экспоненциальную, нормальную);

2.1.6. методы оценки воздействия в медицине посредством DID-анализа (разности разностей).

Умение (способность к деятельности)

2.2.1. проверять гипотезы о различиях выборок в медицине, используя параметрические и непараметрические критерии;

2.2.2. строить адекватные модели бинарной регрессии и проводить интерпретацию результатов моделирования на основе маржинальных эффектов влияния факторов на результат;

2.2.3. проводить ROC-анализ, рассчитывать показатели специфичности и чувствительности;

2.2.4. строить таблицы времен жизни, интерпретировать результаты анализа таблицы времен жизни, анализировать графики функции выживаемости, и функции мгновенного риска, определять оценки выживаемости методом Каплана-Майера, а также разделять и проводить анализ выживаемости по подгруппам;

2.2.5. строить различные регрессионные модели выживаемости, проверять их адекватность реальному моделируемому процессу, интерпретировать результаты сточки зрения оценки мультипликативного эффекта для оценки риска недожития;

2.2.6. уметь оценить эффективность воздействия в медицине, используя DiD-модель.

Навык (использование конкретных инструментов)

2.3.1. интеллектуальной обработки медицинских данных, используя RStudio;

2.3.2. построения адекватных моделей бинарной регрессии и качественной интерпретации результатов моделирования;

2.3.3. проведения анализа выживаемости, DID-анализа;

2.3.4. применения RStudio для проведения анализа медико-биологических данных.

1. **Требования к слушателям**

3.1 Образование: высшее, средне-специальное

3.2 Квалификация: медицинский работник, врач, математик

3.3 Наличию опыта профессиональной деятельности: работа в Excel.

3.4 Предварительное освоение иных дисциплин/курсов /модулей: теория вероятностей и математическая статистика, общая теория медицинской статистики

1. **Учебный план программы «Машинное обучение для обработки медико-биологических данных»**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль** | **Всего, час** | **Виды учебных занятий** | | |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| 0 | **Входное тестирование** | 2 |  |  | **2** |
| 2 | Модуль 1 − Первичная обработка данных. Параметрические и непараметрические тесты | 24 | 6 | 9 | 9 |
| 3 | Модуль 2 – Бинарные модели в медицинских исследованиях, ROC-анализ | 18 | 4 | **8** | 6 |
| 4 | Модуль 3 – Модели анализа выживаемости, оценки таблиц времен жизни | 18 | 4 | **8** | 6 |
| 5 | Модуль 4 – DID-анализ | 6 | 2 | 2 | 2 |
| **Итоговая аттестация** | |  | **Указывается вид (экзамен, зачёт, реферат и т.д.)** | | |
| Итоговый тест | | 4 | Зачет - Тест | | |

**5.Календарный план-график реализации образовательной программы**

(дата начала обучения – дата завершения обучения) в текущем календарном году, указания на периодичность набора групп (не менее 1 группы в месяц)

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Наименование учебных модулей** | **Трудоёмкость (час)** | **Сроки обучения** |
| **0** | **Входное тестирование** | **2** | **1.11.2020** |
| 1 | Модуль 1 − Первичная обработка данных. Параметрические и непараметрические тесты | 24 | 2.11.2020-8.11.2020- |
| 2 | Модуль 2 − Бинарные модели в медицинских исследованиях, ROC-анализ | 18 | 9.11.2020-12.11.2020- |
| 3 | Модуль 3 − Модели анализа выживаемости, оценки таблиц времен жизни | 18 | 13.11.2020-16.11.2020 |
| 4 | Модуль 4 – DID-анализ | 6 | 17.11.2020-18.11.2020 |
|  | Итоговое тестирование | 4 | 19.11.2020 |
| **Всего:** | | 72 | 1.11.2020-19.11.2020 |

**6.Учебно-тематический план программы «Машинное обучение для обработки медико-биологических данных»**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Модуль / Тема** | **Всего, час** | **Виды учебных занятий** | | | **Формы контроля** |
| **лекции** | **практические занятия** | **самостоятельная работа** |
| **0** | **Входное тестирование** | 2 |  |  | 2 | **Тест** |
| 1 | Модуль 1 − Первичная обработка данных. Параметрические и непараметрические тесты | 24 | 6 | 9 | 9 | Кейс 1, 2  Вопросы |
| 2 | Модуль 2 – Бинарные модели в медицинских исследованиях, ROC-анализ | 18 | 4 | 8 | 6 | Кейс 3  **Вопросы** |
| 3 | Модуль 3 – Модели анализа выживаемости, оценки таблиц времен жизни | 18 | 4 | 8 | 6 | Кейс 4  **Вопросы** |
| 4 | Модуль 4 – DID-анализ | 6 | 2 | 2 | 2 | Кейс 5  **Вопросы** |
| 8 | Итоговое тестирование | 4 |  |  | 4 | Тест |

**7. Учебная (рабочая) программа повышения квалификации «Машинное обучение для обработки медико-биологических данных»**

* 1. Модуль 1 «Первичная обработка данных. Параметрические и непараметрические тесты» (24 ак.часов)

*Темы*

Типы данных (качественные и количественные). Методы сводки и группировки. Шкалы измерения данных. Виды статистических наблюдений. Принципы группировки, формула Стержесса. Графическое представление данных. Описательные статистики признака: показатели центра распределения для сгруппированных и несгруппированных данных, порядковые характеристики, показатели вариации, асимметрии и эксцесса. Методы формирования выборок и определение их необходимого объема в зависимости от цели проводимого исследования. Простая случайная, стратификационная и серийная выборки. Показатели связи. Коэффициенты корреляции, ассоциации, контингенции, сопряженности Чупрова и Пирсона. Ранговые показатели связи: коэффициенты Спирмена, Кендалла, конкордации. Биссериальные коэффициенты. Статистические гипотезы проверки значимости показателей связи. Непараметрические методы анализа независимых выборок. Непараметрические тесты о сходстве/различии характеристики для двух групп (метод серий Вальда-Вольфовица, метод Колмогорова Смирнова, U-критерий Манна-Уитни). Непараметрические методы анализа независимых выборок. Непараметрические тесты о сходстве/различии характеристики для трех групп (Краскела-Уоллиса, метод Колмогорова Смирнова, U-критерий Манна-Уитни).

*Задания в виде кейсов 1, 2, вопросы*

* 1. Модуль 2 «Бинарные модели в медицинских исследованиях, ROC-анализ» (18 ак. часов)

*Темы*

Представление о бинарных моделях. Спецификация логит, пробит и гомпит моделей на основе информационных критериев Акайке, Шварца и Ханена-Квина. Применение метода максимального правдоподобия для оценки моделей. Процедура Макфаддена проверки адекватности моделей. Критерий Хосмера-Лемешоу. Применение маржинальных эффектов для оценки результатов моделирования. Проведение ROC-анализа для бинарных моделей. Расчет критериев специфичности и чувствительности модели, определение показателя AUC и коэффициента Джини.

*Задания в виде кейса 3, вопросы*

* 1. Модуль 3 «Модели анализа выживаемости, оценки таблиц времен жизни» (18 ак. часов)

*Темы*

Таблицы времен жизни: частотные таблицы. Оценки Каплана-Майера. Кривые выживаемости. Критерий Вилкокона-Гехана. Сравнение выживаемости в подгруппах. Регрессионные модели выживаемости: Логнормальная регрессия – модели ускоренной жизни; модель Кокса – модель пропорциональных рисков, экспоненциальная модель выживаемости. Проверка адекватности моделей выживаемости. Проверка распределения остатков на соответствие распределению Вейбула.

*Задания в виде кейса 4, вопросы*

* 1. Модуль 4 «DID-анализ» (6 ак. часов)

*Темы*

Метод Разностей разности DiD, предпосылки для проведения методов: формирование контрольной и опытной (воздействия) групп. Оценка методом DiD через линейную регрессию, путем введения фиктивных переменных. Ограничения и сложности применения метода DiD. Оценка эффекта воздействия.

*Задания в виде кейса 5, вопросы*

**Описание практико-ориентированных заданий и кейсов**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Номер темы/модуля** | **Наименование практического занятия** | **Описание** |
| 1.1 | 1 | Кейс-задание 1. (Модуль 1) **Установка RStudio** | Цели практического задания: формирование навыков работы с RStudio, установка соответствующих настроек |
| 1.2. | 2. | Кейс-задание 2. (модуль 1) **Непараметрические методы анализа независимых выборок** | Цели практического задания: научится делать выводы о сходстве/различии характеристики для двух/трех групп с помощью непараметрических тестов |
| 1.3 | 3 | Кейс-задание 3. (модуль 2) **Построение регрессионных моделей с бинарной зависимой переменной** | Цели практического задания: построение статистически значимых бинарных моделей, оценка качества моделей, расчет маржинальных эффектов |
| 1.4 | 4 | Кейс-задание 4 (модуль 3). **Построение моделей анализа выживаемости, оценки таблиц времен жизни** | Цели практического задания: построение таблиц выживаемости. Нахождение оценок Каплана-Майера, построение кривых выживаемости, построение регрессионных моделей выживаемости |
| 1.5 | 5 | Кейс-задание 5. (Модуль 4). **Построение моделей измерения эффекта воздействия. Метод DiD.** | Цели практического задания: оценка эффекта воздействия на опытную группу |

**8.Оценочные материалы по образовательной программе**

**8.1. Вопросы тестирования по модулям**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ модуля** | **Вопросы входного тестирования** | **Вопросы промежуточного тестирования** | **Вопросы итогового тестирования** |
| 0 | 1. Доказательная медицина это   1. подход к медицинской практике, при котором решения о применении профилактических, диагностических и лечебных мероприятий принимаются исходя из имеющихся доказательств их эффективности и безопасности, 2. подход к медицинской практике, при котором решения о применении профилактических, диагностических и лечебных мероприятий принимаются исходя из имеющихся доказательств их эффективности, 3. подход к медицинской практике, при котором решения о применении профилактических, диагностических и лечебных мероприятий принимаются исходя из имеющихся доказательств их безопасности,   2. Множество результатов, отобранных из генеральной совокупности, называют   1. Выборкой 2. Вариационным рядом 3. Статистикой критерия 4. Точечными оценками   3. Статистическая гипотеза – это утверждение о свойствах   1. генеральной совокупности 2. выборки 3. конкретного объекта   4. Значение признака, находящееся в середине вариационного ряда наблюдений,   * 1. мода;   2. средняя арифметическая;   3. медиана;   4. частота;   5. частость.   5. Названия гипотезы, противоположной проверяемой:   1. нулевая 2. простая 3. конкурирующая 4. альтернативная 5. Сложная   6. Интервал возможных значений парного коэффициента корреляции при наличии между величинами X и Y отрицательной, но не функциональной связи:   1. (–1; 0) 2. (0; 1) 3. (–1; -0,5) 4. (–0,5; 0) 5. [–1; 0]   7. Долю числа элементов, попавших в данный интервал, в общем объеме выборки называют   1. Частостью 2. Модой 3. Частотой 4. Генеральной долей 5. Данные, для которых может быть получено среднее значение 6. Количественные 7. Порядковые 8. Номинальные 9. Выбор вида критической области при проверке гипотезы H0 против альтернативы H1 определяется 10. Содержанием гипотезы H0 11. Содержанием альтернативы H1 12. Уровнем значимости α 13. Мощностью критерия (1-β) 14. Степень тесноты статистической связи при значении коэффициента корреляции ρ=-1 15. отсутствует 16. слабая 17. средняя 18. высокая 19. функциональная | Модуль 1 − Первичная обработка данных. Параметрические и непараметрические тесты  1. Типы данных, их классификации  2. Шкалы измерения данных  3. Принципы группировки, формула Стерджесса  4. Схема нахождения коэффициента конкордации Кендалла, его интерпретация.  5. Непараметрические тесты анализа независимых выборок (U-критерий Манна-Уитни)  6. Непараметрические тесты анализа независимых выборок (Критерий серий Вальда—Вольфовица)  Модуль 2 – Бинарные модели в медицинских исследованиях, ROC-анализ  1. Общий вид и оценка пробит-моделей.  2. Общий вид и оценка логит-моделей.  3. Оценивание моделей бинарного выбора методом максимального правдоподобия.  4. Проверка адекватности построенной модели бинарного выбора. Показатели качества оценки.  5. Селекция моделей на основе информационных критериев.  6. Проведение Roc-анализа для бинарных моделей.  Модуль 3 – Модели анализа выживаемости, оценки таблиц времен жизни  1. Анализ плотностей распределения смертей и интенсивности отказов, или функции мгновенного риска смерти  2. Оценки Каплана-Майера.  3. Построение кривых выживаемости по оценкам Каплана-Майера.  4. Критерий Вилкокона-Гехана для функции выживаемости.  5. Регрессионные модели выживаемости: модель Кокса – модель пропорциональных рисков,  6. Проверка адекватности моделей выживаемости. Проверка распределения остатков на соответствие распределению Вейбула.  Модуль 4 – DID-анализ  1. Модели оценки эффектов воздействия в медицине: метод Разностей разности DiD,  2. Предпосылки для проведения метода Разностей разности: формирование контрольной и опытной (воздействия) групп.  3. Оценка методом Did через линейную регрессию, путем введения фиктивных переменных.  4. Ограничения и сложности применения метода DiD.  5. Оценка эффекта воздействия на основе модели Разностей разности DiD | 1. Зависимость, отражающая взаимно однозначное соответствие двух величин медико-биологических данных: 2. статистическая 3. корреляционная 4. функциональная 5. регрессионная   2. Логит- модель имеет вид:   1. .       3.При проведении анализа качества статистической медико-биологической модели считают, что:   1. Чувствительность– это и есть доля истинно положительных случаев; 2. Специфичность – это и есть доля истинно положительных случаев; 3. Чувствительность – это и есть доля истинно отрицательных случаев; 4. Специфичность – это и есть доля истинно отрицательных случаев;   4.Модели выживаемости, обладают следующими тремя основными характеристиками:   1. зависимая переменная, или отклик, – это время ожидания до наступления определенного события, 2. зависимая переменная, это бинарная переменная – выживет/ не выживет 3. наблюдения являются цензурированными в том смысле, что для некоторых объектов наблюдения исследуемое событие не наступило на момент анализа данных, 4. отсутствуют предикторы, или объясняющие переменные, для которых имеется воздействие на время ожидания. 5. Неизвестные коэффициенты в логит-модели находятся с помощью 6. МНК 7. Метода максимального правдоподобия 8. ОМНК 9. Рекурсивного метода 10. Пробит- модель имеет вид: 11. .     17. Информационный критерий Акайке имеет вид: 18. Для оценки качества классификатора в анализе медико-биологических данных используют:     1. Анализ коэффициентов детерминации     2. Анализ матриц сопряженности     3. ROC-анализ 19. Оценки Каплана-Майера позволяют:   А. Оценить функцию выживаемости пациентов  В. Оценить качество классификации  С. Исследовать частоты ложноположительных и ложноотрицательных случаев   1. Для оценки различий в двух группах функций выживаемости используют   А. критерий Филипса- Перрона  В. Логранговый критерий  С. Критерий Мана-Уитни  D. критерий Гехана-Вилкоксона.  11. Модель выживаемости Кокса позволяет:  A. Оценить мультипликативный эффект влияния фактора на выживаемость по сравнению с базовым риском  B. Оценить влияние факторов на бинарный результат  C. Оценить влияния фактора на срок выживаемости  D. Оценить влияние фактора на упорядоченную переменную |

**8.2.** Минимальным проходным баллом теста считается 60% верных ответов по результатам суммарно 2 попыток

Кейс в 10 баллов: Максимум 10 баллов - выставляется при выполнении всех требований к отчету, подробном описании всех этапов и представлении выводов, увязывающих выполненный кейс с встречающимися в практике задачами.

8-10 баллов выставляется при выполнении всех требований к отчету, подробном описании всех этапов или представлении выводов, увязывающих выполненный кейс с встречающимися в практике задачами

6-7 баллов выставляется при выполнении всех требований к отчету, подробном описании отдельных этапов или кратких выводов.

Минимально допустимый балл 5 баллов -выставляется при выполнении минимального требования к отчету кейса.

Кейс в 20 баллов: Максимум 20 баллов - выставляется при выполнении всех требований к отчету, подробном описании всех этапов и представлении выводов, увязывающих выполненный кейс с встречающимися в практике задачами.

15-19 баллов выставляется при выполнении всех требований к отчету, подробном описании всех этапов или представлении выводов, увязывающих выполненный кейс с встречающимися в практике задачами

11-14 баллов выставляется при выполнении всех требований к отчету, подробном описании отдельных этапов или кратких выводов.

Минимально допустимый балл 10 баллов -выставляется при выполнении минимального требования к отчету кейса.

**8.3. Итоговое тестирование по всей образовательной программе**

**Цели задания:** оценка сформированности компетенции цифровой экономики (способность управлять информацией и данными медико-биологического характера) на базовом уровне.

Оценка сформированности компетенции проверяется посредством выполнения итогового теста (вопросы приведены в разделе 8.1).

**8.4. Задания-кейсы**

**Кейс-задание 1 (по модулю 1). Установка RStudio**

1. Скачать необходимые программы для своей операционной системы 1.1. R • Для win – https://cran.r-project.org/bin/windows/base/

* + - Для macos – https://cran.r-project.org/bin/macosx/

1.2. RStudio • https://rstudio.com/products/rstudio/download/#download

1.3. Rtools (только для win) • https://cran.r-project.org/bin/windows/Rtools/

2. Установить скаченные программы. 2.1. Порядок установки R -> RStudio -> Rtools

2.2. Для пользователей Windows: • Важно, чтобы путь установки НЕ содержал русских букв

• Нежелательно устанавливать программы в стандартную папку «Program Files», поскольку в дальнейшем возникнут проблемы при установке дополнительных библиотек

• Лучше всего установить R в «C:/R», RStudio в «C:/RStudio», а Rtools в «C:/Rtools»

3. RStudio – это оболочка для языка R. В данном курсе вы будете использовать именно эту программу. После установки запустите RStudio и убедитесь, что она правильно установлена. Также необходимо изменить некоторые стандартные настройки. Для этого откройте раздел Tools – Global options

* + - 3.1. В разделе General установите настройки на UTF8.

**Кейс-задание 2 (по модулю 1). Применение непараметрических методов анализа независимых выборок**

1. Провести предварительный анализ исходных данных. Исключить аномальные наблюдения (если такие есть), заполнить пропуски (если они имеются).
2. Сделать выводы о сходстве/различии характеристики для двух групп с помощью непараметрических тестов: метод серий Вальда-Вольфовица, метод Колмогорова Смирнова, U-критерий Манна-Уитни.
3. Сделать выводы о сходстве/различии характеристик для трех групп с помощью непараметрических тестов: Краскела-Уоллиса, метод Колмогорова Смирнова, U-критерий Манна-Уитни, Q-критерия Кохрена, теста Макнемара.
4. Оформить отчет о выполнении задания с приведением условия задачи, результатов решения и выводов.

Выполнять кейс-задание рекомендуется с использованием среды RStudio. Результатом выполнения кейс-задания является отчет №2.

**Кейс-задание 3 (модуль 2). Построение регрессионных моделей с бинарной зависимой переменной**

1. Провести предварительный анализ исходных данных. Исключить аномальные наблюдения (если такие есть), заполнить пропуски (если они имеются). Провести корреляционный анализ независимых переменных, исключив переменные, значительно коррелирующие с другими переменными (>0,9).

2. Построить статистически значимую модель бинарной регрессии, оценив параметры методом максимального правдоподобия, применяя метод пошагового исключения, в которой все переменные будут статистически значимы. Подобрать функцию распределения, описывающую вероятность положительной альтернативы (например, выживет пациент или умрет) между нор-мальным распределением (пробит), логистическим (логит) и экстремальным (гомпит) на основе минимума информационных критериев.

3. Проверить качество отобранной модели, подтвердив его значениями коэффициентов R2 МакФаддена, тестом отношения правдоподобия (LR-тестом), результатами теста Хосмера-Лемешоу и любым тестом на нормальность распределения остатков (например, Колмогорова-Смирнова или Бера-Жарка).

4. Рассчитать маржинальные эффекты и провести интерпретацию коэффициентов модели.

5. Оформить отчет о выполнении задания с приведением условия задачи, результатов решения и выводов.

Выполнять кейс-задание рекомендуется с использованием среды [R](http://matrunich.com/data-analysis/rstat/)Studio. Результатом выполнения кейс-задания является отчет №3.

**Кейс-задание 4 (модуль 3). Построение моделей анализа выживаемости, оценка таблиц времен жизни**

1. Провести предварительный анализ исходных данных. Исключить аномальные наблюдения (если такие есть), заполнить пропуски (если они имеются).

2. Построить графики функций выживаемости, оцененные методом Каплана–Майера. Проанализировать полученные результаты, сделать предварительный вывод о различии функций выживаемости для разных категорий объектов

3. Провести логранговый тест Мантеля–Хензеля и тест Гехана–Вилкоксона на значимое различие в оценках выживаемости по подгруппам, сформированным по атрибутивному признаку.

4. Построить модель пропорциональных рисков Кокса, для оценки коэффициентов модели использовать метод частичного правдоподобия согласно методикам Эфрона или Брэслоу. Оценить качество моделей на основе показателя R2mer (мера объясненной случайности,) и R2mev (мера объясненного отклонения), выбрать лучшую методику оценки исходя из минимума информационных критериев Акайке и Шварца.

5. Построить модели лог-логистической, Вейбулла, экспоненциальной регрессии. Оценить качество моделей, выбрать лучшую исходя из минимума информационных критериев Акайке и Шварца.

6. Провести интерпретацию коэффициентов отобранной модели.

7. Оформить отчет о выполнении задания с приведением условия задачи, результатов решения и выводов.

Выполнять кейс-задание рекомендуется с использованием среды RStudio. Результатом выполнения кейс-задания является отчет № 4.

**Кейс-задание 5 (модуль 4). моделей измерения эффекта воздействия. Метод DiD.**

1. Сформировать две выборки: опытную группу и группу контроля. Сформировать фиктивную переменную, измеряющую период воздействия. Сформировать фиктивную переменную, отвечающую за принадлежность к группе воздействия или контроля.
2. Оценить модель линейной регрессии, в которой оценивается влияние характеристик группы, за счет фиктивной переменной принадлежности к группе, фиктивной переменной, отвечающей за период воздействия, их пересечения, отвечающего за эффект воздействия и других контрольных переменных.
3. Провести стандартный мониторинг полученной модели методом DiD.
4. Сделать выводы. Оценить значимость эффекта воздействия на опытную группу.
5. Оформить отчет о выполнении задания с приведением условия задачи, результатов решения и выводов.

Выполнять кейс-задание рекомендуется с использованием среды RStudio. Результатом выполнения кейс-задания является отчет №5.

**8.5.**

|  |  |  |  |
| --- | --- | --- | --- |
| Наименовании модуля | Задание | Балл | Критерии оценки |
| Входное тестирование | ТЕСТ | 10 | Проходной балл -5 |
| Модуль 1 − Первичная обработка данных. Параметрические и непараметрические тесты | Кейс 1 | 10 | Система R должна быть загружена (минимально допустимый балл - 8 баллов) |
| Кейс 2 | 10 | Для минимального балла (5) должны быть выполнены все задания, но могут быть допущены неточности в интерпретации результатов анализа данных |
| тест | 10 | Проходной балл - 6 |
| Модуль 2 – Бинарные модели в медицинских исследованиях, ROC-анализ | Кейс 3 | 20 | Для минимального балла (10) должны быть выполнены все задания, но могут быть допущены неточности в интерпретации результатов анализа данных |
| Тест | 10 | Проходной балл - 6 |
| Модуль 3 – Модели анализа выживаемости, оценки таблиц времен жизни | Кейс 4 | 20 | Для минимального балла (10) должны быть выполнены все задания, но могут быть допущены неточности в интерпретации результатов анализа данных |
| Тест | 8 | Проходной балл - 5 |
| Модуль 4 – DID-анализ | Кейс 5 | 10 | Для минимального балла (5) должны быть выполнены все задания, но могут быть допущены неточности в интерпретации результатов анализа данных |
| тест | 8 | Проходной балл - 5 |
| Итоговая аттестация | Тест | 11 | Проходной балл – 7 |
|  | Минимальный балл для получения зачета по КПК - 72, максимальный - 127 | | |

**9.Организационно-педагогические условия реализации программы**

**9.1. Кадровое обеспечение программы**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№**  **п/п** | **Фамилия, имя, отчество (при наличии)** | **Место основной работы и должность, ученая степень и ученое звание (при наличии)** | **Ссылки на веб-страницы с портфолио (при наличии)** | **Фото в формате jpeg** | **Отметка о полученном согласии на обработку персональных данных** |
| **1** | Лакман Ирина Александровна | ФГБОУ ВО «Уфимский государственный авиационный технический университет», доцент кафедры биомедицинский инженерии, к.т.н., доцент |  |  | Согласна |

**9.2.Учебно-методическое обеспечение и информационное сопровождение**

|  |  |
| --- | --- |
| **Учебно-методические материалы** | |
| Методы, формы и технологии | Методические разработки,  материалы курса, учебная литература |
| Методы организации учебно-познавательной деятельности: практический;  Форма: дистанционная;  Технологии:  Информационно – коммуникационная технология;  Кейс технология | 1. Анализ данных : учебник для академического бакалавриата / ГУ - Высшая школа экономики; под ред. В. С. Мхитаряна .— Москва : Юрайт, 2016 .— 490 (13 экз.) 2. Уразбахтина , Ю. О. Лабораторный практикум по дисциплине "Компьютерные технологии в медико-биологических исследованиях" / Ю. О. Уразбахтина , Р. Н. Уразбахтин, Ю. А. Пугина ; УГАТУ .— Уфа : УГАТУ, 2005 .— 71 с. 3. Уразбахтина, Ю. О. Компьютерные технологии в медико-биологических исследованиях : конспект лекций / Ю. О. Уразбахтина ; Уфимский государственный авиационный технический университет (УГАТУ) .— Уфа : УГАТУ, 2005 .— 210 с.   **Дополнительная литература**   * + - 1. Валеев, В. Т. Применение пакета MATLAB для обработки медико-биологических данных. Применение пакета MATLAB для обработки биомедицинских изображений : лабораторный практикум по дисциплине "Методы обработки биомедицинских сигналов и данных" / В. Т. Валеев ; УГАТУ .— Уфа : УГАТУ, 2006 .       2. Гусев, В. Г. Методы и технические средства для медико-биологических исследований : [учебное пособие для студентов, обучающихся по специальности 190500 - Биотехнические и медицинские аппараты и системы] / В. Г. Гусев ; УГАТУ .— Уфа : УГАТУ, 2001.       3. Цыганкова, И. А. Метод интеллектуальной обработки медико-биологических данных / И. А. Цыганкова // Программные продукты и системы .— 2009 .— N 3 .— С. 120-123 : |

|  |  |
| --- | --- |
| **Информационное сопровождение** | |
| Электронные  образовательные ресурсы | Электронные  информационные ресурсы |
| <https://sdo.ugatu.su/course/view.php?id=5615> | https://www.kaggle.com/ |

**9.3.Материально-технические условия реализации программы**

|  |  |
| --- | --- |
| Вид занятий | Наименование оборудования,  программного обеспечения |
| Лекции, практические занятия | **Аппаратные требования**  Intel Pentium или аналогичный процессор с тактовой частотой 300MHz и выше.  SVGA монитор, с разрешением экрана, как минимум, 800х600 точек и глубиной цвета 16 bit (рекомендуемое разрешение экрана — 1024х768).  Звуковая карта, акустическая система или наушники.  Доступ в Интернет со скоростью 56 кбит/с и выше.  **Программное обеспечение**  Операционная система: Windows 7 или более продвинутая, Macintosh, Linux  Браузер: Internet Explorer 7 или более продвинутый, Mozilla Firefox (скачать бесплатно: http://www. mozilla.org/download.html) и т.п.  **Для просмотра электронных версий учебных курсов необходимо наличие установленных программ:**  Microsoft Internet Explorer 7.0 и выше ([Загрузить с сайта www.microsoft.com](http://www.microsoft.com/rus/windows/internet-explorer/))  Adobe Flash Player версии 7.0 и выше ([Загрузить с сайта http://www.adobe.com/](http://get.adobe.com/flashplayer/)) |

Федеральное государственное бюджетное образовательное учреждение

высшего образования «УФИМСКИЙ ГОСУДАРСТВЕННЫЙ

АВИАЦИОННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Центр дополнительного образования

УТВЕРЖДАЮ

И.о. проректора по учебной работе

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Р.Д. Еникеев

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г.

**ПАСПОРТ КОМПЕТЕНЦИЙ**

**«СПОСОБНОСТЬ УПРАВЛЯТЬ ИНФОРМАЦИЕЙ И ДАННЫМИ МЕДИКО-**

**БИОЛОГИЧЕСКОГО ХАРАКТЕРА»**

Дополнительной профессиональной программы повышения

квалификации для системы предоставления персональных

цифровых сертификатов от государства на развитие

у трудоспособного населения компетенций цифровой экономики

**«Машинное обучение для обработки медико-биологических данных»**

**Уфа 2020**

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | Наименование компетенции | Пояснения | Способность управлять информацией и данными медико-биологического характера |
| 2 | Указание типа компетенции | Профессиональная | Профессиональная |
| 3 | Определение, содержание и основные сущностные характеристики компетенции |  | Под компетенцией понимается способность управлять информацией и данными медико-биологического характера для реализации принципов доказательной медицины  Слушатель должен: знать:   * методы дизайна эксперимента в медицине, формирование контрольных групп и групп воздействия; * методы проверки параметрических и непараметрических критериев для медицинских исследований; * методы построения бинарных регрессий, предпосылки для получения достоверных оценок регрессии различными методами; * методы построения таблиц выживаемости, нахождения оценок Каплана Майера, критерии на разделении выживаемости в подгруппах; * регрессионные модели выживаемости (Кокса, логнормальную, экспоненциальную, нормальную); * методы оценки воздействия в медицине посредством DID-анализа (разности разностей).   уметь:   * проверять гипотезы о различиях выборок в медицине, используя параметрические и непараметрические критерии; * строить адекватные модели бинарной регрессии и проводить интерпретацию результатов моделирования на основе маржинальных эффектов влияния факторов на результат; * проводить ROC-анализ, рассчитывать показатели специфичности и чувствительности; * строить таблицы времён жизни, интерпретировать результаты анализа таблицы времён жизни, анализировать графики функции выживаемости, и функции мгновенного риска, определять оценки выживаемости методом Каплана-Майера, а также разделять и проводить анализ выживаемости по подгруппам; * строить различные регрессионные модели выживаемости, проверять их адекватность реальному моделируемому процессу, интерпретировать результаты сточки зрения оценки мультипликативного эффекта для оценки риска недожития; * уметь оценить эффективность воздействия в медицине, используя DiD-модель.   владеть:   * навыками интеллектуальной обработки медицинских данных, используя RStudio; * навыками построения адекватных моделей бинарной регрессии и качественной интерпретации результатов моделирования; * навыками проведения анализа выживаемости, DID-анализа; * навыками применения RStudio для проведения анализа медико-биологических данных. |
| 4 | Дескриптор знаний, умений и навыков по уровням | Уровни сформированности компетенции обучающегося начальный/базовый | индикаторы сформированности компетенции (знать, уметь, владеть) обучающегося в зависимости от уровня начальный/базовый |
|  |  | Начальный уровень (Компетенция недостаточно развита. Частично проявляет навыки, входящие в состав компетенции. Пытается, стремится проявлять нужные навыки, понимает их необходимость, но у него не всегда получается) | знать:   * методы дизайна эксперимента в медицине, формирование контрольных групп и групп воздействия; * методы проверки параметрических и непараметрических критериев для медицинских исследований;   уметь:   * проверять гипотезы о различиях выборок в медицине, используя параметрические и непараметрические критерии;   владеть:   * навыками интеллектуальной обработки медицинских данных, используя RStudio; * навыками применения RStudio для проведения анализа медико-биологических данных. |
|  |  | Базовый уровень (Уверенно владеет навыками, способен, проявлять соответствующие навыки в ситуациях с элементами неопределённости сложности) | знать:   * методы построения бинарных регрессий, предпосылки для получения достоверных оценок регрессии различными методами; * методы построения таблиц выживаемости, нахождения оценок Каплана Майера, критерии на разделении выживаемости в подгруппах; * регрессионные модели выживаемости (Кокса, логнормальную, экспоненциальную, нормальную); * методы оценки воздействия в медицине посредством DID-анализа (разности разностей).   уметь:   * строить адекватные модели бинарной регрессии и проводить интерпретацию результатов моделирования на основе маржинальных эффектов влияния факторов на результат; * проводить ROC-анализ, рассчитывать показатели специфичности и чувствительности; * строить таблицы времён жизни, интерпретировать результаты анализа таблицы времён жизни, анализировать графики функции выживаемости, и функции мгновенного риска, определять оценки выживаемости методом Каплана-Майера, а также разделять и проводить анализ выживаемости по подгруппам; * строить различные регрессионные модели выживаемости, проверять их адекватность реальному моделируемому процессу, интерпретировать результаты сточки зрения оценки мультипликативного эффекта для оценки риска недожития; * уметь оценить эффективность воздействия в медицине, используя DiD-модель.   владеть:   * построения адекватных моделей бинарной регрессии и качественной интерпретации результатов моделирования; * проведения анализа выживаемости, DID-анализа. |
| 5 | Характеристика взаимосвязи данной компетенции с другими компетенциями/ необходимость владения другими компетенциями для формирования данной компетенции |  | Компетенции цифровой грамотности |
| 6 | Средства и технологии оценки |  | Кейсы-задания, тесты |

**IV.Иная информация о качестве и востребованности образовательной программы** (результаты профессионально-общественной аккредитации образовательной программы, включение в системы рейтингования, призовые места по результатам проведения конкурсов образовательных программ и др.) (при наличии)

Общественная аккредитация программы не проводилась

**V.Рекомендаций к программе от работодателей**: наличие не менее двух писем и/или подтверждения на цифровой платформе Государственной системы предоставления ПЦС от работодателей о рекомендации образовательной программы для реализации в рамках Государственной системы предоставления ПЦС на формирование у трудоспособного населения компетенций цифровой экономики с указанием востребованности результатов освоения программы в сфере деятельности соответствующих компаний и готовности к рассмотрению заявок наиболее успешно освоивших образовательную программу граждан на прохождение стажировки и (или) собеседования на предмет трудоустройства путем проставления отметки в профиле программы

Документы загружены на платформу

**VI. Указание на возможные сценарии профессиональной траектории граждан** по итогам освоения образовательной программы (в соответствии с приложением)

|  |  |
| --- | --- |
| **Цели получения персонального цифрового сертификата** | |
| **текущий статус** | **цель** |
| **Развитие компетенций в текущей сфере занятости** | |
| работающий по найму в организации, на предприятии | развитие профессиональных качеств |
| работающий по найму в организации, на предприятии | повышение заработной платы |
| работающий по найму в организации, на предприятии | смена работы без изменения сферы профессиональной деятельности |
| **Переход в новую сферу занятости** | |

|  |  |
| --- | --- |
| освоение смежных профессиональных областей | повышение уровня дохода, расширение профессиональной деятельности |

**VII.Дополнительная информация**

**VIII.Приложенные Скан-копии**

Утверждённая рабочая программа (подпись, печать, в формате pdf) загружена на платформу