

Паспорт расчетно-графического задания (работы)

по дисциплине «Компьютерные технологии анализа и обработки данных», 1 семестр

1. Методика оценки

Выполнение расчетно-графического задания (работы) (далее - РГЗ(Р)) является формой текущей аттестации (контроля) по дисциплине «Компьютерные технологии анализа и обработки данных», предусмотренной учебным планом.

Цель РГЗ(Р): студенты должны освоить и научиться применять основные методы и алгоритмы, компьютерные технологии решения задач анализа и обработки данных.

Обязательным элементом РГЗ(Р) являются решение поставленной задачи анализа данных средствами R и R-Studio, анализ и интерпретация полученных результатов.

Номер задания (вариант РГЗ) назначается преподавателем.

РГЗ(Р) выполняется индивидуально.

Количество заданий достаточно для обеспечения каждого обучающегося индивидуальным заданием РГЗ(Р).

Замена задания РГЗ(Р) осуществляется по согласованию с преподавателем из числа резервных (не занятых) заданий.

Перед выполнением задания студент должен ознакомиться с учебными материалами к лекционным занятиям, рекомендованной литературой.

Преподаватель осуществляет руководство по выполнению задания, оказывает консультационную помощь и принимает отчет по РГЗ(Р).

По результатам выполнения РГЗ(Р) выполняется отчет, который состоит из следующих частей:

1. Титульный лист
2. Задание на РГЗ(Р)
3. Теоретическая часть, представляющая собой описание методов, моделей, алгоритмов решения заданной задачи анализа данных, особенностей их применения, ограничения использования
4. Практическая часть (описание хода решения заданной задачи анализа данных в среде R и R-Studio, анализ и интерпретация полученных результатов).
5. Выводы по работе в произвольной форме (сравнение R с другими средствами анализа данных, оценка рассматриваемого метода анализа данных, анализ результатов и т.п.).
6. Список использованной литературы, интернет-источников и программных средств
7. Приложение (полный код на R)

Требования к оформлению:

Объем РГЗ(З) до 25 страниц машинописного текста формата А4. Шрифт Times New Roman, 12. Формулы набираются в редакторе Math Type. Размещение сканированных формул не допускается. Нумерация страниц сквозная, в нижней части листа по центру арабскими цифрами. Работа должна быть отредактирована, не содержать

орфографических, синтаксических и стилистических ошибок.

Отчет в установленные сроки сдается на кафедру для проверки. Преподаватель оценивает качество работы, отмечает положительные стороны и недостатки работы и определяет, допускается ли она к защите. При необходимости преподаватель возвращает РГЗ(Р) студенту для доработки и устанавливает сроки повторного предоставления для проверки. До защиты работы студентом должны быть сделаны необходимые исправления и дополнения по всем замечаниям преподавателя.

При положительном результате оценивания РГЗ(Р) студент её распечатывает, передает на кафедру и защищает до сессии в назначенное преподавателем время.

Защита РГЗ(Р) состоит в индивидуальном устном собеседовании студента с преподавателем. В процессе защиты выявляется уровень знаний студента, степень его самостоятельности при выполнении работы. По результатам защиты студенту выставляется оценка в соответствии с критериями, приведенными в п. 2 настоящего Паспорта.

2. Уровни сформированности компетенций и критерии оценки

РГЗ(Р) считается выполненной **на продвинутом** уровне, если все необходимые расчеты произведены самостоятельно и без ошибок; все разделы РГЗ(Р) выполнены правильно и в полном объеме; работа оформлена в соответствии с требованиями; сдана преподавателю в указанные сроки и не возвращалась для доработки; даны полные и развернутые выводы и рекомендации; на защите студентом даны уверенные и аргументированные ответы, что свидетельствует об уровне сформированности компетенций и соотнесенных с ними индикаторов на продвинутом уровне. Оценка составляет *от 18 до 20 баллов*.

РГЗ(Р) считается выполненной **на базовом** уровне, если все необходимые расчеты произведены самостоятельно и без существенных ошибок; все разделы РГЗ(Р) выполнены правильно, но есть замечания к полноте предоставления информации; работа оформлена в соответствии с требованиями; сдана преподавателю в указанные сроки и однократно возвращалась студенту для незначительной доработки; в заключении даны выводы и рекомендации; на защите студентом допущены не принципиальные ошибки, что свидетельствует об уровне сформированности компетенций и соотнесенных с ними индикаторов на базовом уровне. Оценка составляет *от 14 до 17 баллов*.

РГЗ(Р) считается выполненной **на пороговом** уровне, если все необходимые расчеты произведены самостоятельно, но с ошибками, часть из которых носит принципиальный характер; есть замечания к полноте предоставления информации; работа оформлена в соответствии с требованиями; сдана преподавателю, но неоднократно возвращалась студенту для доработки; в заключении даны краткие выводы; защита РГЗ(Р) вызывает у студента серьезные затруднения, что свидетельствует об уровне сформированности компетенций и соотнесенных с ними индикаторов на пороговом уровне. Оценка составляет *от 10 до 13 баллов*.

РГЗ(Р) считается **не выполненной**, если расчеты произведены с серьезными ошибками; есть замечания к полноте предоставления информации и оформлению; РГЗ(Р) была сдана преподавателю, но неоднократно возвращалась студенту для доработки, что не привело к улучшению ее качества; РГЗ(Р) не допущена до защиты, что свидетельствует об не сформированности компетенций и соотнесенных с ними индикаторов. Оценка составляет *менее 10 баллов*.

3. Шкала оценки

В общей оценке по дисциплине баллы за РГЗ(Р) учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе

дисциплины.

РГЗ(Р) как форма текущей аттестации (контроля) по дисциплине считается успешно выполненной, если сумма полученных баллов по всем его заданиям составляет от 10 до 20 баллов включительно.

4. Примерный перечень тем

1. Проверка гипотез однородности (тест Стьюдента, тест Манна-Уитни) для независимых переменных
2. Проверка гипотез однородности (тест Стьюдента, тест Вилкоксона) для зависимых переменных
3. Многомерный регрессионный анализ
4. Пуассоновская регрессия
5. Логистическая регрессия
6. Пошаговый регрессионный анализ
7. Однофакторный дисперсионный анализ
8. Многофакторный дисперсионный анализ
9. Тесты однородности дисперсий (тест Барлетта, Флигнера-Форсайта, Брауна-Форсайта).
10. Анализ мощности
11. Визуализация данных в R (мозаичные диаграммы, пузырьковые диаграммы, ...)
12. Визуализация данных в R средствами пакета lattice
13. Перестановочные тесты
14. Бутстреп-анализ
15. Анализ главных компонент
16. Факторный анализ
17. Анализ пропущенных значений
18. Построение ансамбля моделей классификации (например, случайные леса)
19. Построение ансамбля моделей кластеризации
20. Построение ансамбля моделей прогнозирования временных рядов
21. Ассоциативный анализ
22. Дискриминантный анализ
23. Самоорганизующиеся карты Кохонена
24. Методы сравнения точности моделей
25. Нейросетевые методы для решения задач прогнозирования
26. Нейросетевые методы для решения задач классификации
27. Нейросетевые методы для решения задач кластеризации
28. Обработка пропущенных наблюдений во временных рядах
29. Выявление структурных изменений во временных рядах
30. Байесовские структурные модели временных рядов
31. Кластеризация временных рядов
32. Сравнение моделей. Оценка эффективности моделей. Lift и Profit кривые.
33. Сравнение моделей. Оценка эффективности моделей. ROC-анализ.
34. Деревья регрессии с многомерным откликом
35. Бинарные деревья решений (алгоритм ID3)
36. Анализ последовательностей событий
37. Метод опорных векторов
38. Критерии согласия, исследование законов распределения данных
39. Анализ текстов
40. Интеграция Statistica со средой R

Допускается выбор собственной темы по согласованию с преподавателем.

Задание (разделы) на РГЗ(Р):

1. Теоретическое описание метода анализа данных (задача, для решения которой разработан метод, основные формулы, алгоритм применения метода, требования к структуре и типу исходных данных, ограничения (условия использования) метода, достоинства, недостатки).
2. Постановка задачи, для решения которой будет использоваться рассматриваемый метод, выбор и описание исходных данных (данные должны быть реальными, взятыми из репозитория данных по машинному обучению, например, ресурс Kaggle, репозиторий UCI (UCI Machine Learning Repository), либо собственные данные).
3. Решение задачи средствами R и R-Studio. Анализ и интерпретация полученных результатов, выводы по результатам решения задачи в терминах предметной области.
4. Подготовка и оформление отчета по РГЗ(Р).