# Программа курса «Введение в анализ данных»[1]

## 1. Тематический план

| № | Тема | Занятия (ч.) | |
|---|---|---|---|
| | | лекции | практич. |
| 1 | Введение в анализ данных. Предварительный анализ данных. Описательная статистика. Классификация статистических данных. | 2 | 0 |
| 2 | Одномерный анализ количественных, категориальных и бинарых признаков. | 2 | 0 |
| 3 | Двумерный анализ: суммаризация и корреляция двух признаков. Количественные признаки: линейная регрессия, нелинейная и линеаризованная регрессии. | 2 | 0 |
| 4 | Двумерный анализ. Случай смешанных шкал: номинальный и количественный признаки | 2 | 0 |
| 5 | Двумерный анализ. Случай двух номинальных признаков | 2 | 0 |
| 6 | Корреляция и суммаризация для многомерных данных: Классы решающих правил. | 2 | 0 |
| 7 | Байесовское решающее правило. | 4 | 0 |
| 8 | Меры качества классификатора. | 2 | 0 |
| 9 | Задача кластеризации. | 2 | 0 |
| 10 | Кластеризация методом К-средних | 4 | 0 |
| 11 | Модель множественной регрессии | 4 | 0 |
| 12 | Нелинейные модели регрессии и их линеаризация | 2 | 0 |
| 13 | Регрессионные модели с фиктивными переменными | 2 | 0 |
| 14 | Снижение размерности признакового пространства. Компонентный анализ | 2 | 0 |
| 15 | Снижение размерности признакового пространства. Факторный анализ. | 2 | 0 |
| | ИТОГО | 36 | 0 |

[1] *Курс разработан при поддержке Министерства науки и высшего образования РФ* (Соглашение № 075-02-2021-1552).

## 2. Оценочные средства

### 2.1.Примерный вариант 1 контрольной работы (15 баллов)

1. Оценка параметров генеральной совокупности по выборке. Описательные статистики. Визуализация данных. (3 балла)
2. Кластеризация методом К-средних. (3 балла)
3. В задании требуется объяснить принцип работы медианного фильтра и реализовать программу на языке Python без использования библиотек, содержащих готовые методы. Допускается использование библиотек numpy и PIL. В качестве решения загрузить текст программы. Усредненное фильтрование использует значения элементов, содержащихся в области примыкания, для определения нового значения. Фильтр располагает элементы области примыкания в отсортированном порядке и отбирает среднее значение. Фильтр требуется применить к изображению Lenna. (9 баллов)

### 2.2. Примерный вариант 2 контрольной работы (15 баллов)

1. Анализ главных компонент. (3 балла)
2. Факторный анализ. (3 балла)
3. На указанном датасете построить линейную регрессию и вычислить коэффициенты уравнения регрессии методом наименьших квадратов. Найти коэффициент детерминации R^2R2 и интерпретировать его значение. Решение оформить в Google Colab и в ответе дать ссылку на ноутбук.

   Датасет: https://www.kaggle.com/spittman1248/cdc-data-nutrition-physical-activity-obesity

## 3. Вопросы к зачету

1. Предварительный анализ данных. Описательная статистика.
2. Классификация статистических данных.
3. Генеральная и выборочная совокупности.
4. Одномерный анализ количественных признаков.
5. Одномерный анализ категориальных и бинарых признаков.
6. Нормирование (стандартизация) и унификация данных.
7. Двумерный анализ. Количественные признаки: линейная регрессия,
8. Двумерный анализ. Количественные признаки: нелинейная и линеаризованная регрессии.
9. Двумерный анализ. Случай смешанных шкал: номинальный и количественный признаки. Целевой количественный признак.
10. Двумерный анализ. Случай смешанных шкал: номинальный и количественный признаки. Номинальный целевой признак.
11. Двумерный анализ. Случай двух номинальных признаков.
12. Канонические корреляции и канонические величины генеральной совокупности.
13. Оценка канонических корреляций и канонических величин.
14. Аномальные значения. Методы обнаружения засорения выборки.
15. Корреляция и суммаризация для многомерных данных: Классы решающих правил.
16. Байесовское решающее правило.
17. Меры качества классификатора.
18. Задача кластеризации.
19. Кластеризация методом К-средних.

20. Модель множественной регрессии.
21. Нелинейные модели регрессии и их линеаризация.
22. Регрессионные модели с фиктивными переменными.
23. Снижение размерности признакового пространства. Компонентный анализ.
24. Снижение размерности признакового пространства. Факторный анализ.

## 4. Литература

1. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — М.: Издательство Юрайт, 2018. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/413060.

2. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — М.: Издательство Юрайт, 2018. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/412967.

**б) дополнительная литература:**

3. Крутиков, В.Н. Анализ данных: учебное пособие / В.Н. Крутиков, В.В. Мешечкин; Кемеровский государственный университет. – Кемерово: Кемеровский государственный университет, 2014. – 138 с. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=278426. – Текст: электронный.

4. Каган, Е.С. Прикладной статистический анализ данных: учебное пособие / Е.С. Каган; Кемеровский государственный университет. – Кемерово: Кемеровский государственный университет, 2018. – 235 с. : ил., табл. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=573550. – Текст: электронный.

5. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980. 456с. – URL: https://scask.ru/h_book_lra.php?id=1.

6. Кремер, Н. Ш. Эконометрика: учебник и практикум для академического бакалавриата / Н. Ш. Кремер, Б. А. Путко; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 354 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-02760-0. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/401922.

7. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel: учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 353 с. — (Университеты России). — ISBN 978-5-534-01672-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/400278.