

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 12.08.2026 12:44:46
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

Приложение 4
к образовательной программе

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.05 Анализ данных

(индекс, наименование дисциплины в соответствии с учебным планом)

38.03.05 Бизнес-информатика

(код, наименование направления подготовки/специальности)

Бизнес-аналитика

(наименование образовательной программы)

Очная форма обучения

(форма обучения)

Год набора - 2026

Санкт-Петербург

Автор(ы)-составитель(и) РПД:

Кадура Елена Вячеславовна, к.пс.н., доцент кафедры бизнес-информатики

Заведующий кафедрой:

Наумов Владимир Николаевич, доктор военных наук, профессор, заведующий кафедрой бизнес-информатики

Рабочая программа дисциплины Б1.В.05 Анализ данных одобрена на заседании кафедры бизнес-информатики факультета экономики и финансов СЗИУ РАНХиГС.

протокол № 6 от «26» марта 2026 г.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Типы оценочных материалов, показатели и критерии их оценивания
5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам
6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине
7. Методические материалы по освоению дисциплины
8. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет»
9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Дисциплина Б1.В.05 Анализ данных обеспечивает формирование у обучающихся следующих универсальных, общепрофессиональных и профессиональных компетенций*:

ОТФ/ ТФ и реквизиты ПС (при наличии) **	Код компетенции **	Наименование Компетенции **	Код индикатора достижения компетенций **	Наименование индикатора достижения компетенций **	Образовательный результат **
	УК ОС-9	Способен использовать основы экономических знаний для принятия экономически обоснованных решений в различных сферах деятельности	УК ОС-9.3	Приводит экономическое обоснование принимаемых решений в различных сферах деятельности	УК-9.3. 3-1. Знает принципы экономического обоснования принимаемых решений в различных сферах деятельности; УК-9.3. 3-2. Знает теоретические и прикладные вопросы анализа данных с целью анализа, обоснования и выбора решений. УК-9.3. У-1. Умеет проводить экономическое обоснование принимаемых решений в различных сферах деятельности; УК-9.3. У-2. Умеет обрабатывать эмпирические и экспериментальные данные, осуществлять предобработку и очистку данных, выполнять разведывательный анализ.

* Дисциплина может формировать компетенцию полностью или частично.

** Должно соответствовать Приложению 1 к образовательной программе

2. Объем и место дисциплины в структуре образовательной программы

Общий объем дисциплины:

3,00 з.е., 108 ак.час

Контактная работа обучающихся с преподавателем по видам учебных занятий: 51 ак. час на контактную работу с преподавателем, из них 16 ак.час на лекции и 24 ак.час на практические занятия, 2 ак.часа на консультацию. 39 ак. час на самостоятельную работу обучающихся, 9 ак. часа на каттэк.

Б1.В.05 «Анализ данных» реализуется в 5-м семестре 3-го курса. Преподавание дисциплины «Анализ данных» опирается на знание общеобразовательных дисциплин, в первую очередь, Б1.О.07.05 «Теория вероятностей и математическая статистика», Б1.О.07.01 - «Математический анализ», Б1.В.14 «Введение в науку о данных. SQL и Python», Б1.В.16 «Язык программирования R».

«Анализ данных» предшествует таким дисциплинам, как: Б1.В.ДЭ.03.01 «Методы прогнозирования», Б1.В.ДЭ.03.02 «Прогнозирование временных рядов», а также при выполнении научно-исследовательской и выпускной квалификационной работы.

Объем дисциплины, реализуемый с применением СДО: количество академических часов, выделенных на самостоятельную работу обучающихся: всего с применением СДО – 36 а.ч.

Знания, умения и навыки, полученные при изучении дисциплины, используются студентами при подготовке и сдаче государственного экзамена.

Формой промежуточной аттестации в соответствии с учебным планом является экзамен.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Очная форма обучения

№ п/п	Наименование тем и (или) разделов	ВСЕГО	Объем дисциплины, ак.час											Форма текущего контроля успеваемости, промежуточной аттестации	
			Контактная работа обучающихся с преподавателем по видам учебных занятий								Самостоятельная работа				
			Период теоретического обучения				Период промежуточной аттестации (сессия)								
			Занятия лекционного типа		Занятия семинарского типа		ИК	КСР	КЭ	Каттэк	Контроль	СРкр	СРэк		СР
			Л	ВЛ	ЛР	ПЗ									
Тема 1	Основы анализа данных	8	2			4							2	Тестирование, Практическое контрольное задание	
Тема 2	Предобработка, очистка и разведочный анализ данных	14	2			4							8	Тестирование, Практическое контрольное задание	
Тема 3	Кластерный анализ	14	2			4							8	Тестирование, Практическое контрольное задание Контрольная работа	
Тема 4	Задачи классификации	14	2			4							8	Тестирование, Практическое контрольное задание	
Тема 5	Элементы факторного анализа	16	4			4							8	Тестирование, Практическое контрольное задание	
Тема 6	Ассоциативные	13	4			4							5	Практическое	

	правила и рекомендательные системы													контрольное задание, Контрольная работа
Промежуточная аттестация		29						2	9		18		экзамен	
Итого		108	16			24		2	9		18	39		

Используемые сокращения:

Л – лекции - занятия, предусматривающие преимущественную передачу учебной информации обучающимся педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях,).

ВЛ – видео лекции.

ЛР – лабораторные работы.

ПЗ – практические занятия (за исключением лабораторных работ).

ИК – индивидуальные консультации.

КСР – контроль самостоятельной работы

КЭ – консультации перед экзаменом

Каттэк – контактная работа на аттестацию в период экзаменационных сессий

Контроль - контактная работа на аттестацию в период экзаменационных сессий для заочной формы обучения

СРкр – самостоятельная работа на подготовку курсовой работы/ курсового проекта.

СРэк – самостоятельная работа на подготовку к экзамену.

СР – самостоятельная работа в семестре на подготовку к учебным занятиям.

Т – тестирование.

ПКЗ – практические контрольные задания.

ПИЗ – профессионально-исследовательские задания.

В процессе обучения применяются следующие интерактивные формы: лекция-диалог, работа в малых группах, спарринг-партнерство.

Темы 1-6 могут быть освоены с применением ЭО и ДОТ с контролем в системе электронного обучения Академии.

3.2. Содержание дисциплины

Тема 1. Основы анализа данных. УК ОС-9.3.

Введение. Понятие анализа данных. Задачи систем поддержки принятия решений. Жизненный цикл процесса анализа данных. Классификация методов Data Mining. Модели Data Mining. Понятие данные и знания. Типы и шкалы данных. Репозитории данных. Классические наборы данных. Задачи анализа данных. Описательная аналитика, Диагностическая аналитика. Предсказательная и предписывающая аналитика. Процесс обнаружения знаний. Машинное обучение. Основные проблемы машинного обучения. Переобучение и недообучение модели. Проблема смещения и большой дисперсии. Классификация задач Data Mining. Методы анализа данных. Задачи классификации и регрессии. Средства анализа данных и машинного обучения. Использование статистических пакетов для интеллектуального анализа данных.

Тема 2. Предобработка, очистка и разведочный анализ данных. УК ОС- 9.3.

Методология KDD. Задачи предобработки данных. Технология ETL.

Разведочный анализ данных. Модель анализа данных Тьюки. Набор данных Анскомба. Просмотр данных. Очистка данных. Оценка качества данных. Заполнение пропущенных данных. Аномальные и предельные данные. Использование ящичной диаграммы. Выявление дубликатов и противоречий. Корреляционный анализ. Трансформация данных. Квантование данных (биннинг). Сэмплинг. Повторные выборки. Бутстреп. Кроссвалидация. Перестановочный тест. Метод «складного ножа». Группировка данных. Решение задач предобработки и очистки данных в R (Python).

Решение задач описательной статистики. Графический анализ данных. Статистические диаграммы.

Основные положения непараметрической и нечисловой статистики. Таблицы сопряженности. Таблица сопряженности 2x2. Таблицы флагов и заголовков. Непараметрические и нечисловые критерии. Канонический анализ. Корреляционная матрица. Коэффициенты канонической корреляции. Меры избыточности переменных. Задачи ассоциации.

Тема 3. Кластерный анализ. УК ОС- 9.3.

Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Дендрограммы. Метод К-средних. Профили кластеров. Оценка качества кластерного анализа. Диаграмма «локтя». Диаграмма силуэтов. Кластеризация на графах. Метод DBSCAN. Использование пакета Logitom для решения задач кластерного

анализа. Кластерный анализ в средствах интеллектуального анализа Microsoft Office (на R, JASP, Python).

Тема 4. Задачи классификации. УК ОС-9.3.

Формулировка задачи классификации. Классификационный анализ с обучением. Метод k-ближайших соседей. Наивный байесовский классификатор. Логистическая регрессия. Деревья решений. Метод опорных векторов. Использование нейронных сетей для решения задач классификации. Ансамблевые методы классификации. Сравнение результатов классификации различными методами. Верификация. Оценка качества классификации. ROC-кривая. Показатель AUC. Таблица сопряженности (матрица путаницы). Показатели точности: accuracy, recall, precision, F1. Проблема сбалансированности данных.

Примеры алгоритмов построения деревьев решений. Использование статистических пакетов Loginom, JASP, Excel (R, Python) для построения деревьев решений.

Тема 5. Элементы факторного анализа. УК ОС-9.3.

Агрегирование и группирование данных. Понятие проекции. Принцип ординации наблюдений. Задачи снижения размерности данных. Метод главных компонент. Организация решения задачи методом главных компонент. Матрица факторных нагрузок. Собственные числа. Критерий КМО. Критерий Кайзера, Критерий каменистой осыпи. Объясненная совокупная дисперсия. Факторный анализ. Основные этапы факторного анализа. Общность и характерность переменных. График факторных нагрузок. Методы вращения. Решение задач факторного анализа с помощью Rstudio, Anaconda navigator (Jupyter Notebook). Разведочный и подтверждающий методы факторного анализа.

Тема 6. Ассоциативные правила и рекомендательные системы. УК ОС-9.3.

Ассоциативные правила. Поддержка и достоверность ассоциативных правил. Лифт. Алгоритмы построения ассоциативных правил. Рекомендации по генерации правил. Алгоритм apriori. Использование пакета Deductor, Loginom, Orange для построения ассоциативных правил. Рекомендательные системы. Коллаборационные рекомендательные системы. Item-based и User-based.

4. Типы оценочных материалов, показатели и критерии оценивания

4.1. Оценочные материалы по дисциплине Б1.В.05 Анализ данных входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам образовательной программы составляет фонд оценочных средств (далее – ФОС). ФОС используется при проведении текущего контроля успеваемости и промежуточной аттестации обучающихся с целью оценивания достижения обучающимися планируемых результатов обучения.

4.2. ФОС разработан как комплекс проверочных заданий различного типа и уровня сложности, включает критерии и шкалы оценивания, а также «ключи» правильных ответов. ФОС формируется как отдельный документ и хранится в электронном виде, доступ к ФОС предоставлен ограниченному кругу лиц.

4.3. Для самостоятельной работы обучающихся при подготовке к текущему контролю успеваемости и промежуточной аттестации в рабочих программах дисциплин размещены типовые проверочные задания, которые можно условно разделить на задания закрытого, комбинированного и открытого типов.

Задания закрытого типа — это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных.

Задания комбинированного типа – это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных и обосновать свой выбор.

Задания открытого типа — это задания, в которых на каждый вопрос должен быть предложен развернутый обоснованный ответ.

В зависимости от типа задания рекомендованы определенная последовательность выполнения и система оценивания выполнения заданий.

4.4. Типы заданий, сценарии выполнения, критерии оценивания

ТИП ЗАДАНИЯ	ИНСТРУКЦИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	КРИТЕРИИ ОЦЕНИВАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	Прочитайте текст, выберите правильный ответ	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	Ответ считается верным, если правильно указана цифра или буква
Задание закрытого типа на установление соответствия	Прочитайте текст и установите соответствие	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	Ответ считается верным, если правильно указаны цифры или буквы
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	Прочитайте текст, выберите правильные ответы	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г).	Ответ считается верным, если правильно установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого)
Задание закрытого типа на установление последовательности	Прочитайте текст и установите последовательность	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов.	Ответ считается верным, если правильно указана вся последовательность цифр

		2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).	
Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).	Ответ считается верным, если правильно указана цифра или буква и приведены корректные аргументы, используемые при выборе ответа
Задание открытого типа с развернутым ответом	Прочитайте текст и запишите развернутый обоснованный ответ	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ	Ответ считается верным: 1. Отсутствие фактических ошибок. 2. Раскрытие объема используемых понятий (полнота ответа). 3. Обоснованность ответа (наличие аргументов). 4. Логическая последовательность излагаемого материала.

4.5. Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Итоговая балльная оценка	Традиционная система	Бинарная система	ECTS	
			Для традиционной системы	Для бинарной системы
95-100	Отлично	Зачтено	A	P/ Passed
85-94			B	P/ Passed
75-84	Хорошо		C	P/ Passed
65-74			D	P/ Passed
55-64	Удовлетворительно		E	P/ Passed
0-54	Неудовлетворительно	Не зачтено	F	F/Failed

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
60 баллов	40 баллов	100 баллов	100 баллов

5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

5.1. В ходе реализации дисциплины используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

Т – тестирование, ПКЗ – практические контрольные задания, КР – контрольная работа.

Тема 1. Основы анализа данных.

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
2. Внимательно прочитать предложенные вариант-ты ответа.
3. Выбрать один верный ответ.
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В)..

1. OLAP-кубы. Для чего используют OLAP-кубы?

1. Для снижения размерности задачи
2. Для исключения избыточности
3. Для хранения многомерных данных
4. Для хранения агрегированных данных
5. Для хранения нормализованных данных
6. Для работы с потоком транзакций
7. Для работы с SQL-запросами

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитайте предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

2. Магический квадрант Гартнера. Выберите квадранты, которые входят в магический квадрант Гартнера.

1. Лидеры
2. Претенденты
3. участники (активисты)
4. провидцы (визуализаторы)
5. нишевые игроки
6. конкуренты (активисты)
7. вендоры
8. кандидаты (инноваторы)

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитайте предложенные варианты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

3. Модели данных OLAP

Какие схемы используются при построении многомерной модели данных OLAP?

1. Иерархическая
2. индексных списков
3. звезда
4. нормализованная схема "сущность -отношение"
5. снежинка
6. цепочка
7. сетевая

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитайте предложенные варианты-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

4. Операции над размерностями. Какие операции выполняются с размерностями OLAP-куба?

1. Сечение

2. Транспонирование

3. Срез

4. детализация и консолидация

5. фильтрация

6. запрос

7. запрос на обновление

8. удаление

ПКЗ по теме 1:

Использование пакета QlikView и Power BI для решения задач анализа данных о демографической ситуации в России. Для каждого варианта приведены таблицы с указанием вида исходных данных, которые будут анализироваться средствами бизнес-аналитики.

Вариант	год	область	регион	городское население	заплата	миграция	Млад. Смертность	Рождаемость	Смертность	население	осн. Фонды	преступления	сельское хозяйство	Трудоспособное население	безработные
1	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
2	+	+	+	-	+	+	-	+	-	+	+	-	-	+	+
3	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
4	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
5	+	+	+	+	-	+	-	-	+	+	-	+	-	+	-
6	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
7	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
8	+	+	+	-	+	-	+	+	-	+	+	-	+	+	+
9	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
10	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
11	+	+	+	+	-	-	+	-	+	+	-	+	+	+	-
12	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
13	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
14	+	+	+	-	+	+	-	+	-	+	+	-	-	+	+
15	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
16	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
17	+	+	+	+	-	+	-	-	+	+	-	+	-	+	-
18	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-

19	+	+	+	+	-	-	+	+	+	+	-	+	+	+	+
20	+	+	+	-	+	-	+	+	-	+	+	-	+	+	+
21	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+
22	+	+	+	-	+	+	-	-	-	+	+	-	-	+	-
23	+	+	+	+	-	-	+	-	+	+	-	+	+	+	-
24	+	+	+	-	+	-	+	-	-	+	+	-	+	+	-
25	+	+	+	+	-	+	-	+	+	+	-	+	-	+	+

Тема 2. Предобработка, очистка и разведочный анализ данных.

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные вариант-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

1. Гистограммы распределения

Поставить соответствие между заданными последовательностями и гистограммами распределения, если последовательности представляют собой случайные числа, полученные с помощью различных генераторов с разными законами распределения:

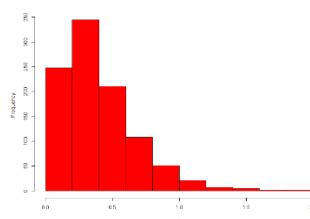
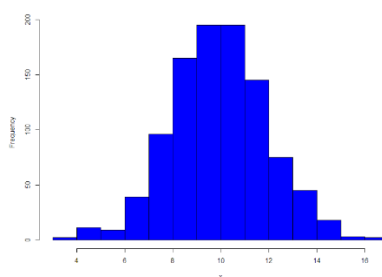
`set.seed(1234)`

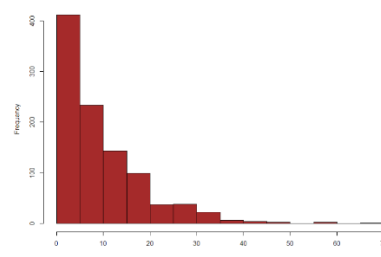
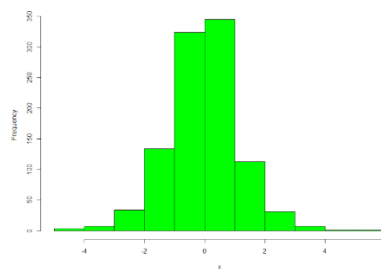
`x1<-rnorm(1000,10,2)`

`x2<-rgamma(1000,2,5)`

`x3<-rt(1000,8)`

`x4<-rexp(1000,0.1)`





Варианты ответов:

- a. X1
- b. X2
- c. X3
- d. X4

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитайте предложенные варианты-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или B).

2. Критерий проверки гипотезы о нормальном законе распределения

Выберите из предложенного списка статистических критериев критерий, который позволяет проверить гипотезу о нормальном законе распределения

A. Колмогорова-Смирнова

B. Пирсона (Chi квадрат)

C. Стьюдента

D. Фишера

E. Шапиро

F. Шапиро-Пятецкого

G. Манна-Уитни

H. Андерсона-Дарлинга

ПКЗ по теме 2:

Задание 1. Решить задачу разведочного анализа для набора данных Boston в dataset Orange данный набор данных называется Housing.

Для решения задачи в excel, R, JASP имеются данные Boston. В python использовать `sklearn.datasets.load_boston`

Задание 2. Набор данных имеет пропуски и аномальные значения

X	Y
1	1
-1	-3
3	4
2	2
5	7
2	2
3	4
5	7
1	1
3	4
4	25
4	5
5	7
1	1
3	4
3	4
2	2
4	6
0	-1
3	4
4	5
7	10
1	1
7	10
3	4
3	4
3	3
4	5
4	19
4	6
1	0
-4	-23
4	6
1	1
2	2
5	7
1	39
4	5
5	7
7	11
4	6
6	8
4	5
2	3
4	6
1	1
5	7
5	7
5	7

3	4
-1	-3
4	6
8	12
3	4
6	9
6	9
4	5
5	7
8	12
3	4
4	6
3	4
3	4
6	9
1	1
4	
4	6
4	5
5	7
5	7
1	1
2	2
2	2
6	9
4	5
2	2
5	
1	1
2	2
1	1
5	7
5	7
2	2
-1	
3	3
3	4
4	5
7	10
6	8
5	7
3	4
4	6
4	5
6	9
5	7
5	
4	5
4	5
1	0

1	1
---	---

- Проверить гипотезу о нормальном законе распределения для переменных X, Y
- Решить задачу восстановления пропущенных данных;
- Найти аномальные значения.
- Проанализировать корреляцию между переменными. Проверить гипотезу о значимости коэффициента корреляции

3. Значимость коэффициента корреляции

Чему равен уровень значимости при принятии гипотезы о наличии корреляции Пирсона для двух случайных величин/ Ответ дать с точностью до двух знаков после запятой

Таблица
наблюдений

x	y
0	3
2	4
3	5
1	3
4	4
5	3

4. Интерквартильный размах

Последовательность задана с помощью предложения

$x < -c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)$

Чему равен интерквартильный размах для данной последовательности, если все пропуски заменены средним значением имеемых данных. Ответ дать с точностью до целых

5. Корреляционный анализ

Чему равен коэффициент корреляции Спирмена для случайных величин X, Y

Таблица
наблюдений

x	y
0	3
2	4
3	5
1	3
4	4
5	3

Ответ дать с точностью до двух знаков после запятой

6. коэффициент корреляции

Чему равен коэффициент корреляции Пирсона для случайных величин

Таблица
наблюдений

x	y
0	3
2	4
3	5
1	3
4	4
5	3

7. Критерий Шапиро-Уилка

Сгенерировать случайную выборку, размером в 500 наблюдений, используя генератор нормально распределенных случайных чисел с математическим ожиданием 10, дисперсией 16. При этом начальное значение генератора случайных чисел равно 4321. Чему равен уровень значимости при проверке гипотезы о нормальном законе распределения с помощью критерия Шапиро-Уилка. Ответ дать с точностью до двух знаков после запятой. В качестве альтернативной использовать двустороннюю гипотезу.

8. Проверка гипотезы о законе распределения

Сгенерировать случайную выборку, размером в 500 наблюдений, используя генератор нормально распределенных случайных чисел с математическим ожиданием 10, дисперсией 16. При этом начальное значение генератора случайных чисел равно 4321. Чему равно наблюдаемое значение критерия Колмогорова-Смирнова. Ответ дать с точностью до двух знаков после запятой. В качестве альтернативной использовать двустороннюю гипотезу.

9. Проверка гипотезы о значимости отличий

Имеется выборка значений двух случайных величин

x	y
0	3
2	4
3	5
1	3
4	4
5	3

Проверить гипотезу о равенстве математических ожиданий при допущении о неравных дисперсии при допущении о независимости данных величин.

Чему равно наблюдаемое значение критерия Стьюдента. Ответ дать с точностью до одного знака после запятой.

10. Проверка гипотезы о равенстве дисперсии

Данные заданы двумя последовательностями

$x \leftarrow (0, 2, 3, 1, 4, 5)$

$y \leftarrow (3, 4, 5, 3, 4, 3)$

Чему равно наблюдаемое значение критерия Фишера при проверке

гипотезы о равенстве дисперсий? Ответ дать с точностью до двух знаков после запятой

11. Пропуски данных

Последовательность задана с помощью предложения

`x<-c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)`

Чему равно среднее данной последовательности, если все пропуски заменены медианным значением. Ответ дать с точностью до двух знаков после запятой

12. Пропуски данных

Последовательность задана с помощью предложения

`x<-c(2,3,4,NA,3,4,5,NA,5,7,13,NA,26,33, 17, NA, NA, NA, 14,28,36)`

Чему равна медиана данной последовательности, если все пропуски заменены средним значением имеемых данных. Ответ дать с точностью до двух знаков после запятой

Тема 3. Кластерный анализ

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитайте предложенные варианты-ты ответа.

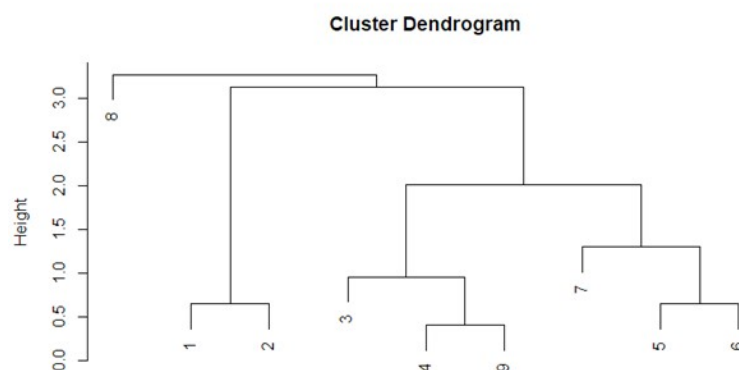
3. Выбрать один верный ответ.

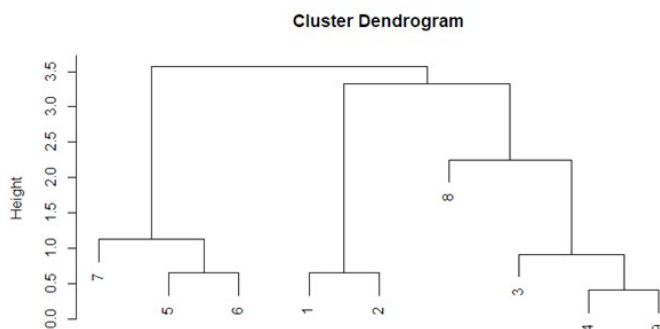
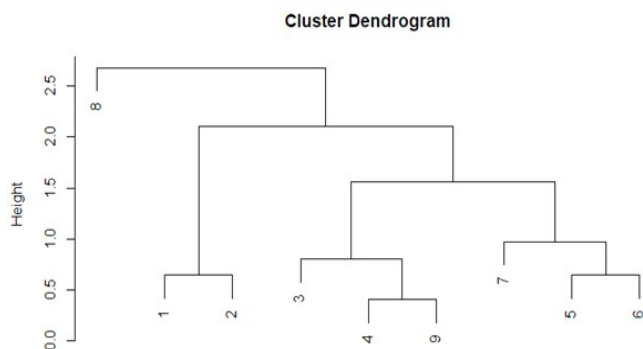
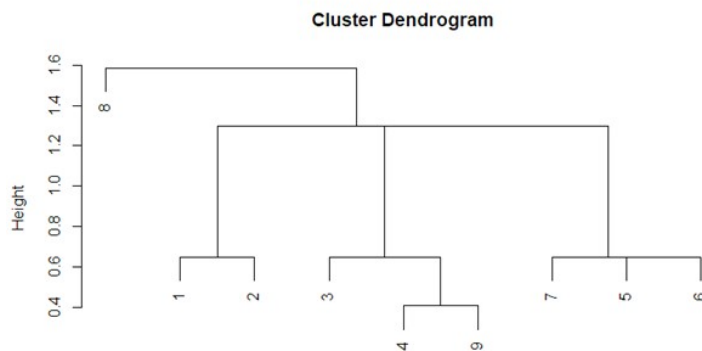
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

1. Дендрограммы. Поставить соответствие между дендрограммами и используемыми методами для наблюдений, признаки которых заданы последовательностями

`x<-c(2,3,4,5,6,7,8,3,5)`

`y<-c(3,2,5,6,3,4,5,10,7).`





Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

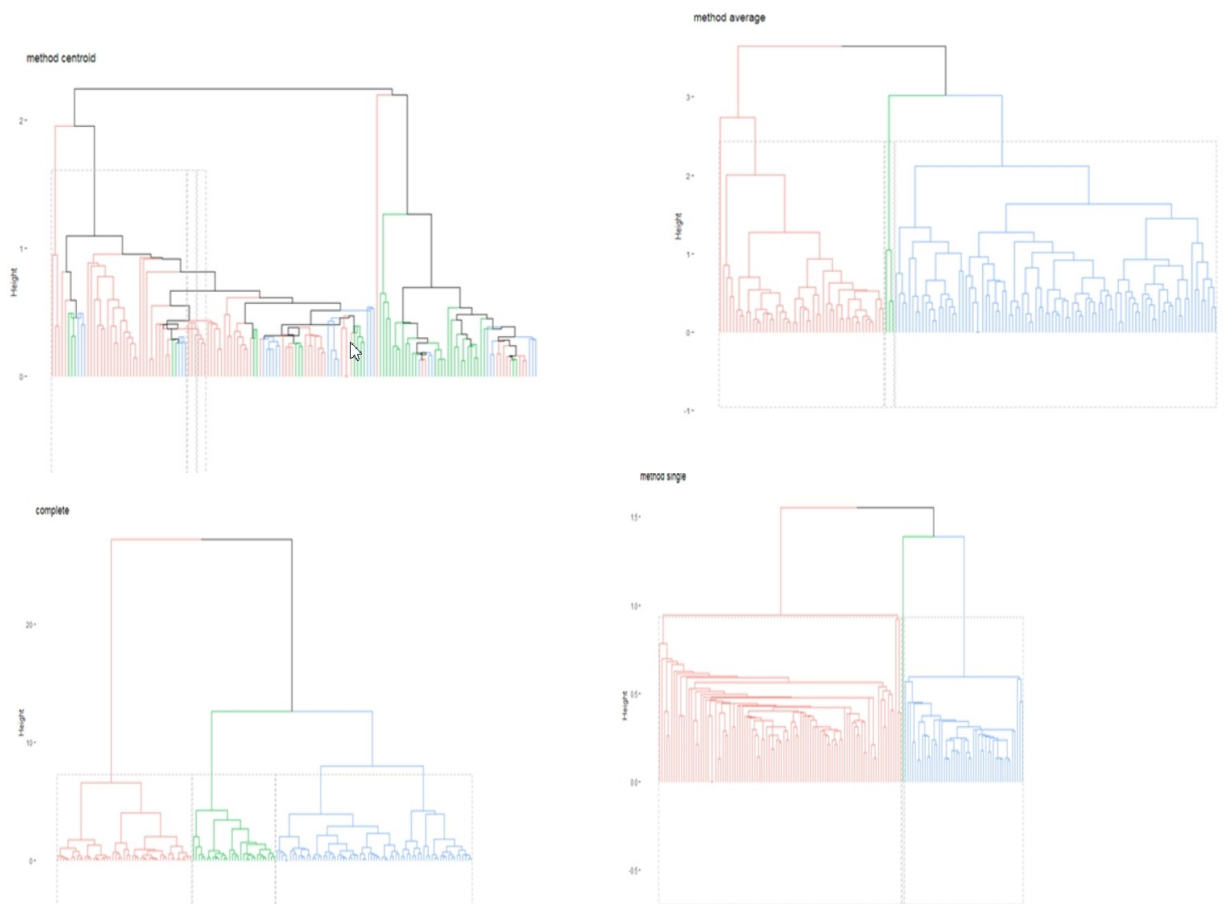
1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные вариант-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

2. Иерархический кластерный анализ. При решении задачи кластерного анализа для набора данных iris различными методами построены дендрограммы с выделением кластеров прямоугольниками для заданного числа кластеров равного трем. Выберите лучший из методов, если дендрограммы имеют следующий вид



- ближнего соседа
- дальнего соседа
- центроидный
- средней связи
- взвешенной средней связи

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитать предложенные вариант-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

3. Методы иерархической кластеризации. Какой из методов кластерного анализа при анализе кандидатов на включения в кластер на текущем шаге использует результаты оценки дисперсий?

1. метод Варда
2. метод ближайшего соседа
3. метод полной связи
4. центроидный метод
5. метод невзвешенного попарного среднего

ПКЗ по теме 3:

Задание 1.

Имеется выборка данных о 6 предприятиях. Найти расстояние между объектами с помощью различных метрик

	X1	X2	X3
1	120	9100	11000
2	180	8400	16000
3	840	13000	20000
4	410	11300	16000
5	460	12000	15000
6	560	11500	13000

С помощью принципа "ближайшего соседа" для метрики Евклида построить дендрограмму. Задачу решить в различных приложениях. Построить workflow-diagram Orange

Задание 2.

На предприятии существуют 5 отделов. Поскольку в них имеется разное число сотрудников, разные виды деятельности и др. решено сгруппировать отделы. Решить задачу группирования иерархическим и методом k-средних.

	Стоимость производственных фондов, X1	Среднемесячный объем работ, X2
1	699	190
2	510	210
3	340	110
4	290	95
5	310	130

Задачу решить в R, JASP, Python

Задание 3.

Решить задачу кластерного анализа для предложенного набора данных. Имеется 30 наборов данных, размещенных в Moodle. Каждый студент решает задачу кластерного анализа для своего варианта исходных данных. При решении задачи кластерного анализа следует решать задачи иерархического кластерного анализа и кластерного анализа, решенного с помощью метода k-средних. В Moodle находится пример решения задачи и оформления отчета. В примере приведены скрипты, позволяющие решить данные задачи на языке R.

Контрольная работа:

Вариант 1

1. Внутрикластерная дисперсия. Используя RStudio решить задачу кластерного анализа методом k-средних для набора данных iris. При решении задачи использовать начальное значение генератора случайных чисел 1234. Задать максимальное число итераций равное 10. Число кластеров задать равным трем. Перед решением задачи выполнить стандартизацию значений параметров с помощью функции scale. Кластеризацию выполнять по четырем переменным набора данных (длине и ширине чашелистика и лепестка). Чему равно суммарное значение внутрикластерной дисперсии для всех трех кластеров?

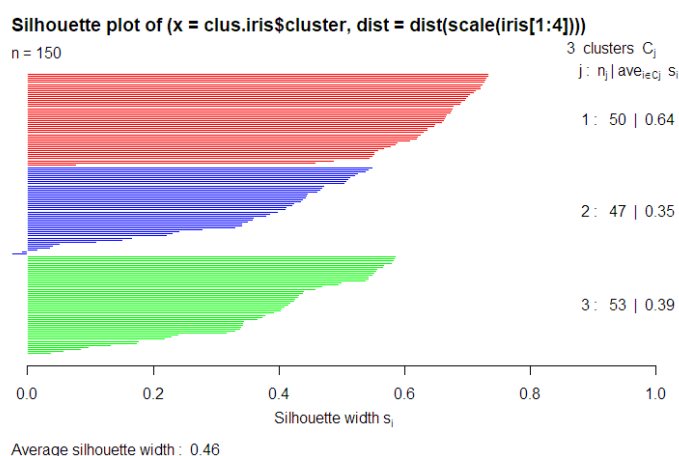
2. Дендрограммы. Набор данных задан двумя последовательностями значений признаков.

$x \leftarrow c(2, 3, 4, 5, 6, 7, 8, 3, 5)$

$y \leftarrow c(3, 2, 5, 6, 3, 4, 5, 10, 7)$.

Чему равно значение расстояния между четвертым и девятым наблюдениями, если все данными были стандартизованы с помощью функции `scale`, а также при построении дендрограммы использовать метод полной связи (дальнего соседа)? Ответ дать с точностью до двух знаков после запятой

3. Диаграмма силуэтов. После решения задачи кластерного анализа для набора данных `iris` получена диаграмма силуэтов, которая имеет вид:



Укажите номер кластера, качество формирования которого наилучшее при ее оценке методом силуэтов? Номер лучшего кластера указать числом

Тема 4. Задачи классификации

Тестовые задания:

Задание открытого типа с написанием правильного ответа.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внести верный ответ в поле ответа.

3. Записать значение цифрами, если в ответе вещественное число, целую часть от дробной отделить точкой, или запятой.

1. Деревья решений. В результате построения дерева решений при решении задачи классификации ирисов получена такая характеристика узлов:

```
library(tree)
```

```
iris.tree <- tree(iris[,5] ~ ., iris[, -5])
```

```
iris.tree
```

```
1) root 150 329.600 setosa ( 0.33333 0.33333 0.33333 )
```

```
2) Petal.Length < 2.45 50 0.000 setosa ( 1.00000 0.00000 0.00000 ) *
```

```
3) Petal.Length > 2.45 100 138.600 versicolor ( 0.00000 0.50000 0.50000 )
```

- 6) Petal.Width < 1.75 54 33.320 versicolor (0.00000 0.90741 0.09259)
 12) Petal.Length < 4.95 48 9.721 versicolor (0.00000 0.97917 0.02083)
 24) Sepal.Length < 5.15 5 5.004 versicolor (0.00000 0.80000 0.20000) *
 25) Sepal.Length > 5.15 43 0.000 versicolor (0.00000 1.00000 0.00000) *
 13) Petal.Length > 4.95 6 7.638 virginica (0.00000 0.33333 0.66667) *
 7) Petal.Width > 1.75 46 9.635 virginica (0.00000 0.02174 0.97826)
 14) Petal.Length < 4.95 6 5.407 virginica (0.00000 0.16667 0.83333) *
 15) Petal.Length > 4.95 40 0.000 virginica (0.00000 0.00000 1.00000) *

Для шестого узла поставьте соответствие значений параметров узла и их названий

- 54
- 33.320
- 0.00000
- 0.90741
- 0.09259

Варианты ответов;

- Вероятность третьего класса в узле;
- Вероятность первого класса в узле;
- Количество элементов в узле;
- Вероятность второго класса в узле;
- Доля правильно классифицированных в узле;
- Доля неверной классификации в узле

Задание открытого типа с написанием правильного ответа.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внести верный ответ в поле ответа.

3. Записать значение цифрами, если в ответе вещественное число, целую часть от дробной отделить точкой, или запятой.

2. Выбор классификатора. Использование различных методов классификации позволяет получить следующие таблицы сопряженности.

Метод линейного дискриминантного анализа

69535	442
4546	477

Метод k ближайших соседей при k=9

69938	39
4973	50

Логистическая регрессия

69770	207
5005	18

Метод опорных векторов

69882	103
4890	125

Выбрать лучший метод по показателю precision

- метод опорных векторов
- метод k-ближайших соседей
- логистическая регрессия
- линейный дискриминантный анализ

Задание открытого типа с написанием правильного ответа.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внести верный ответ в поле ответа.

3. Записать значение цифрами, если в ответе вещественное число, целую часть от дробной отделить точкой, или запятой.

3. Использование различных методов классификации позволяет получить следующие таблицы сопряженности.

Метод линейного дискриминантного анализа

69535	442
.	477

Метод k ближайших соседей при k=9

69938	39
4973	50

Логистическая регрессия

69770	207
5005	18

Метод опорных векторов

69882	103
4890	125

Выбрать лучший метод по показателю accuracy

- метод опорных векторов
- метод k-ближайших соседей
- логистическая регрессия
- линейный дискриминантный анализ
- Деревья решений.

ПКЗ по теме 4:

Задание 1. Построить дерево решений по данным, приведенным в таблице.

Рейтинг	Возраст	Уровень Дохода	Образование
0	35	3000	0

0	25	5000	1
0	31	7000	1
1	56	1000	0
1	62	1100	1
1	49	1500	0

Тема 5. Элементы факторного анализа.

Тестовые задания:

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

2. Внимательно прочитайте предложенные варианты-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

1. Этапы факторного анализа

Поставьте соответствие между номерами этапов факторного анализа и их содержанием

1. Выбор метода факторного анализа

2. Выбор числа компонент (факторов)

3. Построение матриц и определение значений компонент (факторов) для каждого наблюдения

4. Вращение компонент (факторов)

5. Интерпретация полученных компонент (факторов)

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.

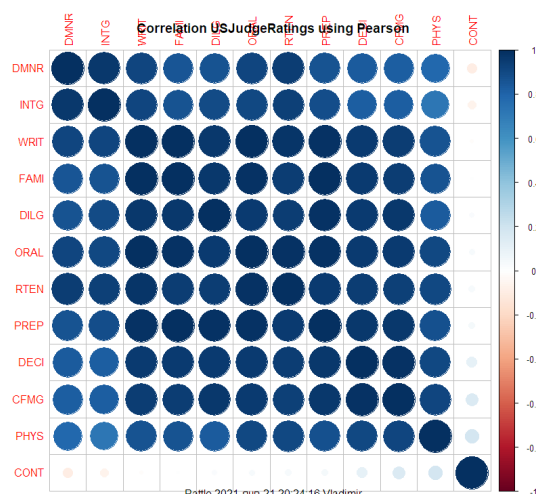
2. Внимательно прочитайте предложенные варианты-ты ответа.

3. Выбрать один верный ответ.

4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).

2. Корреляционная матрица

Диаграмма корреляционной матрицы Пирсона имеет вид



Выберите правильный ответ

1. эффективность факторного анализа невысока
2. следует выполнять факторный анализ. Однако до этого необходимо выбрать подмножество анализируемых переменных
3. Целесообразно решить задачу факторного анализа
4. Для принятия решения о целесообразности факторного анализа необходимо использовать непараметрические критерии корреляции
5. Факторный анализ не целесообразен. Нужно попробовать решить задачу методом главных компонент

ПКЗ по теме 5:

Задание 1. Решить задачу факторного анализа для набора данных ausland.sav. Данные представляют собой результаты анкетирования, проведенного Институтом Социологии Университета Марбург. На основе разработанной анкеты на двух гессенских металлургических предприятиях было произведено исследование отношения к иностранцам. Опрашиваемым предложили высказать свое отношение к следующим пятнадцати положениям:

- 1. Необходимо улучшить интеграцию иностранцев.
- 2. Необходимо мягче относиться к беженцам.
- 3. Деньги Германии должны быть потрачены на нужды страны.
- 4. Германия — это не служба социальной помощи для всего мира.
- 5. Необходимо стараться налаживать хорошие отношения друг с другом.
- 6. Права беженцев следует ограничить.
- 7. Немцы станут меньшинством.
- 8. Право беженцев необходимо охранять во всей Европе.
- 9. Враждебность к иностранцам наносит вред экономике Германии.
- 10. Сначала необходимо создать нормальные жилищные условия для немцев.
- 11. Мы ведь тоже практически везде являемся иностранцами.
- 12. Мультикультура означает мультикриминал.
- 13. В лодке нет свободных мест.
- 14. Иностранцы вон.
- 15. Интеграция иностранцев — это убийство нации.

Оценки ставились по семибальной шкале: от полного несогласия (1) до полного согласия (7). Результаты опроса для 90 человек хранятся в файле [ausland.sav](#) в переменных *a1* - *a15*. Пример приведен в самоучителе JASP.

Отметим, что данные приведены на немецком языке. При выполнении задания следует данные представить на русском языке.

Тема 6. Ассоциативные правила и рекомендательные системы.

ПКЗ по теме 6:

Задание 1

Построить ассоциативные правила по имеемым транзакциям. Рассчитать характеристики для каждого правила.

Транзакционная база данных	
TID	Приобретенные покупки
100	ремень, женская сумка, портмоне
200	женская сумка, косметичка
300	женская сумка, ремень, ключница, портмоне
400	дамский зонт, ключница, косметичка
500	ремень, женская сумка, портмоне, ключница
600	косметичка, портмоне
700	ремень, портфель

Контрольная работа:

Вариант 1

Построить дендрограмму, используя Евклидово расстояние и метод "дальнего соседа". Перед построением кластеров выполнить стандартизацию значений атрибутов

Номер объекта	x1	x2
1	3,00	10,00
2	4,00	11,00
3	6,00	10,00
4	10,00	9,00
5	11,00	9,00
6	10,00	7,00

5.2. Типовые оценочные материалы для текущего контроля успеваемости обучающихся (вне контрольных точек):
приведены в п.6.2.

5.3. Один или несколько тематических блоков дисциплины завершаются контрольной точкой (далее – КТ). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КТ в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КТ составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КТ в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной точки	Максимальное количество баллов за работу в рамках КТ, которое может набрать студент	Коэффициент веса контрольной точки	Результат контрольной точки, участвующий в формировании итоговой балльной оценки по дисциплине (отражается в журнале БРС в СДО)
КТ - 1	100	0,12	12
КТ - 2	100	0,12	12
КТ- 3	100	0,12	12
КТ - 4	100	0,12	12
КТ - 5	100	0,12	12
КТ - 6	100	0,12	12
Итого:	x	0,6	60

Формула расчета результата контрольной точки:

Результат контрольной точки = Количество баллов за работу в рамках КТ X Коэффициент веса контрольной точки.

5.4.Формы текущего контроля успеваемости обучающихся в рамках КТ и типовые оценочные материалы:

КТ-1

Тема 1.

Тестирование.

КТ-2

Тема 2.

Практическое-контрольное задание (ПКЗ).

КТ-3

Тема 3.

Практическое-контрольное задание (ПКЗ).

Контрольная работа

КТ-4

Тема 4.

Тестирование.

Практическое-контрольное задание (ПКЗ).

КТ-5

Тема 5.

Тестирование.

Практическое-контрольное задание (ПКЗ).

КТ-6

Тема 6.

Практическое-контрольное задание (ПКЗ).

Контрольная работа.

Для каждой формы текущего контроля успеваемости обучающихся в рамках КТ определены критерии оценивания результатов выполнения задания.

1. Критерии оценивания тестирования:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Количество правильных ответов</i>	<i>0</i>	<i>Количество правильных ответов менее 55%</i>
	<i>25</i>	<i>Количество правильных ответов от 55% до 64%</i>
	<i>50</i>	<i>Количество правильных ответов от 65% до 74%</i>
	<i>75</i>	<i>Количество правильных ответов от 75% до 84%</i>
	<i>100</i>	<i>Количество правильных ответов от 85% до 100%</i>
Итого максимально:	100	

2. Критерии оценивания ПКЗ:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Содержание и раскрытие выбранных понятий</i>	<i>41-70</i>	<i>Детальное, последовательное описание всех понятий на примере выбранной системы</i>
	<i>21-40</i>	<i>Поверхностное описание без привязки к выбранной системе</i>
	<i>0-20</i>	<i>Понятия раскрыты минимально или не раскрыты вовсе</i>
<i>Количество выполненных заданий</i>	<i>30</i>	<i>Количество выполненных заданий от 85% до 100%</i>
	<i>15</i>	<i>Количество выполненных заданий от 55% до 84%</i>
	<i>0</i>	<i>Количество выполненных заданий менее 55%</i>
Итого максимально:	100	

3. Критерии оценивания КР:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Содержание и раскрытие выбранных понятий</i>	<i>31-50</i>	<i>Детальное, последовательное описание всех понятий на примере выбранной системы</i>
	<i>16-30</i>	<i>Поверхностное описание без привязки к выбранной системе</i>

	0-15	Понятия раскрыты минимально или не раскрыты вовсе
Достоверность и актуальность информации	16-20	Представленная информация подтверждена ссылками на источники
	0-15	Представленная информация частично подтверждена ссылками на источники или не подтверждена
Количество выполненных заданий	30	Количество выполненных заданий от 85% до 100%
	15	Количество выполненных заданий от 55% до 84%
	0	Количество выполненных заданий менее 55%
Итого максимально:	100	

5.5. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Для решения задач открытого типа (ПКЗ, КР), тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных. Для построения интеллект-карты и моделей в различных нотациях студенту можно использовать любой соответствующий онлайн-инструмент.

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

6.1. Промежуточная аттестация проводится в форме экзамена.

Экзамен проводится в письменной форме. Обучающийся получает экзаменационный билет с вариантами 3-х заданий различного типа. На выполнение заданий дается 40-60 минут. По завершении подготовки необходимо представить ответы в письменном виде, подробно изложив ход выполнения задания, сделать выводы (*при необходимости*).

6.2. Типовые оценочные материалы промежуточной аттестации.

Вопросы для подготовки к экзамену.

1. Дать характеристику систем поддержки принятия решений, хранилищ данных.
2. Сформулировать свойства OLAP и OLTP-систем, найти их отличия.
3. Дать определение технологии KDD. Охарактеризовать этапы анализа данных KDD.
4. Объяснить содержание основных элементов математической статистики, используемых для анализа данных. Дать характеристику операций агрегирования данных.

5. Охарактеризовать содержание начальных этапов KDD, предобработки, очистка и трансформации данных.
6. Сделать обзор основного содержания разведывательного анализа, содержания модели Тьюки. Указать основные задачи разведывательного анализа.
7. Дать определение понятия аномалии. Выполнить характеристику методов борьбы с аномалиями. Дать характеристику ящичных диаграмм. Привести примеры.
8. Дать определение повторной выборки. Указать методы повторной выборки и организации их использования.
9. Назвать основные графические средства анализа. Характеризовать организацию построения гистограмм и вероятностных графиков, а также их использования при разведывательном анализе.
10. Описать организацию проверки гипотез о законах распределения. Привести примеры проверки гипотез в R.
11. Описать организацию проверки гипотез с использованием Т-критерия. Привести примеры проверки гипотез в R.
12. Определить гистограмму распределения и диаграмма «ящик с усами». Описать их использование при проверке гипотезы о законе распределения.
13. Охарактеризовать язык R. Выполнить обзор его основных возможностей.
14. Охарактеризовать графическую среду RStudio. Привести примеры решения простейших задач с помощью данной графической среды
15. Дать общую характеристика JASP. Привести примеры решения задач описательной статистики.
16. Определить задачу кластерного анализа. Дать общую характеристику методов кластерного анализа.
17. Привести примеры метрик кластерного анализа.
18. Дать характеристику методов определения близости между кластерами. Привести примеры решения задач определения близости.
19. Объяснить содержание иерархической кластеризации. Охарактеризовать агломеративный и дивизимный методы. Привести примеры в R.
20. Характеризовать метод к-средних. Привести примеры решения задач в JASP и в R.
21. Дать характеристику метода к-средних, методов определения числа кластеров, метода локтя.
22. Определить структуру ассоциативных правил, понятия антимонотонности.
23. Определить метрики построения ассоциативных правил.
24. Характеризовать алгоритм построения ассоциативных правил a’priori, указать на параметры, используемые при построении правил.
25. Дать определение деревьев решений. Дать общую характеристика деревьев решений.
26. Сделать обзор алгоритмов построения деревьев решений. Характеризовать алгоритм CARD, C4.5.

27. Характеризовать задачи классификации. Дать определение ROC-кривой. Описать организацию оценки качества классификации с помощью AUC. Объяснить организацию решения задачи классификации в Deductor. Привести пример построения ROC-кривой.
28. Привести характеристику построения деревьев решения в R.
29. Характеризовать метод random forest. Дать характеристику, привести примеры решения задач классификации с помощью метода случайного леса.
30. Охарактеризовать таблицу сопряженности (conclusion). Описать организацию построения таблиц сопряженности в R. Уточнить содержание ошибок первого и второго рода.
31. Дать определение логистической регрессии. Привести примеры решения задач бинарной классификации различными методами. Определить понятие ансамбля методов.
32. Дать определение нейронной сети. Классифицировать нейронные сети.
33. Характеризовать активизационные функции нейрона.
34. Привести примеры архитектура нейронной сети. Построить нейронные сети в Deductor.
35. Дать характеристику основ синтаксиса языка R, структуры данных языка.
36. Характеризовать средства импорта и экспорта данных. Привести примеры.
37. Классифицировать графику в R. Привести примеры.
38. Характеризовать средства анализ выборки в R. Привести примеры.
39. Продемонстрировать организацию проверки статистических гипотез в R, в JASP. Описать содержание и организацию использования критерия Стьюдента и критерия Манна-Уитни.
40. Продемонстрировать решение задач корреляционного анализа в R. Сделать обзор средств корреляционного анализа.
41. Характеризовать корреляции Пирсона, Спирмена, Кендалла, частной корреляции. Показать примеры их использования.
42. Дать характеристику задачи факторного анализа
43. Сформулировать этапы решения задач факторного анализа.
44. Дать общую характеристику ассоциативных правил
45. Дать общую характеристику алгоритма a'priori
46. Характеризовать рекомендательные системы.

Типовые задания для экзамена.

Задача 1. Проверить гипотезу о значимом отличии среднего балла за экзамены в десятом и одиннадцатом классах, используя критерий Стьюдента и критерий Манна-Уитни. Построить диаграммы «ящик с усами» для школьников, имеющих разные хобби. Построить диаграмму «дерево-листья». Данные находятся в файле тесты Школа.txt. Задачу решить в R и в JASP.

Построить задачу классификации хобби в зависимости от результатов тестирования. Задачу классификации решить с помощью деревьев решений в R.

Задача 2. Создать случайную последовательность размером в 500 наблюдений с использованием генератора равномерно распределенных чисел в диапазоне от 0 до 10. Проверить статистическую гипотезу о числовых значениях параметров:

$$1 \quad H_0 : a = 0,5; H_1 : a \neq 0,5$$

$$2 \quad H_0 : a = 5; H_1 : a > 5$$

Построить гистограмму распределения в R. Построить гистограмму частот и гистограмму относительных частот. При построении гистограммы оценить и задать число интервалов. Указать название осей и название гистограммы, а также заливку синего цвета. На диаграмму поместить кривую ядерной плотности, а также аппроксимацию равномерным законом распределения. При построении кривой регулировать ее гладкость.

- Оценить статистические характеристики.

- При проверке гипотезы: использовать одно выборочный Т-критерий. Задать уровень значимости 0,05. Использовать одностороннюю и двухстороннюю проверки гипотезы.

- Проверить гипотезу о равномерном законе распределения с помощью критерия Колмогорова-Смирнова.

В R использовать функцию t.test

Типовые проверочные задания для самоподготовки обучающегося к промежуточной аттестации:

ТИП ЗАДАНИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	ТИПОВЫЕ ЗАДАНИЯ	
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	1. Что означает термин «центроид кластера»? А) Количество объектов в кластере Б) Среднее значение признаков всех объектов внутри кластера В) Максимальное расстояние между объектами внутри кластера Г) Расстояние между двумя ближайшими кластерами	
		2. Какой метод кластеризации относится к иерархическим алгоритмам? А) Метод k-средних Б) DBSCAN В) Agglomerative Clustering Г) EM-кластеризация	
Задание закрытого типа на установление соответствия	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и	Понятие	Определение
		1. Фактор	А) Показатель, характеризующий степ
		2. Нагрузка	Б) Скрытая латентная величина, объяс
		3. Оборот	В) Число факторов, используемых для
		4. Размерность	Г) Совокупность операций по прео
		1. Установить соответствие между понятием и определением.	
		2. Установить соответствия между понятиями и определениями:	

	цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	<table><tr><th>Термин</th><th>Характеристика</th></tr><tr><td>1. Экстрактор</td><td>А) Матрица, отражающая связь каждого признака</td></tr><tr><td>2. Ортогональный поворот</td><td>Б) Способ извлечения главных компонент</td></tr><tr><td>3. Факторы</td><td>В) Метод вращения факторов, сохраняющий информацию</td></tr><tr><td>4. Ротация</td><td>Г) Латентные величины, выделяемые в результате</td></tr></table>	Термин	Характеристика	1. Экстрактор	А) Матрица, отражающая связь каждого признака	2. Ортогональный поворот	Б) Способ извлечения главных компонент	3. Факторы	В) Метод вращения факторов, сохраняющий информацию	4. Ротация	Г) Латентные величины, выделяемые в результате
Термин	Характеристика											
1. Экстрактор	А) Матрица, отражающая связь каждого признака											
2. Ортогональный поворот	Б) Способ извлечения главных компонент											
3. Факторы	В) Метод вращения факторов, сохраняющий информацию											
4. Ротация	Г) Латентные величины, выделяемые в результате											
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитайте предложенные варианты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или А Г).	1. Какие методы относятся к плоскостным (неиерархическим) методам кластеризации? Выберите все правильные варианты: А) K-means Б) DBSCAN В) Hierarchical clustering Г) Spectral clustering Д) Mean shift 2. Укажите верные утверждения относительно кластерного анализа: Выберите все правильные варианты: А) Цель кластерного анализа заключается в группировке объектов таким образом, чтобы объекты одной группы были похожи друг на друга больше, чем на объекты других групп. Б) Методы кластеризации позволяют разделить выборку на заранее заданное число классов. В) Качество кластеризации можно оценить с помощью метрик, таких как индекс Силуэта или коэффициент Дэйва-Болдина. Г) Результат кластеризации зависит исключительно от выбранного алгоритма и не зависит от масштаба исходных данных.										
Задание закрытого типа на установление последовательности	1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитайте предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).	1. Установите правильную последовательность этапов построения модели классификации: А. Сбор и подготовка данных В. Выбор метода классификации С. Обучение модели Д. Оценка качества модели Е. Применение модели для новых данных 2. Определите порядок шагов процесса кросс-валидации (метод проверки качества классификационной модели): А. Разделение данных на обучающую и тестовую выборки В. Обучение модели на каждом подмножестве данных С. Повторение процедуры для разных комбинаций разделения данных Д. Вычисление средней оценки точности моделей Е. Использование оставшегося подмножества для тестирования										
Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора	1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитайте предложенные варианты ответа. 3. Выбрать один верный ответ.	1. Какое утверждение верно описывает цель факторного анализа? Поясните свой выбор. Выбор варианта ответа: А) Анализ структуры корреляционных связей между признаками для выявления скрытых общих										

	4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).	факторов. Б) Проверка гипотез о значимости различий средних значений переменных. В) Нахождение наилучших предикторов целевого показателя регрессии. Г) Поиск аномалий и выбросов среди исследуемых данных.
		2. Чем отличается вращение Varimax от Promax в рамках факторного анализа? Поясните свой выбор. Выбор варианта ответа: А) Varimax сохраняет ортогональность факторов, тогда как Promax допускает наклонные факторы. Б) Varimax используется для уменьшения размерности пространства признаков, а Promax — для повышения интерпретируемости результатов. С) Varimax подходит для непрерывных данных, а Promax — для категориальных. D) Varimax применим только в статистическом пакете SPSS, а Promax доступен лишь в программе R.
Задание открытого типа с развернутым ответом	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ	1. Объясните подробно этапы очистки данных перед проведением анализа. Приведите конкретные примеры методов обработки пропущенных значений, дубликатов записей и некорректных значений. 2. Что такое разведочный анализ данных (EDA)? Опишите три основных инструмента EDA и приведите пример ситуации, когда каждый из них полезен.

6.3. Критерии и шкала оценивания на основе БРС.

Критерии и балльная шкала определяются преподавателем

КРИТЕРИИ ОЦЕНИВАНИЯ	РЕЗУЛЬТАТ В БАЛЛАХ
Дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где он продемонстрировал знания предмета в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину, самостоятельно, и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса	40
Дан развернутый ответ на поставленный вопрос, где студент демонстрирует знания, приобретенные на лекционных и семинарских занятиях, а также полученные посредством изучения обязательных учебных материалов по курсу, дает аргументированные ответы, приводит примеры, в ответе присутствует свободное владение монологической речью, логичность и последовательность ответа. Однако допускается неточность в ответе.	30-39
Дан ответ, свидетельствующий в основном о знании процессов изучаемой дисциплины, отличающийся недостаточной глубиной и полнотой раскрытия темы, знанием основных вопросов теории, слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать	20-29

аргументированные ответы и приводить примеры, недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа	
Дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Выводы поверхностны. Студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.	0-19

6.4. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Для решения задач открытого типа (кейсов, ПКЗ, ПИЗ), тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных. Для построения интеллект-карты и моделей в различных нотациях студенту можно использовать любой соответствующий онлайн-инструмент.

7. Методические материалы по освоению дисциплины (модуля)

Для изучения основных вопросов образовательной программы необходимо конспектировать материалы лекций, работать с рекомендованной преподавателем литературой, а также ресурсами информационно-телекоммуникационной сети «Интернет». Для приобретения навыков активного использования знаний полезно обсуждать плановые и возникающие вопросы, а также решаемые задачи на практических занятиях. Чтобы легче и прочнее усвоить материал следует постоянно использовать конкретные примеры, сравнения из уже полученных областей наук.

Для закрепления изученного материала даны вопросы по каждой теме дисциплины, на которые следует самостоятельно найти ответы.

Важной составной частью учебного процесса в вузе являются практические занятия. Практические занятия проводятся главным образом по дисциплинам, требующим закрепления навыков решения задач, и помогают студентам глубже усвоить учебный материал, приобрести умения применять принципы системного подхода к решению разнообразных задач, определять и оценивать ресурсы и существующие ограничения разного рода проектов.

При подготовке к практическим занятиям необходимо проанализировать конспект лекции, ознакомиться с рекомендованной литературой по соответствующей теме, осуществить подготовку по рекомендованным в рабочей программе вопросам для обсуждения темы, выполнить домашнее задание (при необходимости).

Необходимо помнить, что на лекции обычно рассматривается не весь материал, а только его часть. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. В процессе этой работы студент должен стремиться понять и запомнить основные положения рассматриваемого материала, примеры, поясняющие его, а также разобраться в иллюстративном материале. В процессе подготовки к занятиям рекомендуется взаимное обсуждение материала, во время которого закрепляются знания, а также приобретается практика в изложении и разъяснении полученных знаний, развивается речь. При необходимости следует обращаться за консультацией к преподавателю (в том числе по электронной почте). Планируя консультацию, необходимо хорошо продумать вопросы, которые требуют разъяснения. Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам. Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику. Кроме того, ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд методических материалов для быстрого повторения изученных вопросов, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

После изучения базовых тем курса проводится текущий контроль знаний студентов в виде опроса или письменного тестирования. Типовые тесты и задания по темам дисциплины приведены в специальном разделе данной рабочей программы.

Подготовка к текущему и промежуточному контролю предполагает изучение представленных вопросов к зачету, работу над тестами, представленными в данной рабочей программе, выполнение семестровой проектной работы по применению системного подхода и методов системного анализа к выбранной системе.

Работа в малых группах – это одна из самых популярных форм проведения занятий, так как она дает всем обучающимся (в том числе и стеснительным) возможность участвовать в работе, практиковать навыки сотрудничества, межличностного общения (в частности, умение активно слушать, вырабатывать общее мнение, разрешать возникающие разногласия). Цель данной формы проведения занятий: продемонстрировать сходство или различия определенных явлений, выработать стратегию или разработать план, выяснить отношение различных групп участников к одному и тому же вопросу. В ходе этой работы дополнительно решаются следующие задачи: развитие навыков общения и взаимодействия в группе, формирование ценностно-

ориентационного единства группы, поощрение к гибкой смене социальных ролей в зависимости от ситуации.

Группа студентов делится на несколько малых групп. Количество групп определяется числом творческих заданий, которые будут обсуждаться в процессе занятия. Малые группы формируются либо по желанию студентов, либо по родственной тематике для обсуждения. Каждая малая группа обсуждает творческое задание в течение отведенного времени. Основным этапом – проведение обсуждения творческого задания. Заслушиваются суждения, предлагаемые каждой малой группой по творческому заданию. Преподаватель дает оценочное суждение и работе малых групп, по решению творческих заданий, и эффективности предложенных путей решения.

В качестве самостоятельной работы студентами выполняется семестровая работа по применению системного подхода и методов системного анализа к выбранной системе по всем темам. Рекомендуется выбрать организационно-техническую систему. Перед выполнением задания по теме 1 выбранную систему необходимо согласовать с преподавателем. При выполнении заданий по темам могут использоваться представленные студентом материалы по предыдущим темам. Выполненная семестровая работа представляется студентом на открытой защите на промежуточной аттестации.

8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

8.1. Основная литература

1. Бурнашев, Р. А. Анализ и прогнозирование временных рядов: путь от классических моделей к современным решениям. Часть 1. Базовые инструменты и методы : учебное пособие для вузов / Р. А. Бурнашев, М. К. Арабов, М. Д. Миссаров. — Санкт-Петербург : Лань, 2025. — 216 с. — ISBN 978-5-507-52440-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/488951> (дата обращения: 01.05.2026). — Режим доступа: для авториз. пользователей.

2. Наумов, В. Н. Методы прогнозирования временных рядов : учебное пособие для вузов / В. Н. Наумов. — 2-е изд., стер. — Санкт-Петербург : Лань, 2025. — 196 с. — ISBN 978-5-507-53013-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/464216> (дата обращения: 01.05.2026). — Режим доступа: для авториз. пользователей.

8.2. Дополнительная литература

1. Гусарова Н.Ф, Анализ социальных сетей. Основные понятия и метрики. – СПб.: Университет ИТМО, 2016 – 67 с.
2. Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow : Концепции, инструменты и техники для создания интеллектуальных систем : полноцветное издание : перевод с английского - ПрМ.:Диалектика. -684 с.
3. Ланц Б. – Машинное обучение на R/ - СПб. :Питер. – 2020 – 464 с.
4. Principles of Econometrics with R [Электронный ресурс] – URL: <https://bookdown.org/ccolonescu/RPoE4/>
5. Люк Д. Анализ сетей (графов) в среде R. Руководство пользователя- М.: ДМК Пресс, 2017 – 250 с.
6. Маккинни У. Python и анализ данных. – М.: ДМК, 2015 – 482 с.
7. Нильсен Эйлин. Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение. –М.: ООО Диалектика – 2021 – 544 с.
8. Наумов Владимир Николаевич. Средства бизнес-аналитики: учеб. пособие / В. Н. Наумов ; Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2016. - 107 с.
9. Наумов В.Н. Анализ данных и машинное обучение: методы и инструментальные средства. Федер. гос. бюджет. образоват. учреждение высш. образования "Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации", Сев.-Зап. ин-т упр. - СПб. : СЗИУ - фил. РАНХиГС, 2020. - 260 с.
10. Шолле Ф. Глубокое обучение на Python. – СПб.: Питер. 2018. -400 с.
11. Шолле Ф. Глубокое обучение на R. – СПб.: Питер. 2018. -400 с.

8.3. Нормативные правовые документы и иная правовая информация

Не используются

8.4 Интернет-ресурсы

Обучающимся обеспечен доступ к материалам курса в СДО Академии <http://lms.ranepa.ru>, а так же через сайт научной библиотеки к следующим подписным электронным ресурсам:

Русскоязычные ресурсы

- Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»
- Электронные учебники электронно-библиотечной системы (ЭБС) «Лань»
- Электронные учебники электронно-библиотечной системы (ЭБС) «ZNANIUM.COM»
- Электронные учебники электронно-библиотечной системы (ЭБС) «BOOK.RU»
- Электронные учебники электронно-библиотечной системы (ЭБС) «IPRSMART»

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

№ п/п	Наименование
1.	Специализированные залы для проведения лекций, оснащенные персональным компьютером/ноутбуком и мультимедийным проектором
2.	Аудитории и компьютерные классы, оборудованные посадочными местами и персональными компьютерами с выходом в Интернет для проведения практических занятий
3.	«МТС Линк» — российская платформа для онлайн-коммуникаций и совместной работы команд ; «Яндекс Телемост» — сервис для видеоконференций от Яндекса; Я-мессенджер
4.	Технические средства обучения: персональные компьютеры; программные средства, обеспечивающие просмотр видеофайлов в форматах AVI, MPEG-4, DivX, RMVB, WMV; программы для работы с электронными таблицами для обработки, анализа и визуализации данных; соответствующие онлайн-инструменты для построения интеллект-карты и моделей в различных нотациях
5.	Научная библиотека (в т.ч. электронные информационные ресурсы научной библиотеки)
6.	СДО Академии https://lms.ranepa.ru/