

Структура экзамена по курсу “Введение в Data Science”

Вопросы экзамена разбиваются на три блока.

1. Тестовая часть (Часть А)

Короткие вопросы с вариантами ответов (множественный выбор: один или несколько правильных ответов) и на соответствие.

Баллы: 15 вопросов по 0.2 балла

2. Задачи (Часть В)

Короткие задачи на расчёт стандартных статистических величин (численный ответ или множественный выбор).

Баллы: 4 вопроса по 0.85 балла

3. Работа с набором данных (Часть С)

Баллы: 9 вопросов по 0.4 балла

Темы экзамена

Статистика

1. Сбор данных. Выборка и генеральная совокупность. Репрезентативность выборки.
2. Частотные таблицы и распределения. Непрерывные распределения. Функция плотности.
3. Корреляция Пирсона.
4. Тестирование гипотез. Статистическая значимость. p-value.
5. Ошибки I и II рода.
6. Доверительные интервалы, основанные на Z- и t-распределениях, для среднего и доли.
7. Z-тест и t-тест для одной выборки и для двух независимых выборок с одинаковыми дисперсиями.
8. Хи-квадрат критерий согласия Пирсона.

Анализ данных

1. Работа с данными: агрегирование, фильтрация, создание новых переменных, работа со сводными таблицами.

2. Типы переменных. Меры центральной тенденции и разброса. Обнаружение выбросов. Работа с выбросами.
3. Визуализация.
4. Обнаружение пропущенных значений. Работа с пропущенными значениями: удаление и стратегии замены.

Машинное обучение

1. Постановка задачи машинного обучения. Виды задач машинного обучения.
2. Метод k ближайших соседей.
3. Линейная регрессия. Интерпретация оценок коэффициентов в линейной регрессии.
4. Метрики качества: MSE, MAE, R2.
5. Логистическая регрессия.
6. ROC-кривая.

Вопрос	Тип	Тема	Компетенция
A1	Множественный выбор (один или несколько правильных ответов)	Визуализация	Умение выбрать корректный тип визуализации для заданной ситуации
A2	Множественный выбор (один или несколько правильных ответов)	Типы переменных. Меры центральной тенденции и разброса. Выбросы	Умение интерпретировать описательные статистики и особенности распределения переменных
A3	Множественный выбор (один правильный ответ)	Логистическая регрессия	Умение получить предсказания в модели логистической регрессии
A4	Множественный выбор (один или несколько правильных ответов)	Частотные таблицы и распределения. Основные виды распределений	Умение определить пропорцию выборки в нормальном распределении
A5	Множественный выбор (один или несколько правильных ответов)	Сбор данных. Выборка и генеральная совокупность	Умение оценить репрезентативность выборки
A6	Множественный выбор (один или несколько правильных ответов)	Корреляция Пирсона	Умение интерпретировать выборочный коэффициент корреляции Пирсона
A7	Множественный выбор (один правильный ответ)	Метод k ближайших соседей	Умение получить предсказания в модели kNN
A8	Множественный выбор (один правильный ответ)	Частотные таблицы и распределения. Основные виды распределений.	Умение определить распределение переменной
A9	На соответствие	Основные виды распределений. Меры центральной тенденции и разброса	Умение соотнести меры центральной тенденции и разброса относительно типа распределения
A10	Множественный выбор (один или несколько правильных ответов)	Проверка гипотез. Z-тест и t-тест для одной выборки и для двух независимых выборок с одинаковыми дисперсиями. p-value. Статистическая значимость	Умение проверять гипотезы при помощи t-теста и Z-теста. Умение сделать вывод о принятии/опровержении нулевой гипотезы
A11	На соответствие	Проверка гипотез. Виды ошибок.	Умение определять тип ошибки при тестировании гипотезы

A12	Множественный выбор (один правильный ответ)	Постановка задачи машинного обучения. Виды задач машинного обучения	Умение определить тип задачи машинного обучения
A13	Множественный выбор (один или несколько правильных ответов)	Линейная регрессия. Интерпретация оценок коэффициентов в линейной регрессии	Умение интерпретировать оценки коэффициентов в модели линейной регрессии
A14	Множественный выбор (один или несколько правильных ответов)	Линейная регрессия. Метрики качества: MSE, MAE, R2	Умение интерпретировать метрики качества в линейной регрессии
A15	Множественный выбор (один или несколько правильных ответов)	Логистическая регрессия. ROC-кривая	Умение визуально интерпретировать ROC-кривую
B1	Численный ответ	Доверительные интервалы, основанные на Z- и t-распределениях, для одного параметра	Умение построить доверительный интервал для одного параметра
B2	Множественный выбор (один правильный ответ)	Z-тест и t-тест для одной выборки и для двух независимых выборок с одинаковыми дисперсиями	Умение проверить гипотезу при помощи t-теста и Z-теста
B3	Множественный выбор (один правильный ответ)	Корреляция Пирсона	Умение рассчитать выборочный коэффициент корреляции Пирсона
B4	Численный ответ	Хи-квадрат критерий согласия Пирсона	Умение рассчитать хи-квадрат критерий согласия Пирсона
C1.1	Численный ответ	Меры центральной тенденции и разброса.	Умение вычислить базовые описательные статистики
C1.2	Численный ответ	Меры центральной тенденции и разброса.	Умение вычислить базовые описательные статистики
C1.3	Численный ответ	Меры центральной тенденции и разброса.	Умение вычислить базовые описательные статистики
C2	Численный ответ ИЛИ Короткий ответ	Работа с данными: агрегирование, фильтрация, создание новых переменных	Умение вычислить значение признака или описательные статистики с группировкой по другому признаку
C3	Численный ответ	Работа с данными: агрегирование, фильтрация, создание новых переменных	Умение создавать новые переменные. Умение вычислить базовые описательные

			статистики
C4	Численный ответ	Работа с данными: агрегирование, фильтрация, создание новых переменных	Умение вычислить описательные статистики с группировкой по другому признаку
C5.1	Численный ответ	Работа с пропущенными значениями	Умение работать с пропущенными значениями. Умение заполнить пропущенные значения
C5.2	Численный ответ	Работа с пропущенными значениями	Умение работать с пропущенными значениями. Умение заполнить пропущенные значения
C6	Множественный выбор (один или несколько правильных ответов)	Визуализация	Умение построить визуализацию заданного типа. Умение интерпретировать визуализацию
C7	Множественный выбор (один или несколько правильных ответов)	Визуализация	Умение построить визуализацию заданного типа. Умение интерпретировать визуализацию
C8	Численный ответ	Корреляция Пирсона	Умение построить корреляционную матрицу
C9.1	Численный ответ	Линейная регрессия. Логистическая регрессия	Умение оценить логистическую и линейную регрессию. Умение рассчитать метрику качества оценки модели. Умение интерпретировать коэффициенты модели
C9.2	Численный ответ	Линейная регрессия. Логистическая регрессия	Умение оценить логистическую и линейную регрессию. Умение рассчитать метрику качества оценки модели. Умение интерпретировать коэффициенты модели