

Программа курса „Методы современной прикладной статистики“

лектор — доц. И.В. Родионов

Весна 2016

1. Введение.

- Парадигмы математической статистики: параметрическая, непараметрическая и байесовская. Достоинства и недостатки парадигм.
- Семейства распределений и их простые свойства.
- Виды оценок: несмещённые, состоятельные и асимптотически нормальные оценки.
- Методы получения оценок: метод моментов, метод максимального правдоподобия, метод выборочной квантили. Их свойства.
- Доверительные интервалы. Метод центральной статистики. Асимптотические доверительные интервалы.
- Байесовские оценки, сопряжённые распределения, неинформативные априорные распределения.

2. Проверка статистических гипотез.

- Основные понятия: критическое множество, уровень значимости, альтернативы, ошибки первого и второго рода, функция мощности, p -значение.
- Сравнения критериев: наиболее мощные и равномерно наиболее мощные критерии.
- Лемма Неймана–Пирсона. Построение с ее помощью наиболее мощных критериев.
- Теорема о монотонном отношении правдоподобия. Построение равномерно наиболее мощных критериев для односторонних альтернатив.
- Проверка статистических гипотез с помощью предельных теорем и доверительных интервалов.
- Проверка гипотез в байесовском подходе.

3. Критерии согласия и проверка нормальности.

- Критерий Колмогорова-Смирнова.

- Критерий хи-квадрат Пирсона.
- Критерий омега-квадрат: критерии Крамера-Мизеса и Андерсона-Дарлинга.
- Проверка нормальности: гистограмма, qq-plots, критерии асимметрии и эксцесса.

4. Ядерные оценки плотности.

- Оценка Розенблатта-Парзена, её асимптотическое поведение.
- Выбор оптимального окна пропускания, наилучшая скорость сходимости ядерной оценки.
- Виды ядерных функций: ядро Епанечникова, квадратичное ядро, треугольное ядро, ядро Гаусса, прямоугольное ядро.

5. Дисперсионный анализ.

- Две независимые нормальные выборки. Критерий Фишера проверки равенства дисперсий. Критерий Стьюдента.
- Две независимые выборки. Критерий Смирнова. Критерий омега-квадрат Розенблатта. Критерий Манна-Уитни.
- Парные повторные наблюдения. Критерий знаков. Критерий знаковых рангов Уилкоксона. Поправка на средние ранги в критерии Уилкоксона.
- Много независимых нормальных выборок. Критерий Бартлетта. F-критерий для проверки однородности нормальных выборок.
- Однофакторная модель. Критерий Краскела-Уоллиса.
- Двухфакторная модель. Критерий Фридмана.

6. Корреляционный анализ.

- Коэффициенты корреляции Пирсона, Спирмэна, Кэндалла и их свойства. Обобщённый коэффициент корреляции.
- Ранговый коэффициент конкордации Кэндалла, его свойства.
- Частные корреляции.

7. Регрессионный анализ.

- Линейная регрессионная модель. Оценка наименьших квадратов, ее основные свойства. Оценка σ^2 .
- Гауссовская линейная модель. Оптимальные оценки параметров в гауссовской линейной модели, их распределения.
- Распределения хи-квадрат, Стьюдента и Фишера, их свойства. Теорема об ортогональном разложении гауссовского вектора. Доверительные интервалы для параметров гауссовской линейной модели.
- F-критерий проверки линейных гипотез в гауссовской линейной модели.

- Метод Тейла.
- Подбор полинома в линейной регрессионной модели. Парадоксы регрессии. Взвешенный метод наименьших квадратов.
- Коэффициент детерминации R^2 и его связь с коэффициентом корреляции Пирсона. Решение задачи о значимости факторов с помощью F-критерия. Информационные критерии. Обнаружение мультиколлинеарности и ридж-регрессия.
- Анализ остатков. Гетероскедастичность. Критерий Голдфелда-Квандта. Критерий Дарбина-Уотсона. Критерий Харке-Бера. Критерий RESET Рэмси.
- Непараметрическая регрессия. Оценки Надарая-Ватсона и Гассера-Мюллера. Скорость сходимости оценки Гассера-Мюллера.
- Метод локальных полиномов в непараметрической регрессии.

8. Кластерный анализ и классификация.

- Методы нормировки данных. Методы задания межобъектных расстояний.
- Эвристические методы классификации: метод связанных компонент, метод кратчайшего незамкнутого пути, метод k -средних, метод “Форель”, метод потенциальных ям.
- Иерархические процедуры: метод ближайшего соседа, метод дальнего соседа, метод средней связи, метод центров масс, метод Уорда.
- Функционалы качества и сравнение методов кластерного анализа.
- Логистическая регрессия.

9. Факторный анализ и снижение размерности.

- Метод главных компонент.
- Степенной метод вычисления главных компонент.
- Нелинейные методы понижения размерности. Метод Сэммона. Применение метода сопряжённых градиентов для минимизации функционала Сэммона.
- Методы заполнения пропусков в многомерных данных. Метод локального заполнения.

10. Робастные методы.

- Асимптотическая толерантность робастного метода.
- L -оценки, M -оценки, R -оценки, условия их асимптотической нормальности и асимптотической эффективности.
- Определение робастности, качественная и количественная робастность.
- Изучение свойств статистик с помощью функции влияния.
- Робастные оценки параметров в линейной регрессионной модели.

СПИСОК ЛИТЕРАТУРЫ

1. *Лагутин М. Б.* Наглядная математическая статистика. — 2-е изд. — М.: БИНОМ. Лаборатория знаний, 2009.
2. *Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.* Классификация и снижение размерности. М.: “Финансы и статистика”, 1989.
3. *Friedman J., Hastie T., Tibshirani R.* The Elements of Statistic Learning. 2008.
4. *Хьюбер П.* Робастность в статистике. М.: Мир, 1984.
5. *Носко В.П.* Эконометрика. М.: Дело, 2011.