

УЧЕБНАЯ ПРОГРАММА.

Статистический анализ биомедицинских данных

СОСТАВИТЕЛЬ.

Скакун Виктор Васильевич, заведующий кафедрой системного анализа и компьютерного моделирования Белорусского государственного университета, кандидат физико-математических наук, доцент.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Основные положения теории вероятностей

Тема 1.1. Вероятность, дискретные и непрерывные случайные величины, функция распределения вероятности, функция плотности вероятности, гистограмма, функция накопленных частот.

Тема 1.2. Характеристики популяции: математическое ожидание, дисперсия и стандартное отклонение. Понятие моментов случайных величин.

Тема 1.3. Центральная предельная теорема. Закон больших чисел.

Тема 1.4. Наиболее известные распределения вероятностей: Бернулли, биномиальное, Пуассона, равномерное, показательное и степенное. Нормальное распределение.

Раздел 2. Введение в статистику.

Тема 2.1. Основная терминология. Переменные (признаки), генеральная совокупность и выборка, параметры и статистика. Неопределенность в измерениях, точность и достоверность, систематические и случайные ошибки.

Тема 2.2. Описательная статистика: вариационный ряд, гистограмма, мера центральной тенденции (мода, медиана, выборочное среднее), мера разброса (межквартильный диапазон, среднее абсолютное отклонение, среднеквадратичная ошибка, выборочная дисперсия), коэффициент вариации, коэффициенты асимметрии и эксцесса, графическое представление данных (коробчатая диаграмма, диаграмма рассеяния).

Тема 2.3. Нормировка и стандартизация данных. Анализ и устранение выбросов, аномальных и экстремальных значений. Заполнение пропусков. Усреднение данных. Стратегии вычисления выборочного среднего. Стандартная ошибка.

Раздел 3. Оценка параметров и проверка гипотез

Тема 3.1. Выборка, стратегии выборки, точечные оценки параметров, доверительные интервалы для выборочного среднего и дисперсии.

Тема 3.2. Проверка гипотез. Ошибки первого и второго рода. Связь проверки гипотез с вычислением доверительных интервалов, проверка гипотез для среднего при известной и неизвестной дисперсии, проверка гипотез для дисперсии, проверка гипотезы о равенстве средних значений двух популяций (t-критерий), парный t-критерий, гипотеза о равенстве дисперсий двух популяций (F-тест), t-статистика, F-статистика.

Тема 3.3. Статистические критерии. Критерий χ^2 , критерий Колмогорова-Смирнова.

Тема 3.4. Тесты нормальности распределения данных.

Тема 3.5. Непараметрические тесты. Тест Манна-Уитни, тест Уилкоксона.

Раздел 4. Дисперсионный анализ

Тема 4.1. Однофакторный и двухфакторный дисперсионный анализ. Многовариантный дисперсионный анализ.

Раздел 5. Корреляционный анализ

Тема 5.1. Мера связи между двумя переменными: ковариация и коэффициент корреляции, интерпретация коэффициента корреляции.

Тема 5.2. Проверка наличия статистической связи между двумя переменными. Коэффициент корреляции Пирсона и оценка его статистической значимости. Ранговая корреляция. Коэффициент ранговой корреляции Кендалла и коэффициент корреляции Спирмена. Проверка гипотез о наличии корреляции.

Раздел 6. Регрессия

Тема 6.1. Линейная регрессия. Связь коэффициента регрессии с коэффициентом корреляции. Нелинейная регрессия. Метод наименьших квадратов.

Тема 6.2. Оценка качества аппроксимации экспериментальных данных. Анализ остатков.

Раздел 7. Моделирование псевдослучайных величин

Тема 7.1. Моделирование непрерывных случайных величин. Метод обратных функций. Метод Неймана. Моделирование нормальной случайной величины. Моделирование коррелированных случайных величин.

Тема 7.2. Моделирование дискретных случайных величин. Моделирование распределения Бернулли. Моделирование распределения Пуассона.

ТЕМАТИКА ЛЕКЦИЙ И ЛАБОРАТОРНЫХ ЗАНЯТИЙ

Тематика лекций (22 часа)

- 1) Основные положения теории вероятностей (2 часа).
- 2) Центральная предельная теорема. Закон больших чисел. Наиболее известные распределения вероятностей (2 часа).
- 3) Введение в статистику. Основная терминология (2 часа).
- 4) Описательная статистика (2 часа).
- 5) Нормировка и стандартизация данных. Заполнение пропусков и устранение выбросов. Усреднение данных (2 часа).
- 6) Выборка, точечные оценки параметров, построение доверительных интервалов. Проверка гипотез. Связь проверки гипотез с вычислением доверительных интервалов (2 часа).
- 7) Проверка гипотез для среднего и дисперсии, проверка гипотезы о равенстве средних значений и дисперсий двух популяций, t-статистика, F-статистика (2 часа).
- 8) Статистические критерии. Тесты на нормальность. Непараметрические тесты (2 часа).
- 9) Дисперсионный анализ. Корреляционный анализ (2 часа).
- 10) Регрессия. Метод наименьших квадратов. Оценка качества аппроксимации экспериментальных данных (2 часа).
- 11) Моделирование псевдослучайных величин (2 часа).

Тематика лабораторных занятий (28 часов)

- 1) Описательная статистика Нормировка и стандартизация данных. Заполнение пропусков и устранение выбросов. Усреднение данных (4 часа).
- 2) Построение доверительных интервалов. Проверка гипотез о равенстве средних значений и дисперсий двух популяций (4 часа).
- 3) Проверка гипотез с помощью статистических критериев. Тесты на нормальность. Непараметрические тесты (4 часа).
- 4) Корреляционный анализ (4 часа).
- 5) Дисперсионный анализ (4 часа).
- 6) Регрессия. Метод наименьших квадратов (4 часа).
- 7) Моделирование непрерывных и дискретных случайных величин (4 часа).

ИНФОРМАЦИОННАЯ ЧАСТЬ

Перечень основной литературы

- 1) Гланц С. Медико-биологическая статистика / С. Гланц. – М: Практика.– 1999. – 459 с.
- 2) Петри, А. Наглядная медицинская статистика: учеб. пособие / А. Петри, К. Сэбин; пер. с англ. под ред. В. П. Леонова. – 3-е изд., перераб. и доп. – М: ГЭОТАР-Медиа, 2015. – 216 с.
- 3) Шипунов, А.Б. Наглядная статистика. ИспользуемR! / А.Б. Шипунов, Е.М. Балдин, П.А. Волкова, А.И. Коробейников, С.А. Назарова, С.В. Петров, В.Г. Суфиянов. – E-book. – 2014.
- 4) van Belle, G. Biostatistics: A Methodology for the Health Sciences, 2nd Edition / G. van Belle, L.D. Fisher, P.J. Heagerty, T. Lumley. – Wiley, 2004.– 896 p.
- 5) Le, C.T. Introductory Biostatistics, 2nd Edition / C.T. Le, L.E. Eerly. – Wiley, 2016. – 616 p.
- 6) Gareth, J. An Introduction to Statistical Learning with Applications in R / J. Gareth, Witten D, Trevor H.R. Tibshirani. – Springer, 2017. – 426 p.

Перечень дополнительной литературы

- 7) Мастицкий С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. Хайдельберг.– Лондон.– Тольятти, 2014.– 401 с.
- 8) Schumacker R. Understanding Statistics Using R / R. Schumacker, S. Tomek – Springer, 2013. – 289 p.