

Оглавление

1.	Моделирование, оценка параметров и проверка гипотез согласия эмпирического распределения с теоретическим . . .	2
2.	Описательная статистика	3
3.	Статистический анализ категориальных признаков	4
4.	Критерии однородности для независимых выборок	4
5.	Критерии однородности для зависимых выборок	5
6.	Корреляционный анализ	6
7.	Факторный анализ	6
8.	Дискриминантный анализ	6

Практические задания

1. Моделирование, оценка параметров, проверка гипотез согласия эмпирического распределения с теоретическим.
2. Описательная статистика.
3. Статистический анализ качественных признаков
4. Критерии однородности
5. Метод наименьших квадратов в задаче линейной и нелинейной регрессии.
6. Непараметрические коэффициенты корреляции. Значимость частных коэффициентов регрессии.
7. Многомерная статистика: кластерный, дискриминантный, факторный анализ, случайный лес.

1. Моделирование, оценка параметров и проверка гипотез согласия эмпирического распределения с теоретическим

- Про моделировать распределение

1. биномиальное $\beta(n, p)$
2. Пуассона $P(\lambda)$
3. гамма $\Gamma(\alpha, \lambda)$
4. Вейбулла $W(k, \lambda)$
5. равномерное $U(a, b)$

6. нормальное $\mathcal{N}(\mu, \sigma)$.

- Оценить параметры распределения по методу моментов и по методу максимального правдоподобия.
- Проверить согласие эмпирического и теоретического распределения по критерию хи-квадрат Пирсона.

2. Описательная статистика

Данные об ожирении у детей dataX.csv

Код	Расшифровка
1 Sex	пол; 1-м, 2-ж
2 Age	возраст
3 liver	заболевания печени (0 нет, 1 есть)
4 vision	ухудшение зрения (0 нет, 1 есть)
5 sss	заболевания сердечно-сосудистой системы (0 нет, 1 есть)
6 kidneys	заболевания почек (0 нет, 1 есть)
7 OL	ожирение -1 дефицит, 0 норма, 1 изб.масса, 2,3,4 ож.
8 height	рост
9 weight	вес
10 bmi	индекс массы тела
11 Zscore_HeightageWHO	индекс МТ детский
12 LVDS	диастолический размер левого желудочка
13 EF	фракция выброса

Таблица 1. Соответствие между кодировкой и расшифровкой названий признаков.

- Для метрических признаков построить гистограммы, найти математические ожидания, дисперсии, медианы, минимумы, максимумы, квартили и интерквартильные размахи, асимметрию, эксцесс. Какие распределения можно считать унимодальными, симметричными, нормально распределенными?

3. Статистический анализ категориальных признаков

Данные (табл.1) об ожирении у детей dataX.csv

Категориальные переменные упорядочить по значимости для фактора ожирения. Привести соответствующие частоты, значимости точного критерия Фишера, коэффициенты неопределенности.

4. Критерии однородности для независимых выборок

Данные (табл.1) об ожирении у детей dataX.csv

Варианты метрической переменной (X), категориальной с двумя градациями (Y_1), категориальной с числом градаций более двух Y_2 (OL для всех вариантов).

- 1) рост (X) и заболевания почек (Y_1),
 - 2) вес (X) и заболевания печени (Y_1),
 - 3) фракция выброса (X) и заболевания сердечно-сосудистой системы (Y_1),
 - 4) индекс (X) ИМТ (детский) и ухудшение зрения (Y_1),
 - 5) диастолический размер левого желудочка (X) и пол (Y_1).
- Проверить на предмет однородности независимых выборок данные по переменной X в зависимости от Y_1 по критерию 1) Вилкоксона, 2) Фишера равенства дисперсий, 3) Стьюдента равенства средних.

Привести значения средних с ошибками среднего, медиан с интерквартильным размахом, значимости соответствующих критериев.

- Проверить на предмет однородности данные по переменной X в зависимости от Y_2 1) по критерию Краскела-Уоллиса, 2) при помощи однофакторного дисперсионного анализа. Построить бокс-плот.
- Применить критерий Стьюдента для множественных сравнений с поправкой Бонферони и критерий Тьюки.
- Выполнить двухфакторный дисперсионный анализ по переменной X с двумя факторами Y_1, Y_2 . Сравнить результаты использования моделей с фиксированными и случайными эффектами.

5. Критерии однородности для зависимых выборок

Данные о динамике иммунологических показателей среди больных ВИЧ. Рассматриваются три группы: 1) больные вне зависимости от пола, не приверженные диспансерному наблюдению и не получающие антивиральную терапию (АВРТ) 43% смертность; 2) мужчины: приверженные, но не получающую АВРТ или неприверженные, но получающие АВРТ 12% смертность; 3) приверженные мужчины с АВРТ и женщины: или приверженные, или получающую АВРТ.

По ранговому критерию Вилкоксона для зависимых выборок найти пары последовательных наблюдений с наибольшим отличием. По критерию Фридмана три последовательных наблюдения с наибольшими изменениями в динамике.

Применить метод ANAVA Repeated Measures и проверить значимость эффектов взаимодействия факторов времени и группы.

6. Корреляционный анализ

Данные (табл.1) об ожирении у детей dataX.csv

Построить корреляционную матрицу по метрическим показателям и найти наиболее значимую корреляцию. Для этого случая построить двумерную диаграмму.

Применить модель множественной регрессии для зависимой переменной индекс МТ детский и независимыми переменными: возраст, рост, вес, имт, диастолический размер левого желудочка, фракция выброса. Определить значимость прогноза и наиболее значимые переменные.

7. Факторный анализ

Данные (табл.1) об ожирении у детей dataX.csv

Проинтерпретировать главные компоненты, определить вклад первых двух в общую дисперсию и построить двумерную диаграмму первых двух факторов.

8. Дискриминантный анализ

Данные (табл.1) об ожирении у детей dataX.csv

Выяснить возможности классификации детей с разными формами ожирения.