Перечень контрольных вопросов или тестовых заданий

**Задания закрытого типа**

**Задания альтернативного выбора**

*Выберите* ***один*** *правильный ответ*

**Простые (1 уровень)**

1. ANOVA - это …

А) дисперсионный анализ

Б) регрессионный анализ

В) факторный анализ

1. Процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью – это …

А) планирование эксперимента

Б) принятие решений

В) оценка рисков

1. Изучение предметной области, сбор данных, предварительная обработка данных, анализ данных, интерпретация найденных паттернов, использование новых знаний – это этапы …

А) интеллектуального анализа данных

Б) составления долгосрочного прогноза

В) принятия решений

1. Оценка коэффициентов регрессионной модели и коэффициента корреляции, проверка гипотезы о значимости коэффициентов, проверка гипотезы об адекватности модели – это основные этапы …

А) регрессионного анализа

Б) дисперсионного анализа

В) спектрального анализа

1. Вычисление корреляционной матрицы для всех переменных, участвующих в анализе, извлечение факторов, выбор факторов и вращение факторов для создания упрощенной структуры, интерпретация факторов – это этапы …

А) факторного анализа

Б) регрессионного анализа

В) дисперсионного анализа

**Средне –сложные (2 уровень)**

1. Обнаружение ошибок, удаление ошибок, обнаружение несоответствий в данных – это процессы…

А) очистки данных

Б) преобразования данных

В) хранения данных

1. Система интеграции данных – это …

А) объединение данных, находящихся в различных источниках, и предоставление данных пользователям в унифицированном виде

Б) выборка данных, находящихся в различных источниках

В) сортировка данных, находящихся в различных источниках

1. Интеграция данных создает новый единый источник информации, который представляет собой …

А) информационные ресурсы всей новой интегрированной совокупности источников

Б) ресурс, обеспечивающий единый унифицированный интерфейс

В) общий ресурс, объединяющий данные из группы источников

1. Содержание источников данных должно быть:

А) неизменным или обновляемым

Б) не обновляемым

В) изменяемым или не обновляемым

1. Анализируя соотношение среднего значения, моды и медианы можно сделать вывод о …

А) симметрии данных

Б) чистоте данных

В) корреляции

1. Метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях – это …

А) дисперсионный анализ

Б) факторный анализ

В) регрессионный анализ

1. Задачи, решаемые методами data mining, принято разделять на…

А) описательные и предсказательные

Б) статистические и оптимизационные

В) подготовку и сортировку

1. Для обработки результатов эксперимента применяется метод…

А) наименьших квадратов

Б) Рунге-Кутта

В) Монте-Карло

1. В задачах классификации используется логистическая ….

А) регрессия

Б) прогрессия

В) оптимизация

1. Логистическая регрессия применяется для прогнозирования, если отклик …

А) бинарный

Б) отрицательный

В) целый

1. Статистическую взаимосвязь между двумя случайными величинами показывает …

А) коэффициент корреляции

Б) дисперсия

В) коэффициент Фишера

1. Адекватность регрессионной модели может быть оценена с помощью …

А) коэффициента Фишера

Б) коэффициента корреляции

В) дисперсии

1. Исследуемые данные измеряются в единицах номинальной шкалы. Невозможно определить:

А) медиану

Б) моду

В) распределение частот

1. Исследуемые данные измеряются в единицах порядковой шкалы. Невозможно определить:

А) среднее значение

Б) моду

В) распределение частот

1. Мода, медиана и среднее значение для набора исследуемых данных совпадают. Это означает, что распределение этой случайной величины …

А) симметрично

Б) дискретно

В) асимметрично

1. Для вычисления дисперсии адекватности требуется знать …

А) расчётные и экспериментальные значения функции отклика

Б) результаты наблюдений в параллельных опытах

В) число степеней свободы

1. Критерий инверсий применяется для проверки гипотезы о

А) наличии тренда

Б) равенстве дисперсий

В) достоверности данных

**Сложные (3 уровень)**

1. Для вычисления дисперсии воспроизводимости требуется знать …

А) результаты наблюдений в параллельных опытах

Б) результаты наблюдений в одном опыте

В) значения факторов в параллельных опытах

1. В изучаемом наборе данных есть поля (столбцы) с информацией о времени оплаты проезда; сумме оплаты; способе оплаты (банковская карта транспортная карта, другое); координаты геопозиционирования транспортного средства; наименование банка, через который прошла оплата. Цель анализа прогноз изменения пассажиропотока. Дубликатами являются две записи (строки), если в них совпадает только сумма оплаты проезда.

А) не верно

Б) верно

В) при данных условиях нельзя дать точный ответ

1. Если статистическое значение критерия Фишера больше критического, то относительно исследуемого параметра делается вывод о статистически значимом ….

А) отличии в разных опытах

Б) совпадении в разных опытах

В) отличии в одном опыте

**Задания на установление соответствия**

*Установите соответствие между левым и правым столбцами.*

**Простые (1 уровень)**

|  |  |
| --- | --- |
| 26 Установите соответствие: |  |
| 1. Алгоритм кластеризации | А) K-means |
| 2. Алгоритм классификации | Б) «наивный Байес» |
|  | В) метод наименьших квадратов |

|  |  |
| --- | --- |
| 27 Установите соответствие: |  |
| 1. регрессионный анализ | А) метод наименьших квадратов |
| 2. дисперсионный анализ | Б) критерий Фишера |
|  | В) метод Монте-Карло |

**Средне-сложные (2 уровень)**

|  |  |
| --- | --- |
| 28 Установите соответствие: |  |
| 1. Логистическая регрессия | А) бинарный отклик |
| 2. Нормальное распределение | Б) симметричность |
|  | В) рост энтропии |

|  |  |
| --- | --- |
| 29 Установите соответствие: |  |
| 1. Алгоритм кластеризации | А) KNN |
| 2. Алгоритм классификации | Б) «наивный Байес» |
|  | В) метод Монте-Карло |

|  |  |
| --- | --- |
| 30 Установите соответствие: |  |
| 1. поиск минимума непрерывной функции | А) градиентные методы |
| 2. поиск минимума дискретной функции | Б) симплекс-метод |
|  | В) «наивный Байес» |

|  |  |
| --- | --- |
| 31 Установите соответствие: |  |
| 1. генетические алгоритмы | А) оптимизация |
| 2. «наивный Байес» | Б) глубокое обучение |
|  | В) дискретный анализ |

|  |  |
| --- | --- |
| 32 Установите соответствие: |  |
| 1. регрессия | А) метод наименьших квадратов |
| 2. кластеризация | Б) алгоритмы машинного обучения |
|  | В) метод Монте-Карло |

|  |  |
| --- | --- |
| 33 Установите соответствие: |  |
| 1. алгоритм «К ближайших соседей» | А) KNN |
| 2. алгоритм «К средних» | Б) K-means |
|  | В) ANOVA |

|  |  |
| --- | --- |
| 34 Установите соответствие: |  |
| 1 дисперсионный анализ | А) ANOVA |
| 2 кластеризация | Б) K-means |
|  | В) KNN |

**Сложные (3 уровень)**

|  |  |
| --- | --- |
| 35 Установите соответствие: |  |
| 1. эвристические алгоритмы | А) генетический алгоритм классификации |
| 2. алгоритмы быстрого спуска | Б) градиентный спуск |
|  | В) регрессионный анализ |

**Задания открытого типа**

**Задания на дополнение**

*Напишите пропущенное слово.*

**Простые (1 уровень)**

36 Обозначение дисперсионного анализа, принятое в международных источниках и программных продуктах: …(латиница)

37 Термин, введенный Пятецким-Шапиро, иногда используемый для обозначения интеллектуального анализа данных: … (латиница)

1. Объединение данных, находящихся в различных источниках, и предоставление данных пользователям в унифицированном виде – это система интеграции …

39 Метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях – это дисперсионный …

40 Статистическую взаимосвязь между двумя случайными величинами показывает коэффициент …

1. Адекватность регрессионной модели может быть оценена с помощью критерия …
2. Для нахождения доверительных границ к коэффициентам регрессии используется распределение …

**Средне-сложные (2 уровень)**

1. Обнаружение ошибок, удаление ошибок, обнаружение несоответствий в данных – это процессы очистки …
2. Для обработки результатов эксперимента применяется метод наименьших …
3. Для построения доверительного интервала к коэффициенту регрессии нужно задать доверительную …
4. В задачах кластеризации используется алгоритм K-….
5. Задачи исследования наличия взаимосвязей между отдельными группами переменных относятся к корреляционному …
6. Задачи, в которых переменные x1, x2, ..., xn имеют качественный характер, а исследуется и устанавливается степень их влияния на переменное y относятся к дисперсионному …
7. В задачах кластеризации используется алгоритм K ближайших ….
8. Алгоритм «наивный Байес» основывается на вычислении условной …
9. Алгоритм K-means требует формирования матрицы …
10. Для проверки гипотезы о наличии тренда применяется критерий …
11. Для вычисления дисперсии воспроизводимости требуется знать результаты наблюдений в параллельных …
12. Для вычисления дисперсии адекватности требуется знать расчётные и экспериментальные значения функции …
13. Для проверки гипотезы о равенстве дисперсий применяется критерий ….
14. Для проверки гипотезы о значимости коэффициента корреляции применяется критерий ….
15. Алгоритм KNN относится к алгоритмам глубокого …

1. Коэффициенты регрессионной модели находят с помощью метода наименьших ….
2. При анализе невозможно определить среднее значение, если исследуемые данные измеряются в единицах … шкалы
3. Отклонение от ожидаемых значений, которые могут указывать на ошибки, выбросы или необычные события называются …
4. Для проверки гипотезы о принадлежности двух наборов данных одной совокупности применяется критерий…
5. График плотности вероятности нормального распределения симметричен относительно линии, проходящей через математическое …
6. Распределение Стьюдента используется для построение доверительных ….
7. Коэффициент Стьюдента зависит от доверительной вероятности и числа степеней …
8. Условная вероятность вычисляется в алгоритме наивный …
9. Обучающий набор данных необходим для реализации алгоритмов глубокого …
10. Матрица центроидов необходима для реализации алгоритма кластеризации K- …

**Сложные (3 уровень)**

1. Функция excel, возвращающая среднее арифметическое значение всех ячеек, которые соответствуют нескольким условиям: … (кириллица)
2. Функция excel, которая применяет критерии к ячейкам в нескольких диапазонах и вычисляет количество соответствий всем критериям:… (кириллица)
3. Высокоуровневый язык программирования, на котором написаны большинство нейросетей называется …(латиница)