

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a76e51317ff8d391

**Школа вычислительных социальных наук**

Протокол УС № 2 от 26.02.2025 г.



## Санкт-Петербург

**Автор:**

**Тенишева К.А.**, кандидат социологических наук, доцент направления Социология, директор программ по направлению Социология Школы Вычислительных социальных наук

**Рецензент:** Аркатов Д.А., канд. полит.наук, Доцент по направлению Социология, Школа вычислительных социальных наук

Рабочая программа дисциплины «**Анализ данных-1**», входящей в образовательную программу уровня магистратуры «Вычислительная социология», утверждена на заседании Совета Школы вычислительных социальных наук

Протокол заседания № 04 от 25.02.2025 года

## АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «**Анализ данных-1**» является дисциплиной базовой части основной профессиональной образовательной программы высшего образования «Вычислительная

Дисциплина «**Анализ данных-1**» представляет собой введение в количественный анализ данных. При этом он будет интересен как новичкам, так и тем, кто уже знает основы анализа данных: мы сразу разберем принципы симуляции данных для оценки неопределенности (uncertainty) и внедрим этот подход в анализ данных. Также, мы уделим большое внимание power analysis и тому, для каких целей можно использовать предельные эффекты. Ключевые навыки, которым будут обучаться студенты, помимо собственно методов анализа включают способность определять уместность и необходимость изученных методов для решения конкретных задач, а также умение давать содержательную интерпретацию результатов анализа. Ключевые темы: распределения; типы гипотезы и простые тесты; симуляция данных для проверки гипотез; линейная регрессия; логистическая регрессия; методы снижения размерности (факторный анализ, кластерный анализ).

Общая трудоемкость освоения дисциплины составляет 6 зачетных единиц, 216 часов.

## Содержание

1. НАИМЕНОВАНИЕ, ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....	5
2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....	5
3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	6
4. ОБЪЕМ ДИСЦИПЛИНЫ.....	6
5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ .....	6
5.1 Содержание дисциплины .....	6
5.2 Структура дисциплины.....	8
6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ.....	9
6.1 Общие положения .....	9
6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины .....	9
6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине .....	9
6.4 Перечень литературы для самостоятельной работы обучающегося: .....	11
6.5 Перечень учебно-методического обеспечения для самостоятельной работы.....	11
7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ .....	11
7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации.....	11
7.2 Контрольные задания для текущей аттестации.....	13
7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации .....	13
7.4 Типовые задания к промежуточной аттестации.....	18
7.5 Средства оценки индикаторов достижения компетенций.....	21
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	22
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.....	22
9.1 Программное обеспечение .....	23
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины: .....	23
9.3 Лицензионные электронные ресурсы библиотеки Университета .....	23
9.4 Электронная информационно-образовательная среда Университета.....	24
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.....	24
ПРИЛОЖЕНИЕ 1 .....	26

## 1. НАИМЕНОВАНИЕ, ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

**Цель** освоения дисциплины «Анализ данных-1» состоит в том, чтобы познакомить слушателей с различными методами статистического анализа данных, которые широко используются в социальных науках, привить им практические навыки работы с количественными методами в рамках своих собственных исследовательских проектов.

### Задачи:

1. Рассмотреть основные методы статистического анализа данных.
2. Развить навыки работы в программной среде R.
3. Сформировать умение формулировать исследовательские задачи в привязке к статистической методологии.

## 2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: общепрофессиональными (ОПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Таблица 1

### Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-3 (С) Способен прогнозировать социальные явления и процессы, выявлять социально значимые проблемы и вырабатывать пути их решения на основе использования научных теорий, концепций, подходов и социальных технологий	ИД.ОПК-3.1. Анализирует проблемы развития социальных явлений и процессов с использованием статистических процедур для обработки социологических данных	Знать: актуальные научные теории, концепции выявления и прогнозирования социальных явлений, процессов, проблем, современные подходы и социальные технологий решения социально значимых проблем З (ОПК-3)
	ИД.ОПК-3.2. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов	Уметь: составлять прогноз социальных явлений и процессов, выявлять социально значимые проблемы и вырабатывать пути их решения на основе использования научных теорий, концепций, подходов и социальных технологий У (ОПК-3)
	ИД.ОПК-3.3. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования ИД.ОПК-3.4. Научно обосновывает постановку фундаментальных и прикладных социо-логических исследований для решения социально значимых проблем ИД.ОПК-3.5. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий ИД.ОПК-3.6. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов	Владеть: навыками прогнозирования социальных явлений и процессов с целью выявления и решения социально значимых проблем общества на основе использования научных теорий, концепций, подходов и социальных технологий В (ОПК-3)

В результате освоения дисциплины магистрант должен:

ЗНАТЬ: методы социологического исследования

УМЕТЬ: применять методы математического анализа и моделирования;

ВЛАДЕТЬ: навыками обобщения и анализа профессиональной информации, навыками оценки достоверности научного знания

### 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Анализ данных-1» является дисциплиной обязательной части Блока 1 «Дисциплины (модули)» образовательной программы «Вычислительная социология». Курс читается во втором и третьем модулях, форма промежуточной аттестации – 2 модуль – зачет, 3 модуль - зачет с оценкой.

Для успешного освоения данной дисциплины требуется параллельное изучения дисциплин: «Введение в казуальность».

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения учебной и производственной практики, выполнения выпускной квалификационной работы.

### 4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 6 зачетных единиц, 216 часов.

Таблица 2

Объем дисциплины

Типы учебных занятий и самостоятельная работа		Объем дисциплины										
		Всего	Модуль									
			1	2	3	4	5	6	7	8	9	10
Контактная работа обучающихся с преподавателем в соответствии с УП:		56	28	28	-	-	-	-	-	-	-	-
Лекции (Л)		28	14	14	-	-	-	-	-	-	-	-
Семинарские занятия (СЗ)		28	14	14	-	-	-	-	-	-	-	-
Самостоятельная работа (СР)		160	116	44	-	-	-	-	-	-	-	-
Промежуточная аттестация	форма	Зачет, зачет с оценкой	Зачет	Зачет с оценкой	-	-	-	-	-	-	-	-
	час.	-	-	-	-	-	-	-	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)		216/6	144/4	72/2	-	-	-	-	-	-	-	-

### 5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине: через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

#### 5.1 Содержание дисциплины

Таблица 3

Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)
1	Введение. Данные. Типы распределений.	Данные, переменные, векторы, шкалы. Типы распределений. Меры центральности. Генеральная совокупность и выборка. Центральная предельная теорема. Визуализация базовых характеристик данных.	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)
2	Вероятность	Выборочное распределение. Точечная и интервальная оценка. Доверительный интервал. Понятие статистической значимости	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)
3	Базовые тесты	Биномиальный тест, хи-квадрат, T-test и z- стандартизация. ANOVA. Корреляция	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)
4	Введение в регрессию	Метод наименьших квадратов. Простая линейная регрессия. Разные типы предикторов: интервальный, бинарный, категориальный	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)
5-7	Множественная регрессия	Множественная регрессия. Стандартизация коэффициентов. Визуализация результатов. Простая диагностика: мультиколлинеарност ь, выбросы. Интерактивные эффекты.	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)
8-10	Логистическая регрессия	Отношение шансов. Максимальное правдоподобие. Бинарная логистическая регрессия. Интерпретация результатов логистической регрессии. Предсказанные вероятности. Предельные эффекты.	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)
		Визуализация результатов. Диагностика.			
11-12	Факторный анализ	Введение в факторный анализ. Эксплораторный и подтверждающий факторный анализ.	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)
13-14	Продвинутая классификация данных	Продвинутая классификация данных. Кластерный анализ. Деревья решений	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)

## 5.2 Структура дисциплины

Таблица 4

Структура дисциплины

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП			СР	
			Л	СЗ	СРП		
Очная форма обучения							
Тема 1	Введение. Данные. Типы распределений.	20	2	2		16	Опрос
Тема 2	Вероятность	20	2	2		16	Опрос, домашнее задание
Тема 3	Базовые тесты	20	2	2		16	Опрос, решение и анализ практических заданий, домашнее задание
Тема 4	Введение в регрессию	20	2	2		16	Опрос, решение и анализ практических заданий, домашнее задание
Тема 5-7	Множественная регрессия	20	6	6		52	Опрос, решение и анализ практических заданий, домашнее задание
Промежуточная аттестация 2 модуль							Зачет
Тема 8-10	Логистическая регрессия	18	4	4		10	Опрос, решение и анализ практических заданий, домашнее задание

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП			СР	
			Л	СЗ	СРП		
Очная форма обучения							
Тема 11-12	Факторный анализ	18	4	4		10	Опрос, решение и анализ практических заданий, домашнее задание
Тема 13-14	Продвинутая классификация данных	36	6	6		24	Опрос, домашнее задание
Промежуточная аттестация 3 модуль		-	-	-	-	-	Зачет с оценкой
Всего:		216/6	28	28	-	160	-

*\*Примечание: формы текущего контроля успеваемости: опрос (О), домашнее задание (ДЗ), решение и анализ практических заданий (РАПЗ)*

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

### 6.1 Общие положения

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения текстов и исследовательской литературы (из списков основной и дополнительной литературы) и их анализа.

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение вспомогательных учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

### 6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

#### Тема 1: Введение. Данные. Типы распределений

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 12 часов. Итого: 16 часов.

#### Тема 2: Вероятность

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 12 часов. Итого: 16 часов.

### **Тема 3. Базовые тесты.**

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 12 часов. Итого: 16 часов.

### **Тема 4. Введение в регрессию**

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 12 часов. Итого: 16 часов.

### **Тема 5-7. Множественная регрессия**

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 12 часов.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 40 часов. Итого: 52 часа.

### **Тема 8-10. Логистическая регрессия**

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 8 часов. Итого: 10 часов.

### **Тема 11-12. Факторный анализ.**

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 8 часов. Итого: 10 часов.

### **Тема 13-14. Продвинутая классификация данных**

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 6 часов.
2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 18 часов. Итого: 24 часа.

## **6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине**

1. Примеры использования статистического анализа данных в социальных науках.
2. Основные источники данных и требования к ним.

3. Этапы статистического исследования.
4. Понятие случайности. Функция распределения.
5. Числовые характеристики распределения.
6. Описание данных с помощью дескриптивной статистики и статистические графики: график плотности распределения, гистограммы, диаграммы рассеивания.
7. Точечные и интервальные оценки параметров генеральной совокупности.
8. Двусторонние и односторонние гипотезы.
9. Коэффициент корреляции Пирсона и его свойства.
10. Статистический контроль в социальных науках.
11. Стандартная ошибка регрессии, Коэффициент детерминации R-квадрат.
12. Допущения моделей множественной регрессии.
13. Модели линейной вероятности и метод максимального правдоподобия.
14. Интерпретация регрессионных коэффициентов.
15. Вероятности, типы распределений и центральная предельная теорема
16. Тесты на сравнение средних с непрерывными переменными: t-test и ANOVA
17. Корреляции и линейная регрессия
18. Диагностика линейной регрессии, интерактивные эффекты в регрессионных моделях
19. Вероятности, отношения шансов и биномиальная логистическая регрессия
20. Метод главных компонент, факторный анализ, кластеризация

#### **6.4 Перечень литературы для самостоятельной работы обучающегося:**

1. Agresti et.al. "Statistics: The Art and Science of Learning from Data". 4<sup>th</sup> edition (2018)
2. Field A., Miles J., Field Z. Discovering statistics using R. – Sage publications, 2012.
3. Gelman A., Hill J., Vehtari A. Regression and other stories. – Cambridge University Press, 2020.

#### **6.5 Перечень учебно-методического обеспечения для самостоятельной работы**

Для обеспечения самостоятельной работы магистрантов по дисциплине «Анализ данных-1» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).
2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).
3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).
4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

### **7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ**

#### **7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации**

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому семинарскому занятию, активное слушание на лекциях, участие в опросах, выполнение магистрантами домашних заданий, решения и анализа магистрантами практических задач. Магистрант

должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по темам обсуждения.

Текущий контроль проводится в форме оценивания участия магистрантов в опросах, дискуссиях, выполнения магистрантами письменных работ, демонстрирующих степень знакомства магистрантов с дополнительной литературой.

Таблица 5

**Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации**

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
1. Введение. Данные. Типы распределений.	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос	зачтено/ не зачтено
2. Вероятность	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, домашнее задание	зачтено/ не зачтено
3. Базовые тесты	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
4. Введение в регрессию	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
5-7. Множественная регрессия	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
8-10. Логистическая регрессия	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
11-12. Факторный анализ	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
13-14. Продвинутая классификация данных	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, домашнее задание	зачтено/ не зачтено

Таблица 6

### Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Опрос	ответ отсутствует или является односложным, или содержит существенные ошибки – не зачтено магистрант в ответах демонстрирует знание всех теоретических положений, (развернуто) отвечает на все поставленные вопросы, предлагает обоснования при ответе на все или большинство поставленных вопросов; несущественные ошибки не снижают качество ответа — зачтено
Решение и анализ практических задач	Магистрант выполняет работу частично или с существенными недочетами – не зачтено, Полное и правильное выполнение заданий работы – зачтено
Домашнее задание	Магистрант выполняет работу частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные социальные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, Полное и правильное выполнение заданий работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

## 7.2 Контрольные задания для текущей аттестации

### Примерный материал дискуссий, домашних заданий, решения практических задач

#### Тема 1: Введение. Данные. Типы распределений.

##### Опрос 1:

1. Знакомство с R: базовые команды и функции. Типы переменных: непрерывные переменные, упорядоченные и неупорядоченные категории, вывод результатов; преобразования переменных; сортировка наблюдений и отбор по условию.
2. Процедуры для описания данных. Графическое представление данных. Дескриптивные статистики. Показатели центральной тенденции и разброса в случае симметричного и несимметричного распределения.
3. Методы сбора данных. Опрос: плюсы и минусы метода.

#### Тема 2: Вероятность.

### **Опрос 2:**

1. Выборочное распределение: Чем отличается выборочное распределение среднего от распределения генеральной совокупности? Как на него влияет объем выборки?
2. Точечные и интервальные оценки: В каких случаях предпочтительнее использовать доверительный интервал вместо точечной оценки (например, среднего)?
3. Статистическая значимость: Как связаны уровень значимости ( $\alpha$ ) и доверительный интервал? Почему  $p\text{-value} < 0.05$  — не абсолютный критерий истины?

### **Домашнее задание 1**

Сгенерируйте в R 100 выборок объема  $n=30$  из нормального распределения ( $\mu=50$ ,  $\sigma=10$ ), постройте выборочное распределение средних и сравните его с теоретическим  $N(50, \sigma/\sqrt{n})$ . Рассчитайте 95% доверительный интервал для одной из выборок и проверьте, покрывает ли он истинное  $\mu$ .

### **Тема 3. Базовые тесты.**

#### **Опрос 3:**

1. Поиск связи между переменными.
2. Отсутствие и наличие связи между признаками.
3. Типы простых тестов.
4. Тесты для сравнения средних.
5. Тесты для оценки связи двух интервальных переменных.

#### **Решение и анализ практических заданий 1.**

Для заданных пар переменных выберите верные парные тесты (t-test, хи-квадрат, ANOVA, корреляция). Для каждой пары проинтерпретируйте результаты теста с точки зрения значимости и размера эффекта.

### **Домашнее задание 2**

Загрузите в R датасет `mtcars`. Проведите t-test для сравнения `mpg` у авто с `am=0` и `am=1`. Корреляцию Пирсона между `hp` и `mpg`. Хи-квадрат для `vs` и `gear`. Оформите выводы в таблице: тест, p-value, интерпретация.

### **Тема 4. Введение в регрессию**

#### **Опрос 4:**

1. Построение модели парной линейной регрессии. Обсуждение интерпретации регрессионных коэффициентов
2. Интерпретация коэффициентов для факторных и интервальных переменных
3. Построение модели линейной регрессии с несколькими предикторами. Сравнение интерпретации регрессионных коэффициентов с моделью парной линейной регрессии
4. Интерпретация коэффициента детерминации для оценки качества модели;

#### **Решение и анализ практических заданий 2.**

Для заданной зависимой переменной рассчитайте простую регрессионную модель с одним предиктором. Проинтерпретируйте результаты теста с точки зрения значимости и размера эффекта, а также качества модели. Добавьте еще один предиктор: как изменилась модель? Аргументируйте, почему. Добавьте порядковый предиктор. Дайте интерпретацию эффекта.

### **Домашнее задание 3**

На данных `mtcars` постройте простую регрессию `mpg ~ wt`, интерпретируйте коэффициент. Добавьте предиктор `hp`, объясните изменения. Преобразуйте `cyl` в фактор, добавьте в модель. Сравните  $R^2$ .

### **Тема 5-7. Множественная регрессия**

#### **Опрос 5:**

1. Множественная регрессия. Понятие контрольных переменных

2. Стандартизованные и нестандартизованные коэффициенты
3. Типы регрессионных графиков. Интерпретация.
4. Типы регрессионных предпосылок. Подходы к диагностике.
5. Мультиколлинеарность: диагностика, подходы к исправлению
6. Гетероскедастичность: диагностика, подходы к исправлению
7. Нелинейность: диагностика, подходы к исправлению
8. Выбросы: диагностика, подходы к исправлению

### **Решение и анализ практических заданий 3.**

Для заданной зависимой переменной и набора предикторов, постройте регрессионную модель методом последовательного добавления. Какие переменные вы используете как контрольные, что это значит? Дайте интерпретацию с помощью стандартизованных и нестандартизованных коэффициентов. В чем разница? Определите наиболее сильный предиктор. Постройте регрессионные графики. Проведите диагностику качества модели, исправьте выявленные проблемы.

### **Домашнее задание 4**

Для `mtcars` создайте модель `mpg ~ wt + hp + factor(cyl)`. Проверьте мультиколлинеарность (VIF). Постройте график остатков vs предсказанные значения. Удалите выбросы (критерий Кука), пересчитайте модель.

## **Тема 8-10. Логистическая регрессия**

### **Опрос 6:**

1. Отношение шансов. Максимальное правдоподобие
2. Бинарная логистическая регрессия. Интерпретация результатов логистической регрессии.
3. Предсказанные вероятности.
4. Предельные эффекты. Средние предельные эффекты (AME)
5. Визуализация результатов. Типы графиков. Интерпретация.
6. Диагностика: таблица классификации, чувствительность, специфичность.
7. Площадь под кривой. AUC ROC.

### **Решение и анализ практических заданий 4.**

Для заданной зависимой переменной и набора предикторов, постройте логистическую регрессионную модель методом последовательного добавления. Дайте интерпретацию с помощью нестандартизованных коэффициентов и отношения шансов. В чем разница? Посчитайте и интерпретируйте предельные эффекты. Определите наиболее сильный предиктор. Постройте регрессионные графики. Проведите диагностику качества модели, исправьте выявленные проблемы.

### **Домашнее задание 5**

На данных Titanic (из `titanic::titanic_train`). Постройте модель `Survived ~ Age + Pclass`, интерпретируйте OR. Рассчитайте AME для Pclass. Постройте ROC-кривую, вычислите AUC.

## **Тема 11-12. Факторный анализ.**

### **Опрос 7:**

1. Обсуждение примеров латентных признаков
2. Эксплораторный факторный анализ.
3. Разработка наблюдаемых индикаторов для этих латентных признаков
4. Реализация в R факторного анализа для измерения латентного признака
5. Сравнение моделей факторного анализа с разным количеством факторов, разными видами вращений
6. Конфирматорный факторный анализ. Пакет `lavaan`. Критерии качества модели.

### **Решение и анализ практических заданий 5.**

Рассмотрите данный набор переменных. Какое количество факторов вы ожидаете для них обнаружить? Проведите эксплораторный факторный анализ. Сколько факторов выделяется? Коррелируют ли они? Какую долю дисперсии объясняет каждый фактор? Дайте содержательную интерпретацию полученным факторам.

### Домашнее задание 6

Для psych::bfi (5 личностных черт). Проведите EFA с вращением varimax, выберите 2-3 фактора. Назовите факторы на основе нагрузок ( $>0.5$ ). Сравните с CFA (lavaan), оцените CFI/RMSEA.

### Тема 13-14. Продвинутая классификация данных

#### Опрос 8:

1. Обсуждение примеров реальных задач, требующих построения классификации объектов
2. Кластерный анализ. Основные алгоритмы.
3. Построение классификации объектов в R с помощью алгоритма k- средних
4. Сравнение моделей с разным количеством кластеров
5. Обсуждение критериев оценки качества полученной классификации
6. Деревья решений

#### Решение и анализ практических заданий 6.

Для заданной базы данных, выберите переменные, на основе которых можно кластеризовать наблюдения. Выберите алгоритм. Определите количество кластеров. Проведите кластеризацию и дайте интерпретацию полученным кластерам. Оцените качество.

### Домашнее задание 7

Для iris выполните кластеризацию k-means ( $k=3$ ), сравните с Species. Постройте дендрограмму (иерархическая кластеризация). Оцените точность через матрицу сопряженности.

## 7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

### Форма промежуточной аттестации:

**2 модуль – зачет**, который проходит в форме тестирования.

**3 модуль – зачет с оценкой**, который проходит в форме итогового тестирования.

Перед зачетом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине.

Таблица 7

### Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
<b>2 модуль – зачет-тестирование</b>	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	100-41% правильных ответов	Зачтено

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
		ИД.ОПК-3.5. ИД.ОПК-3.6.		40-0% правильных ответов	Не зачтено
<b>3 модуль - зачет оценкой / тестирование</b>	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	100-81% правильных ответов	Зачтено, отлично
				80-61% правильных ответов	Зачтено, хорошо
				60-41% правильных ответов	Зачтено, удовлетворительно
				40-0% правильных ответов	Не зачтено, неудовлетворительно

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 7а.

Таблица 7а

#### Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценках «зачтено» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценке «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценке «не зачтено, неудовлетворительно», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

## 7.4 Типовые задания к промежуточной аттестации

### Требования к тестам

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

### Задания закрытого типа

*Базовый уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте - 1 балл; ответ отличен от эталонного - 0 баллов.

*Повышенный уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балла; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

### Комбинированные задания

*Базовый уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

*Повышенный уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутствует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

### Задания открытого типа

*Повышенный уровень сложности:* ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

*Высокий уровень сложности:* магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left( \frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

$k_n$  – максимально возможное количество баллов за вопросы по компетенции,

$x_n$  – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

## Примеры тестовых заданий для промежуточной аттестации 2 модуль

Задания для проверки компетенции ОПК-3 (С) (Способен прогнозировать социальные явления и процессы, выявлять социально значимые проблемы и вырабатывать пути их решения на основе использования научных теорий, концепций, подходов и социальных технологий)

### Задания закрытого типа (базовый уровень сложности)

#### Задание 1

Инструкция: Прочитайте задание и установите соответствие:

Вопрос: При анализе данных опроса о потребительских предпочтениях были выявлены следующие проблемы. Установите соответствие между проблемами и оптимальными методами их решения:

	Проблема данных		Метод решения
1	30% респондентов не указали уровень дохода	А	Множественная импутация методом MICE
2	Распределение возраста имело пики на круглых цифрах (30, 40, 50 лет)	В	Взвешивание по возрастным группам
3	В выборке было только 5% респондентов старше 60 лет	С	Исключение систематических ошибок округления

Поле для ответа:

1	2	3

### Задания комбинированного типа (повышенный уровень сложности)

#### Задание 1

Инструкция: Выберите все правильные утверждения и обоснуйте выбор.

Вопрос: Какие утверждения демонстрируют критическое влияние методов обработки данных на результаты? В исследовании факторов миграции:

1. После применения бутстрапа доверительный интервал для коэффициента регрессии сократился на 15%
2. Дендрограмма выявила 4 кластера по мотивам переезда
3. При замене пропущенных значений медианой корреляция между переменными изменилась на 0.3
4. После взвешивания по образованию доля экономических мигрантов увеличилась с 42% до 51%

Поле для ответа:

--	--	--	--

Обоснование: \_\_\_\_\_

### Задания открытого типа (повышенный уровень сложности)

#### Задание 1

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ.

Вопрос: Сравните возможности байесовских сетей и деревьев решений для прогнозирования результатов выборов. Приведите конкретные примеры ограничений каждого метода.

Поле для ответа: \_\_\_\_\_

### **Задания открытого типа (высокий уровень сложности)**

#### ***Задание 1***

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ.

Вопрос: В рамках исследования социального неравенства анализируется опрос, содержащий данные о доходах, образовании и уровне жизни 5000 респондентов. Около 15% значений в переменной «доход» пропущены, причем пропуски не случайны: чаще отсутствуют у респондентов с низким уровнем образования. Как определить механизм пропусков (MCAR, MAR, MNAR) в данном случае? Какой метод импутации вы выберете? Обоснуйте выбор между множественной регрессионной импутацией (MICE), импутацией на основе машинного обучения (например, Random Forest) и методами, учитывающими MNAR (например, Heckman-коррекция). Как проверить, что импутация не исказила распределение доходов?

Поле для ответа: \_\_\_\_\_

### **Примеры тестовых заданий для промежуточной аттестации 3 модуль**

**Задания для проверки компетенции ОПК-3 (С) (Способен прогнозировать социальные явления и процессы, выявлять социально значимые проблемы и вырабатывать пути их решения на основе использования научных теорий, концепций, подходов и социальных технологий)**

### **Задания закрытого типа (базовый уровень сложности)**

#### ***Задание 1***

Инструкция: Прочитайте задание и установите соответствие:

Вопрос: В исследовании уровня бедности в регионе столкнулись с тремя проблемами данных. Установите соответствие между проблемами и методами их коррекции:

	Проблема данных		Метод решения
	40% респондентов отказались указывать точный доход, предоставив только диапазон		Квантильная регрессия
	Данные о расходах имели выраженный правый скос		Стратифицированное взвешивание
	В выборке отсутствовали жители отдаленных сельских районов		Импутация интервальных значений

Поле для ответа:

1	2	3

### **Задания комбинированного типа (повышенный уровень сложности)**

#### ***Задание 1***

Инструкция: Выберите два правильных утверждения и обоснуйте выбор.

Вопрос: При анализе эффекта образовательной программы методом Difference in Differences (DiD) были получены следующие результаты. Какие утверждения критически важны для валидности результатов?

1. Коэффициент DiD составил 0.15 ( $p < 0.01$ )
2. График параллельных трендов показал расхождение линий до вмешательства
3. После включения ковариат эффект уменьшился до 0.10
4. Тест на баланс характеристик между группами дал  $p\text{-value} = 0.8$

Поле

для

ответа:

--	--

Обоснование: \_\_\_\_\_

### **Задания открытого типа (повышенный уровень сложности)**

#### ***Задание 1***

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ.

Вопрос: Объясните, почему GAM (Generalized Additive Models) предпочтительнее линейной регрессии для анализа влияния возраста на зарплату. Приведите пример интерпретации нелинейного эффекта из исследования.

Поле для ответа: \_\_\_\_\_

### **Задания открытого типа (высокий уровень сложности)**

#### ***Задание 1***

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ.

Вопрос: В исследовании социальных детерминант здоровья анализируются данные 10.000 респондентов. Зависимая переменная: Наличие хронического заболевания (бинарная: 1 = да, 0 = нет). Если модель показывает  $OR=2.5$  для курения (да/нет), как это преобразовать в понятное сообщение для министерства здравоохранения? Какие ограничения логистической регрессии могут привести к ошибочным выводам (например, при анализе редких событий)? В каких случаях стоит использовать методы машинного обучения (например, Random Forest) вместо логистической регрессии? Как совместить логистическую регрессию с кластерным анализом для выявления групп риска?

## **7.5 Средства оценки индикаторов достижения компетенций**

Таблица 8

**Средства оценки индикаторов достижения компетенций**

<b>Коды компетенций</b>	<b>Индикаторы компетенций (в соот.с Таблицей 1)</b>	<b>Средства оценки (в соот. с Таблицами 5, 7)</b>
ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	Опрос (О), домашнее задание (ДЗ), решение и анализ практических заданий (РАПЗ), тест

Таблица 9

**Описание средств оценки индикаторов достижения компетенций**

<b>Средства оценки (в соот. с Таблицами 5, 7)</b>	<b>Рекомендованный план выполнения работы</b>
Опрос	Магистрант в ходе подготовки и участия в опросах по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получившую развитие в рамках данной дисциплины: 1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Домашнее задание	<p>Магистрант должен быть готовым в ходе подготовки и представления домашнего задания по темам дисциплины, выполнять следующие действия:</p> <p>1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов</p>
Решение и анализ практических заданий	<p>Магистрант должен быть готовым в ходе подготовки и решения практических заданий по темам дисциплины, выполнять следующие действия:</p> <p>1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов</p>
Тест	<p>Магистрант должен быть готовым в ходе подготовки к тесту, выполнять следующие действия:</p> <p>1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов</p>

## 8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

### 8.1. Основная литература

1. Колданов, А. П. Теория вероятностей и математическая статистика : учебник / А. П. Колданов, П. А. Колданов ; Нац. исслед. ун-т «Высшая школа экономики». — М. : Изд. дом Высшей школы экономики, 2023. — 245, [3] с. — (Учебники Высшей школы экономики). — ISBN 978-5-7598-2544-9 (в пер.). — ISBN 978-5-7598-2829-7 (e-book). — URL: <http://176.9.74.196/book.html?currBookId=45113>.

2. Дятлов, А. В. Методы многомерного статистического анализа данных в социологии : учебник / А. В. Дятлов, П. Н. Лукичев ; Южный федеральный университет. - Ростов-на-Дону : Издательство Южного федерального университета, 2023. - 236 с. - ISBN 978-5-9275-4265-9. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2135836>. – Режим доступа: по подписке.

### 8.2. Дополнительная литература

1. Agresti et.al. “Statistics: The Art and Science of Learning from Data”. 4<sup>th</sup> edition (2018)

2. Field A., Miles J., Field Z. Discovering statistics using R. – Sage publications, 2012.
3. Gelman A., Hill J., Vehtari A. Regression and other stories. – Cambridge University Press, 2020.

## **9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

### **9.1 Программное обеспечение**

1. OS Microsoft Windows (OVS OS Platform)
2. MS Office (OVS Office Platform)
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. ABBYY FineReader 11 Corporate Edition
6. ABBYY Lingvo x5
7. Adobe Acrobat Reader DC /Pro – бесплатно
8. Google Chrome – бесплатно
9. Opera – бесплатно
10. Mozilla – бесплатно
11. VLC – бесплатно
12. Яндекс Браузер

### **9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:**

#### **Информационно-справочные системы**

1. Гарант.Ру. Информационно-правовой портал: <http://www.garant.ru>
2. Информационная система «Единое окно доступа к образовательным ресурсам»: <http://window.edu.ru/>
3. Открытое образование. Ассоциация «Национальная платформа открытого образования»: <http://npoed.ru>
4. Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
5. Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
6. Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
7. Российское образование. Федеральный портал: <http://www.edu.ru>

#### **Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:**

1. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
2. Президентская библиотека: <http://www.prilib.ru>
3. Российская государственная библиотека: <http://www.rsl.ru/>
4. Российская национальная библиотека: <http://www.nlr.ru/poisk/>

### **9.3 Лицензионные электронные ресурсы библиотеки Университета**

#### **Профессиональные базы данных:**

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета <https://eusp.org/library/electronic-resources>, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: <http://elibrary.ru>;
2. **Университетская информационная система РОССИЯ** — база

электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: <http://www.uisrussia.msu.ru/>;

3. Электронные журналы по подписке (текущие номера научных зарубежных журналов).

**Электронные библиотечные системы:**

1. **Znaniium.com** – Электронная библиотечная система (ЭБС) – <http://znaniium.com/>;

2. Университетская библиотека онлайн – Электронная библиотечная система (ЭБС) – <http://biblioclub.ru/>

#### **9.4 Электронная информационно-образовательная среда Университета**

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета, официальный сайт Университета (Европейский университет в Санкт-Петербурге [<https://eusp.org/>]), локальную сеть и корпоративную электронную почту Университета, и обеспечивает:

- доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;
- фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;
- формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;
- взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине

### **10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа (практических занятий, лабораторных работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для

самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к электронным ресурсам библиотеки Университета с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ**

**«Анализ данных-1»**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И  
ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ**

**1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации**

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому семинарскому занятию, активное слушание на лекциях, участие в опросах, выполнение магистрантами домашних заданий, решения и анализа магистрантами практических задач. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по темам обсуждения.

Текущий контроль проводится в форме оценивания участия магистрантов в опросах, дискуссиях, выполнения магистрантами письменных работ, демонстрирующих степень знакомства магистрантов с дополнительной литературой.

Таблица 1

**Показатели, критерии и оценивание компетенций и индикаторов их  
достижения в процессе текущей аттестации**

<b>Наименование тем (разделов)</b>	<b>Коды компетенций</b>	<b>Индикаторы компетенций</b>	<b>Коды ЗУВ (в соот. с Таблицей 1)</b>	<b>Формы текущего контроля успеваемости</b>	<b>Результаты текущего контроля</b>
1. Введение. Данные. Типы распределений.	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос	зачтено/ не зачтено
2. Вероятность	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, домашнее задание	зачтено/ не зачтено
3. Базовые тесты	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
4. Введение в регрессию	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
5-7. Множественная регрессия	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
8-10. Логистическая регрессия	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
11-12. Факторный анализ	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, решение и анализ практических заданий, домашнее задание	зачтено/ не зачтено
13-14. Продвинутая классификация данных	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	Опрос, домашнее задание	зачтено/ не зачтено

Таблица 2

**Критерии оценивания**

Формы текущего контроля успеваемости	Критерии оценивания
Опрос	ответ отсутствует или является односложным, или содержит существенные ошибки – не зачтено магистрант в ответах демонстрирует знание всех теоретических положений, (развернуто) отвечает на все поставленные вопросы, предлагает обоснования при ответе на все или большинство поставленных вопросов; несущественные ошибки не снижают качество ответа — зачтено
Решение и анализ практических задач	Магистрант выполняет работу частично или с существенными недочетами – не зачтено, Полное и правильное выполнение заданий работы – зачтено
Домашнее задание	Магистрант выполняет работу частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные социальные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, Полное и правильное выполнение заданий работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

## 2 Контрольные задания для текущей аттестации

### Материал дискуссий, домашних заданий, практических заданий

#### Тема 1: Введение. Данные. Типы распределений.

##### Опрос 1:

1. Знакомство с R: базовые команды и функции. Типы переменных: непрерывные переменные, упорядоченные и неупорядоченные категории, вывод результатов; преобразования переменных; сортировка наблюдений и отбор по условию.
2. Процедуры для описания данных. Графическое представление данных. Дескриптивные статистики. Показатели центральной тенденции и разброса в случае симметричного и несимметричного распределения.
3. Методы сбора данных. Опрос: плюсы и минусы метода.

#### Тема 2: Вероятность.

##### Опрос 2:

1. Выборочное распределение: Чем отличается выборочное распределение среднего от распределения генеральной совокупности? Как на него влияет объем выборки?
2. Точечные и интервальные оценки: В каких случаях предпочтительнее использовать доверительный интервал вместо точечной оценки (например, среднего)?
3. Статистическая значимость: Как связаны уровень значимости ( $\alpha$ ) и доверительный интервал? Почему  $p\text{-value} < 0.05$  — не абсолютный критерий истины?

#### Домашнее задание 1

Сгенерируйте в R 100 выборок объема  $n=30$  из нормального распределения ( $\mu=50$ ,  $\sigma=10$ ), постройте выборочное распределение средних и сравните его с теоретическим  $N(50, \sigma/\sqrt{n})$ . Рассчитайте 95% доверительный интервал для одной из выборок и проверьте, покрывает ли он истинное  $\mu$ .

#### Тема 3. Базовые тесты.

##### Опрос 3:

1. Поиск связи между переменными.
2. Отсутствие и наличие связи между признаками.
3. Типы простых тестов.
4. Тесты для сравнения средних.
5. Тесты для оценки связи двух интервальных переменных.

#### Решение и анализ практических заданий 1.

Для заданных пар переменных выберите верные парные тесты (t-test, хи-квадрат, ANOVA, корреляция). Для каждой пары проинтерпретируйте результаты теста с точки зрения значимости и размера эффекта.

#### Домашнее задание 2

Загрузите в R датасет `mtcars`. Проведите t-test для сравнения `mpg` у авто с `am=0` и `am=1`. Корреляцию Пирсона между `hp` и `mpg`. Хи-квадрат для `vs` и `gear`. Оформите выводы в таблице: тест, p-value, интерпретация.

#### Тема 4. Введение в регрессию

##### Опрос 4:

1. Построение модели парной линейной регрессии. Обсуждение интерпретации регрессионных коэффициентов
2. Интерпретация коэффициентов для факторных и интервальных переменных
3. Построение модели линейной регрессии с несколькими предикторами. Сравнение интерпретации регрессионных коэффициентов с моделью парной линейной регрессии
4. Интерпретация коэффициента детерминации для оценки качества модели;

### **Решение и анализ практических заданий 2.**

Для заданной зависимой переменной рассчитайте простую регрессионную модель с одним предиктором. Проинтерпретируйте результаты теста с точки зрения значимости и размера эффекта, а также качества модели. Добавьте еще один предиктор: как изменилась модель? Аргументируйте, почему. Добавьте порядковый предиктор. Дайте интерпретацию эффекта.

### **Домашнее задание 3**

На данных `mtcars` постройте простую регрессию  $\text{mpg} \sim \text{wt}$ , интерпретируйте коэффициент. Добавьте предиктор `hp`, объясните изменения. Преобразуйте `cyl` в фактор, добавьте в модель. Сравните  $R^2$ .

## **Тема 5-7. Множественная регрессия**

### **Опрос 5:**

1. Множественная регрессия. Понятие контрольных переменных
2. Стандартизованные и нестандартизованные коэффициенты
3. Типы регрессионных графиков. Интерпретация.
4. Типы регрессионных предпосылок. Подходы к диагностике.
5. Мультиколлинеарность: диагностика, подходы к исправлению
6. Гетероскедастичность: диагностика, подходы к исправлению
7. Нелинейность: диагностика, подходы к исправлению
8. Выбросы: диагностика, подходы к исправлению

### **Решение и анализ практических заданий 3.**

Для заданной зависимой переменной и набора предикторов, постройте регрессионную модель методом последовательного добавления. Какие переменные вы используете как контрольные, что это значит? Дайте интерпретацию с помощью стандартизованных и нестандартизованных коэффициентов. В чем разница? Определите наиболее сильный предиктор. Постройте регрессионные графики. Проведите диагностику качества модели, исправьте выявленные проблемы.

### **Домашнее задание 4**

Для `mtcars` создайте модель  $\text{mpg} \sim \text{wt} + \text{hp} + \text{factor}(\text{cyl})$ . Проверьте мультиколлинеарность (VIF). Постройте график остатков vs предсказанные значения. Удалите выбросы (критерий Кука), пересчитайте модель.

## **Тема 8-10. Логистическая регрессия**

### **Опрос 6:**

1. Отношение шансов. Максимальное правдоподобие
2. Бинарная логистическая регрессия. Интерпретация результатов логистической регрессии.
3. Предсказанные вероятности.
4. Предельные эффекты. Средние предельные эффекты (AME)
5. Визуализация результатов. Типы графиков. Интерпретация.
6. Диагностика: таблица классификации, чувствительность, специфичность.

## 7. Площадь под кривой. AUC ROC.

### Решение и анализ практических заданий 4.

Для заданной зависимой переменной и набора предикторов, постройте логистическую регрессионную модель методом последовательного добавления. Дайте интерпретацию с помощью нестандартизованных коэффициентов и отношения шансов. В чем разница? Посчитайте и интерпретируйте предельные эффекты. Определите наиболее сильный предиктор. Постройте регрессионные графики. Проведите диагностику качества модели, исправьте выявленные проблемы.

### Домашнее задание 5

На данных Titanic (из `titanic::titanic_train`). Постройте модель  $\text{Survived} \sim \text{Age} + \text{Pclass}$ , интерпретируйте OR. Рассчитайте AME для Pclass. Постройте ROC-кривую, вычислите AUC.

## Тема 11-12. Факторный анализ.

### Опрос 7:

1. Обсуждение примеров латентных признаков
2. Эксплораторный факторный анализ.
3. Разработка наблюдаемых индикаторов для этих латентных признаков
4. Реализация в R факторного анализа для измерения латентного признака
5. Сравнение моделей факторного анализа с разным количеством факторов, разными видами вращений
6. Конфирматорный факторный анализ. Пакет `lavaan`. Критерии качества модели.

### Решение и анализ практических заданий 5.

Рассмотрите данный набор переменных. Какое количество факторов вы ожидаете для них обнаружить? Проведите эксплораторный факторный анализ. Сколько факторов выделяется? Коррелируют ли они? Какую долю дисперсии объясняет каждый фактор? Дайте содержательную интерпретацию полученным факторам.

### Домашнее задание 6

Для `psych::bfi` (5 личностных черт). Проведите EFA с вращением `varimax`, выберите 2-3 фактора. Назовите факторы на основе нагрузок ( $>0.5$ ). Сравните с CFA (`lavaan`), оцените CFI/RMSEA.

## Тема 13-14. Продвинутая классификация данных

### Опрос 8:

1. Обсуждение примеров реальных задач, требующих построения классификации объектов
2. Кластерный анализ. Основные алгоритмы.
3. Построение классификации объектов в R с помощью алгоритма k- средних
4. Сравнение моделей с разным количеством кластеров
5. Обсуждение критериев оценки качества полученной классификации
6. Деревья решений

### Решение и анализ практических заданий 6.

Для заданной базы данных, выберите переменные, на основе которых можно кластеризовать наблюдения. Выберите алгоритм. Определите количество кластеров. Проведите кластеризацию и дайте интерпретацию полученным кластерам. Оцените качество.

### Домашнее задание 7

Для iris выполните кластеризацию k-means (k=3), сравните с Species. Постройте дендрограмму (иерархическая кластеризация). Оцените точность через матрицу сопряженности.

### 3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

#### Форма промежуточной аттестации:

**2 модуль – зачет**, который проходит в форме тестирования.

**3 модуль – зачет с оценкой**, который проходит в форме итогового тестирования.

Перед зачетом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине.

Таблица 3

#### Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
<b>2 модуль – зачет-тестирование</b>	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	100-41% правильных ответов	Зачтено
				40-0% правильных ответов	Не зачтено
<b>3 модуль - зачет с оценкой / тестирование</b>	ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	З (ОПК-3) У (ОПК-3) В (ОПК-3)	100-81% правильных ответов	Зачтено, отлично
				80-61% правильных ответов	Зачтено, хорошо
				60-41% правильных ответов	Зачтено, удовлетворительно
				40-0% правильных ответов	Не зачтено, неудовлетворительно

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 3а.

Таблица 3а

#### Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценках «зачтено» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценке «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценке «не зачтено, неудовлетворительно», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

#### **4 Задания к промежуточной аттестации**

Требования к тестам

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

##### **Задания закрытого типа**

*Базовый уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте - 1 балл; ответ отличен от эталонного - 0 баллов.

*Повышенный уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балла; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

##### **Комбинированные задания**

*Базовый уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

*Повышенный уровень сложности:* задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутствует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

### **Задания открытого типа**

*Повышенный уровень сложности:* ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

*Высокий уровень сложности:* магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left( \frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

$k_n$  – максимально возможное количество баллов за вопросы по компетенции,

$x_n$  – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

### **Задания к промежуточной аттестации**

#### **Тестирование**

**ОПК-3 (С) Способен прогнозировать социальные явления и процессы, выявлять социально значимые проблемы и вырабатывать пути их решения на основе использования научных теорий, концепций, подходов и социальных технологий**

#### **Задания закрытого типа (базовый уровень сложности)**

##### **Задание 1**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какая мера центральной тенденции наиболее устойчива к выбросам?

А) Среднее арифметическое

Б) Медиана

В) Мода

Г) Дисперсия

Правильный ответ:

##### **Задание 2**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой график лучше всего подходит для визуализации распределения непрерывной переменной?

А) Столбчатая диаграмма

Б) Круговая диаграмма

В) Гистограмма

Г) Линейный график

Правильный ответ:

### **Задание 3**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.  
Вопрос: Какой тест следует использовать для сравнения средних значений в двух независимых группах?

- А) t-тест
- Б) Хи-квадрат
- В) Биномиальный тест
- Г) ANOVA

Правильный ответ:

### **Задание 4**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.  
Вопрос: Что показывает доверительный интервал?

- А) Диапазон наиболее вероятных значений параметра
- Б) Разброс данных
- В) Стандартное отклонение
- Г) Уровень значимости

Правильный ответ:

### **Задание 5**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.  
Вопрос: Какой коэффициент показывает силу линейной связи между переменными?

- А) Коэффициент детерминации
- Б) Коэффициент корреляции Пирсона
- В) Стандартная ошибка
- Г) F-статистика

Правильный ответ:

### **Задание 6**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.  
Вопрос: Какой метод используется для оценки параметров в линейной регрессии?

- А) Метод максимального правдоподобия
- Б) Метод наименьших квадратов
- В) Бутстреп-метод
- Г) Метод моментов

Правильный ответ:

### **Задание 7**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.  
Вопрос: Какой тип переменной представляет собой "уровень образования" (начальное, среднее, высшее)?

- А) Интервальная
- Б) Порядковая
- В) Номинальная
- Г) Дихотомическая

Правильный ответ:

### **Задание 8**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.  
Вопрос: Какой тест используется для проверки нормальности распределения?

- А) t-тест
- Б) Тест Шапиро-Уилка
- В) Тест Левена

Г) Тест Бартлетта

Правильный ответ:

### **Задание 9**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Что означает  $p\text{-value} < 0.05$ ?

- А) Нулевая гипотеза верна
- Б) Результаты статистически значимы
- В) Эффект очень сильный
- Г) Выборка слишком мала

Правильный ответ:

### **Задание 10**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой метод используется для анализа категориальных зависимых переменных?

- А) Линейная регрессия
- Б) Логистическая регрессия
- В) Факторный анализ
- Г) Кластерный анализ

Правильный ответ:

### **Задание 11**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой график используется для проверки линейности в регрессии?

- А) Q-Q plot
- Б) Scatter plot
- В) Box plot
- Г) Bar plot

Правильный ответ:

### **Задание 12**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой показатель интерпретируется в логистической регрессии?

- А) Коэффициент корреляции
- Б) Отношение шансов
- В) R-квадрат
- Г) F-статистика

Правильный ответ:

### **Задание 13**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой метод используется для уменьшения размерности данных в отношении переменных?

- А) Регрессионный анализ
- Б) Факторный анализ
- В) Кластерный анализ
- Г) Дисперсионный анализ

Правильный ответ:

### **Задание 14**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой метод используется для группировки похожих наблюдений?

- А) Регрессионный анализ
- Б) Кластерный анализ

- В) Факторный анализ  
Г) Дисперсионный анализ  
Правильный ответ:

#### **Задание 15**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Какой показатель измеряет долю объясненной дисперсии в регрессии?

- А) Коэффициент корреляции  
Б) R-квадрат  
В) Стандартная ошибка  
Г) p-value

Правильный ответ:

#### **Задания закрытого типа (повышенный уровень сложности)**

#### **Задание 16**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: Исследователь получил коэффициент корреляции  $r = 0.15$  ( $p < 0.05$ ) между временем учебы и оценками. Как интерпретировать этот результат?

- А) Сильная значимая связь  
Б) Слабая, но статистически значимая связь  
В) Связь отсутствует  
Г) Отрицательная связь

Правильный ответ:

#### **Задание 17**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В регрессионной модели с двумя предикторами получены  $VIF = 1.2$  и  $4.8$ . Что это означает?

- А) Нет мультиколлинеарности  
Б) Умеренная мультиколлинеарность для одного предиктора  
В) Сильная мультиколлинеарность для обоих предикторов  
Г) Недостаточно данных для вывода

Правильный ответ:

#### **Задание 18**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: При анализе данных о доходах обнаружено, что 95% ДИ для среднего дохода составляет [50000, 60000]. Как интерпретировать этот результат?

- А) 95% населения имеют доход в этом диапазоне  
Б) С вероятностью 95% истинное среднее находится в этом диапазоне  
В) Разброс доходов составляет 10000  
Г) Средний доход точно равен 55000

Правильный ответ:

#### **Задание 19**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В логистической регрессии для переменной "пол" (1=мужчины) получено  $OR = 2.5$ . Как это интерпретировать?

- А) Мужчины имеют в 2.5 раза больше шансов на исход  
Б) Женщины имеют в 2.5 раза больше шансов на исход  
В) Связь незначима  
Г) Модель объясняет показатели мужчин в 2.5 раза лучше

Правильный ответ:

### **Задание 20**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: При анализе данных ANOVA показала  $F(2, 97) = 5.67$ ,  $p = 0.005$ . Какой вывод?

- А) Нет различий между группами
- Б) Есть статистически значимые различия между группами
- В) Все группы одинаковы
- Г) Недостаточно данных для вывода

Правильный ответ:

### **Задание 21**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В факторном анализе получено, что 3 фактора объясняют 65% дисперсии. Какой вывод?

- А) Модель плохая, нужно больше факторов
- Б) Модель удовлетворительная
- В) Нужно уменьшить число факторов
- Г) Нельзя интерпретировать

Правильный ответ:

### **Задание 22**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В кластерном анализе дендрограмма показывает 2 четких группы. Что это означает?

- А) Данные плохого качества
- Б) В данных есть 2 естественных кластера
- В) Нужно больше кластеров
- Г) Метод не подходит

Правильный ответ:

### **Задание 23**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: При анализе интерактивного эффекта в регрессии  $p < 0.01$  для произведения предикторов. Какой вывод?

- А) Эффект незначим
- Б) Есть статистически значимое взаимодействие
- В) Модель неадекватна
- Г) Нужно удалить предикторы

Правильный ответ:

### **Задание 24**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R команда `lm(y ~ x1 + x2, data=df)` означает:

- А) Кластерный анализ
- Б) Линейную регрессию
- В) Логистическую регрессию
- Г) Факторный анализ

Правильный ответ:

### **Задание 25**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R команда `glm(y ~ x, family=binomial, data=df)` означает:

- А) Линейную регрессию
- Б) Логистическую регрессию
- В) Пуассоновскую регрессию

Г) Факторный анализ

Правильный ответ:

### **Задание 26**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R после `fit <- kmeans(df, centers=3)`, `fit$cluster` содержит:

- А) Коэффициенты регрессии
- Б) Назначения кластеров для наблюдений
- В) р-значения
- Г) Факторные нагрузки

Правильный ответ:

### **Задание 27**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R `summary(lm(y~x))$r.squared` возвращает:

- А) Коэффициент корреляции
- Б) R-квадрат модели
- В) р-значение
- Г) Стандартную ошибку

Правильный ответ:

### **Задание 28**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R `confint(lm(y~x))` возвращает:

- А) Коэффициенты регрессии
- Б) Доверительные интервалы для коэффициентов
- В) р-значения
- Г) Остатки

Правильный ответ:

### **Задание 29**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R `factanal(df, factors=3)` выполняет:

- А) Линейную регрессию
- Б) Факторный анализ с 3 факторами
- В) Кластерный анализ
- Г) ANOVA

Правильный ответ:

### **Задание 30**

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос.

Вопрос: В R `plot(lm(y~x))` создает:

- А) Матрицу корреляций
- Б) Диагностические графики для регрессии
- В) Дендрограмму
- Г) Box plot

Правильный ответ:

## **Задания комбинированного типа (повышенный уровень сложности)**

### **Задание 31**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Центральная предельная теорема применима только к нормально распределённым переменным.
2. По мере увеличения объёма выборки выборочное среднее стремится к среднему генеральной совокупности.
3. Генеральная совокупность всегда ограничена конкретной выборкой.
4. Центральная предельная теорема позволяет использовать нормальное приближение даже при неизвестном распределении переменной.
5. Центральная предельная теорема не применима к интервальным шкалам.

Правильные ответы:

Эталонный пример обоснования:

### **Задание 32**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Доверительный интервал даёт границы, в которых обязательно содержится значение в конкретном исследовании.
2. Более широкий доверительный интервал соответствует большему уровню уверенности.
3. Доверительный интервал зависит только от объёма выборки, но не от дисперсии.
4. При увеличении размера выборки доверительный интервал сужается.
5. Доверительный интервал применяется только в логистической регрессии.

Правильные ответы:

Эталонный пример обоснования:

### **Задание 33**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. ANOVA используется для сравнения более двух групп по среднему значению.
2. Биномиальный тест применим только к непрерывным переменным.
3. Хи-квадрат позволяет оценить связь между двумя категориальными переменными.
4. T-test не может быть использован при сравнении двух средних.
5. Z-стандартизация преобразует данные в значения между 0 и 1.

Правильные ответы:

Эталонный пример обоснования:

### **Задание 34**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. В простой линейной регрессии предиктор может быть как количественным, так и категориальным (при соответствующем кодировании).
2. Метод наименьших квадратов минимизирует сумму абсолютных отклонений.
3. Коэффициент при предикторе показывает изменение зависимой переменной при изменении предиктора на единицу.
4. Линейная регрессия не допускает использование бинарных предикторов.
5. Остатки в модели должны иметь линейную зависимость от предикторов.

Правильные ответы:

Эталонный пример обоснования:

### **Задание 35**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Множественная регрессия позволяет оценить влияние каждого предиктора при контроле других.
2. Мультиколлинеарность означает сильную корреляцию между зависимой и независимой переменной.
3. Стандартизированные коэффициенты регрессии позволяют сравнивать вклад разных предикторов.
4. Наличие выбросов не влияет на результаты линейной регрессии.
5. Интерактивные эффекты показывают, как влияние одного предиктора изменяется при разных значениях другого.

Правильные ответы:

Эталонный пример обоснования:

### **Задания комбинированного типа (высокий уровень сложности)**

#### **Задание 36**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. В модели логистической регрессии интерпретация коэффициента происходит через логарифм отношения шансов.
2. Предельные эффекты показывают абсолютное изменение вероятности при изменении предиктора.
3. Коэффициенты логистической регрессии можно напрямую сравнивать с коэффициентами линейной регрессии.
4. Логистическая регрессия не позволяет использовать категориальные переменные.
5. Визуализация предсказанных вероятностей помогает интерпретировать модель в прикладном контексте.

Правильные ответы:

Эталонный пример обоснования:

#### **Задание 37**

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Факторный анализ используется для выявления скрытых латентных конструкций на основе набора переменных.
2. В конфирматорном факторном анализе структура факторов заранее задаётся и проверяется на данных.
3. Факторный анализ можно применять к любой категориальной переменной без ограничений.
4. Эксплораторный факторный анализ проверяет точность уже существующей модели.
5. Факторный анализ используется только в биомедицинских исследованиях.

Правильные ответы:

Эталонный пример обоснования:

### Задание 38

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Кластерный анализ позволяет выявлять группы наблюдений с похожими характеристиками.
2. Решения дерева всегда линейны по отношению к предикторам.
3. В иерархической кластеризации возможно заранее не задавать число кластеров.
4. Алгоритмы дерева решений полезны только в визуализации, но не в прогнозировании.
5. В кластерном анализе нельзя использовать стандартные меры расстояния.

Правильные ответы:

Эталонный пример обоснования:

### Задание 39

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. В R команда `lm(y ~ x1 + x2, data = df)` строит модель множественной линейной регрессии.
2. Вызов `summary(model)` возвращает значения  $R^2$  и p-value для предикторов.
3. Команда `predict(model)` в R выдаёт график остатков модели.
4. `ggplot2` используется для визуализации структуры факторной модели.
5. `cor(df)` возвращает коэффициенты корреляции между всеми переменными в `df`.

Правильные ответы:

Эталонный пример обоснования:

### Задание 40

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. При интерпретации регрессионной модели важно учитывать как значения коэффициентов, так и доверительные интервалы.
2. Отрицательный коэффициент в модели всегда означает плохую модель.
3. Графики остатков используются для диагностики соответствия модели требованиям.
4. Стандартизированные коэффициенты применимы только в логистической регрессии.
5. Визуализация предсказанных значений помогает связать модель с прикладными вопросами исследования.

Правильные ответы:

Эталонный пример обоснования:

## **Задания открытого типа (повышенный уровень сложности)**

### **Задание 41**

Инструкция: Объясните, что представляет собой генеральная совокупность и как она соотносится с выборкой.

Эталонный ответ:

### **Задание 42**

Инструкция: Обоснуйте значение центральной предельной теоремы для статистического анализа.

Эталонный ответ:

### **Задание 43**

Инструкция: Объясните различие между точечной и интервальной оценкой параметров.

Эталонный ответ:

### **Задание 44**

Инструкция: Объясните, зачем проводится стандартизация переменных в регрессионном анализе.

Эталонный ответ:

### **Задание 45**

Инструкция: Раскройте назначение бинарной логистической регрессии.

Эталонный ответ:

### **Задание 46**

Инструкция: Объясните, почему доверительный интервал может сужаться при увеличении размера выборки.

Эталонный ответ:

### **Задание 47**

Инструкция: Обоснуйте, зачем проверяется мультиколлинеарность в множественной регрессии.

Эталонный ответ:

### **Задание 48**

Инструкция: Объясните, зачем используется метод наименьших квадратов в регрессии.

Эталонный ответ:

### **Задание 49**

Инструкция: Объясните, чем эксплораторный факторный анализ отличается от конфирматорного.

Эталонный ответ:

### **Задание 50**

Инструкция: Обоснуйте значение кластерного анализа в классификации социальных явлений.

Эталонный ответ:

## Задания открытого типа (высокий уровень сложности)

### Задание 51

Инструкция: На графике распределения переменной `income` наблюдается выраженный положительный скошенный хвост. Объясните, как это влияет на выбор меры центральной тенденции и тип используемой регрессии.

Эталонный ответ:

### Задание 52

Инструкция: Проанализируйте вывод функции `t.test()`: `t.test(support ~ group, data = df)`, где `p-value = 0.07`, `mean in group A = 3.2`, `mean in group B = 2.8`.

Объясните, как интерпретировать результат.

Эталонный ответ:

### Задание 53

Инструкция: Рассмотрите вывод простой линейной регрессии:

```
model <- lm(support ~ income, data = data)
summary(model)
```

Коэффициент при `income` положителен и значим, но значение  $R^2 = 0.04$ . Объясните, как интерпретировать результат.

Эталонный ответ:

### Задание 54

Инструкция: В результате множественной регрессии один из коэффициентов оказался значимым, но заметно растет при включении новых переменных. Объясните возможную причину и как её диагностировать.

Эталонный ответ:

### Задание 55

Инструкция: На графике остатков множественной регрессии вы наблюдаете веерообразное распределение остатков. Объясните, что это значит.

Эталонный ответ:

### Задание 56

Инструкция: Рассмотрите результаты логистической регрессии:

```
glm(voted ~ age + gender + education, family = binomial, data = data)
```

При этом OR (отношение шансов) для образования равно 1.7, но CI включает 1. Объясните, как интерпретировать результат.

Эталонный ответ:

### Задание 57

Инструкция: В модели логистической регрессии получены предельные эффекты (`margins`) для бинарной переменной `gender` (референтная категория мужчины). Значение = 0.08. Объясните, что означает этот результат.

Эталонный ответ:

### Задание 58

Инструкция: Вы провели факторный анализ, и первая компонента объясняет 45%

дисперсии, вторая — 15%. При этом факторные нагрузки переменных  $x_1$ ,  $x_2$ ,  $x_3$  на первую компоненту составляют: 0.78, 0.75 и 0.72. Объясните, что это означает.

Эталонный ответ:

#### Задание 59

Инструкция: При построении дерева решений с использованием `rpart()` вы получили дерево, в котором переменная `age` используется на всех уровнях. Объясните возможные причины и как их интерпретировать.

Эталонный ответ:

#### Задание 60

Инструкция: Вы провели кластеризацию методом `kmeans()` и получили средние значения переменных внутри кластеров. Один из кластеров показывает значительно более низкий уровень удовлетворённости, чем другие. Объясните, как использовать этот результат для выявления социальной проблемы и её анализа.

Эталонный ответ:

### 5 Средства оценки индикаторов достижения компетенций

Таблица 4

**Средства оценки индикаторов достижения компетенций**

Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ОПК-3 (С)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-3.5. ИД.ОПК-3.6.	Опрос (О), домашнее задание (ДЗ), решение и анализ практических заданий (РАПЗ), тест

Таблица 5

**Описание средств оценки индикаторов достижения компетенций**

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Опрос	Магистрант в ходе подготовки и участия в опросах по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получившую развитие в рамках данной дисциплины: 1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов
Домашнее задание	Магистрант должен быть готовым в ходе подготовки и представления домашнего задания по темам дисциплины, выполнять следующие действия: 1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
	социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов
Решение и анализ практических заданий	Магистрант должен быть готовым в ходе подготовки и решения практических заданий по темам дисциплины, выполнять следующие действия: 1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов
Тест	Магистрант должен быть готовым в ходе подготовки к тесту, выполнять следующие действия: 1. Анализирует проблемы развития социальных явлений и процессов с использованием процедур качественного анализа для обработки социологических данных. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для анализа и прогнозирования социальных явлений и процессов. Выявляет социально значимые проблемы и предлагает пути их решения на основе социологической теории и социологических методов исследования. Научно обосновывает постановку фундаментальных и прикладных социологических исследований для решения социально значимых проблем. Предлагает описательные, объяснительные и прогнозные модели социальных явлений и процессов на основе социологических теорий. Разрабатывает предложения по совершенствованию социологических концепций описания и объяснения социальных явлений и процессов