

**КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)
АВТОНОМНОЙ НЕКОММЕРЧЕСКОЙ ОБРАЗОВАТЕЛЬНОЙ
ОРГАНИЗАЦИИ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Специальность: 10.05.01 Компьютерная безопасность

Специализация: «Разработка защищенного программного обеспечения»

Формы обучения: очная

Объем дисциплины:

в зачетных единицах: 5 з.е.

в академических часах: 180 ак.ч.

Форма промежуточной аттестации: экзамен

Оценочные материалы по дисциплине «Многомерный статистический анализ»

обсуждены и рекомендованы к утверждению решением кафедры естественных дисциплин, сервиса и туризма от 25 марта 2025 г., протокол № 6

одобрены Научно-методическим советом Казанского кооперативного института (филиала) от 2 апреля 2025 г., протокол № 3

СОДЕРЖАНИЕ

1. Паспорт оценочных материалов по дисциплине
2. Оценочные материалы по дисциплине:
 - 2.1. Оценочные материалы для проведения текущего контроля.
 - 2.2. Оценочные материалы для проведения промежуточной аттестации.

Обновление оценочных материалов дисциплины

Целью оценочных материалов по дисциплине (модулю) (далее – ОМ) является комплексная оценка результатов обучения по дисциплине Многомерный статистический анализ и установление уровня сформированности компетенций обучающегося.

ОМ предназначены для проведения текущего контроля успеваемости и промежуточной аттестации.

ОМ включают оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации.

Оценочные материалы разработаны в соответствии с рабочей программой дисциплины Многомерный статистический анализ.

1. Паспорт оценочных материалов по дисциплине «Многомерный статистический анализ»

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения	Контролируемые разделы, темы дисциплины	Наименование оценочного средства ¹	
				Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6
ОПК-3. Способен на основании совокупности математических методов разрабатывать, обосновывать и реализовывать процедуры решения задач профессиональной деятельности	ОПК-3.4 Демонстрирует умение обосновывать и реализовывать процедуры решения стандартных задач профессиональной деятельности с применением методов анализа данных	Знать: основных математических методов и их модификаций для обоснования процедур решения задач профессиональной деятельности Уметь: реализовывать процедуры решения многомерных статистических задач с применением основных математических методов Владеть: навыками теоретического и экспериментального исследования совокупности математических методов для разработки, обоснования и реализации процедур решения задач профессиональной деятельности	Тема 1. Элементарные понятия статистики Тема 2. Измерительные шкалы. Работа с данными. Основные операции над случаями и переменными Тема 3. Основные статистики, корреляционный анализ Тема 4. Сравнение средних величин	Реферат (Тема 1) Тест (тема 1) Реферат (Тема 2) Тест (тема 2) Реферат (Тема 3) Тест (тема 3) Реферат (Тема 4) Тест (тема 4)	Тест (1-26 тестового задания)

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения	Контролируемые разделы, темы дисциплины	Наименование оценочного средства ¹	
				Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6
			критерием Стьюдента Тема 5. Непараметрическая статистика. Сравнение величин методами непараметрической статистики	Реферат (Тема 5) Тест (тема 5)	
ОПК-3.1 Способен проводить анализ программного кода с целью поиска потенциальных уязвимостей и недокументированных возможностей	ОПК-3.1.1 Знает методы и средства защиты информации в компьютерных сетях, операционных системах и системах управления базами данных	Знать: основные понятия и современные принципы многомерного анализа. Уметь: строить прогностические модели для прогнозирования новых угроз, строить закономерности на данных о ранее совершенных атаках Владеть: математическим аппаратом для построения моделей	Тема 6. Группировка и однофакторная ANOVA Тема 7. Дисперсионный анализ Тема 8. Таблицы частот, сопряженности, флагов и заголовков	Реферат (Тема 6) Тест (тема 6) Реферат (Тема 7) Тест (тема 7) Реферат (Тема 8) Тест (тема 8)	Тест (27-50 тестового задания)

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения	Контролируемые разделы, темы дисциплины	Наименование оценочного средства ¹	
				Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6
			Тема 9. Канонический анализ Тема 10. Линейное и нелинейное моделирование взаимосвязей	Реферат (Тема 9) Тест (тема 9) Реферат (Тема 10) Тест (тема 10)	
	ОПК-3.1.2 Демонстрирует умение применять методы анализа защищенности информации от атак на компьютерные системы и сети	Знать: основные понятия и современные принципы многомерного анализа. Уметь: логически верно, аргументировано и ясно строить устную и письменную речь Владеть: культурой мышления, навыками обобщенного анализа и восприятия информации.	Тема 11. Дискриминантный анализ Тема 12. Кластерный анализ Тема 13. Деревья классификации. Тема 14. Факторный анализ	Реферат (Тема 11) Тест (тема 11) Реферат (Тема 12) Тест (тема 12) Реферат (Тема 13) Тест (тема 13) Реферат (Тема 14) Тест (тема 14)	Тест (51-78 тестового задания)

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения	Контролируемые разделы, темы дисциплины	Наименование оценочного средства ¹	
				Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6
			Тема 15. Многомерное шкалирование Тема 16. Обзор пройденного материала	Реферат (Тема 15) Тест (тема 15) Реферат (Тема 16) Тест (тема 16)	

2. Оценочные материалы по дисциплине «Многомерный статистический анализ»

2.1 Оценочные материалы для проведения текущего контроля успеваемости

Тематика рефератов

Тема 1. Элементарные понятия статистики

1. История развития статистики как науки: от древних времен до современности
2. Основные задачи и направления статистики: описание, анализ и прогнозирование данных
3. Сравнение описательной и аналитической статистики: задачи и методы
4. Виды статистических данных: количественные и качественные данные
5. Методы сбора и обработки статистических данных: выборка, группировка, агрегирование
6. Статистические распределения: нормальное, биномиальное, Пуассона и другие
7. Центральные тенденции в статистике: среднее, медиана, мода
8. Меры разброса данных: дисперсия, стандартное отклонение, размах
9. Корреляция и регрессия: взаимосвязь между переменными
10. Графическое представление данных: гистограммы, диаграммы рассеяния, круговые диаграммы
11. Статистические гипотезы и их проверка: нулевая и альтернативная гипотезы
12. Критерии значимости в статистике: t-критерий, F-критерий, критерий хи-квадрат
13. Применение статистики в различных областях: экономика, медицина, социология
14. Проблемы и ограничения статистического анализа: предвзятость, ошибки выборки
15. Этика в статистике: конфиденциальность данных и непредвзятость анализа

Тема 2. Измерительные шкалы. Работа с данными. Основные операции над случаями и переменными

1. Типы измерительных шкал: номинальная, порядковая, интервальная и относительная
2. Применение измерительных шкал в различных областях: социология, экономика, медицина
3. Преобразование данных между различными типами шкал: методы и примеры
4. Кодирование и декодирование данных в статистическом анализе
5. Работа с пропущенными значениями в данных: методы заполнения и удаления
6. Методы нормализации и стандартизации данных: z-score, min-max scaling
7. Агрегирование данных: методы и примеры применения
8. Группировка данных: критерии и алгоритмы
9. Фильтрация данных: методы и примеры применения
10. Сортировка данных: алгоритмы и примеры применения
11. Трансформация данных: логарифмическое, степенное и другие преобразования
12. Работа с категориальными переменными: методы кодирования и интерпретации
13. Работа с временными рядами: методы анализа и прогнозирования
14. Проблемы и ограничения работы с данными: предвзятость, ошибки выборки
15. Этика в работе с данными: конфиденциальность и защита персональных данных

Тема 3. Основные статистики, корреляционный анализ

1. История развития корреляционного анализа: от зарождения до современных методов
2. Основные задачи и направления корреляционного анализа: выявление взаимосвязей

между переменными

3. Виды корреляции: положительная, отрицательная и нулевая корреляция
4. Коэффициент корреляции Пирсона: принципы, методы и примеры применения
5. Коэффициент корреляции Спирмена: принципы, методы и примеры применения
6. Коэффициент корреляции Кендалла: принципы, методы и примеры применения
7. Корреляционный анализ в экономике: прогнозирование и анализ данных
8. Корреляционный анализ в медицине: диагностика и прогнозирование заболеваний
9. Корреляционный анализ в социологии: анализ социальных процессов и явлений
10. Корреляционный анализ в маркетинге: сегментация клиентов и прогнозирование продаж
11. Корреляционный анализ в финансах: оценка рисков и прогнозирование рынков
12. Проблемы и ограничения корреляционного анализа: причинно-следственные связи и ложные корреляции
13. Методы визуализации корреляций: корреляционные матрицы и диаграммы рассеяния
14. Корреляционный анализ в больших данных: методы и примеры применения
15. Этика в корреляционном анализе: конфиденциальность и защита персональных данных

Тема 4. Сравнение средних величин критерием Стьюдента

1. История и развитие критерия Стьюдента: от Уильяма Госсета до современных применений
2. Основные задачи и направления применения критерия Стьюдента: сравнение средних величин
3. Виды критерия Стьюдента: для независимых и зависимых выборок
4. Критерий Стьюдента для двух независимых выборок: принципы, методы и примеры применения
5. Критерий Стьюдента для двух зависимых выборок: принципы, методы и примеры применения
6. Критерий Стьюдента в экономике: сравнение доходов, цен и других экономических показателей
7. Критерий Стьюдента в медицине: сравнение эффективности лечения и диагностики
8. Критерий Стьюдента в социологии: сравнение социальных показателей и явлений
9. Критерий Стьюдента в маркетинге: сравнение продаж и потребительского поведения
10. Критерий Стьюдента в финансах: сравнение доходности и рисков инвестиций
11. Критерий Стьюдента в образовании: сравнение успеваемости и результатов обучения
12. Проблемы и ограничения критерия Стьюдента: нормальность распределения и равенство дисперсий
13. Альтернативы критерию Стьюдента: непараметрические методы и методы для малых выборок
14. Критерий Стьюдента в больших данных: методы и примеры применения
15. Этика в применении критерия Стьюдента: конфиденциальность и защита персональных данных

Тема 5. Непараметрическая статистика. Сравнение величин методами непараметрической статистики.

1. История и развитие непараметрической статистики: от зарождения до современных методов
2. Основные задачи и направления непараметрической статистики: сравнение величин без предположений о распределении
3. Сравнение двух независимых выборок: U-критерий Манна-Уитни
4. Сравнение двух зависимых выборок: критерий Вилкоксона

5. Сравнение нескольких независимых выборок: критерий Краскела-Уоллиса
6. Сравнение нескольких зависимых выборок: критерий Фридмана
7. Непараметрическая статистика в экономике: сравнение доходов, цен и других экономических показателей
8. Непараметрическая статистика в медицине: сравнение эффективности лечения и диагностики
9. Непараметрическая статистика в социологии: сравнение социальных показателей и явлений
10. Непараметрическая статистика в маркетинге: сравнение продаж и потребительского поведения
11. Непараметрическая статистика в финансах: сравнение доходности и рисков инвестиций
12. Непараметрическая статистика в образовании: сравнение успеваемости и результатов обучения
13. Проблемы и ограничения непараметрической статистики: мощность тестов и чувствительность к выбросам
14. Преимущества непараметрической статистики: устойчивость к нарушениям нормальности и равенства дисперсий
15. Этика в применении непараметрической статистики: конфиденциальность и защита персональных данных

Тема 6. Группировка и однофакторная ANOVA.

1. История и развитие однофакторного дисперсионного анализа (ANOVA): от Фишера до современных методов
2. Основные задачи и направления однофакторного ANOVA: сравнение средних величин в нескольких группах
3. Принципы и методы однофакторного ANOVA: разбиение общей дисперсии на компоненты
4. Критерий Фишера в однофакторном ANOVA: принципы, методы и примеры применения
5. Постфактум тесты в однофакторном ANOVA: методы множественных сравнений (Тьюки, Шеффе, Бонферрони)
6. Однофакторный ANOVA в экономике: сравнение доходов, цен и других экономических показателей
7. Однофакторный ANOVA в медицине: сравнение эффективности лечения и диагностики
8. Однофакторный ANOVA в социологии: сравнение социальных показателей и явлений
9. Однофакторный ANOVA в маркетинге: сравнение продаж и потребительского поведения
10. Однофакторный ANOVA в финансах: сравнение доходности и рисков инвестиций
11. Однофакторный ANOVA в образовании: сравнение успеваемости и результатов обучения
12. Проблемы и ограничения однофакторного ANOVA: нормальность распределения и равенство дисперсий
13. Альтернативы однофакторному ANOVA: непараметрические методы и методы для малых выборок
14. Однофакторный ANOVA в больших данных: методы и примеры применения
15. Этика в применении однофакторного ANOVA: конфиденциальность и защита персональных данных

Тема 7. Дисперсионный анализ

1. История и развитие дисперсионного анализа: от Фишера до современных методов
2. Основные задачи и направления дисперсионного анализа: сравнение средних величин в нескольких группах
3. Принципы и методы однофакторного дисперсионного анализа (ANOVA): разбиение общей дисперсии на компоненты
4. Принципы и методы двухфакторного дисперсионного анализа: учет взаимодействия факторов
5. Критерий Фишера в дисперсионном анализе: принципы, методы и примеры применения
6. Постфактум тесты в дисперсионном анализе: методы множественных сравнений (Тьюки, Шеффе, Бонферрони)
7. Дисперсионный анализ в экономике: сравнение доходов, цен и других экономических показателей
8. Дисперсионный анализ в медицине: сравнение эффективности лечения и диагностики
9. Дисперсионный анализ в социологии: сравнение социальных показателей и явлений
10. Дисперсионный анализ в маркетинге: сравнение продаж и потребительского поведения
11. Дисперсионный анализ в финансах: сравнение доходности и рисков инвестиций
12. Дисперсионный анализ в образовании: сравнение успеваемости и результатов обучения
13. Проблемы и ограничения дисперсионного анализа: нормальность распределения и равенство дисперсий
14. Альтернативы дисперсионному анализу: непараметрические методы и методы для малых выборок
15. Этика в применении дисперсионного анализа: конфиденциальность и защита персональных данных

Тема 8. Таблицы частот, сопряженности, флагов и заголовков

1. История и развитие таблиц частот и сопряженности: от зарождения до современных методов
2. Основные задачи и направления использования таблиц частот и сопряженности: анализ данных и выявление взаимосвязей
3. Принципы построения таблиц частот: одномерные и многомерные таблицы
4. Принципы построения таблиц сопряженности: анализ взаимосвязей между категориальными переменными
5. Методы анализа таблиц сопряженности: критерий хи-квадрат и другие статистические тесты
6. Использование таблиц флагов и заголовков: структурирование и визуализация данных
7. Таблицы частот и сопряженности в экономике: анализ доходов, цен и других экономических показателей
8. Таблицы частот и сопряженности в медицине: анализ заболеваемости и эффективности лечения
9. Таблицы частот и сопряженности в социологии: анализ социальных процессов и явлений
10. Таблицы частот и сопряженности в маркетинге: анализ потребительского поведения и сегментации
11. Таблицы частот и сопряженности в финансах: анализ доходности и рисков инвестиций

12. Таблицы частот и сопряженности в образовании: анализ успеваемости и результатов обучения
13. Проблемы и ограничения таблиц частот и сопряженности: предвзятость, ошибки выборки
14. Методы визуализации данных на основе таблиц частот и сопряженности: диаграммы, графики, тепловые карты
15. Этика в использовании таблиц частот и сопряженности: конфиденциальность и защита персональных данных

Тема 9. Канонический анализ

1. История и развитие канонического анализа: от зарождения до современных методов
2. Основные задачи и направления канонического анализа: выявление взаимосвязей между множествами переменных
3. Принципы и методы канонического корреляционного анализа: максимизация корреляции между линейными комбинациями переменных
4. Канонический дискриминантный анализ: классификация и разделение групп на основе канонических переменных
5. Канонический анализ в экономике: анализ взаимосвязей между экономическими показателями
6. Канонический анализ в медицине: выявление взаимосвязей между клиническими и лабораторными показателями
7. Канонический анализ в социологии: анализ взаимосвязей между социальными переменными
8. Канонический анализ в маркетинге: анализ взаимосвязей между потребительскими предпочтениями и характеристиками продуктов
9. Канонический анализ в финансах: анализ взаимосвязей между финансовыми показателями и рисками
10. Канонический анализ в образовании: анализ взаимосвязей между успеваемостью и факторами обучения
11. Проблемы и ограничения канонического анализа: мультиколлинеарность, размерность данных
12. Методы визуализации результатов канонического анализа: графики, диаграммы, тепловые карты
13. Канонический анализ в больших данных: методы и примеры применения
14. Сравнение канонического анализа с другими методами многомерного анализа: факторный анализ, дискриминантный анализ
15. Этика в применении канонического анализа: конфиденциальность и защита персональных данных

Тема 10. Линейное и нелинейное моделирование взаимосвязей

1. История и развитие методов линейного и нелинейного моделирования: от простых моделей до современных подходов
2. Основные задачи и направления линейного и нелинейного моделирования: прогнозирование и анализ взаимосвязей
3. Линейная регрессия: принципы, методы и примеры применения
4. Нелинейная регрессия: принципы, методы и примеры применения
5. Сравнение линейных и нелинейных моделей: преимущества и ограничения
6. Методы оценки параметров линейных и нелинейных моделей: метод наименьших квадратов и его модификации

7. Линейное и нелинейное моделирование в экономике: прогнозирование и анализ данных
8. Линейное и нелинейное моделирование в медицине: диагностика и прогнозирование заболеваний
9. Линейное и нелинейное моделирование в социологии: анализ социальных процессов и явлений
10. Линейное и нелинейное моделирование в маркетинге: прогнозирование продаж и потребительского поведения
11. Линейное и нелинейное моделирование в финансах: прогнозирование цен на акции и анализ рисков
12. Линейное и нелинейное моделирование в образовании: прогнозирование успеваемости и результатов обучения
13. Проблемы и ограничения линейного и нелинейного моделирования: мультиколлинеарность, выбросы, нелинейность данных
14. Методы визуализации результатов линейного и нелинейного моделирования: графики, диаграммы, поверхности отклика
15. Этика в применении линейного и нелинейного моделирования: конфиденциальность и защита персональных данных

Тема 11. Дискриминантный анализ

1. История и развитие дискриминантного анализа: от Фишера до современных методов
2. Основные задачи и направления дискриминантного анализа: классификация и разделение групп
3. Принципы и методы дискриминантного анализа: линейный дискриминантный анализ (LDA)
4. Квадратичный дискриминантный анализ (QDA): принципы, методы и примеры применения
5. Регуляризованный дискриминантный анализ (RDA): принципы, методы и примеры применения
6. Дискриминантный анализ в экономике: классификация и прогнозирование экономических показателей
7. Дискриминантный анализ в медицине: диагностика заболеваний и прогнозирование исходов
8. Дискриминантный анализ в социологии: анализ социальных процессов и явлений
9. Дискриминантный анализ в маркетинге: сегментация клиентов и прогнозирование продаж
10. Дискриминантный анализ в финансах: оценка кредитоспособности и выявление мошенничества
11. Дискриминантный анализ в образовании: оценка успеваемости и прогнозирование результатов обучения
12. Проблемы и ограничения дискриминантного анализа: нормальность распределения, равенство дисперсий
13. Методы визуализации результатов дискриминантного анализа: графики, диаграммы, канонические переменные
14. Сравнение дискриминантного анализа с другими методами классификации: логистическая регрессия, деревья решений
15. Этика в применении дискриминантного анализа: конфиденциальность и защита персональных данных

Тема 12. Кластерный анализ

1. История и развитие кластерного анализа: от зарождения до современных методов
2. Основные задачи и направления кластерного анализа: группировка и классификация данных
3. Иерархические методы кластерного анализа: агломеративные и дивизимные методы
4. Нехирерические методы кластерного анализа: метод k-средних, DBSCAN
5. Кластерный анализ в экономике: сегментация рынков и анализ данных
6. Кластерный анализ в медицине: классификация заболеваний и анализ медицинских данных
7. Кластерный анализ в социологии: анализ социальных процессов и явлений
8. Кластерный анализ в маркетинге: сегментация клиентов и анализ потребительского поведения
9. Кластерный анализ в финансах: анализ финансовых данных и выявление аномалий
10. Кластерный анализ в образовании: анализ успеваемости и результатов обучения
11. Проблемы и ограничения кластерного анализа: выбор числа кластеров, чувствительность к выбросам
12. Методы визуализации результатов кластерного анализа: дендрограммы, кластерные карты
13. Сравнение кластерного анализа с другими методами классификации: дискриминантный анализ, деревья решений
14. Кластерный анализ в больших данных: методы и примеры применения
15. Этика в применении кластерного анализа: конфиденциальность и защита персональных данных

Тема 13. Деревья классификации

1. История и развитие деревьев классификации: от зарождения до современных методов
2. Основные задачи и направления деревьев классификации: классификация и прогнозирование
3. Принципы и методы построения деревьев классификации: ID3, C4.5, CART
4. Критерии разделения узлов в деревьях классификации: энтропия, информационный выигрыш, Gini-индекс
5. Деревья классификации в экономике: прогнозирование и анализ данных
6. Деревья классификации в медицине: диагностика заболеваний и прогнозирование исходов
7. Деревья классификации в социологии: анализ социальных процессов и явлений
8. Деревья классификации в маркетинге: сегментация клиентов и прогнозирование продаж
9. Деревья классификации в финансах: оценка кредитоспособности и выявление мошенничества
10. Деревья классификации в образовании: оценка успеваемости и прогнозирование результатов обучения
11. Проблемы и ограничения деревьев классификации: переобучение, чувствительность к выбросам
12. Методы визуализации деревьев классификации: графическое представление и интерпретация
13. Сравнение деревьев классификации с другими методами классификации: логистическая регрессия, дискриминантный анализ
14. Деревья классификации в больших данных: методы и примеры применения
15. Этика в применении деревьев классификации: конфиденциальность и защита персональных данных

Тема 14. Факторный анализ

1. История и развитие факторного анализа: от Спирмена до современных методов
2. Основные задачи и направления факторного анализа: выявление латентных факторов и сокращение размерности данных
3. Принципы и методы факторного анализа: метод главных компонент (РСА), метод максимального правдоподобия
4. Вращение факторов в факторном анализе: ортогональное и косоугольное вращение
5. Факторный анализ в экономике: анализ финансовых данных и выявление факторов роста
6. Факторный анализ в медицине: анализ медицинских данных и выявление факторов риска заболеваний
7. Факторный анализ в социологии: анализ социальных процессов и явлений
8. Факторный анализ в маркетинге: анализ потребительского поведения и сегментация рынка
9. Факторный анализ в финансах: анализ финансовых рынков и выявление факторов доходности
10. Факторный анализ в образовании: анализ успеваемости и факторов успеваемости
11. Проблемы и ограничения факторного анализа: интерпретация факторов, выбор числа факторов
12. Методы визуализации результатов факторного анализа: факторные нагрузки, диаграммы рассеяния
13. Сравнение факторного анализа с другими методами сокращения размерности: метод главных компонент, канонический анализ
14. Факторный анализ в больших данных: методы и примеры применения
15. Этика в применении факторного анализа: конфиденциальность и защита персональных данных

Тема 15. Многомерное шкалирование

1. История и развитие многомерного шкалирования: от Торгерсона до современных методов
2. Основные задачи и направления многомерного шкалирования: визуализация и анализ данных в многомерном пространстве
3. Методы многомерного шкалирования: метрическое и неметрическое многомерное шкалирование
4. Многомерное шкалирование в экономике: анализ и визуализация экономических данных
5. Многомерное шкалирование в медицине: анализ и визуализация медицинских данных
6. Многомерное шкалирование в социологии: анализ и визуализация социальных процессов и явлений
7. Многомерное шкалирование в маркетинге: анализ и визуализация потребительского поведения
8. Многомерное шкалирование в финансах: анализ и визуализация финансовых данных
9. Многомерное шкалирование в образовании: анализ и визуализация успеваемости и результатов обучения
10. Проблемы и ограничения многомерного шкалирования: интерпретация результатов, выбор метрики расстояния
11. Методы визуализации результатов многомерного шкалирования: диаграммы рассеяния, карты Саммона
12. Сравнение многомерного шкалирования с другими методами визуализации данных: факторный анализ, многомерное шкалирование
13. Многомерное шкалирование в больших данных: методы и примеры применения

14. Многомерное шкалирование в анализе текстовых данных: визуализация семантических пространств
15. Этика в применении многомерного шкалирования: конфиденциальность и защита персональных данных

Тема 16. Обзор пройденного материала

1. Основные понятия и методы многомерного статистического анализа: обзор и сравнение
2. Корреляционный анализ и линейное моделирование: взаимосвязь и применение
3. Дискриминантный и кластерный анализ: классификация и группировка данных
4. Факторный анализ и многомерное шкалирование: сокращение размерности и визуализация данных
5. Сравнение методов многомерного анализа: преимущества и ограничения
6. Применение многомерного анализа в экономике: прогнозирование и анализ данных
7. Применение многомерного анализа в медицине: диагностика и прогнозирование заболеваний
8. Применение многомерного анализа в социологии: анализ социальных процессов и явлений
9. Применение многомерного анализа в маркетинге: сегментация клиентов и прогнозирование продаж
10. Применение многомерного анализа в финансах: оценка рисков и прогнозирование рынков
11. Применение многомерного анализа в образовании: анализ успеваемости и результатов обучения
12. Проблемы и ограничения многомерного анализа: предвзятость, надежность, интерпретация результатов
13. Методы визуализации результатов многомерного анализа: графики, диаграммы, карты
14. Этика в применении многомерного анализа: конфиденциальность и защита персональных данных
15. Перспективы и направления развития многомерного статистического анализа: большие данные, искусственный интеллект

Критерии оценки выполнения рефератов по учебной дисциплине

Оценка «отлично» выставляется студенту, если тема реферата раскрыта в полной мере и все требования по написанию реферата соблюдены.

Оценка «хорошо» выставляется студенту, если недостаточно раскрыта тема реферата, но все требования по написанию реферата соблюдены.

Оценка «удовлетворительно» выставляется студенту, если недостаточно раскрыта тема реферата, не все требования по написанию реферата соблюдены, присутствуют ошибки и неточности в работе.

Оценка «неудовлетворительно» выставляется студенту, если отсутствует реферат.

Рекомендации по написанию реферата:

Реферат — письменная работа, выполняемая обучающимся в течение определенного срока.

Реферат — краткое точное изложение сущности какого-либо вопроса, темы на основе одной или нескольких книг, монографий или других первоисточников. Реферат должен содержать основные фактические сведения и выводы по рассматриваемому вопросу.

Реферат отвечает на вопрос — что содержится в данной публикации (публикациях). Реферат — не механический пересказ работы, а изложение ее существа.

Кроме реферирования прочитанной литературы, от обучающегося требуется аргументированное изложение собственных мыслей по рассматриваемому вопросу. Тему реферата предлагает преподаватель или сам обучающийся, в последнем случае она должна быть согласована с преподавателем.

В реферате нужны развернутые аргументы, рассуждения, сравнения. Материал подается не столько в развитии, сколько в форме констатации или описания.

Содержание реферируемого произведения излагается объективно от имени автора. Если в первичном документе главная мысль сформулирована недостаточно четко, в реферате она должна быть конкретизирована и выделена.

Структура реферата:

1. Титульный лист.
2. После титульного листа на отдельной странице следует оглавление (план, содержание), в котором указаны названия всех разделов (пунктов плана) реферата и номера страниц, указывающие начало этих разделов в тексте реферата.
3. После оглавления следует введение. Объем введения составляет 1,5-2 страницы.
4. Основная часть реферата может иметь одну или несколько глав, состоящих из 2-3 параграфов (подпунктов, разделов) и предполагает осмысленное и логичное изложение главных положений и идей, содержащихся в изученной литературе. В тексте обязательны ссылки на первоисточники. В том случае если цитируется или используется чья-либо неординарная мысль, идея, вывод, приводится какой-либо цифрой материал, таблицу - обязательно сделайте ссылку на того автора у кого вы взяли данный материал.
5. Заключение содержит главные выводы, и итоги из текста основной части, в нем отмечается, как выполнены задачи и достигнуты ли цели, сформулированные во введении.
6. Приложение может включать графики, таблицы, расчеты.
7. Библиография (список литературы) здесь указывается реально использованная для написания реферата литература. Список составляется согласно правилам библиографического описания.

Этапы работы над рефератом.

Работу над рефератом можно условно подразделить на три этапа:

1. Подготовительный этап, включающий изучение предмета исследования;
2. Изложение результатов изучения в виде связного текста;
3. Устное сообщение по теме реферата.

1. Подготовительный этап работы.

Формулировка темы. Подготовительная работа над рефератом начинается с формулировки темы. Тема в концентрированном виде выражает содержание будущего текста, фиксируя как предмет исследования, так и его ожидаемый результат. Для того чтобы работа над рефератом была успешной, необходимо, чтобы тема заключала в себе проблему, скрытый вопрос (даже если наука уже давно дала ответ на этот вопрос, студент, только знакомящийся с соответствующей областью знаний, будет вынужден искать ответ заново, что даст толчок к развитию проблемного, исследовательского мышления).

Поиск источников. Грамотно сформулированная тема зафиксировала предмет изучения; задача обучающегося — найти информацию, относящуюся к данному предмету и разрешить поставленную проблему. Выполнение этой задачи начинается с поиска источников. На этом этапе необходимо вспомнить, как работать с энциклопедиями и энциклопедическими словарями.

Работа с источниками. Работу с источниками надо начинать с ознакомительного чтения, т. е. просмотреть текст, выделяя его структурные единицы. При ознакомительном чтении закладками отмечаются те страницы, которые требуют более внимательного изучения. В зависимости от результатов ознакомительного чтения выбирается дальнейший способ работы с источником. Если для разрешения поставленной задачи требуется изучение

некоторых фрагментов текста, то используется метод выборочного чтения. Если в книге нет подробного оглавления, следует обратить внимание на предметные и именные указатели.

Избранные фрагменты или весь текст (если он целиком имеет отношение к теме) требуют вдумчивого, неторопливого чтения с «мысленной проработкой» материала. Такое чтение предполагает выделение: 1) главного в тексте; 2) основных аргументов; 3) выводов. Особое внимание следует обратить на то, вытекает тезис из аргументов или нет. Необходимо также проанализировать, какие из утверждений автора носят проблематичный, гипотетический характер и уловить скрытые вопросы.

Понятно, что умение таким образом работать с текстом приходит далеко не сразу. Наилучший способ научиться выделять главное в тексте, улавливать проблематичный характер утверждений, давать оценку авторской позиции — это сравнительное чтение, в ходе которого студент знакомится с различными мнениями по одному и тому же вопросу, сравнивает весомость и доказательность аргументов сторон и делает вывод о наибольшей убедительности той или иной позиции.

Создание конспектов для написания реферата.

Подготовительный этап работы завершается созданием конспектов, фиксирующих основные тезисы и аргументы. Здесь важно вспомнить, что конспекты пишутся на одной стороне листа, с полями и достаточным для исправления и ремарок межстрочным расстоянием (эти правила соблюдаются для удобства редактирования). Если в конспектах приводятся цитаты, то непременно должно быть дано указание на источник (автор, название, выходные данные, № страницы).

По завершении предварительного этапа можно переходить непосредственно к созданию текста реферата.

2. Создание текста.

Общие требования к тексту.

Текст реферата должен подчиняться определенным требованиям: он должен раскрывать тему, обладать связностью и цельностью.

Раскрытие темы предполагает, что в тексте реферата излагается относящийся к теме материал и предлагаются пути решения содержащейся в теме проблемы; связность текста предполагает смысловую соотносительность отдельных компонентов, а цельность — смысловую законченность текста.

С точки зрения связности все тексты делятся на тексты-констатации и тексты-рассуждения. Тексты-констатации содержат результаты ознакомления с предметом и фиксируют устойчивые и несомненные суждения. В текстах-рассуждениях одни мысли извлекаются из других, некоторые ставятся под сомнение, дается им оценка, выдвигаются различные предположения.

План реферата.

Универсальный план реферата — введение, основной текст и заключение.

Требования к введению.

Во введении аргументируется актуальность исследования, — т. е. выявляется практическое и теоретическое значение данного исследования. Далее констатируется, что сделано в данной области предшественниками; перечисляются положения, которые должны быть обоснованы. Введение может также содержать обзор источников или экспериментальных данных, уточнение исходных понятий и терминов, сведения о методах исследования. Во введении обязательно формулируются цель и задачи реферата.

Объем введения — в среднем около 10% от общего объема реферата.

Основная часть реферата.

Основная часть реферата раскрывает содержание темы. Она наиболее значительна по объему, наиболее значима и ответственна. В ней обосновываются основные тезисы реферата, приводятся развернутые аргументы, предполагаются гипотезы, касающиеся существа обсуждаемого вопроса. Важно проследить, чтобы основная часть не имела форму монолога. Аргументируя собственную позицию, можно и должно анализировать, и оценивать позиции

различных исследователей, с чем-то соглашаться, чему-то возражать, кого-то опровергать. Текст основной части делится на главы, параграфы, пункты. План основной части может быть составлен с использованием различных методов группировки материала: классификации (эмпирические исследования), типологии (теоретические исследования), периодизации (исторические исследования).

Заключение.

Заключение — последняя часть научного текста. В ней краткой и сжатой форме излагаются полученные результаты, представляющие собой ответ на главный вопрос исследования. Здесь же могут намечаться и дальнейшие перспективы развития темы. Небольшое по объему сообщение также не может обойтись без заключительной части - пусть это будут две-три фразы. Но в них должен подводиться итог проделанной работы.

Список использованной литературы.

Реферат любого уровня сложности обязательно сопровождается списком используемой литературы. Названия книг в списке располагают по алфавиту с указанием выходных данных использованных книг.

Требования, предъявляемые к оформлению реферата

Объемы рефератов – от 10-18 машинописных страниц. Работа выполняется на одной стороне листа стандартного формата. По обеим сторонам листа оставляются поля размером 35 мм. слева и 15 мм. справа, рекомендуется шрифт 12-14, интервал - 1,5. Все листы реферата должны быть пронумерованы. Каждый вопрос в тексте должен иметь заголовок в точном соответствии с наименованием в плане-оглавлении. При написании и оформлении реферата следует избегать типичных ошибок, например, таких:

- поверхностное изложение основных теоретических вопросов выбранной темы, когда автор не понимает, какие проблемы в тексте являются главными, а какие второстепенными,

- в некоторых случаях проблемы, рассматриваемые в разделах, не раскрывают основных аспектов выбранной для реферата темы,

- дословное переписывание книг, статей, заимствования рефератов из интернета и т.

д.

При проверке реферата преподавателем оцениваются:

1. Знания и умения на уровне требований стандарта конкретной дисциплины: знание фактического материала, усвоение общих представлений, понятий, идей.

2. Характеристика реализации цели и задач исследования (новизна и актуальность поставленных в реферате проблем, правильность формулирования цели, определения задач исследования, правильность выбора методов решения задач и реализации цели; соответствие выводов решаемым задачам, поставленной цели, убедительность выводов).

3. Степень обоснованности аргументов и обобщений (полнота, глубина, всесторонность раскрытия темы, логичность и последовательность изложения материала, корректность аргументации и системы доказательств, характер и достоверность примеров, иллюстративного материала, широта кругозора автора, наличие знаний интегрированного характера, способность к обобщению).

4. Качество и ценность полученных результатов (степень завершенности реферативного исследования, спорность или однозначность выводов).

5. Использование литературных источников.

6. Культура письменного изложения материала.

7. Культура оформления материалов работы.

Типовые тесты

Тема 1. Элементарные понятия статистики

Вопрос 1 (выбор одного правильного ответа):

Что такое среднее арифметическое?

- A) Сумма всех значений, деленная на количество значений.
- B) Значение, которое встречается чаще всего.
- C) Значение, которое делит упорядоченный набор данных на две равные части.
- D) Разница между максимальным и минимальным значениями.

Ответ: A

Обоснование: Среднее арифметическое — это сумма всех значений, деленная на количество значений. Это один из основных показателей центральной тенденции.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих величин являются мерами разброса данных?

- A) Дисперсия.
- B) Стандартное отклонение.
- C) Медиана.
- D) Размах.

Ответ: A, B, D

Обоснование: Меры разброса данных включают:

Дисперсию — среднее квадратическое отклонение от среднего.

Стандартное отклонение — квадратный корень из дисперсии.

Размах — разница между максимальным и минимальным значениями.

Вопрос 3 (установление соответствия):

Установите соответствие между типами шкал и примерами переменных:

Номинальная шкала

Порядковая шкала

Интервальная шкала

Относительная шкала

Примеры переменных:

- A) Цвет глаз.
- B) Оценка по пятибалльной шкале.
- C) Температура в градусах Цельсия.
- D) Рост человека в сантиметрах.

Ответ:

Номинальная шкала — A (Цвет глаз).

Порядковая шкала — B (Оценка по пятибалльной шкале).

Интервальная шкала — C (Температура в градусах Цельсия).

Относительная шкала — D (Рост человека в сантиметрах).

Обоснование:

Номинальная шкала используется для классификации данных без упорядочения (например, цвет глаз).

Порядковая шкала используется для упорядочения данных без точного измерения

разницы (например, оценка по пятибалльной шкале).

Интервальная шкала используется для измерения с равными интервалами, но без абсолютного нуля (например, температура в градусах Цельсия).

Относительная шкала используется для измерения с абсолютным нулем и равными интервалами (например, рост человека в сантиметрах).

Вопрос 4 (вопрос с открытым ответом):

Что такое медиана в статистике?

Ответ: Медиана — это значение, которое делит упорядоченный набор данных на две равные части. Половина данных находится ниже медианы, а другая половина — выше.

Обоснование: Медиана — это один из показателей центральной тенденции, который особенно полезен при наличии выбросов в данных, так как она менее чувствительна к экстремальным значениям.

Вопрос 5 (вопрос с кратким ответом):

Как называется мера разброса данных, которая показывает среднее квадратическое отклонение от среднего?

Ответ: Дисперсия

Обоснование: Дисперсия — это мера разброса данных, которая показывает среднее квадратическое отклонение от среднего. Она используется для оценки вариации данных вокруг среднего значения.

Вопрос 5:

Что такое переобучение в машинном обучении?

A) Когда модель слишком хорошо обучается на тренировочных данных и плохо обобщает на новых данных.

B) Когда модель недостаточно обучается на тренировочных данных.

C) Когда модель обучается только на малых данных.

D) Когда модель обучается только на больших данных.

Ответ: A

Обоснование: Переобучение происходит, когда модель машинного обучения слишком хорошо подстраивается под тренировочные данные, включая их шум и случайные колебания. В результате модель плохо обобщает на новых, ранее не виденных данных, что снижает ее предсказательную способность.

Тема 2. Измерительные шкалы. Работа с данными. Основные операции над случаями и переменными.

Вопрос 1 (выбор одного правильного ответа):

Что такое номинальная шкала?

A) Шкала, которая упорядочивает данные по рангу.

B) Шкала, которая классифицирует данные без упорядочения.

C) Шкала, которая измеряет данные с равными интервалами.

D) Шкала, которая измеряет данные с абсолютным нулем.

Ответ: B

Обоснование: Номинальная шкала используется для классификации данных без упорядочения. Примеры включают такие переменные, как цвет глаз, пол или национальность.

Вопрос 2 (выбор нескольких правильных ответов):

Какие операции можно выполнять с данными, измеренными по интервальной шкале?

- A) Сложение и вычитание.
- B) Умножение и деление.
- C) Сравнение на равенство и неравенство.
- D) Вычисление среднего арифметического.

Ответ: A, C, D

Обоснование: С данными, измеренными по интервальной шкале, можно выполнять следующие операции:

Сложение и вычитание — для вычисления разницы между значениями.

Сравнение на равенство и неравенство — для определения, равны ли значения или нет.

Вычисление среднего арифметического — для нахождения центральной тенденции.

Умножение и деление не имеют смысла для интервальной шкалы, так как она не имеет абсолютного нуля.

Вопрос 3 (установление соответствия):

Установите соответствие между типами шкал и примерами переменных:

Номинальная шкала

Порядковая шкала

Интервальная шкала

Относительная шкала

Примеры переменных:

- A) Цвет глаз.
- B) Оценка по пятибалльной шкале.
- C) Температура в градусах Цельсия.
- D) Рост человека в сантиметрах.

Ответ:

Номинальная шкала — A (Цвет глаз).

Порядковая шкала — B (Оценка по пятибалльной шкале).

Интервальная шкала — C (Температура в градусах Цельсия).

Относительная шкала — D (Рост человека в сантиметрах).

Обоснование:

Номинальная шкала используется для классификации данных без упорядочения (например, цвет глаз).

Порядковая шкала используется для упорядочения данных без точного измерения разницы (например, оценка по пятибалльной шкале).

Интервальная шкала используется для измерения с равными интервалами, но без абсолютного нуля (например, температура в градусах Цельсия).

Относительная шкала используется для измерения с абсолютным нулем и равными интервалами (например, рост человека в сантиметрах).

Вопрос 4 (вопрос с открытым ответом):

Что такое порядковая шкала?

Ответ: Порядковая шкала — это шкала, которая упорядочивает данные по рангу, но не позволяет точно измерить разницу между значениями. Примеры включают такие переменные, как оценка по пятибалльной шкале или место в соревновании.

Обоснование: Порядковая шкала позволяет ранжировать данные, но не предоставляет информации о точных различиях между значениями. Например, разница между первым и вторым местом может быть больше, чем между вторым и третьим, но порядковая шкала этого не отражает.

Вопрос 5 (вопрос с кратким ответом):

Какая шкала используется для измерения температуры в градусах Цельсия?

Ответ: Интервальная шкала

Обоснование: Температура в градусах Цельсия измеряется по интервальной шкале, так как она имеет равные интервалы между значениями, но не имеет абсолютного нуля. Это означает, что разница между 10°C и 20°C такая же, как между 20°C и 30°C , но 0°C не означает полное отсутствие температуры.

Тема 3. Основные статистики, корреляционный анализ

Вопрос 1 (выбор одного правильного ответа):

Что такое корреляция в статистике?

- A) Мера зависимости между двумя переменными.
- B) Мера разброса данных.
- C) Мера центральной тенденции.
- D) Мера вероятности события.

Ответ: A

Обоснование: Корреляция — это мера зависимости между двумя переменными. Она показывает, насколько сильно связаны изменения одной переменной с изменениями другой.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих величин являются мерами центральной тенденции?

- A) Среднее арифметическое.
- B) Медиана.
- C) Мода.
- D) Дисперсия.

Ответ: A, B, C

Обоснование: Меры центральной тенденции включают:

Среднее арифметическое — сумма всех значений, деленная на количество значений.
Медиана — значение, которое делит упорядоченный набор данных на две равные части.

Мода — значение, которое встречается чаще всего.

Вопрос 3 (установление соответствия):

Установите соответствие между типами корреляции и примерами:

Положительная корреляция

Отрицательная корреляция

Нулевая корреляция

Примеры:

А) Рост и вес человека.

В) Температура и количество людей на пляже.

С) Цвет глаз и рост человека.

Ответ:

Положительная корреляция — А (Рост и вес человека).

Отрицательная корреляция — В (Температура и количество людей на пляже).

Нулевая корреляция — С (Цвет глаз и рост человека).

Обоснование:

Положительная корреляция — когда увеличение одной переменной сопровождается увеличением другой (например, рост и вес).

Отрицательная корреляция — когда увеличение одной переменной сопровождается уменьшением другой (например, температура и количество людей на пляже в зимний период).

Нулевая корреляция — когда между переменными нет никакой связи (например, цвет глаз и рост).

Вопрос 4 (вопрос с открытым ответом):

Что такое коэффициент корреляции Пирсона?

Ответ: Коэффициент корреляции Пирсона — это мера линейной зависимости между двумя переменными. Он принимает значения от -1 до 1, где -1 указывает на сильную отрицательную корреляцию, 1 — на сильную положительную корреляцию, а 0 — на отсутствие корреляции.

Обоснование: Коэффициент корреляции Пирсона используется для оценки силы и направления линейной связи между двумя переменными. Он помогает определить, насколько сильно изменения одной переменной связаны с изменениями другой.

Вопрос 5 (вопрос с кратким ответом):

Как называется мера разброса данных, которая показывает среднее квадратическое отклонение от среднего?

Ответ: Дисперсия

Обоснование: Дисперсия — это мера разброса данных, которая показывает среднее квадратическое отклонение от среднего. Она используется для оценки вариации данных вокруг среднего значения.

Тема 4. Сравнение средних величин критерием Стьюдента

Вопрос 1 (выбор одного правильного ответа):

Что такое критерий Стьюдента?

А) Метод для сравнения средних величин двух выборок.

В) Метод для сравнения дисперсий двух выборок.

С) Метод для сравнения медиан двух выборок.

D) Метод для сравнения частот двух выборок.

Ответ: А

Обоснование: Критерий Стьюдента (t-критерий) используется для сравнения средних величин двух выборок. Он позволяет определить, есть ли статистически значимая разница между средними значениями двух групп.

Вопрос 2 (выбор нескольких правильных ответов):

Какие предположения должны выполняться для применения критерия Стьюдента?

A) Нормальное распределение данных.

B) Равенство дисперсий в двух группах.

C) Независимость наблюдений.

D) Одинаковый размер выборок.

Ответ: A, B, C

Обоснование: Для применения критерия Стьюдента должны выполняться следующие предположения:

Нормальное распределение данных — данные в обеих группах должны быть распределены нормально.

Равенство дисперсий — дисперсии в двух группах должны быть примерно равны.

Независимость наблюдений — наблюдения в каждой группе должны быть независимыми друг от друга.

Размер выборок не обязательно должен быть одинаковым.

Вопрос 3 (установление соответствия):

Установите соответствие между типами критерия Стьюдента и примерами:

Критерий Стьюдента для независимых выборок

Критерий Стьюдента для зависимых выборок

Примеры:

A) Сравнение среднего дохода мужчин и женщин.

B) Сравнение результатов теста до и после обучения.

Ответ:

Критерий Стьюдента для независимых выборок — А (Сравнение среднего дохода мужчин и женщин).

Критерий Стьюдента для зависимых выборок — В (Сравнение результатов теста до и после обучения).

Обоснование:

Критерий Стьюдента для независимых выборок используется, когда две группы не связаны друг с другом (например, мужчины и женщины).

Критерий Стьюдента для зависимых выборок используется, когда две группы связаны (например, один и тот же человек до и после обучения).

Вопрос 4 (вопрос с открытым ответом):

Что такое р-значение в контексте критерия Стьюдента?

Ответ: Р-значение — это вероятность получить наблюдаемый результат (или более экстремальный) при условии, что нулевая гипотеза верна. Если р-значение меньше

выбранного уровня значимости (например, 0.05), то нулевая гипотеза отвергается, и считается, что разница между средними значениями статистически значима.

Обоснование: Р-значение помогает определить, является ли разница между средними значениями двух групп случайной или статистически значимой. Чем меньше р-значение, тем сильнее доказательства против нулевой гипотезы.

Вопрос 5 (вопрос с кратким ответом):

Как называется нулевая гипотеза в критерии Стьюдента?

Ответ: Нулевая гипотеза в критерии Стьюдента утверждает, что нет статистически значимой разницы между средними значениями двух групп. Обычно она обозначается как $H_0: \mu_1 = \mu_2$, где μ_1 и μ_2 — средние значения двух групп.

Обоснование: Нулевая гипотеза (H_0) в критерии Стьюдента предполагает, что средние значения двух групп равны. Альтернативная гипотеза (H_1) утверждает, что средние значения различны.

Тема 5. Непараметрическая статистика. Сравнение величин методами непараметрической статистики

Вопрос 1 (выбор одного правильного ответа):

Что такое непараметрическая статистика?

- A) Методы, которые не требуют предположений о распределении данных.
- B) Методы, которые требуют предположений о нормальном распределении данных.
- C) Методы, которые используются только для больших выборок.
- D) Методы, которые применяются только для категориальных данных.

Ответ: A

Обоснование: Непараметрическая статистика — это методы, которые не требуют предположений о распределении данных. Они особенно полезны, когда данные не соответствуют нормальному распределению или когда размер выборки мал.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к непараметрическим методам?

- A) U-критерий Манна-Уитни.
- B) Критерий Вилкоксона.
- C) Критерий хи-квадрат.
- D) F-критерий.

Ответ: A, B, C

Обоснование: К непараметрическим методам относятся:

U-критерий Манна-Уитни — для сравнения двух независимых выборок.

Критерий Вилкоксона — для сравнения двух зависимых выборок.

Критерий хи-квадрат — для анализа категориальных данных.

F-критерий является параметрическим методом и используется для сравнения дисперсий.

Вопрос 3 (установление соответствия):

Установите соответствие между типами непараметрических тестов и примерами:

U-критерий Манна-Уитни

Критерий Вилкоксона

Критерий хи-квадрат

Примеры:

А) Сравнение среднего дохода мужчин и женщин.

В) Сравнение результатов теста до и после обучения.

С) Анализ связи между полом и предпочтениями в еде.

Ответ:

U-критерий Манна-Уитни — А (Сравнение среднего дохода мужчин и женщин).

Критерий Вилкоксона — В (Сравнение результатов теста до и после обучения).

Критерий хи-квадрат — С (Анализ связи между полом и предпочтениями в еде).

Обоснование:

U-критерий Манна-Уитни используется для сравнения двух независимых выборок (например, мужчины и женщины).

Критерий Вилкоксона используется для сравнения двух зависимых выборок (например, один и тот же человек до и после обучения).

Критерий хи-квадрат используется для анализа связи между двумя категориальными переменными (например, пол и предпочтения в еде).

Вопрос 4 (вопрос с открытым ответом):

Что такое ранг в контексте непараметрической статистики?

Ответ: Ранг в контексте непараметрической статистики — это порядковый номер, который присваивается каждому наблюдению в упорядоченном наборе данных. Ранговые методы используются, когда данные не соответствуют нормальному распределению или когда важно учитывать только порядок значений, а не их точные величины.

Обоснование: Ранговые методы, такие как U-критерий Манна-Уитни и критерий Вилкоксона, основаны на ранжировании данных. Это позволяет сравнивать выборки без предположений о конкретном распределении данных.

Вопрос 5 (вопрос с кратким ответом):

Как называется нулевая гипотеза в U-критерии Манна-Уитни?

Ответ: Нулевая гипотеза в U-критерии Манна-Уитни утверждает, что две выборки происходят из одного и того же распределения. Это означает, что нет статистически значимой разницы между двумя группами.

Обоснование: Нулевая гипотеза (H_0) в U-критерии Манна-Уитни предполагает, что две выборки не различаются. Альтернативная гипотеза (H_1) утверждает, что одна выборка имеет тенденцию к более высоким значениям, чем другая.

Тема 6. Группировка и однофакторная ANOVA

Вопрос 1 (выбор одного правильного ответа):

Что такое однофакторный дисперсионный анализ (ANOVA)?

А) Метод для сравнения средних величин двух выборок.

- В) Метод для сравнения средних величин более чем двух выборок.
- С) Метод для сравнения дисперсий двух выборок.
- Д) Метод для сравнения медиан двух выборок.

Ответ: В

Обоснование: Однофакторный дисперсионный анализ (ANOVA) используется для сравнения средних величин более чем двух выборок. Он позволяет определить, есть ли статистически значимая разница между средними значениями нескольких групп.

Вопрос 2 (выбор нескольких правильных ответов):

Какие предположения должны выполняться для применения однофакторного ANOVA?

- А) Нормальное распределение данных в каждой группе.
- В) Равенство дисперсий в группах.
- С) Независимость наблюдений.
- Д) Одинаковый размер выборок.

Ответ: А, В, С

Обоснование: Для применения однофакторного ANOVA должны выполняться следующие предположения:

Нормальное распределение данных — данные в каждой группе должны быть распределены нормально.

Равенство дисперсий — дисперсии в группах должны быть примерно равны.

Независимость наблюдений — наблюдения в каждой группе должны быть независимыми друг от друга.

Размер выборок не обязательно должен быть одинаковым.

Вопрос 3 (установление соответствия):

Установите соответствие между типами ANOVA и примерами:

Однофакторный ANOVA

Двухфакторный ANOVA

Примеры:

А) Сравнение среднего дохода в трех разных регионах.

В) Сравнение среднего дохода в трех разных регионах и двух разных профессиях.

Ответ:

Однофакторный ANOVA — А (Сравнение среднего дохода в трех разных регионах).

Двухфакторный ANOVA — В (Сравнение среднего дохода в трех разных регионах и двух разных профессиях).

Обоснование:

Однофакторный ANOVA используется, когда есть один фактор (например, регион).

Двухфакторный ANOVA используется, когда есть два фактора (например, регион и профессия).

Вопрос 4 (вопрос с открытым ответом):

Что такое нулевая гипотеза в однофакторном ANOVA?

Ответ: Нулевая гипотеза в однофакторном ANOVA утверждает, что нет

статистически значимой разницы между средними значениями всех групп. Это означает, что все группы происходят из одного и того же распределения.

Обоснование: Нулевая гипотеза (H_0) в однофакторном ANOVA предполагает, что средние значения всех групп равны. Альтернативная гипотеза (H_1) утверждает, что хотя бы одна группа имеет среднее значение, отличное от других.

Вопрос 5 (вопрос с кратким ответом):

Как называется статистика, используемая в однофакторном ANOVA?

Ответ: Статистика, используемая в однофакторном ANOVA, называется F-статистикой. Она вычисляется как отношение дисперсии между группами к дисперсии внутри групп.

Обоснование: F-статистика в однофакторном ANOVA используется для оценки значимости различий между средними значениями групп. Если F-статистика велика, это указывает на то, что различия между группами статистически значимы.

Тема 7. Дисперсионный анализ

Вопрос 1 (выбор одного правильного ответа):

Что такое дисперсионный анализ (ANOVA)?

- A) Метод для сравнения средних величин двух выборок.
- B) Метод для сравнения средних величин более чем двух выборок.
- C) Метод для сравнения дисперсий двух выборок.
- D) Метод для сравнения медиан двух выборок.

Ответ: B

Обоснование: Дисперсионный анализ (ANOVA) используется для сравнения средних величин более чем двух выборок. Он позволяет определить, есть ли статистически значимая разница между средними значениями нескольких групп.

Вопрос 2 (выбор нескольких правильных ответов):

Какие предположения должны выполняться для применения дисперсионного анализа?

- A) Нормальное распределение данных в каждой группе.
- B) Равенство дисперсий в группах.
- C) Независимость наблюдений.
- D) Одинаковый размер выборок.

Ответ: A, B, C

Обоснование: Для применения дисперсионного анализа должны выполняться следующие предположения:

Нормальное распределение данных — данные в каждой группе должны быть распределены нормально.

Равенство дисперсий — дисперсии в группах должны быть примерно равны.

Независимость наблюдений — наблюдения в каждой группе должны быть независимыми друг от друга.

Размер выборок не обязательно должен быть одинаковым.

Вопрос 3 (установление соответствия):

Установите соответствие между типами дисперсионного анализа и примерами:

Однофакторный ANOVA

Двухфакторный ANOVA

Примеры:

А) Сравнение среднего дохода в трех разных регионах.

В) Сравнение среднего дохода в трех разных регионах и двух разных профессиях.

Ответ:

Однофакторный ANOVA — А (Сравнение среднего дохода в трех разных регионах).

Двухфакторный ANOVA — В (Сравнение среднего дохода в трех разных регионах и двух разных профессиях).

Обоснование:

Однофакторный ANOVA используется, когда есть один фактор (например, регион).

Двухфакторный ANOVA используется, когда есть два фактора (например, регион и профессия).

Вопрос 4 (вопрос с открытым ответом):

Что такое нулевая гипотеза в дисперсионном анализе?

Ответ: Нулевая гипотеза в дисперсионном анализе утверждает, что нет статистически значимой разницы между средними значениями всех групп. Это означает, что все группы происходят из одного и того же распределения.

Обоснование: Нулевая гипотеза (H_0) в дисперсионном анализе предполагает, что средние значения всех групп равны. Альтернативная гипотеза (H_1) утверждает, что хотя бы одна группа имеет среднее значение, отличное от других.

Вопрос 5 (вопрос с кратким ответом):

Как называется статистика, используемая в дисперсионном анализе?

Ответ: Статистика, используемая в дисперсионном анализе, называется F-статистикой. Она вычисляется как отношение дисперсии между группами к дисперсии внутри групп.

Обоснование: F-статистика в дисперсионном анализе используется для оценки значимости различий между средними значениями групп. Если F-статистика велика, это указывает на то, что различия между группами статистически значимы.

Тема 8. Таблицы частот, сопряженности, флагов и заголовков

Вопрос 1 (выбор одного правильного ответа):

Что такое таблица сопряженности?

А) Таблица, которая показывает распределение одной переменной.

В) Таблица, которая показывает взаимосвязь между двумя категориальными переменными.

С) Таблица, которая показывает взаимосвязь между двумя непрерывными переменными.

Д) Таблица, которая показывает взаимосвязь между тремя и более переменными.

Ответ: В

Обоснование: Таблица сопряженности — это таблица, которая показывает взаимосвязь между двумя категориальными переменными. Она используется для анализа связи между двумя переменными, представленными в виде категорий.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих величин можно использовать для анализа таблиц сопряженности?

- A) Коэффициент корреляции Пирсона.
- B) Критерий хи-квадрат.
- C) Коэффициент сопряженности.
- D) Коэффициент детерминации.

Ответ: B, C

Обоснование: Для анализа таблиц сопряженности можно использовать:

Критерий хи-квадрат — для проверки независимости двух категориальных переменных.

Коэффициент сопряженности — для измерения силы связи между двумя категориальными переменными.

Коэффициент корреляции Пирсона и коэффициент детерминации используются для анализа связи между непрерывными переменными.

Вопрос 3 (установление соответствия):

Установите соответствие между типами таблиц и примерами:

Таблица частот

Таблица сопряженности

Таблица флагов

Примеры:

- A) Таблица, показывающая количество студентов по факультетам.
- B) Таблица, показывающая взаимосвязь между полом и предпочтениями в еде.
- C) Таблица, показывающая наличие или отсутствие определенного признака у каждого объекта.

Ответ:

Таблица частот — A (Таблица, показывающая количество студентов по факультетам).

Таблица сопряженности — B (Таблица, показывающая взаимосвязь между полом и предпочтениями в еде).

Таблица флагов — C (Таблица, показывающая наличие или отсутствие определенного признака у каждого объекта).

Обоснование:

Таблица частот используется для представления распределения одной переменной (например, количество студентов по факультетам).

Таблица сопряженности используется для анализа взаимосвязи между двумя категориальными переменными (например, пол и предпочтения в еде).

Таблица флагов используется для представления наличия или отсутствия определенного признака у каждого объекта (например, наличие или отсутствие заболевания).

Вопрос 4 (вопрос с открытым ответом):

Что такое таблица флагов?

Ответ: Таблица флагов — это таблица, которая показывает наличие или отсутствие определенного признака у каждого объекта. Обычно в такой таблице используются бинарные значения (например, 0 или 1), где 0 означает отсутствие признака, а 1 — его наличие.

Обоснование: Таблицы флагов используются для представления категориальных данных, где интерес представляет наличие или отсутствие определенного признака. Это может быть полезно для анализа данных, где важно определить, присутствует ли определенный атрибут у объекта.

Вопрос 5 (вопрос с кратким ответом):

Как называется статистика, используемая для проверки независимости в таблицах сопряженности?

Ответ: Статистика, используемая для проверки независимости в таблицах сопряженности, называется критерием хи-квадрат (χ^2). Она позволяет определить, есть ли статистически значимая связь между двумя категориальными переменными.

Обоснование: Критерий хи-квадрат используется для проверки гипотезы о независимости двух категориальных переменных. Если значение критерия хи-квадрат велико, это указывает на то, что переменные связаны.

Тема 9. Канонический анализ

Вопрос 1 (выбор одного правильного ответа):

Что такое канонический анализ?

- A) Метод для выявления взаимосвязей между двумя множествами переменных.
- B) Метод для классификации данных на основе их сходства.
- C) Метод для уменьшения размерности данных.
- D) Метод для прогнозирования одной переменной на основе другой.

Ответ: A

Обоснование: Канонический анализ — это метод, который используется для выявления взаимосвязей между двумя множествами переменных. Он позволяет найти линейные комбинации переменных, которые максимизируют корреляцию между двумя наборами данных.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к каноническому анализу?

- A) Канонический корреляционный анализ.
- B) Канонический дискриминантный анализ.
- C) Факторный анализ.
- D) Кластерный анализ.

Ответ: A, B

Обоснование: К каноническому анализу относятся:

Канонический корреляционный анализ — для выявления взаимосвязей между двумя множествами переменных.

Канонический дискриминантный анализ — для классификации и разделения групп на

основе канонических переменных.

Факторный и кластерный анализ относятся к другим методам многомерного анализа.

Вопрос 3 (установление соответствия):

Установите соответствие между типами канонического анализа и примерами:

Канонический корреляционный анализ

Канонический дискриминантный анализ

Примеры:

А) Анализ взаимосвязи между набором экономических показателей и набором социальных показателей.

В) Классификация пациентов на основе набора клинических и лабораторных показателей.

Ответ:

Канонический корреляционный анализ — А (Анализ взаимосвязи между набором экономических показателей и набором социальных показателей).

Канонический дискриминантный анализ — В (Классификация пациентов на основе набора клинических и лабораторных показателей).

Обоснование:

Канонический корреляционный анализ используется для анализа взаимосвязи между двумя множествами переменных (например, экономические и социальные показатели).

Канонический дискриминантный анализ используется для классификации объектов на основе нескольких переменных (например, классификация пациентов по клиническим и лабораторным показателям).

Вопрос 4 (вопрос с открытым ответом):

Что такое канонические переменные в каноническом анализе?

Ответ: Канонические переменные — это линейные комбинации исходных переменных, которые максимизируют корреляцию между двумя множествами переменных. Они используются для упрощения анализа и интерпретации взаимосвязей между множествами переменных.

Обоснование: Канонические переменные создаются таким образом, чтобы они были максимально коррелированы между собой и минимально коррелированы внутри каждого множества. Это позволяет выявить основные взаимосвязи между двумя множествами переменных.

Вопрос 5 (вопрос с кратким ответом):

Как называется статистика, используемая для оценки значимости канонических корреляций?

Ответ: Статистика, используемая для оценки значимости канонических корреляций, называется критерием Уилкса (Wilks' Lambda). Она позволяет определить, являются ли канонические корреляции статистически значимыми.

Обоснование: Критерий Уилкса используется для проверки гипотезы о том, что канонические корреляции равны нулю. Если значение критерия Уилкса значительно, это указывает на наличие статистически значимых взаимосвязей между множествами переменных.

Тема 10. Линейное и нелинейное моделирование взаимосвязей

Вопрос 1 (выбор одного правильного ответа):

Что такое линейная регрессия?

- A) Метод для моделирования нелинейных взаимосвязей между переменными.
- B) Метод для моделирования линейных взаимосвязей между переменными.
- C) Метод для классификации данных на основе их сходства.
- D) Метод для уменьшения размерности данных.

Ответ: B

Обоснование: Линейная регрессия — это метод, который используется для моделирования линейных взаимосвязей между переменными. Она позволяет предсказать значение зависимой переменной на основе одной или нескольких независимых переменных.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к нелинейному моделированию?

- A) Полиномиальная регрессия.
- B) Логистическая регрессия.
- C) Линейная регрессия.
- D) Деревья решений.

Ответ: A, B, D

Обоснование: К нелинейному моделированию относятся:

Полиномиальная регрессия — используется для моделирования нелинейных взаимосвязей с помощью полиномов.

Логистическая регрессия — используется для моделирования нелинейных взаимосвязей в задачах классификации.

Деревья решений — используются для моделирования нелинейных взаимосвязей и классификации данных.

Линейная регрессия относится к линейному моделированию.

Вопрос 3 (установление соответствия):

Установите соответствие между типами моделей и примерами:

Линейная регрессия

Нелинейная регрессия

Примеры:

- A) Прогнозирование цены на недвижимость на основе площади и количества комнат.
- B) Прогнозирование вероятности заболевания на основе возраста и индекса массы тела.
- C) Прогнозирование роста растения на основе времени и количества удобрений.

Ответ:

Линейная регрессия — A (Прогнозирование цены на недвижимость на основе площади и количества комнат).

Нелинейная регрессия — B (Прогнозирование вероятности заболевания на основе возраста и индекса массы тела), C (Прогнозирование роста растения на основе времени и количества удобрений).

Обоснование:

Линейная регрессия используется для прогнозирования цены на недвижимость, так как взаимосвязь между ценой, площадью и количеством комнат может быть аппроксимирована линейной моделью.

Нелинейная регрессия используется для прогнозирования вероятности заболевания и роста растения, так как взаимосвязь между переменными может быть нелинейной.

Вопрос 4 (вопрос с открытым ответом):

Что такое полиномиальная регрессия?

Ответ: Полиномиальная регрессия — это метод нелинейного моделирования, который используется для аппроксимации взаимосвязей между переменными с помощью полиномов. Она позволяет моделировать нелинейные взаимосвязи, добавляя в модель степени независимых переменных.

Обоснование: Полиномиальная регрессия расширяет линейную регрессию, добавляя члены с более высокими степенями независимых переменных. Это позволяет моделировать более сложные взаимосвязи между переменными.

Вопрос 5 (вопрос с кратким ответом):

Как называется метод, который используется для оценки значимости коэффициентов в линейной регрессии?

Ответ: Метод, который используется для оценки значимости коэффициентов в линейной регрессии, называется t-тестом. Он позволяет определить, являются ли коэффициенты статистически значимыми, то есть отличаются ли они от нуля.

Обоснование: T-тест используется для проверки гипотезы о том, что коэффициент регрессии равен нулю. Если значение t-статистики значительно, это указывает на то, что коэффициент статистически значим и не равен нулю.

Тема 11. Дискриминантный анализ.

Вопрос 1 (выбор одного правильного ответа):

Что такое дискриминантный анализ?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: D

Обоснование: Дискриминантный анализ — это метод, который используется для разделения данных на группы на основе линейных комбинаций переменных. Он позволяет классифицировать объекты в одну из нескольких групп.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к дискриминантному анализу?

- A) Линейный дискриминантный анализ (LDA).
- B) Квадратичный дискриминантный анализ (QDA).
- C) Канонический дискриминантный анализ.

D) Кластерный анализ.

Ответ: A, B, C

Обоснование: К дискриминантному анализу относятся:

Линейный дискриминантный анализ (LDA) — используется для классификации данных с предположением о нормальном распределении и равных дисперсиях.

Квадратичный дискриминантный анализ (QDA) — используется для классификации данных без предположения о равных дисперсиях.

Канонический дискриминантный анализ — используется для выявления линейных комбинаций переменных, которые максимизируют различия между группами.

Кластерный анализ относится к другим методам многомерного анализа.

Вопрос 3 (установление соответствия):

Установите соответствие между типами дискриминантного анализа и примерами:

Линейный дискриминантный анализ (LDA)

Квадратичный дискриминантный анализ (QDA)

Примеры:

A) Классификация пациентов на основе набора клинических и лабораторных показателей.

B) Классификация изображений на основе набора пикселей.

Ответ:

Линейный дискриминантный анализ (LDA) — A (Классификация пациентов на основе набора клинических и лабораторных показателей).

Квадратичный дискриминантный анализ (QDA) — B (Классификация изображений на основе набора пикселей).

Обоснование:

Линейный дискриминантный анализ (LDA) часто используется в медицине для классификации пациентов на основе клинических и лабораторных показателей.

Квадратичный дискриминантный анализ (QDA) может быть полезен для классификации изображений, где дисперсии могут различаться.

Вопрос 4 (вопрос с открытым ответом):

Что такое дискриминантная функция в дискриминантном анализе?

Ответ: Дискриминантная функция — это линейная комбинация переменных, которая используется для разделения данных на группы. Она максимизирует различия между группами и минимизирует различия внутри групп, позволяя классифицировать новые объекты в одну из групп.

Обоснование: Дискриминантная функция строится таким образом, чтобы она наилучшим образом разделяла данные на группы. Это достигается путем максимизации межгрупповой дисперсии и минимизации внутригрупповой дисперсии.

Вопрос 5 (вопрос с кратким ответом):

Как называется статистика, используемая для оценки значимости дискриминантной функции?

Ответ: Статистика, используемая для оценки значимости дискриминантной функции,

называется F-статистикой. Она вычисляется как отношение дисперсии между группами к дисперсии внутри групп и используется для проверки гипотезы о том, что дискриминантная функция значимо различает группы.

Обоснование: F-статистика в дискриминантном анализе используется для проверки гипотезы о том, что различия между группами значимы. Если значение F-статистики велико, это указывает на то, что дискриминантная функция статистически значима.

Тема 12. Кластерный анализ

Тестовые задания на Тему 12. Кластерный анализ

Вопрос 1 (выбор одного правильного ответа):

Что такое кластерный анализ?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: A

Обоснование: Кластерный анализ — это метод, который используется для классификации данных на основе их сходства. Он позволяет группировать объекты в кластеры, где объекты внутри одного кластера более похожи друг на друга, чем на объекты из других кластеров.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к кластерному анализу?

- A) Иерархическая кластеризация.
- B) Метод k-средних.
- C) Линейный дискриминантный анализ.
- D) Квадратичный дискриминантный анализ.

Ответ: A, B

Обоснование: К кластерному анализу относятся:

Иерархическая кластеризация — метод, который создает иерархию кластеров, начиная с отдельных объектов и объединяя их в более крупные кластеры.

Метод k-средних — метод, который разделяет данные на k кластеров, минимизируя сумму квадратов расстояний от объектов до центров кластеров.

Линейный и квадратичный дискриминантный анализ относятся к другим методам многомерного анализа.

Вопрос 3 (установление соответствия):

Установите соответствие между типами кластерного анализа и примерами:

Иерархическая кластеризация

Метод k-средних

Примеры:

- A) Построение дендрограммы для анализа сходства видов растений.
- B) Разделение клиентов на сегменты на основе их покупательского поведения.

Ответ:

Иерархическая кластеризация — А (Построение дендрограммы для анализа сходства видов растений).

Метод k-средних — В (Разделение клиентов на сегменты на основе их покупательского поведения).

Обоснование:

Иерархическая кластеризация часто используется для построения дендрограмм, которые показывают иерархию сходства между объектами (например, виды растений).

Метод k-средних часто применяется в маркетинге для сегментации клиентов на основе их покупательского поведения.

Вопрос 4 (вопрос с открытым ответом):

Что такое кластер в кластерном анализе?

Ответ: Кластер в кластерном анализе — это группа объектов, которые более похожи друг на друга, чем на объекты из других кластеров. Кластеры формируются на основе меры сходства или расстояния между объектами.

Обоснование: Кластеры создаются таким образом, чтобы объекты внутри одного кластера были максимально похожи друг на друга, а объекты из разных кластеров — максимально различны. Это позволяет выявить структуры и закономерности в данных.

Вопрос 5 (вопрос с кратким ответом):

Как называется метод, который используется для определения оптимального числа кластеров в методе k-средних?

Ответ: Метод, который используется для определения оптимального числа кластеров в методе k-средних, называется методом локтя (elbow method). Он основан на анализе графика суммы квадратов расстояний от объектов до центров кластеров и выборе точки, где график резко меняет свой наклон.

Обоснование: Метод локтя помогает определить оптимальное число кластеров, выбирая точку, где дальнейшее увеличение числа кластеров не приводит к значительному улучшению качества кластеризации.

Тема 13. Деревья классификации

Тестовые задания на Тему 13. Деревья классификации

Вопрос 1 (выбор одного правильного ответа):

Что такое дерево классификации?

А) Метод для классификации данных на основе их сходства.

В) Метод для уменьшения размерности данных.

С) Метод для прогнозирования одной переменной на основе другой.

Д) Метод для разделения данных на группы на основе иерархической структуры.

Ответ: D

Обоснование: Дерево классификации — это метод, который используется для разделения данных на группы на основе иерархической структуры. Он позволяет классифицировать объекты в одну из нескольких категорий, используя набор правил, основанных на значениях переменных.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к деревьям классификации?

- A) ID3.
- B) C4.5.
- C) CART.
- D) Линейный дискриминантный анализ.

Ответ: A, B, C

Обоснование: К деревьям классификации относятся:

ID3 — алгоритм, который использует энтропию для выбора признаков.

C4.5 — улучшенная версия ID3, которая учитывает пропущенные значения и непрерывные переменные.

CART — метод, который использует критерий Gini для выбора признаков и может строить как классификационные, так и регрессионные деревья.

Линейный дискриминантный анализ относится к другим методам многомерного анализа.

Вопрос 3 (установление соответствия):

Установите соответствие между типами деревьев классификации и примерами:

ID3

C4.5

CART

Примеры:

- A) Дерево, построенное на основе энтропии.
- B) Дерево, которое учитывает пропущенные значения.
- C) Дерево, которое использует критерий Gini.

Ответ:

ID3 — A (Дерево, построенное на основе энтропии).

C4.5 — B (Дерево, которое учитывает пропущенные значения).

CART — C (Дерево, которое использует критерий Gini).

Обоснование:

ID3 использует энтропию для выбора признаков.

C4.5 является улучшенной версией ID3 и учитывает пропущенные значения.

CART использует критерий Gini для выбора признаков.

Вопрос 4 (вопрос с открытым ответом):

Что такое критерий разделения узлов в деревьях классификации?

Ответ: Критерий разделения узлов в деревьях классификации — это правило, которое используется для выбора признака и порога, по которому будет производиться разделение данных на два или более подмножества. Критерии разделения узлов выбирают признак и порог таким образом, чтобы максимизировать чистоту (однородность) полученных подмножеств.

Обоснование: Критерии разделения узлов, такие как энтропия, информационный выигрыш и Gini-индекс, помогают выбрать признак и порог, которые наилучшим образом разделяют данные на подмножества, где объекты внутри одного подмножества более похожи

друг на друга, чем на объекты из других подмножеств.

Вопрос 5 (вопрос с кратким ответом):

Как называется метод, который используется для предотвращения переобучения в деревьях классификации?

Ответ: Метод, который используется для предотвращения переобучения в деревьях классификации, называется обрезкой (pruning). Обрезка включает удаление ветвей дерева, которые не вносят значительного вклада в точность классификации, чтобы упростить модель и повысить ее обобщаемость.

Обоснование: Обрезка помогает избежать переобучения, когда дерево становится слишком сложным и начинает учитывать шум и случайные колебания в данных. Удаление лишних ветвей позволяет улучшить способность модели обобщать на новые данные.

Тема 14. Факторный анализ

Вопрос 1 (выбор одного правильного ответа):

Что такое факторный анализ?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных путем выявления латентных факторов.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: B

Обоснование: Факторный анализ — это метод, который используется для уменьшения размерности данных путем выявления латентных факторов. Он позволяет выявить скрытые структуры в данных и представить их в более компактной форме.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к факторному анализу?

- A) Метод главных компонент (PCA).
- B) Канонический корреляционный анализ.
- C) Линейный дискриминантный анализ.
- D) Факторный анализ методом максимального правдоподобия.

Ответ: A, D

Обоснование: К факторному анализу относятся:

Метод главных компонент (PCA) — метод, который используется для уменьшения размерности данных путем выделения главных компонент.

Факторный анализ методом максимального правдоподобия — метод, который используется для оценки параметров факторной модели на основе максимального правдоподобия.

Канонический корреляционный анализ и линейный дискриминантный анализ относятся к другим методам многомерного анализа.

Вопрос 3 (установление соответствия):

Установите соответствие между типами факторного анализа и примерами:

Метод главных компонент (РСА)

Факторный анализ методом максимального правдоподобия

Примеры:

А) Анализ взаимосвязи между набором экономических показателей и набором социальных показателей.

В) Уменьшение размерности набора данных, описывающего характеристики пациентов.

Ответ:

Метод главных компонент (РСА) — В (Уменьшение размерности набора данных, описывающего характеристики пациентов).

Факторный анализ методом максимального правдоподобия — А (Анализ взаимосвязи между набором экономических показателей и набором социальных показателей).

Обоснование:

Метод главных компонент (РСА) используется для уменьшения размерности данных, например, при анализе характеристик пациентов.

Факторный анализ методом максимального правдоподобия может быть использован для анализа взаимосвязи между экономическими и социальными показателями.

Вопрос 4 (вопрос с открытым ответом):

Что такое факторная нагрузка в факторном анализе?

Ответ: Факторная нагрузка — это коэффициент, который показывает силу связи между исходной переменной и латентным фактором. Она указывает на то, насколько сильно каждая переменная связана с каждым фактором.

Обоснование: Факторные нагрузки помогают интерпретировать факторы, показывая, какие переменные наиболее сильно связаны с каждым фактором. Это позволяет понять, какие переменные вносят наибольший вклад в каждый фактор.

Вопрос 5 (вопрос с кратким ответом):

Как называется метод, который используется для оценки значимости факторов в факторном анализе?

Ответ: Метод, который используется для оценки значимости факторов в факторном анализе, называется тестом Бартлетта. Он позволяет определить, являются ли факторы статистически значимыми и вносят ли они существенный вклад в объяснение вариации данных.

Обоснование: Тест Бартлетта используется для проверки гипотезы о том, что все факторы незначимы. Если тест показывает, что факторы значимы, это означает, что они вносят существенный вклад в объяснение вариации данных.

Тема 15. Многомерное шкалирование

Вопрос 1 (выбор одного правильного ответа):

Что такое многомерное шкалирование?

А) Метод для классификации данных на основе их сходства.

В) Метод для уменьшения размерности данных путем визуализации их в пространстве

меньшей размерности.

С) Метод для прогнозирования одной переменной на основе другой.

Д) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: В

Обоснование: Многомерное шкалирование — это метод, который используется для уменьшения размерности данных путем визуализации их в пространстве меньшей размерности. Он позволяет представить данные в виде точек в двумерном или трехмерном пространстве, сохраняя при этом расстояния между объектами.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к многомерному шкалированию?

А) Метрическое многомерное шкалирование.

В) Неметрическое многомерное шкалирование.

С) Линейный дискриминантный анализ.

Д) Факторный анализ.

Ответ: А, В

Обоснование: К многомерному шкалированию относятся:

Метрическое многомерное шкалирование — метод, который сохраняет точные расстояния между объектами.

Неметрическое многомерное шкалирование — метод, который сохраняет только порядок расстояний между объектами.

Линейный дискриминантный анализ и факторный анализ относятся к другим методам многомерного анализа.

Вопрос 3 (установление соответствия):

Установите соответствие между типами многомерного шкалирования и примерами:

Метрическое многомерное шкалирование

Неметрическое многомерное шкалирование

Примеры:

А) Визуализация данных, где расстояния между точками точно соответствуют исходным расстояниям.

В) Визуализация данных, где сохраняется только порядок расстояний между точками.

Ответ:

Метрическое многомерное шкалирование — А (Визуализация данных, где расстояния между точками точно соответствуют исходным расстояниям).

Неметрическое многомерное шкалирование — В (Визуализация данных, где сохраняется только порядок расстояний между точками).

Обоснование:

Метрическое многомерное шкалирование сохраняет точные расстояния между объектами.

Неметрическое многомерное шкалирование сохраняет только порядок расстояний между объектами.

Вопрос 4 (вопрос с открытым ответом):

Что такое стресс в контексте многомерного шкалирования?

Ответ: Стресс в контексте многомерного шкалирования — это мера, которая показывает, насколько хорошо визуализация в пространстве меньшей размерности сохраняет исходные расстояния между объектами. Чем меньше значение стресса, тем лучше визуализация отражает исходные расстояния.

Обоснование: Стресс используется для оценки качества многомерного шкалирования. Он вычисляется как разность между исходными и восстановленными расстояниями, и его минимизация является целью многомерного шкалирования.

Вопрос 5 (вопрос с кратким ответом):

Как называется метод, который используется для визуализации данных в многомерном шкалировании?

Ответ: Метод, который используется для визуализации данных в многомерном шкалировании, называется многомерным шкалированием (MDS). Он позволяет представить данные в виде точек в двумерном или трехмерном пространстве, сохраняя при этом расстояния между объектами.

Обоснование: Многомерное шкалирование (MDS) — это основной метод, используемый для визуализации данных в пространстве меньшей размерности, сохраняя при этом расстояния между объектами.

Тема 16. Обзор пройденного материала

Вопрос 1 (выбор одного правильного ответа):

Что такое многомерный статистический анализ?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных путем выявления латентных факторов.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для анализа взаимосвязей между несколькими переменными.

Ответ: D

Обоснование: Многомерный статистический анализ — это метод, который используется для анализа взаимосвязей между несколькими переменными. Он позволяет исследовать сложные структуры данных и выявлять скрытые закономерности.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к многомерному статистическому анализу?

- A) Линейная регрессия.
- B) Факторный анализ.
- C) Кластерный анализ.
- D) Дискриминантный анализ.

Ответ: B, C, D

Обоснование: К многомерному статистическому анализу относятся:

Факторный анализ — метод для уменьшения размерности данных путем выявления

латентных факторов.

Кластерный анализ — метод для классификации данных на основе их сходства.

Дискриминантный анализ — метод для разделения данных на группы на основе линейных комбинаций переменных.

Линейная регрессия относится к другим методам анализа данных.

Вопрос 3 (установление соответствия):

Установите соответствие между типами многомерного анализа и примерами:

Факторный анализ

Кластерный анализ

Дискриминантный анализ

Примеры:

А) Уменьшение размерности набора данных, описывающего характеристики пациентов.

В) Классификация пациентов на основе набора клинических и лабораторных показателей.

С) Анализ взаимосвязи между набором экономических показателей и набором социальных показателей.

Ответ:

Факторный анализ — А (Уменьшение размерности набора данных, описывающего характеристики пациентов).

Кластерный анализ — В (Классификация пациентов на основе набора клинических и лабораторных показателей).

Дискриминантный анализ — С (Анализ взаимосвязи между набором экономических показателей и набором социальных показателей).

Обоснование:

Факторный анализ используется для уменьшения размерности данных.

Кластерный анализ используется для классификации данных.

Дискриминантный анализ используется для анализа взаимосвязи между переменными.

Вопрос 4 (вопрос с открытым ответом):

Что такое многомерное шкалирование?

Ответ: Многомерное шкалирование — это метод, который используется для уменьшения размерности данных путем визуализации их в пространстве меньшей размерности. Он позволяет представить данные в виде точек в двумерном или трехмерном пространстве, сохраняя при этом расстояния между объектами.

Обоснование: Многомерное шкалирование помогает визуализировать данные и выявить скрытые структуры, сохраняя при этом расстояния между объектами.

Вопрос 5 (вопрос с кратким ответом):

Как называется метод, который используется для оценки значимости факторов в факторном анализе?

Ответ: Метод, который используется для оценки значимости факторов в факторном анализе, называется тестом Бартлетта. Он позволяет определить, являются ли факторы статистически значимыми и вносят ли они существенный вклад в объяснение вариации

данных.

Обоснование: Тест Бартлетта используется для проверки гипотезы о том, что все факторы незначимы. Если тест показывает, что факторы значимы, это означает, что они вносят существенный вклад в объяснение вариации данных.

Критерии оценивания тестовых заданий для текущего контроля

Максимальное количество баллов за решение одного теста составляет 5 баллов, которые переводятся в академические оценки следующим образом:

- оценка «отлично» (5 баллов) выставляется при условии правильного ответа студента не менее чем на 85% контрольных заданий
- оценка «хорошо» (4 балла) выставляется при условии правильного ответа студента не менее чем на 70% контрольных заданий
- оценка «удовлетворительно» (3 балла) выставляется при условии правильного ответа студента не менее чем на 51% контрольных заданий
- оценка «неудовлетворительно» (1-2 балла) выставляется при условии правильного ответа студента менее чем на 50% контрольных заданий

2.2 Оценочные материалы для проведения промежуточной аттестации

Вопрос 1 (выбор одного правильного ответа):

Что такое среднее арифметическое?

- A) Сумма всех значений, деленная на количество значений.
- B) Значение, которое встречается чаще всего.
- C) Значение, которое делит упорядоченный набор данных на две равные части.
- D) Разница между максимальным и минимальным значениями.

Ответ: A

Обоснование: Среднее арифметическое — это сумма всех значений, деленная на количество значений. Это один из основных показателей центральной тенденции.

Вопрос 2 (выбор нескольких правильных ответов):

Какие из следующих величин являются мерами разброса данных?

- A) Дисперсия.
- B) Стандартное отклонение.
- C) Медиана.
- D) Размах.

Ответ: A, B, D

Обоснование: Меры разброса данных включают:

Дисперсию — среднее квадратическое отклонение от среднего.

Стандартное отклонение — квадратный корень из дисперсии.

Размах — разница между максимальным и минимальным значениями.

Вопрос 3 (установление соответствия):

Установите соответствие между типами шкал и примерами переменных:

Номинальная шкала

Порядковая шкала
Интервальная шкала
Относительная шкала
Примеры переменных:

- A) Цвет глаз.
- B) Оценка по пятибалльной шкале.
- C) Температура в градусах Цельсия.
- D) Рост человека в сантиметрах.

Ответ:

Номинальная шкала — A (Цвет глаз).
Порядковая шкала — B (Оценка по пятибалльной шкале).
Интервальная шкала — C (Температура в градусах Цельсия).
Относительная шкала — D (Рост человека в сантиметрах).
Обоснование:

Номинальная шкала используется для классификации данных без упорядочения (например, цвет глаз).

Порядковая шкала используется для упорядочения данных без точного измерения разницы (например, оценка по пятибалльной шкале).

Интервальная шкала используется для измерения с равными интервалами, но без абсолютного нуля (например, температура в градусах Цельсия).

Относительная шкала используется для измерения с абсолютным нулем и равными интервалами (например, рост человека в сантиметрах).

Вопрос 4 (вопрос с открытым ответом):
Что такое медиана в статистике?

Ответ: Медиана — это значение, которое делит упорядоченный набор данных на две равные части. Половина данных находится ниже медианы, а другая половина — выше.

Обоснование: Медиана — это один из показателей центральной тенденции, который особенно полезен при наличии выбросов в данных, так как она менее чувствительна к экстремальным значениям.

Вопрос 5 (вопрос с кратким ответом):

Как называется мера разброса данных, которая показывает среднее квадратическое отклонение от среднего?

Ответ: Дисперсия

Обоснование: Дисперсия — это мера разброса данных, которая показывает среднее квадратическое отклонение от среднего. Она используется для оценки вариации данных вокруг среднего значения.

Вопрос 6:

Что такое переобучение в машинном обучении?

- A) Когда модель слишком хорошо обучается на тренировочных данных и плохо обобщает на новых данных.
- B) Когда модель недостаточно обучается на тренировочных данных.

- C) Когда модель обучается только на малых данных.
D) Когда модель обучается только на больших данных.
Ответ: A

Обоснование: Переобучение происходит, когда модель машинного обучения слишком хорошо подстраивается под тренировочные данные, включая их шум и случайные колебания. В результате модель плохо обобщает на новых, ранее не виденных данных, что снижает ее предсказательную способность.

Вопрос 7 (выбор одного правильного ответа):
Что такое номинальная шкала?

- A) Шкала, которая упорядочивает данные по рангу.
B) Шкала, которая классифицирует данные без упорядочения.
C) Шкала, которая измеряет данные с равными интервалами.
D) Шкала, которая измеряет данные с абсолютным нулем.
Ответ: B

Обоснование: Номинальная шкала используется для классификации данных без упорядочения. Примеры включают такие переменные, как цвет глаз, пол или национальность.

Вопрос 8 (выбор нескольких правильных ответов):
Какие операции можно выполнять с данными, измеренными по интервальной шкале?

- A) Сложение и вычитание.
B) Умножение и деление.
C) Сравнение на равенство и неравенство.
D) Вычисление среднего арифметического.
Ответ: A, C, D

Обоснование: С данными, измеренными по интервальной шкале, можно выполнять следующие операции:

Сложение и вычитание — для вычисления разницы между значениями.
Сравнение на равенство и неравенство — для определения, равны ли значения или нет.
Вычисление среднего арифметического — для нахождения центральной тенденции.
Умножение и деление не имеют смысла для интервальной шкалы, так как она не имеет абсолютного нуля.

Вопрос 9 (установление соответствия):
Установите соответствие между типами шкал и примерами переменных:

Номинальная шкала
Порядковая шкала
Интервальная шкала
Относительная шкала
Примеры переменных:

- A) Цвет глаз.
B) Оценка по пятибалльной шкале.

С) Температура в градусах Цельсия.

Д) Рост человека в сантиметрах.

Ответ:

Номинальная шкала — А (Цвет глаз).

Порядковая шкала — В (Оценка по пятибалльной шкале).

Интервальная шкала — С (Температура в градусах Цельсия).

Относительная шкала — D (Рост человека в сантиметрах).

Обоснование:

Номинальная шкала используется для классификации данных без упорядочения (например, цвет глаз).

Порядковая шкала используется для упорядочения данных без точного измерения разницы (например, оценка по пятибалльной шкале).

Интервальная шкала используется для измерения с равными интервалами, но без абсолютного нуля (например, температура в градусах Цельсия).

Относительная шкала используется для измерения с абсолютным нулем и равными интервалами (например, рост человека в сантиметрах).

Вопрос 10 (вопрос с открытым ответом):

Что такое порядковая шкала?

Ответ: Порядковая шкала — это шкала, которая упорядочивает данные по рангу, но не позволяет точно измерить разницу между значениями. Примеры включают такие переменные, как оценка по пятибалльной шкале или место в соревновании.

Обоснование: Порядковая шкала позволяет ранжировать данные, но не предоставляет информации о точных различиях между значениями. Например, разница между первым и вторым местом может быть больше, чем между вторым и третьим, но порядковая шкала этого не отражает.

Вопрос 11 (вопрос с кратким ответом):

Какая шкала используется для измерения температуры в градусах Цельсия?

Ответ: Интервальная шкала

Обоснование: Температура в градусах Цельсия измеряется по интервальной шкале, так как она имеет равные интервалы между значениями, но не имеет абсолютного нуля. Это означает, что разница между 10°C и 20°C такая же, как между 20°C и 30°C , но 0°C не означает полное отсутствие температуры.

Вопрос 12 (выбор одного правильного ответа):

Что такое корреляция в статистике?

А) Мера зависимости между двумя переменными.

В) Мера разброса данных.

С) Мера центральной тенденции.

Д) Мера вероятности события.

Ответ: А

Обоснование: Корреляция — это мера зависимости между двумя переменными. Она показывает, насколько сильно связаны изменения одной переменной с изменениями другой.

Вопрос 13 (выбор нескольких правильных ответов):

Какие из следующих величин являются мерами центральной тенденции?

А) Среднее арифметическое.

В) Медиана.

С) Мода.

Д) Дисперсия.

Ответ: А, В, С

Обоснование: Меры центральной тенденции включают:

Среднее арифметическое — сумма всех значений, деленная на количество значений.

Медиана — значение, которое делит упорядоченный набор данных на две равные части.

Мода — значение, которое встречается чаще всего.

Вопрос 14 (установление соответствия):

Установите соответствие между типами корреляции и примерами:

Положительная корреляция

Отрицательная корреляция

Нулевая корреляция

Примеры:

А) Рост и вес человека.

В) Температура и количество людей на пляже.

С) Цвет глаз и рост человека.

Ответ:

Положительная корреляция — А (Рост и вес человека).

Отрицательная корреляция — В (Температура и количество людей на пляже).

Нулевая корреляция — С (Цвет глаз и рост человека).

Обоснование:

Положительная корреляция — когда увеличение одной переменной сопровождается увеличением другой (например, рост и вес).

Отрицательная корреляция — когда увеличение одной переменной сопровождается уменьшением другой (например, температура и количество людей на пляже в зимний период).

Нулевая корреляция — когда между переменными нет никакой связи (например, цвет глаз и рост).

Вопрос 15 (вопрос с открытым ответом):

Что такое коэффициент корреляции Пирсона?

Ответ: Коэффициент корреляции Пирсона — это мера линейной зависимости между двумя переменными. Он принимает значения от -1 до 1, где -1 указывает на сильную отрицательную корреляцию, 1 — на сильную положительную корреляцию, а 0 — на отсутствие корреляции.

Обоснование: Коэффициент корреляции Пирсона используется для оценки силы и направления линейной связи между двумя переменными. Он помогает определить,

насколько сильно изменения одной переменной связаны с изменениями другой.

Вопрос 16 (вопрос с кратким ответом):

Как называется мера разброса данных, которая показывает среднее квадратическое отклонение от среднего?

Ответ: Дисперсия

Обоснование: Дисперсия — это мера разброса данных, которая показывает среднее квадратическое отклонение от среднего. Она используется для оценки вариации данных вокруг среднего значения.

Вопрос 17 (выбор одного правильного ответа):

Что такое критерий Стьюдента?

- A) Метод для сравнения средних величин двух выборок.
- B) Метод для сравнения дисперсий двух выборок.
- C) Метод для сравнения медиан двух выборок.
- D) Метод для сравнения частот двух выборок.

Ответ: A

Обоснование: Критерий Стьюдента (t-критерий) используется для сравнения средних величин двух выборок. Он позволяет определить, есть ли статистически значимая разница между средними значениями двух групп.

Вопрос 18 (выбор нескольких правильных ответов):

Какие предположения должны выполняться для применения критерия Стьюдента?

- A) Нормальное распределение данных.
- B) Равенство дисперсий в двух группах.
- C) Независимость наблюдений.
- D) Одинаковый размер выборок.

Ответ: A, B, C

Обоснование: Для применения критерия Стьюдента должны выполняться следующие предположения:

Нормальное распределение данных — данные в обеих группах должны быть распределены нормально.

Равенство дисперсий — дисперсии в двух группах должны быть примерно равны.

Независимость наблюдений — наблюдения в каждой группе должны быть независимыми друг от друга.

Размер выборок не обязательно должен быть одинаковым.

Вопрос 19 (установление соответствия):

Установите соответствие между типами критерия Стьюдента и примерами:

Критерий Стьюдента для независимых выборок

Критерий Стьюдента для зависимых выборок

Примеры:

- A) Сравнение среднего дохода мужчин и женщин.

В) Сравнение результатов теста до и после обучения.

Ответ:

Критерий Стьюдента для независимых выборок — А (Сравнение среднего дохода мужчин и женщин).

Критерий Стьюдента для зависимых выборок — В (Сравнение результатов теста до и после обучения).

Обоснование:

Критерий Стьюдента для независимых выборок используется, когда две группы не связаны друг с другом (например, мужчины и женщины).

Критерий Стьюдента для зависимых выборок используется, когда две группы связаны (например, один и тот же человек до и после обучения).

Вопрос 20 (вопрос с открытым ответом):

Что такое р-значение в контексте критерия Стьюдента?

Ответ: Р-значение — это вероятность получить наблюдаемый результат (или более экстремальный) при условии, что нулевая гипотеза верна. Если р-значение меньше выбранного уровня значимости (например, 0.05), то нулевая гипотеза отвергается, и считается, что разница между средними значениями статистически значима.

Обоснование: Р-значение помогает определить, является ли разница между средними значениями двух групп случайной или статистически значимой. Чем меньше р-значение, тем сильнее доказательства против нулевой гипотезы.

Вопрос 21 (вопрос с кратким ответом):

Как называется нулевая гипотеза в критерии Стьюдента?

Ответ: Нулевая гипотеза в критерии Стьюдента утверждает, что нет статистически значимой разницы между средними значениями двух групп. Обычно она обозначается как $H_0: \mu_1 = \mu_2$, где μ_1 и μ_2 — средние значения двух групп.

Обоснование: Нулевая гипотеза (H_0) в критерии Стьюдента предполагает, что средние значения двух групп равны. Альтернативная гипотеза (H_1) утверждает, что средние значения различны.

Вопрос 22 (выбор одного правильного ответа):

Что такое непараметрическая статистика?

А) Методы, которые не требуют предположений о распределении данных.

В) Методы, которые требуют предположений о нормальном распределении данных.

С) Методы, которые используются только для больших выборок.

Д) Методы, которые применяются только для категориальных данных.

Ответ: А

Обоснование: Непараметрическая статистика — это методы, которые не требуют предположений о распределении данных. Они особенно полезны, когда данные не соответствуют нормальному распределению или когда размер выборки мал.

Вопрос 23 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к непараметрическим методам?

А) U-критерий Манна-Уитни.

В) Критерий Вилкоксона.

С) Критерий хи-квадрат.

Д) F-критерий.

Ответ: А, В, С

Обоснование: К непараметрическим методам относятся:

U-критерий Манна-Уитни — для сравнения двух независимых выборок.

Критерий Вилкоксона — для сравнения двух зависимых выборок.

Критерий хи-квадрат — для анализа категориальных данных.

F-критерий является параметрическим методом и используется для сравнения дисперсий.

Вопрос 24 (установление соответствия):

Установите соответствие между типами непараметрических тестов и примерами:

U-критерий Манна-Уитни

Критерий Вилкоксона

Критерий хи-квадрат

Примеры:

А) Сравнение среднего дохода мужчин и женщин.

В) Сравнение результатов теста до и после обучения.

С) Анализ связи между полом и предпочтениями в еде.

Ответ:

U-критерий Манна-Уитни — А (Сравнение среднего дохода мужчин и женщин).

Критерий Вилкоксона — В (Сравнение результатов теста до и после обучения).

Критерий хи-квадрат — С (Анализ связи между полом и предпочтениями в еде).

Обоснование:

U-критерий Манна-Уитни используется для сравнения двух независимых выборок (например, мужчины и женщины).

Критерий Вилкоксона используется для сравнения двух зависимых выборок (например, один и тот же человек до и после обучения).

Критерий хи-квадрат используется для анализа связи между двумя категориальными переменными (например, пол и предпочтения в еде).

Вопрос 4 (вопрос с открытым ответом):

Что такое ранг в контексте непараметрической статистики?

Ответ: Ранг в контексте непараметрической статистики — это порядковый номер, который присваивается каждому наблюдению в упорядоченном наборе данных. Ранговые методы используются, когда данные не соответствуют нормальному распределению или когда важно учитывать только порядок значений, а не их точные величины.

Обоснование: Ранговые методы, такие как U-критерий Манна-Уитни и критерий Вилкоксона, основаны на ранжировании данных. Это позволяет сравнивать выборки без предположений о конкретном распределении данных.

Вопрос 25 (вопрос с кратким ответом):

Как называется нулевая гипотеза в U-критерии Манна-Уитни?

Ответ: Нулевая гипотеза в U-критерии Манна-Уитни утверждает, что две выборки происходят из одного и того же распределения. Это означает, что нет статистически значимой разницы между двумя группами.

Обоснование: Нулевая гипотеза (H_0) в U-критерии Манна-Уитни предполагает, что две выборки не различаются. Альтернативная гипотеза (H_1) утверждает, что одна выборка имеет тенденцию к более высоким значениям, чем другая.

Вопрос 26 (выбор одного правильного ответа):

Что такое однофакторный дисперсионный анализ (ANOVA)?

- A) Метод для сравнения средних величин двух выборок.
- B) Метод для сравнения средних величин более чем двух выборок.
- C) Метод для сравнения дисперсий двух выборок.
- D) Метод для сравнения медиан двух выборок.

Ответ: B

Обоснование: Однофакторный дисперсионный анализ (ANOVA) используется для сравнения средних величин более чем двух выборок. Он позволяет определить, есть ли статистически значимая разница между средними значениями нескольких групп.

Вопрос 27 (выбор нескольких правильных ответов):

Какие предположения должны выполняться для применения однофакторного ANOVA?

- A) Нормальное распределение данных в каждой группе.
- B) Равенство дисперсий в группах.
- C) Независимость наблюдений.
- D) Одинаковый размер выборок.

Ответ: A, B, C

Обоснование: Для применения однофакторного ANOVA должны выполняться следующие предположения:

Нормальное распределение данных — данные в каждой группе должны быть распределены нормально.

Равенство дисперсий — дисперсии в группах должны быть примерно равны.

Независимость наблюдений — наблюдения в каждой группе должны быть независимыми друг от друга.

Размер выборок не обязательно должен быть одинаковым.

Вопрос 28 (установление соответствия):

Установите соответствие между типами ANOVA и примерами:

Однофакторный ANOVA

Двухфакторный ANOVA

Примеры:

- A) Сравнение среднего дохода в трех разных регионах.
- B) Сравнение среднего дохода в трех разных регионах и двух разных профессиях.

Ответ:

Однофакторный ANOVA — А (Сравнение среднего дохода в трех разных регионах).
Двухфакторный ANOVA — В (Сравнение среднего дохода в трех разных регионах и двух разных профессиях).

Обоснование:

Однофакторный ANOVA используется, когда есть один фактор (например, регион).
Двухфакторный ANOVA используется, когда есть два фактора (например, регион и профессия).

Вопрос 29 (вопрос с открытым ответом):

Что такое нулевая гипотеза в однофакторном ANOVA?

Ответ: Нулевая гипотеза в однофакторном ANOVA утверждает, что нет статистически значимой разницы между средними значениями всех групп. Это означает, что все группы происходят из одного и того же распределения.

Обоснование: Нулевая гипотеза (H_0) в однофакторном ANOVA предполагает, что средние значения всех групп равны. Альтернативная гипотеза (H_1) утверждает, что хотя бы одна группа имеет среднее значение, отличное от других.

Вопрос 30 (вопрос с кратким ответом):

Как называется статистика, используемая в однофакторном ANOVA?

Ответ: Статистика, используемая в однофакторном ANOVA, называется F-статистикой. Она вычисляется как отношение дисперсии между группами к дисперсии внутри групп.

Обоснование: F-статистика в однофакторном ANOVA используется для оценки значимости различий между средними значениями групп. Если F-статистика велика, это указывает на то, что различия между группами статистически значимы.

Вопрос 31 (выбор одного правильного ответа):

Что такое дисперсионный анализ (ANOVA)?

- А) Метод для сравнения средних величин двух выборок.
- В) Метод для сравнения средних величин более чем двух выборок.
- С) Метод для сравнения дисперсий двух выборок.
- Д) Метод для сравнения медиан двух выборок.

Ответ: В

Обоснование: Дисперсионный анализ (ANOVA) используется для сравнения средних величин более чем двух выборок. Он позволяет определить, есть ли статистически значимая разница между средними значениями нескольких групп.

Вопрос 32 (выбор нескольких правильных ответов):

Какие предположения должны выполняться для применения дисперсионного анализа?

- А) Нормальное распределение данных в каждой группе.
- В) Равенство дисперсий в группах.

- С) Независимость наблюдений.
D) Одинаковый размер выборок.
Ответ: А, В, С

Обоснование: Для применения дисперсионного анализа должны выполняться следующие предположения:

Нормальное распределение данных — данные в каждой группе должны быть распределены нормально.

Равенство дисперсий — дисперсии в группах должны быть примерно равны.

Независимость наблюдений — наблюдения в каждой группе должны быть независимыми друг от друга.

Размер выборок не обязательно должен быть одинаковым.

Вопрос 33 (установление соответствия):

Установите соответствие между типами дисперсионного анализа и примерами:

Однофакторный ANOVA

Двухфакторный ANOVA

Примеры:

А) Сравнение среднего дохода в трех разных регионах.

В) Сравнение среднего дохода в трех разных регионах и двух разных профессиях.

Ответ:

Однофакторный ANOVA — А (Сравнение среднего дохода в трех разных регионах).

Двухфакторный ANOVA — В (Сравнение среднего дохода в трех разных регионах и двух разных профессиях).

Обоснование:

Однофакторный ANOVA используется, когда есть один фактор (например, регион).

Двухфакторный ANOVA используется, когда есть два фактора (например, регион и профессия).

Вопрос 34 (вопрос с открытым ответом):

Что такое нулевая гипотеза в дисперсионном анализе?

Ответ: Нулевая гипотеза в дисперсионном анализе утверждает, что нет статистически значимой разницы между средними значениями всех групп. Это означает, что все группы происходят из одного и того же распределения.

Обоснование: Нулевая гипотеза (H_0) в дисперсионном анализе предполагает, что средние значения всех групп равны. Альтернативная гипотеза (H_1) утверждает, что хотя бы одна группа имеет среднее значение, отличное от других.

Вопрос 35 (вопрос с кратким ответом):

Как называется статистика, используемая в дисперсионном анализе?

Ответ: Статистика, используемая в дисперсионном анализе, называется F-статистикой. Она вычисляется как отношение дисперсии между группами к дисперсии внутри групп.

Обоснование: F-статистика в дисперсионном анализе используется для оценки

значимости различий между средними значениями групп. Если F-статистика велика, это указывает на то, что различия между группами статистически значимы.

Вопрос 36 (выбор одного правильного ответа):

Что такое таблица сопряженности?

- A) Таблица, которая показывает распределение одной переменной.
 - B) Таблица, которая показывает взаимосвязь между двумя категориальными переменными.
 - C) Таблица, которая показывает взаимосвязь между двумя непрерывными переменными.
 - D) Таблица, которая показывает взаимосвязь между тремя и более переменными.
- Ответ: B

Обоснование: Таблица сопряженности — это таблица, которая показывает взаимосвязь между двумя категориальными переменными. Она используется для анализа связи между двумя переменными, представленными в виде категорий.

Вопрос 37 (выбор нескольких правильных ответов):

Какие из следующих величин можно использовать для анализа таблиц сопряженности?

- A) Коэффициент корреляции Пирсона.
- B) Критерий хи-квадрат.
- C) Коэффициент сопряженности.
- D) Коэффициент детерминации.

Ответ: B, C

Обоснование: Для анализа таблиц сопряженности можно использовать:

Критерий хи-квадрат — для проверки независимости двух категориальных переменных.

Коэффициент сопряженности — для измерения силы связи между двумя категориальными переменными.

Коэффициент корреляции Пирсона и коэффициент детерминации используются для анализа связи между непрерывными переменными.

Вопрос 38 (установление соответствия):

Установите соответствие между типами таблиц и примерами:

Таблица частот

Таблица сопряженности

Таблица флагов

Примеры:

- A) Таблица, показывающая количество студентов по факультетам.
- B) Таблица, показывающая взаимосвязь между полом и предпочтениями в еде.
- C) Таблица, показывающая наличие или отсутствие определенного признака у каждого объекта.

Ответ:

Таблица частот — A (Таблица, показывающая количество студентов по факультетам).

Таблица сопряженности — В (Таблица, показывающая взаимосвязь между полом и предпочтениями в еде).

Таблица флагов — С (Таблица, показывающая наличие или отсутствие определенного признака у каждого объекта).

Обоснование:

Таблица частот используется для представления распределения одной переменной (например, количество студентов по факультетам).

Таблица сопряженности используется для анализа взаимосвязи между двумя категориальными переменными (например, пол и предпочтения в еде).

Таблица флагов используется для представления наличия или отсутствия определенного признака у каждого объекта (например, наличие или отсутствие заболевания).

Вопрос 39 (вопрос с открытым ответом):

Что такое таблица флагов?

Ответ: Таблица флагов — это таблица, которая показывает наличие или отсутствие определенного признака у каждого объекта. Обычно в такой таблице используются бинарные значения (например, 0 или 1), где 0 означает отсутствие признака, а 1 — его наличие.

Обоснование: Таблицы флагов используются для представления категориальных данных, где интерес представляет наличие или отсутствие определенного признака. Это может быть полезно для анализа данных, где важно определить, присутствует ли определенный атрибут у объекта.

Вопрос 40 (вопрос с кратким ответом):

Как называется статистика, используемая для проверки независимости в таблицах сопряженности?

Ответ: Статистика, используемая для проверки независимости в таблицах сопряженности, называется критерием хи-квадрат (χ^2). Она позволяет определить, есть ли статистически значимая связь между двумя категориальными переменными.

Обоснование: Критерий хи-квадрат используется для проверки гипотезы о независимости двух категориальных переменных. Если значение критерия хи-квадрат велико, это указывает на то, что переменные связаны.

Вопрос 41 (выбор одного правильного ответа):

Что такое канонический анализ?

A) Метод для выявления взаимосвязей между двумя множествами переменных.

B) Метод для классификации данных на основе их сходства.

C) Метод для уменьшения размерности данных.

D) Метод для прогнозирования одной переменной на основе другой.

Ответ: A

Обоснование: Канонический анализ — это метод, который используется для выявления взаимосвязей между двумя множествами переменных. Он позволяет найти линейные комбинации переменных, которые максимизируют корреляцию между двумя наборами данных.

Вопрос 42 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к каноническому анализу?

- A) Канонический корреляционный анализ.
- B) Канонический дискриминантный анализ.
- C) Факторный анализ.
- D) Кластерный анализ.

Ответ: A, B

Обоснование: К каноническому анализу относятся:

Канонический корреляционный анализ — для выявления взаимосвязей между двумя множествами переменных.

Канонический дискриминантный анализ — для классификации и разделения групп на основе канонических переменных.

Факторный и кластерный анализ относятся к другим методам многомерного анализа.

Вопрос 43 (установление соответствия):

Установите соответствие между типами канонического анализа и примерами:

Канонический корреляционный анализ

Канонический дискриминантный анализ

Примеры:

A) Анализ взаимосвязи между набором экономических показателей и набором социальных показателей.

B) Классификация пациентов на основе набора клинических и лабораторных показателей.

Ответ:

Канонический корреляционный анализ — A (Анализ взаимосвязи между набором экономических показателей и набором социальных показателей).

Канонический дискриминантный анализ — B (Классификация пациентов на основе набора клинических и лабораторных показателей).

Обоснование:

Канонический корреляционный анализ используется для анализа взаимосвязи между двумя множествами переменных (например, экономические и социальные показатели).

Канонический дискриминантный анализ используется для классификации объектов на основе нескольких переменных (например, классификация пациентов по клиническим и лабораторным показателям).

Вопрос 44 (вопрос с открытым ответом):

Что такое канонические переменные в каноническом анализе?

Ответ: Канонические переменные — это линейные комбинации исходных переменных, которые максимизируют корреляцию между двумя множествами переменных. Они используются для упрощения анализа и интерпретации взаимосвязей между множествами переменных.

Обоснование: Канонические переменные создаются таким образом, чтобы они были максимально коррелированы между собой и минимально коррелированы внутри каждого множества. Это позволяет выявить основные взаимосвязи между двумя множествами

переменных.

Вопрос 45 (вопрос с кратким ответом):

Как называется статистика, используемая для оценки значимости канонических корреляций?

Ответ: Статистика, используемая для оценки значимости канонических корреляций, называется критерием Уилкса (Wilks' Lambda). Она позволяет определить, являются ли канонические корреляции статистически значимыми.

Обоснование: Критерий Уилкса используется для проверки гипотезы о том, что канонические корреляции равны нулю. Если значение критерия Уилкса значительно, это указывает на наличие статистически значимых взаимосвязей между множествами переменных.

Вопрос 46 (выбор одного правильного ответа):

Что такое линейная регрессия?

- A) Метод для моделирования нелинейных взаимосвязей между переменными.
- B) Метод для моделирования линейных взаимосвязей между переменными.
- C) Метод для классификации данных на основе их сходства.
- D) Метод для уменьшения размерности данных.

Ответ: B

Обоснование: Линейная регрессия — это метод, который используется для моделирования линейных взаимосвязей между переменными. Она позволяет предсказать значение зависимой переменной на основе одной или нескольких независимых переменных.

Вопрос 47 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к нелинейному моделированию?

- A) Полиномиальная регрессия.
- B) Логистическая регрессия.
- C) Линейная регрессия.
- D) Деревья решений.

Ответ: A, B, D

Обоснование: К нелинейному моделированию относятся:

Полиномиальная регрессия — используется для моделирования нелинейных взаимосвязей с помощью полиномов.

Логистическая регрессия — используется для моделирования нелинейных взаимосвязей в задачах классификации.

Деревья решений — используются для моделирования нелинейных взаимосвязей и классификации данных.

Линейная регрессия относится к линейному моделированию.

Вопрос 48 (установление соответствия):

Установите соответствие между типами моделей и примерами:

Линейная регрессия

Нелинейная регрессия

Примеры:

- А) Прогнозирование цены на недвижимость на основе площади и количества комнат.
- В) Прогнозирование вероятности заболевания на основе возраста и индекса массы тела.
- С) Прогнозирование роста растения на основе времени и количества удобрений.

Ответ:

Линейная регрессия — А (Прогнозирование цены на недвижимость на основе площади и количества комнат).

Нелинейная регрессия — В (Прогнозирование вероятности заболевания на основе возраста и индекса массы тела), С (Прогнозирование роста растения на основе времени и количества удобрений).

Обоснование:

Линейная регрессия используется для прогнозирования цены на недвижимость, так как взаимосвязь между ценой, площадью и количеством комнат может быть аппроксимирована линейной моделью.

Нелинейная регрессия используется для прогнозирования вероятности заболевания и роста растения, так как взаимосвязь между переменными может быть нелинейной.

Вопрос 49 (вопрос с открытым ответом):

Что такое полиномиальная регрессия?

Ответ: Полиномиальная регрессия — это метод нелинейного моделирования, который используется для аппроксимации взаимосвязей между переменными с помощью полиномов. Она позволяет моделировать нелинейные взаимосвязи, добавляя в модель степени независимых переменных.

Обоснование: Полиномиальная регрессия расширяет линейную регрессию, добавляя члены с более высокими степенями независимых переменных. Это позволяет моделировать более сложные взаимосвязи между переменными.

Вопрос 50 (вопрос с кратким ответом):

Как называется метод, который используется для оценки значимости коэффициентов в линейной регрессии?

Ответ: Метод, который используется для оценки значимости коэффициентов в линейной регрессии, называется t-тестом. Он позволяет определить, являются ли коэффициенты статистически значимыми, то есть отличаются ли они от нуля.

Обоснование: Т-тест используется для проверки гипотезы о том, что коэффициент регрессии равен нулю. Если значение t-статистики значительно, это указывает на то, что коэффициент статистически значим и не равен нулю.

Вопрос 51 (выбор одного правильного ответа):

Что такое дискриминантный анализ?

- А) Метод для классификации данных на основе их сходства.
- В) Метод для уменьшения размерности данных.
- С) Метод для прогнозирования одной переменной на основе другой.
- Д) Метод для разделения данных на группы на основе линейных комбинаций

переменных.

Ответ: D

Обоснование: Дискриминантный анализ — это метод, который используется для разделения данных на группы на основе линейных комбинаций переменных. Он позволяет классифицировать объекты в одну из нескольких групп.

Вопрос 52 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к дискриминантному анализу?

- A) Линейный дискриминантный анализ (LDA).
- B) Квадратичный дискриминантный анализ (QDA).
- C) Канонический дискриминантный анализ.
- D) Кластерный анализ.

Ответ: A, B, C

Обоснование: К дискриминантному анализу относятся:

Линейный дискриминантный анализ (LDA) — используется для классификации данных с предположением о нормальном распределении и равных дисперсиях.

Квадратичный дискриминантный анализ (QDA) — используется для классификации данных без предположения о равных дисперсиях.

Канонический дискриминантный анализ — используется для выявления линейных комбинаций переменных, которые максимизируют различия между группами.

Кластерный анализ относится к другим методам многомерного анализа.

Вопрос 53 (установление соответствия):

Установите соответствие между типами дискриминантного анализа и примерами:

Линейный дискриминантный анализ (LDA)

Квадратичный дискриминантный анализ (QDA)

Примеры:

A) Классификация пациентов на основе набора клинических и лабораторных показателей.

B) Классификация изображений на основе набора пикселей.

Ответ:

Линейный дискриминантный анализ (LDA) — A (Классификация пациентов на основе набора клинических и лабораторных показателей).

Квадратичный дискриминантный анализ (QDA) — B (Классификация изображений на основе набора пикселей).

Обоснование:

Линейный дискриминантный анализ (LDA) часто используется в медицине для классификации пациентов на основе клинических и лабораторных показателей.

Квадратичный дискриминантный анализ (QDA) может быть полезен для классификации изображений, где дисперсии могут различаться.

Вопрос 54 (вопрос с открытым ответом):

Что такое дискриминантная функция в дискриминантном анализе?

Ответ: Дискриминантная функция — это линейная комбинация переменных, которая

используется для разделения данных на группы. Она максимизирует различия между группами и минимизирует различия внутри групп, позволяя классифицировать новые объекты в одну из групп.

Обоснование: Дискриминантная функция строится таким образом, чтобы она наилучшим образом разделяла данные на группы. Это достигается путем максимизации межгрупповой дисперсии и минимизации внутригрупповой дисперсии.

Вопрос 55 (вопрос с кратким ответом):

Как называется статистика, используемая для оценки значимости дискриминантной функции?

Ответ: Статистика, используемая для оценки значимости дискриминантной функции, называется F-статистикой. Она вычисляется как отношение дисперсии между группами к дисперсии внутри групп и используется для проверки гипотезы о том, что дискриминантная функция значимо различает группы.

Обоснование: F-статистика в дискриминантном анализе используется для проверки гипотезы о том, что различия между группами значимы. Если значение F-статистики велико, это указывает на то, что дискриминантная функция статистически значима.

Вопрос 56 (выбор одного правильного ответа):

Что такое кластерный анализ?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: A

Обоснование: Кластерный анализ — это метод, который используется для классификации данных на основе их сходства. Он позволяет группировать объекты в кластеры, где объекты внутри одного кластера более похожи друг на друга, чем на объекты из других кластеров.

Вопрос 57 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к кластерному анализу?

- A) Иерархическая кластеризация.
- B) Метод k-средних.
- C) Линейный дискриминантный анализ.
- D) Квадратичный дискриминантный анализ.

Ответ: A, B

Обоснование: К кластерному анализу относятся:

Иерархическая кластеризация — метод, который создает иерархию кластеров, начиная с отдельных объектов и объединяя их в более крупные кластеры.

Метод k-средних — метод, который разделяет данные на k кластеров, минимизируя сумму квадратов расстояний от объектов до центров кластеров.

Линейный и квадратичный дискриминантный анализ относятся к другим методам

многомерного анализа.

Вопрос 58 (установление соответствия):

Установите соответствие между типами кластерного анализа и примерами:

Иерархическая кластеризация

Метод k-средних

Примеры:

А) Построение дендрограммы для анализа сходства видов растений.

В) Разделение клиентов на сегменты на основе их покупательского поведения.

Ответ:

Иерархическая кластеризация — А (Построение дендрограммы для анализа сходства видов растений).

Метод k-средних — В (Разделение клиентов на сегменты на основе их покупательского поведения).

Обоснование:

Иерархическая кластеризация часто используется для построения дендрограмм, которые показывают иерархию сходства между объектами (например, виды растений).

Метод k-средних часто применяется в маркетинге для сегментации клиентов на основе их покупательского поведения.

Вопрос 4 (вопрос с открытым ответом):

Что такое кластер в кластерном анализе?

Ответ: Кластер в кластерном анализе — это группа объектов, которые более похожи друг на друга, чем на объекты из других кластеров. Кластеры формируются на основе меры сходства или расстояния между объектами.

Обоснование: Кластеры создаются таким образом, чтобы объекты внутри одного кластера были максимально похожи друг на друга, а объекты из разных кластеров — максимально различны. Это позволяет выявить структуры и закономерности в данных.

Вопрос 59 (вопрос с кратким ответом):

Как называется метод, который используется для определения оптимального числа кластеров в методе k-средних?

Ответ: Метод, который используется для определения оптимального числа кластеров в методе k-средних, называется методом локтя (elbow method). Он основан на анализе графика суммы квадратов расстояний от объектов до центров кластеров и выборе точки, где график резко меняет свой наклон.

Обоснование: Метод локтя помогает определить оптимальное число кластеров, выбирая точку, где дальнейшее увеличение числа кластеров не приводит к значительному улучшению качества кластеризации.

Вопрос 60 (выбор одного правильного ответа):

Что такое дерево классификации?

А) Метод для классификации данных на основе их сходства.

В) Метод для уменьшения размерности данных.

- C) Метод для прогнозирования одной переменной на основе другой.
D) Метод для разделения данных на группы на основе иерархической структуры.
Ответ: D

Обоснование: Дерево классификации — это метод, который используется для разделения данных на группы на основе иерархической структуры. Он позволяет классифицировать объекты в одну из нескольких категорий, используя набор правил, основанных на значениях переменных.

Вопрос 61 (выбор нескольких правильных ответов):
Какие из следующих методов относятся к деревьям классификации?

- A) ID3.
B) C4.5.
C) CART.
D) Линейный дискриминантный анализ.
Ответ: A, B, C

Обоснование: К деревьям классификации относятся:

ID3 — алгоритм, который использует энтропию для выбора признаков.

C4.5 — улучшенная версия ID3, которая учитывает пропущенные значения и непрерывные переменные.

CART — метод, который использует критерий Gini для выбора признаков и может строить как классификационные, так и регрессионные деревья.

Линейный дискриминантный анализ относится к другим методам многомерного анализа.

Вопрос 62 (установление соответствия):
Установите соответствие между типами деревьев классификации и примерами:

ID3
C4.5
CART
Примеры:

- A) Дерево, построенное на основе энтропии.
B) Дерево, которое учитывает пропущенные значения.
C) Дерево, которое использует критерий Gini.
Ответ:

ID3 — A (Дерево, построенное на основе энтропии).

C4.5 — B (Дерево, которое учитывает пропущенные значения).

CART — C (Дерево, которое использует критерий Gini).

Обоснование:

ID3 использует энтропию для выбора признаков.

C4.5 является улучшенной версией ID3 и учитывает пропущенные значения.

CART использует критерий Gini для выбора признаков.

Вопрос 4 (вопрос с открытым ответом):

Что такое критерий разделения узлов в деревьях классификации?

Ответ: Критерий разделения узлов в деревьях классификации — это правило, которое используется для выбора признака и порога, по которому будет производиться разделение данных на два или более подмножества. Критерии разделения узлов выбирают признак и порог таким образом, чтобы максимизировать чистоту (однородность) полученных подмножеств.

Обоснование: Критерии разделения узлов, такие как энтропия, информационный выигрыш и Gini-индекс, помогают выбрать признак и порог, которые наилучшим образом разделяют данные на подмножества, где объекты внутри одного подмножества более похожи друг на друга, чем на объекты из других подмножеств.

Вопрос 63 (вопрос с кратким ответом):

Как называется метод, который используется для предотвращения переобучения в деревьях классификации?

Ответ: Метод, который используется для предотвращения переобучения в деревьях классификации, называется обрезкой (pruning). Обрезка включает удаление ветвей дерева, которые не вносят значительного вклада в точность классификации, чтобы упростить модель и повысить ее обобщаемость.

Обоснование: Обрезка помогает избежать переобучения, когда дерево становится слишком сложным и начинает учитывать шум и случайные колебания в данных. Удаление лишних ветвей позволяет улучшить способность модели обобщать на новые данные.

Вопрос 64 (выбор одного правильного ответа):

Что такое факторный анализ?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных путем выявления латентных факторов.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: B

Обоснование: Факторный анализ — это метод, который используется для уменьшения размерности данных путем выявления латентных факторов. Он позволяет выявить скрытые структуры в данных и представить их в более компактной форме.

Вопрос 65 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к факторному анализу?

- A) Метод главных компонент (PCA).
- B) Канонический корреляционный анализ.
- C) Линейный дискриминантный анализ.
- D) Факторный анализ методом максимального правдоподобия.

Ответ: A, D

Обоснование: К факторному анализу относятся:

Метод главных компонент (PCA) — метод, который используется для уменьшения размерности данных путем выделения главных компонент.

Факторный анализ методом максимального правдоподобия — метод, который используется для оценки параметров факторной модели на основе максимального правдоподобия.

Канонический корреляционный анализ и линейный дискриминантный анализ относятся к другим методам многомерного анализа.

Вопрос 66 (установление соответствия):

Установите соответствие между типами факторного анализа и примерами:

Метод главных компонент (РСА)

Факторный анализ методом максимального правдоподобия

Примеры:

А) Анализ взаимосвязи между набором экономических показателей и набором социальных показателей.

В) Уменьшение размерности набора данных, описывающего характеристики пациентов.

Ответ:

Метод главных компонент (РСА) — В (Уменьшение размерности набора данных, описывающего характеристики пациентов).

Факторный анализ методом максимального правдоподобия — А (Анализ взаимосвязи между набором экономических показателей и набором социальных показателей).

Обоснование:

Метод главных компонент (РСА) используется для уменьшения размерности данных, например, при анализе характеристик пациентов.

Факторный анализ методом максимального правдоподобия может быть использован для анализа взаимосвязи между экономическими и социальными показателями.

Вопрос 67 (вопрос с открытым ответом):

Что такое факторная нагрузка в факторном анализе?

Ответ: Факторная нагрузка — это коэффициент, который показывает силу связи между исходной переменной и латентным фактором. Она указывает на то, насколько сильно каждая переменная связана с каждым фактором.

Обоснование: Факторные нагрузки помогают интерпретировать факторы, показывая, какие переменные наиболее сильно связаны с каждым фактором. Это позволяет понять, какие переменные вносят наибольший вклад в каждый фактор.

Вопрос 68 (вопрос с кратким ответом):

Как называется метод, который используется для оценки значимости факторов в факторном анализе?

Ответ: Метод, который используется для оценки значимости факторов в факторном анализе, называется тестом Бартлетта. Он позволяет определить, являются ли факторы статистически значимыми и вносят ли они существенный вклад в объяснение вариации данных.

Обоснование: Тест Бартлетта используется для проверки гипотезы о том, что все факторы незначимы. Если тест показывает, что факторы значимы, это означает, что они вносят существенный вклад в объяснение вариации данных.

Вопрос 69 (выбор одного правильного ответа):

Что такое многомерное шкалирование?

- A) Метод для классификации данных на основе их сходства.
- B) Метод для уменьшения размерности данных путем визуализации их в пространстве меньшей размерности.
- C) Метод для прогнозирования одной переменной на основе другой.
- D) Метод для разделения данных на группы на основе линейных комбинаций переменных.

Ответ: B

Обоснование: Многомерное шкалирование — это метод, который используется для уменьшения размерности данных путем визуализации их в пространстве меньшей размерности. Он позволяет представить данные в виде точек в двумерном или трехмерном пространстве, сохраняя при этом расстояния между объектами.

Вопрос 70 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к многомерному шкалированию?

- A) Метрическое многомерное шкалирование.
- B) Неметрическое многомерное шкалирование.
- C) Линейный дискриминантный анализ.
- D) Факторный анализ.

Ответ: A, B

Обоснование: К многомерному шкалированию относятся:

Метрическое многомерное шкалирование — метод, который сохраняет точные расстояния между объектами.

Неметрическое многомерное шкалирование — метод, который сохраняет только порядок расстояний между объектами.

Линейный дискриминантный анализ и факторный анализ относятся к другим методам многомерного анализа.

Вопрос 71 (установление соответствия):

Установите соответствие между типами многомерного шкалирования и примерами:

Метрическое многомерное шкалирование
Неметрическое многомерное шкалирование
Примеры:

A) Визуализация данных, где расстояния между точками точно соответствуют исходным расстояниям.

B) Визуализация данных, где сохраняется только порядок расстояний между точками.

Ответ:

Метрическое многомерное шкалирование — A (Визуализация данных, где расстояния между точками точно соответствуют исходным расстояниям).

Неметрическое многомерное шкалирование — B (Визуализация данных, где сохраняется только порядок расстояний между точками).

Обоснование:

Метрическое многомерное шкалирование сохраняет точные расстояния между объектами.

Неметрическое многомерное шкалирование сохраняет только порядок расстояний между объектами.

Вопрос 72 (вопрос с открытым ответом):

Что такое стресс в контексте многомерного шкалирования?

Ответ: Стресс в контексте многомерного шкалирования — это мера, которая показывает, насколько хорошо визуализация в пространстве меньшей размерности сохраняет исходные расстояния между объектами. Чем меньше значение стресса, тем лучше визуализация отражает исходные расстояния.

Обоснование: Стресс используется для оценки качества многомерного шкалирования. Он вычисляется как разность между исходными и восстановленными расстояниями, и его минимизация является целью многомерного шкалирования.

Вопрос 73 (вопрос с кратким ответом):

Как называется метод, который используется для визуализации данных в многомерном шкалировании?

Ответ: Метод, который используется для визуализации данных в многомерном шкалировании, называется многомерным шкалированием (MDS). Он позволяет представить данные в виде точек в двумерном или трехмерном пространстве, сохраняя при этом расстояния между объектами.

Обоснование: Многомерное шкалирование (MDS) — это основной метод, используемый для визуализации данных в пространстве меньшей размерности, сохраняя при этом расстояния между объектами.

Вопрос 74 (выбор одного правильного ответа):

Что такое многомерный статистический анализ?

А) Метод для классификации данных на основе их сходства.
В) Метод для уменьшения размерности данных путем выявления латентных факторов.

С) Метод для прогнозирования одной переменной на основе другой.

Д) Метод для анализа взаимосвязей между несколькими переменными.

Ответ: D

Обоснование: Многомерный статистический анализ — это метод, который используется для анализа взаимосвязей между несколькими переменными. Он позволяет исследовать сложные структуры данных и выявлять скрытые закономерности.

Вопрос 75 (выбор нескольких правильных ответов):

Какие из следующих методов относятся к многомерному статистическому анализу?

А) Линейная регрессия.

В) Факторный анализ.

С) Кластерный анализ.

Д) Дискриминантный анализ.

Ответ: B, C, D

Обоснование: К многомерному статистическому анализу относятся:

Факторный анализ — метод для уменьшения размерности данных путем выявления латентных факторов.

Кластерный анализ — метод для классификации данных на основе их сходства.

Дискриминантный анализ — метод для разделения данных на группы на основе линейных комбинаций переменных.

Линейная регрессия относится к другим методам анализа данных.

Вопрос 76 (установление соответствия):

Установите соответствие между типами многомерного анализа и примерами:

Факторный анализ

Кластерный анализ

Дискриминантный анализ

Примеры:

А) Уменьшение размерности набора данных, описывающего характеристики пациентов.

В) Классификация пациентов на основе набора клинических и лабораторных показателей.

С) Анализ взаимосвязи между набором экономических показателей и набором социальных показателей.

Ответ:

Факторный анализ — А (Уменьшение размерности набора данных, описывающего характеристики пациентов).

Кластерный анализ — В (Классификация пациентов на основе набора клинических и лабораторных показателей).

Дискриминантный анализ — С (Анализ взаимосвязи между набором экономических показателей и набором социальных показателей).

Обоснование:

Факторный анализ используется для уменьшения размерности данных.

Кластерный анализ используется для классификации данных.

Дискриминантный анализ используется для анализа взаимосвязи между переменными.

Вопрос 77 (вопрос с открытым ответом):

Что такое многомерное шкалирование?

Ответ: Многомерное шкалирование — это метод, который используется для уменьшения размерности данных путем визуализации их в пространстве меньшей размерности. Он позволяет представить данные в виде точек в двумерном или трехмерном пространстве, сохраняя при этом расстояния между объектами.

Обоснование: Многомерное шкалирование помогает визуализировать данные и выявить скрытые структуры, сохраняя при этом расстояния между объектами.

Вопрос 78 (вопрос с кратким ответом):

Как называется метод, который используется для оценки значимости факторов в

факторном анализе?

Ответ: Метод, который используется для оценки значимости факторов в факторном анализе, называется тестом Бартлетта. Он позволяет определить, являются ли факторы статистически значимыми и вносят ли они существенный вклад в объяснение вариации данных.

Обоснование: Тест Бартлетта используется для проверки гипотезы о том, что все факторы незначимы. Если тест показывает, что факторы значимы, это означает, что они вносят существенный вклад в объяснение вариации данных.

Критерии оценки для проведения зачета по дисциплине (тестирование)

Шкала оценивания результатов промежуточной аттестации и критерии выставления оценок.

Оценка	Уровень сформированности компетенции	Критерии
«Отлично» («зачтено»)	Высокий	Выполнено более 80% заданий предложенного теста
«Хорошо» («зачтено»)	Хороший	Выполнено 60-80% заданий предложенного теста
«Удовлетворительно» («зачтено»)	Достаточный	Выполнено 35-60% заданий предложенного теста
«Неудовлетворительно» («не зачтено»)	Недостаточный	Выполнено менее 35% заданий предложенного теста

Обновление оценочных материалов дисциплины

обсуждено и рекомендовано к утверждению решением кафедры
наименование кафедры _____
от ____ ____ 20__ г., протокол № ____

одобрено Научно-методическим советом _____ от ____ ____ 20__ г.,
протокол № ____