

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра автоматизированных систем управления
Кафедра вычислительной техники

Паспорт зачета

по дисциплине «Методы анализа данных», 7 семестр

1. Методика оценки

Зачет проводится в письменной форме, по билетам. Билет формируется по следующему правилу: первые два вопроса выбираются из диапазона вопросов 1-20 (список вопросов приведен ниже), задача выбирается одного из типов 1-6 (типы задач приведены ниже). В ходе экзамена преподаватель вправе задавать студенту дополнительные вопросы из общего перечня (п. 4).

Таким образом, проверяются результаты обучения по дисциплине, соотнесенные с установленными в программе индикаторами достижения компетенций.

На зачете преподаватель вправе задавать студенту уточняющие и дополнительные вопросы из общего перечня (п. 4).

Форма билета для зачета

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет АВТФ

Билет № _____

к зачету по дисциплине «Методы анализа данных»

1. Вопрос 1
2. Вопрос 2.
3. Задача.

Утверждаю: зав. кафедрой _____ зав. кафедрой ВТ, Якименко А.А.
(подпись)

(дата)

2. Критерии оценки результатов обучения, соотнесенных с уровнями освоения индикаторов достижения компетенций

Ответ на билет для зачета засчитывается на **продвинутом** уровне, если студент проводит сравнительный комплексный анализ материала, выявляет проблемы, предлагает

механизмы их решения, представляет количественные характеристики определенных процессов, приводит конкретные примеры, не допускает ошибок и способен обосновать выбор метода решения задачи. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций не содержит пробелов. Установленные в программе компетенции сформированы в полном объеме. Оценка составляет *от 18 до 20 баллов*.

Ответ на билет для зачета засчитывается на **базовом** уровне, если студент при ответе на вопросы формулирует основные понятия, дает характеристику процессов, явлений, проводит анализ причин, условий, способен представить количественные и качественные характеристики процессов, не допускает существенных ошибок при решении задачи. Совокупность результатов обучения по дисциплине (модулю) и соотнесенных с ними индикаторов достижения компетенций содержит несущественные пробелы. Установленные в программе компетенции сформированы на базовом уровне. Оценка составляет *от 14 до 17 баллов*.

Ответ на билет для зачета засчитывается на **пороговом** уровне, если студент при ответе на вопросы дает определение основных понятий, может показать причинно-следственные связи явлений, при решении задачи допускает непринципиальные ошибки, например, вычислительные. Совокупность результатов обучения по дисциплине (модулю) и соотнесенных с ними индикаторов достижения компетенций содержит пробелы. Установленные в программе компетенции сформированы на пороговом уровне. Оценка составляет *от 10 до 13 баллов*.

Ответ на билет для зачета считается **неудовлетворительным** (ниже порогового уровня), если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений, при решении задачи допускает принципиальные ошибки. Совокупность результатов обучения по дисциплине (модулю) и соотнесенных с ними индикаторов достижения компетенций содержит существенные пробелы. Установленные в программе компетенции не сформированы. Оценка составляет *менее 10 баллов*.

3. Шкала оценки

Зачет считается сданным, если сумма баллов по всем заданиям билета составляет от 10 до 20 баллов включительно. Сумма менее 10 баллов признается неудовлетворительным результатом промежуточной аттестации по дисциплине.

В общей оценке по дисциплине баллы за зачет учитываются в соответствии с правилами балльно-рейтинговой системы, установленными в НГТУ.

4. Вопросы к зачету по дисциплине «Методы анализа данных»

1. Основные понятия АД, типы задач АД, примеры задач. Математический аппарат АД.
2. Основные этапы компьютерной технологии анализа данных, особенности.
3. Графический разведочный анализ данных. Методы визуализации данных на примере прикладного пакета Statistica. Достоинства и недостатки визуальных методов анализа данных.
4. Первичный разведочный анализ данных (ПРАД). Основные числовые характеристики данных. Статистическая проверка гипотез.
5. Первичный разведочный анализ данных (ПРАД). Исследование закона распределения данных. Критерии согласия.
6. Корреляционный анализ количественных переменных. Основные характеристики связи. Примеры задач АД.
7. Ранговый корреляционный анализ. Основные характеристики связи. Примеры

задач АД.

8. Анализ таблиц сопряженности. Основные характеристики связи, статистические критерии. Примеры задач АД.
9. Кластерный анализ. Постановка задачи. Примеры задач АД. Методологические этапы проведения кластерного анализа. Используемые меры расстояния между объектами и выбор меры. Достоинства и недостатки различных мер расстояний.
10. Кластерный анализ. Классификация методов кластерного анализа данных. Иерархические агломеративные методы (метод одиночной связи, метод полной связи, метод Варда), метод К-средних. Достоинства и недостатки методов кластеризации.
11. Кластерный анализ. Оценка адекватности полученных решений: кофенетическая корреляция; тесты значимости признаков на основе дисперсионного анализа; метод повторных выборок; процедура Монте-Карло.
12. Дискриминантный анализ. Постановка задачи интерпретации и классификации. Примеры задач.
13. Дискриминантный анализ. Задача классификации. Понятие классификационной функции. Прогноз принадлежности объекта к классу на основе классификационной функции. Примеры задач АД. Оценка адекватности функции классификации выборочным данным.
14. Деревья классификации (деревья решений). Примеры задач АД. Основные понятия, связанные с деревьями классификации, показатели сложности и точности деревьев решений. Методологические этапы построения дерева решений.
15. Деревья классификации (деревья решений). Выбор критерия точности прогноза. Метод CART построения дерева решений.
16. Деревья классификации (деревья решений). Определение момента прекращения ветвления, правила останова. Процедуры проверки прогнозной способности построенного дерева решений.
17. Деревья классификации (деревья решений). Выбор размера дерева, кросс-проверочное отсечение по минимальной цене сложности.
18. Компьютерные технологии решения разных типов задач АД в современном статистическом программном обеспечении на примере реальных пакетов.
19. Технологии Data Mining, основные этапы, особенности.
20. Направления развития методов, технологий и средств решения задач АД.

Типы задач

1. Выполнить кластеризацию объектов, используя метод одиночной связи. Построить дендрограмму. Построить график зависимости коэффициента слияния от количества кластеров. Сделать вывод о количестве кластеров в данных и перечислить объекты, входящие в каждый кластер.
2. Выполнить кластеризацию объектов, используя метод полной связи. Построить дендрограмму. Построить график зависимости коэффициента слияния от количества кластеров. Сделать вывод о количестве кластеров в данных и перечислить объекты, входящие в каждый кластер.
3. Выполнить кластеризацию объектов, используя метод k-средних. Перечислить объекты, относящиеся к каждому кластеру.
4. Выполнить кросс-проверочное отсечение дерева по минимальной цене сложности. Провести тестирование полученного дерева решений.
5. Рассчитать заданный коэффициент корреляции (коэффициент корреляции Спирмена, коэффициент корреляции Кендалла, коэффициент корреляции Пирсона, множественный коэффициент корреляции, коэффициент конкордации, частный коэффициент корреляции) между заданными показателями. Сделать вывод о силе и

направлении корреляционной взаимосвязи.

6. Выполнить анализ таблиц сопряженности. Оценить степень взаимосвязи между показателями на основе критерия хи-квадрат, расчета коэффициента Пирсона, Чупрова.