

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 14.08.2025 17:03:06

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1225917e739a70e315177f88591

**Автономная некоммерческая образовательная организация высшего образования
«Европейский университет в Санкт-Петербурге»**

Школа вычислительных социальных наук

УТВЕРЖДАЮ:

Ректор

В.В. Волков

« 26 »

августа

2025 г.

Протокол УС № 2

от 26.02.2025 г.



**Рабочая программа дисциплины
Анализ данных-2**

**образовательная программа
направление подготовки
39.04.01 Социология**

**направленность (профиль) программы
«Вычислительная социология»
уровень высшего образования – магистратура**

Программа двух квалификаций:

- «магистр» по направлению подготовки 39.04.01 Социология;
- дополнительная квалификация – «магистр» по направлению подготовки 09.04.03 Прикладная информатика

язык обучения – русский
форма обучения - очная

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-2**», 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-2**» является дисциплиной базовой части основной профессиональной образовательной программы высшего образования «Вычислительная

В фокусе курса – техники, позволяющие преодолеть ограничения классических опросов и получить более точные и содержательные результаты. Ключевые модули: Обогащенные опросы (Enriched Surveys), регрессионный анализ с постстратификацией, методы корректировки смещений выборки, эксперименты в опросах, модели выбора (Choice Models). В курсе сделан акцент на сравнительный анализ традиционных и вычислительных методов, валидацию результатов через триангуляцию подходов и критическую работу с ограничениями каждого метода.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы, 144 часа.

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.....	8
6.1 Общие положения	8
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:	10
6.5 Перечень учебно-методического обеспечения для самостоятельной работы.....	10
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.....	12
7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации	13
7.4 Типовые задания к промежуточной аттестации.....	16
7.5 Средства оценки индикаторов достижения компетенций.....	20
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	21
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.....	21
9.1 Программное обеспечение	22
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	22
9.3 Лицензионные электронные ресурсы библиотеки Университета	22
9.4 Электронная информационно-образовательная среда Университета.....	23
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.....	23
ПРИЛОЖЕНИЕ 1	25

1. НАИМЕНОВАНИЕ, ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины «Количественный анализ данных-2» 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. Дать понимание основных концептов математической статистики.
2. Рассмотреть основные методы статистического анализа данных.
3. Развить навыки работы в программной среде R.
4. 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 3 (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	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)

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
	описания и объяснения социальных явлений и процессов	

В результате освоения дисциплины магистрант должен:

- ☐ **знать:** основные концепты математической статистики
- ☐ **уметь:** применять основные методы статистического анализа данных
- ☐ **владеть:** навыками работы в программной среде, формулировать

исследовательские задачи в привязке к статистической методологии.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Анализ данных-2» является дисциплиной обязательной части Блока 1 «Дисциплины (модули)» образовательной программы «Вычислительная социология». Курс читается во втором и третьем модулях, форма промежуточной аттестации – 6 модуль – зачет, 7 модуль - зачет с оценкой.

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 4 зачетных единицы, 144 часа.

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)		88	-	-	-	-	-	80	8	-	-	-
Промежуточная аттестация	форма	Зачет, зачет с оценкой	-	-	-	-	-	Зачет	Зачет с оценкой	-	-	-
	час.	-	-	-	-	-	-	-	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)		144/4	-	-	-	-	-	108/3	36/1	-	-	-

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	Качество данных. Основные проблемы исходных данных	Основные характеристики данных. Типы распределений. Типичные проблемы с данными: скошенные данные, пропуски, цензурированные и усеченные данные.	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	Коррекция выборки. Работа с пропущенными значениями.	Робастные оценки. Типы сдвига. Взвешивание, типы опросных весов. Ресэмплинг. Бутстраппинг. Удаление. Импутация. Типы импутации. Множественная регрессионная импутация. Импутация с помощью деревьев решений; k-means импутация.	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	Кластерный анализ	Кластеризация. Определение числа кластеров. Методы кластерного анализа. Иерархический кластерный анализ. Дендрограммы. Интерпретация результатов.	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	Анализируя принятие решений. Деревья решений. Байесовские сети.	Деревья решений. Основные алгоритмы. Интерпретация результатов. Байесовские сети. Отличия от деревьев решений. Интерпретация.	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	Продвинутые методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)	Регрессионные методы для особых случаев. Квантильная регрессия. Определение количества квантилей. Интерпретация. Общие аддитивные модели GAM для моделирования нелинейности.	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)
6	Введение в моделирование каузальности: discontinuity regression. Difference in difference.	Определение и типы каузальности. Методы для определения каузальности. Метод discontinuity regression: основные положения, условия	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)
	Анализ медиации	применения, интерпретация. Анализ Difference in difference: основные положения, условия применения, интерпретация. Анализ медиации: типы анализа, интерпретация результатов.			

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	Качество данных. Основные проблемы исходных данных	20	2	2		22	Опрос, обсуждение литературы
Тема 2	Коррекция выборки. Работа с пропущенными значениями.	20	4	4		22	
Тема 3	Кластерный анализ	20	4	4		22	Опрос, решение и анализ практических заданий, дискуссия
Тема 4	Анализируя принятие решений. Деревья решений. Байесовские сети.	20	4	4		22	
Промежуточная аттестация 6 модуль							Зачет
Тема 5	Продвинутые методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)	18	8	8		4	Опрос, решение и анализ практических заданий, дискуссия
Тема 6	Введение в моделирование каузальности: discontinuity regression. Difference in difference. Анализ медиации	18	6	6		4	Опрос, решение и анализ практических заданий
Промежуточная аттестация 7 модуль		-	-	-	-	-	Зачет с оценкой
Всего:		144/4	28	28	-	88	-

*Примечание: формы текущего контроля успеваемости: опрос (О), дискуссия (Д), обсуждение литературы (ОЛ), 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. Оценка качества данных. Основные проблемы исходных данных.

1. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 16 часов. Итого: 22 часа.

Тема 2. Коррекция выборки. Работа с пропущенными значениями.

1. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 16 часов. Итого: 22 часа.

Тема 3. Кластерный анализ

1. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 16 часов. Итого: 22 часа.

Тема 4. Анализируя принятие решений. Деревья решений. Байесовские сети.

1. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 16 часов. Итого: 22 часа.

Тема 5. Продвинутое методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)

1. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 6. Введение в моделирование каузальности: discontinuity regression. Difference in difference. Анализ медиации.

1. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

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

1. Типы распределений. Типичные проблемы с данными: скошенные данные, пропуски, цензурированные и усеченные данные.
2. Робастные оценки. Типы сдвига.
3. Взвешивание, типы опросных весов.
4. Ресэмплинг. Бутстраппинг. Удаление.
5. Импутация. Типы импутации. Множественная регрессионная импутация. Импутация с помощью деревьев решений; k-means импутация.
6. Кластеризация. Определение числа кластеров. Методы кластерного анализа. Иерархический кластерный анализ. Дендрограммы. Интерпретация результатов.
7. Деревья решений. Основные алгоритмы. Интерпретация результатов.
8. Байесовские сети. Отличия от деревьев решений. Интерпретация.
9. Квантильная регрессия. Определение количества квантилей. Интерпретация.
10. Общие аддитивные модели GAM для моделирования нелинейности.
11. Методы для определения каузальности.
12. Метод discontinuity regression: основные положения, условия применения, интерпретация.
13. Анализ Difference in difference: основные положения, условия применения, интерпретация.

Анализ медиации: типы анализа, интерпретация результатов.

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Кузнецов И.Н. Основы научных исследований: Учебное пособие [Электронный ресурс]. / И.Н. Кузнецов. – 4-е изд. — М.: Изд-во «Дашков и К», 2018. – 284 с. – URL: <http://znanium.com/bookread2.php?book=415064>

2. Социально-экономическая статистика: Учебное пособие [Электронный ресурс] / Я.С. Мелкумов. - М.: НИЦ Инфра-М, 2013. - 236 с. – URL: <http://znanium.com/bookread2.php?book=366756>

3. Статистика: Учебное пособие / Е.В. Иода. - М.: Вузовский учебник: НИЦ Инфра-М, 2012. - 303 с. – [Электронный ресурс]. – URL: <http://znanium.com/bookread2.php?book=260143>

4. Шафранов-Куцев, Г. Ф. Социология: курс лекций [Электронный ресурс] : учеб. пособие / Г.Ф. Шафранов-Куцев. – 3-е изд., перераб. и доп. – М. : Логос, 2012. – 368 с. — URL: <http://znanium.com/bookread2.php?book=469419>

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Анализ данных-2» 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, участия в дискуссиях, 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
Качество данных. Основные проблемы исходных 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, обсуждение литературы	зачтено/ не зачтено
Коррекция выборки. Работа с пропущенными значениями.	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)		

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Кластерный 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, дискуссия	зачтено/ не зачтено
Анализируя принятие решений. Деревья решений. Байесовские сети.	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)		
Продвинутые методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)	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, дискуссия	зачтено/ не зачтено
Введение в моделирование каузальности: discontinuity regression. Difference in difference. Анализ медиации	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	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
Дискуссия	пассивность, участие без представления аргументов и обоснования точки зрения, несформированность навыков профессиональной коммуникации в группе — не зачтено представление аргументированной научной позиции, обоснование точки зрения в дискуссии, демонстрация навыков профессиональной коммуникации в группе — зачтено
Обсуждение литературы	магистрант не знает или указывает не всех наиболее значимых авторов, писавших по данной проблематике, приводит отдельные работы или не может привести пример литературы по данной проблематике, неуверенно и/или с существенными недочетами, ошибками излагает идеи из книг по обсуждаемому вопросу, не выделяет основные темы и вопросы по прочитанной литературе — не зачтено магистрант указывает авторов, приводит широкий круг работ по данной проблематике, уверенно и профессионально грамотно излагает идеи из книг по обсуждаемому вопросу, выделяет основные темы и вопросы по прочитанной литературе — зачтено
Решение и анализ практических задач	Магистрант выполняет работу частично или с существенными недочетами – не зачтено,

Формы текущего контроля успеваемости	Критерии оценивания
	Полное и правильное выполнение заданий работы – зачтено

7.2 Контрольные задания для текущей аттестации

Примерный материал опросов, дискуссий, обсуждений литературы

Тема 1-2: Оценка качества данных. Основные проблемы исходных данных.

Коррекция выборки. Работа с пропущенными значениями.

Опрос 1:

1. Основные характеристики данных. Типы распределений.
2. Типичные проблемы с данными: скошенные данные, пропуски.
3. Цензурированные и усеченные данные.
4. Робастные оценки. Типы сдвига. Взвешивание, типы опросных весов.

Конструирование весов. Применение весов в анализе.

Обсуждение литературы по теме лекции.

Тема 3. Кластерный анализ.

Опрос 2:

1. Определение числа кластеров.
2. Методы кластерного анализа.
3. Иерархический кластерный анализ. Дендрограммы.
4. Интерпретация результатов.

Тема 4-5. Анализируя принятие решений. Деревья решений. Байесовские сети.

Продвинутые методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)

Опрос 3:

1. Деревья решений. Основные алгоритмы. Интерпретация результатов.
2. Байесовские сети. Отличия от деревьев решений. Интерпретация.
3. Квантильная регрессия.
4. Общие аддитивные модели GAM для моделирования нелинейности.

Дискуссия 1 по теме лекции.

Темы дискуссии:

1. Для каких задач стоит применять деревья решений?
2. В каких случаях байесовские сети будут более применимы?
3. Какую проблему решает квантильная регрессия? В каких ситуациях она неприменима или бесполезна?

Решение и анализ практических заданий 1.

Практическое задание 1:

Постройте два дерева решений: методом `ctree` и `part`. При желании, вы можете добавить дополнительные предикторы. Проведите необходимую пре-/постобрезку дерева. Нарисуйте оба дерева, а также выведите структуру дерева. Дайте интерпретацию результатам. Насколько они различаются для двух алгоритмов?

Практическое задание 2:

Постройте пуассоновскую регрессию, предсказывая количество детей с помощью переменных возраст, пол, брачный статус, доход, религиозная конфессия.

Проинтерпретируйте свою финальную модель: какие переменные значимы? Какой эффект они оказывают? Не забудьте экспонировать коэффициенты любым удобным вам способом, чтобы получить risk ratios.

Тема 6. Введение в моделирование каузальности: discontinuity regression. Difference in difference. Анализ медиации.

Опрос 4:

1. Определение и типы каузальности.
2. Методы для определения каузальности.
3. Discontinuity regression.
4. Difference in difference.
5. Анализ медиации.

Решение и анализ практических заданий 2.

Практическое задание 1:

На данных изменения уровня безработицы в США проверьте, как на безработицу повлияло повышение минимального уровня оплаты труда, с помощью регрессии Difference in Differences. Не забудьте добавить контрольные переменные. Проинтерпретируйте результаты: может ли изменение в оплате труда считаться причиной изменения в уровне безработицы?

Практическое задание 2:

Наша теоретическая модель предполагает, что образование родителей влияет на образование ребенка, которое затем влияет на уровень заработной платы этого выросшего ребенка. На наших данных воспроизведите эту модель с помощью пакета lavaan и проверьте, действительно ли образование ребенка медирует эффект образования родителей на доход ребенка. Каков коэффициент полной медиации?

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

Форма промежуточной аттестации:

6 модуль – зачет, который проходит в форме тестирования.

7 модуль – зачет с оценкой, который проходит в форме 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
6 модуль – зачет-тестирование	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
7 модуль - зачет оценкой / тестирование	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 6 модуль

Задания для проверки компетенции ОПК-3 (С) (Способен прогнозировать социальные явления и процессы, выявлять социально значимые проблемы и вырабатывать пути их решения)

Задания закрытого типа (базовый уровень сложности)

Задание 1

Инструкция: Выберите правильное утверждение:

Вопрос: Какой метод наиболее подходит для обработки данных с пропущенными значениями, если пропуски носят случайный характер (MAR)?

1. Полное удаление наблюдений
2. Среднеарифметическая импутация
3. Множественная регрессионная импутация
4. Замена медианным значением

Поле для ответа:

Задания закрытого типа (повышенный уровень сложности)

Задание 1

Инструкция: Прочитайте задание и установите соответствие:

Вопрос: При анализе социальных данных исследователь столкнулся с различными проблемами. Установите соответствие между типом проблемы и рекомендуемым методом ее решения:

	Тип проблемы данных		Метод решения
1	Данные с правосторонним скосом	A	Логарифмическое преобразование
2	Цензурированные данные	B	Тобит-модель
3	Пропуски в категориальных данных	C	Импутация модой
4	Кластерная структура данных	D	Иерархическая кластеризация

Поле для ответа:

Задания комбинированного типа (повышенный уровень сложности)

Задание 1

Инструкция: Выберите два правильных утверждения и обоснуйте выбор.

Вопрос: Какие утверждения верно описывают особенности работы с социальными данными?

1. Бутстраппинг позволяет оценить устойчивость результатов при малых выборках
2. Для определения числа кластеров достаточно визуального анализа дендрограммы
3. Взвешивание данных устраняет проблему систематических ошибок измерения
4. Деревья решений требуют предварительного предположения о виде функциональной зависимости

Поле для ответа:

--	--

Обоснование: _____

Задания открытого типа (повышенный уровень сложности)

Задание 1

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ.

Вопрос: Разработайте стратегию анализа данных опроса о социальном неравенстве с 30% пропущенных значений. Включите:

1. Диагностику механизма пропусков
2. Обоснование выбора метода импутации
3. Методы проверки качества импутированных данных

Поле для ответа: _____

Задания открытого типа (высокий уровень сложности)

Задание 1

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ.

Вопрос: Постройте алгоритм прогнозирования уровня бедности на основе данных с проблемами:

1. Опишите методы предобработки данных (скос, пропуски)
2. Обоснуйте выбор между деревьями решений и байесовскими сетями
3. Предложите метод оценки устойчивости прогнозной модели

Поле для ответа: _____

Примеры тестовых заданий для промежуточной аттестации 7 модуль

Задания для проверки компетенции ОПК-3 (С) (Способен прогнозировать социальные явления и процессы, выявлять социально значимые проблемы и вырабатывать пути их решения)

Задания закрытого типа (базовый уровень сложности)

Задание 1

Инструкция: Выберите правильное утверждение.

Вопрос: Какой метод регрессионного анализа наиболее подходит для изучения влияния факторов на разные части распределения зависимой переменной (например, не только на среднее, но и на 10-й или 90-й процентиль)?

1. Линейная регрессия (OLS)
2. Логистическая регрессия
3. Квантильная регрессия
4. Ridge-регрессия

Поле для ответа:

☐

Задания закрытого типа (повышенный уровень сложности)

Задание 1

Инструкция: Прочитайте задание и установите соответствие:

Вопрос: Установите соответствие между методом каузального анализа и ситуацией, в которой он применяется.

	Метод анализа		Ситуация применения
--	---------------	--	---------------------

1	Discontinuity Regression	A	Оценка эффекта политики, где воздействие зависит от порогового значения переменной
2	Difference-in-Differences	B	Сравнение изменений в группе с воздействием и контрольной группой до и после
3	Медиационный анализ	C	Определение, как именно (через какие механизмы) X влияет на Y
4	GAM (Generalized Additive Models)	D	Моделирование нелинейных зависимостей без жестких параметрических ограничений

Поле для ответа:

Задания комбинированного типа (повышенный уровень сложности)

Задание 1

Инструкция: Выберите два правильных утверждения и обоснуйте выбор.

Вопрос: Какие утверждения верно описывают особенности методов каузального анализа?

1. Difference-in-Differences требует параллельных трендов до вмешательства.
2. Квантильная регрессия всегда дает более точные прогнозы, чем OLS.
3. Discontinuity Regression применим только при четком пороговом значении.
4. Медиационный анализ не требует контроля конфаундеров.

Поле для ответа:

--	--

Обоснование: _____

Задания открытого типа (повышенный уровень сложности)

Задание 1

Инструкция: Дайте развернутый ответ.

Вопрос: Предложите стратегию анализа влияния новой образовательной программы на успеваемость студентов, используя Difference-in-Differences. Опишите:

1. Как выбрать контрольную группу.
2. Какие переменные включить в модель.
3. Как проверить предположение о параллельных трендах.

Поле для ответа: _____

Задания открытого типа (высокий уровень сложности)

Задание 1

Инструкция: Дайте развернутый ответ.

Вопрос: Разработайте исследовательский дизайн для изучения нелинейного влияния дохода на уровень счастья с помощью GAM-моделей. Включите:

1. Обоснование выбора метода (почему не линейная регрессия?).
2. Как интерпретировать smooth-функции в GAM.
3. Какие диагностические тесты провести для проверки адекватности модели.

Поле для ответа: _____

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

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, дискуссия, обсуждение литературы, решение и анализ практических заданий, тест

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Опрос	<p>Магистрант в ходе подготовки и участия в опросах по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получившую развитие в рамках данной дисциплины:</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>
Обсуждение литературы	<p>Магистрант в ходе подготовки и участия в обсуждениях литературы по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получившую развитие в рамках данной дисциплины:</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>
Дискуссия	<p>Магистрант должен быть готовым в ходе подготовки и участия в дискуссии 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>

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
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. Татарникова, Т. М. Интеллектуальный анализ данных : учебное пособие / Т. М. Татарникова. - Москва ; Вологда : Инфра-Инженерия, 2024. - 172 с. - ISBN 978-5-9729-1772-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2169704>
2. Кабаков, Р. R в действии. Анализ и визуализация данных с использованием R и Tidyverse / Р. Кабаков ; пер. с англ. А. Н. Киселева. - 3-е изд. - Москва : ДМК Пресс, 2023. - 769 с. - ISBN 978-5-93700-173-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2150529>

8.2. Дополнительная литература

1. Дятлов, А.В. Анализ данных в социологии : учебник / А.В.Дятлов, Д.А.Гугуева ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2018. - 226 с. - ISBN 978-5-9275-2690-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1039664>
2. Социально-экономическая статистика: Учебное пособие [Электронный ресурс] / Я.С. Мелкумов. - М.: НИЦ Инфра-М, 2013. - 236 с. - URL: <http://znanium.com/bookread2.php?book=366756>
3. Статистика: Учебное пособие / Е.В. Иода. - М.: Вузовский учебник: НИЦ Инфра-М, 2012. - 303 с. - [Электронный ресурс]. - URL: <http://znanium.com/bookread2.php?book=260143>
4. Мелехина, Т. Л. Методы и средства научного исследования при освоении дисциплины Анализ данных : учебное пособие / Т. Л. Мелехина. - Москва : Прометей, 2023. - 138 с. - ISBN 978-5-00172-413-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2144361>

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. R — бесплатно

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. **Znanium.com** – Электронная библиотечная система (ЭБС) – <http://znanium.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://euspr.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-2»

**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, участия в дискуссиях, 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**

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Качество данных. Основные проблемы исходных 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, обсуждение литературы	зачтено/ не зачтено
Коррекция выборки. Работа с пропущенными значениями.	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)		
Кластерный 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, дискуссия	зачтено/ не зачтено
Анализируя принятие решений. Деревья решений. Байесовские сети.	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)		

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Продвинутые методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)	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, дискуссия	зачтено/ не зачтено
Введение в моделирование каузальности: discontinuity regression. Difference in difference. Анализ медиации	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	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
Дискуссия	пассивность, участие без представления аргументов и обоснования точки зрения, несформированность навыков профессиональной коммуникации в группе — не зачтено представление аргументированной научной позиции, обоснование точки зрения в дискуссии, демонстрация навыков профессиональной коммуникации в группе — зачтено
Обсуждение литературы	магистрант не знает или указывает не всех наиболее значимых авторов, писавших по данной проблематике, приводит отдельные работы или не может привести пример литературы по данной проблематике, неуверенно и/или с существенными недочетами, ошибками излагает идеи из книг по обсуждаемому вопросу, не выделяет основные темы и вопросы по прочитанной литературе — не зачтено магистрант указывает авторов, приводит широкий круг работ по данной проблематике, уверенно и профессионально грамотно излагает идеи из книг по обсуждаемому вопросу, выделяет основные темы и вопросы по прочитанной литературе — зачтено
Решение и анализ практических задач	Магистрант выполняет работу частично или с существенными недочетами – не зачтено, Полное и правильное выполнение заданий работы – зачтено

2 Контрольные задания для текущей аттестации

Материал дискуссий, опросов, практических заданий, обсуждения литературы

Тема 1-2: Оценка качества данных. Основные проблемы исходных данных. Коррекция выборки. Работа с пропущенными значениями.

Опрос 1:

- Основные характеристики данных. Типы распределений.
- Типичные проблемы с данными: скошенные данные, пропуски.
- Цензурированные и усеченные данные.

8. Робастные оценки. Типы скоса. Взвешивание, типы опросных весов. Конструирование весов. Применение весов в анализе.

Обсуждение литературы по теме лекции.

Тема 3. Кластерный анализ.

Опрос 2:

1. Определение числа кластеров.
2. Методы кластерного анализа.
3. Иерархический кластерный анализ. Дендрограммы.
4. Интерпретация результатов.

Тема 4-5. Анализируя принятие решений. Деревья решений. Байесовские сети. Продвинутое методы регрессионного анализа. Квантильная регрессия. Нелинейная регрессия (gam)

Опрос 3:

1. Деревья решений. Основные алгоритмы. Интерпретация результатов.
2. Байесовские сети. Отличия от деревьев решений. Интерпретация.
3. Квантильная регрессия.
4. Общие аддитивные модели GAM для моделирования нелинейности.

Дискуссия 1 по теме лекции.

Темы дискуссии:

1. Для каких задач стоит применять деревья решений?
2. В каких случаях байесовские сети будут более применимы?
3. Какую проблему решает квантильная регрессия? В каких ситуациях она неприменима или бесполезна?

Решение и анализ практических заданий 1.

Практическое задание 1:

Постройте два дерева решений: методом `ctree` и `rpart`. При желании, вы можете добавить дополнительные предикторы. Проведите необходимую пре-/постобрезку дерева. Нарисуйте оба дерева, а также выведите структуру дерева. Дайте интерпретацию результатам. Насколько они различаются для двух алгоритмов?

Практическое задание 2:

Постройте пуассоновскую регрессию, предсказывая количество детей с помощью переменных возраст, пол, брачный статус, доход, религиозная конфессия. Проинтерпретируйте свою финальную модель: какие переменные значимы? Какой эффект они оказывают? Не забудьте экспонировать коэффициенты любым удобным вам способом, чтобы получить `risk ratios`.

Тема 6. Введение в моделирование каузальности: discontinuity regression. Difference in difference. Анализ медиации.

Опрос 4:

1. Определение и типы каузальности.
2. Методы для определения каузальности.

3. Discontinuity regression.
4. Difference in difference.
5. Анализ медиации.

Решение и анализ практических заданий 2.

Практическое задание 1:

На данных изменения уровня безработицы в США проверьте, как на безработицу повлияло повышение минимального уровня оплаты труда, с помощью регрессии Difference in Differences. Не забудьте добавить контрольные переменные. Проинтерпретируйте результаты: может ли изменение в оплате труда считаться причиной изменения в уровне безработицы?

Практическое задание 2:

Наша теоретическая модель предполагает, что образование родителей влияет на образование ребенка, которое затем влияет на уровень заработной платы этого выросшего ребенка. На наших данных воспроизведите эту модель с помощью пакета lavaan и проверьте, действительно ли образование ребенка медирует эффект образования родителей на доход ребенка. Каков коэффициент полной медиации?

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

Форма промежуточной аттестации:

6 модуль – зачет, который проходит в форме тестирования.

7 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
6 модуль – зачет-тестирование	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
7 модуль - зачет с оценкой / тестирование	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

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
				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 балл; 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: Какой из перечисленных методов используется для работы с пропущенными

значениями в данных?

- А) Кластерный анализ
- Б) Импутация
- В) Корреляционный анализ
- Г) Регрессионный анализ

Правильный ответ:

Задание 2

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой тип распределения характеризуется симметричной формой и одинаковыми значениями среднего, медианы и моды?

- А) Нормальное распределение
- Б) Скошенное распределение
- В) Бимодальное распределение
- Г) Равномерное распределение

Правильный ответ:

Задание 3

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод используется для коррекции выборки при наличии систематических ошибок?

- А) Бутстраппинг
- Б) Взвешивание
- В) Удаление выбросов
- Г) Кластеризация

Правильный ответ:

Задание 4

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой из перечисленных методов относится к робастным оценкам?

- А) Метод наименьших квадратов
- Б) Медианная регрессия
- В) Линейная регрессия
- Г) Кластерный анализ

Правильный ответ:

Задание 5

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод кластерного анализа предполагает построение дендрограммы?

- А) K-means
- Б) Иерархический кластерный анализ
- В) DBSCAN
- Г) Метод главных компонент

Правильный ответ:

Задание 6

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой алгоритм используется для построения деревьев решений?

- А) Метод k-ближайших соседей
- Б) Алгоритм CART
- В) Метод главных компонент
- Г) Линейная регрессия

Правильный ответ:

Задание 7

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод регрессии используется для моделирования нелинейных зависимостей?

- А) Линейная регрессия
- Б) Квантильная регрессия
- В) Общие аддитивные модели (GAM)
- Г) Логистическая регрессия

Правильный ответ:

Задание 8

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод используется для анализа каузальности, когда есть естественный разрыв в данных?

- А) Difference in difference
- Б) Discontinuity regression
- В) Анализ медиации
- Г) Кластерный анализ

Правильный ответ:

Задание 9

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой тип данных характеризуется отсутствием некоторых значений?

- А) Цензурированные данные
- Б) Пропущенные данные
- В) Усеченные данные
- Г) Скошенные данные

Правильный ответ:

Задание 10

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод используется для оценки устойчивости статистических результатов?

- А) Бутстраппинг
- Б) Импутация
- В) Взвешивание
- Г) Кластеризация

Правильный ответ:

Задание 11

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод регрессии позволяет анализировать различные квантили распределения?

- А) Линейная регрессия
- Б) Квантильная регрессия
- В) Логистическая регрессия
- Г) Полиномиальная регрессия

Правильный ответ:

Задание 12

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод используется для определения числа кластеров в данных?

- А) Метод локтя
- Б) Метод наименьших квадратов
- В) Метод главных компонент

Г) Линейная регрессия

Правильный ответ:

Задание 13

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой тип скоса в данных характеризуется длинным хвостом вправо?

А) Левый скос

Б) Правый скос

В) Симметричный скос

Г) Бимодальный скос

Правильный ответ:

Задание 14

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод используется для анализа влияния промежуточной переменной на зависимую переменную?

А) Difference in difference

Б) Discontinuity regression

В) Анализ медиации

Г) Кластерный анализ

Правильный ответ:

Задание 15

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод импутации использует множественную регрессию для заполнения пропущенных значений?

А) Импутация средним

Б) Импутация медианой

В) Множественная регрессионная импутация

Г) Импутация с помощью k-means

Правильный ответ:

Задания закрытого типа (повышенный уровень сложности)

Задание 16

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод коррекции выборки следует использовать, если данные имеют значительный дисбаланс по полу и возрасту?

А) Бутстраппинг

Б) Взвешивание

В) Удаление выбросов

Г) Импутация средним

Правильный ответ:

Задание 17

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод кластерного анализа наиболее подходит для данных с неизвестным числом кластеров и наличием шума?

А) K-means

Б) Иерархическая кластеризация

В) DBSCAN

Г) Метод главных компонент

Правильный ответ:

Задание 18

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод регрессии следует выбрать, если зависимая переменная имеет выбросы и гетероскедастичность?

- А) Линейная регрессия
- Б) Квантильная регрессия
- В) Логистическая регрессия
- Г) Полиномиальная регрессия

Правильный ответ:

Задание 19

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод импутации наиболее эффективен, если пропущенные значения зависят от других переменных?

- А) Импутация средним
- Б) Множественная регрессионная импутация
- В) Импутация медианой
- Г) Удаление строк с пропусками

Правильный ответ:

Задание 20

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод анализа каузальности следует использовать, если есть естественный порог, разделяющий группы?

- А) Difference in difference
- Б) Discontinuity regression
- В) Анализ медиации
- Г) Кластерный анализ

Правильный ответ:

Задание 21

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод позволяет оценить влияние промежуточной переменной на зависимую переменную?

- А) Difference in difference
- Б) Discontinuity regression
- В) Анализ медиации
- Г) Квантильная регрессия

Правильный ответ:

Задание 22

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод регрессии позволяет моделировать сложные нелинейные зависимости без задания явной формы связи?

- А) Линейная регрессия
- Б) Общие аддитивные модели (GAM)
- В) Логистическая регрессия
- Г) Полиномиальная регрессия

Правильный ответ:

Задание 23

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод следует использовать для оценки устойчивости статистических результатов при малом объеме данных?

- А) Бутстраппинг
 - Б) Импутация
 - В) Взвешивание
 - Г) Кластеризация
- Правильный ответ:

Задание 24

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод кластерного анализа позволяет визуализировать иерархию кластеров?

- А) K-means
- Б) Иерархическая кластеризация
- В) DBSCAN
- Г) Метод локтя

Правильный ответ:

Задание 25

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод анализа данных позволяет выявить скрытые закономерности в категориальных переменных?

- А) Деревья решений
- Б) Байесовские сети
- В) Линейная регрессия
- Г) Квантильная регрессия

Правильный ответ:

Задание 26

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод регрессии следует использовать, если зависимая переменная имеет мультимодальное распределение?

- А) Линейная регрессия
- Б) Квантильная регрессия
- В) GAM
- Г) Логистическая регрессия

Правильный ответ:

Задание 27

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод анализа данных позволяет оценить эффект политического вмешательства, сравнивая изменения в двух группах?

- А) Discontinuity regression
- Б) Difference in difference
- В) Анализ медиации
- Г) Кластерный анализ

Правильный ответ:

Задание 28

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод работы с пропущенными значениями сохраняет структуру данных, но требует сложных вычислений?

- А) Импутация средним
- Б) Импутация с помощью деревьев решений
- В) Удаление строк с пропусками

Г) Импутация медианой

Правильный ответ:

Задание 29

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод анализа данных позволяет выявить причинно-следственные связи при отсутствии экспериментальных данных?

А) Кластерный анализ

Б) Разведочный анализ данных

В) Difference in difference

Г) Корреляционный анализ

Правильный ответ:

Задание 30

Инструкция: Выберите правильное утверждение в ответ на поставленный вопрос

Вопрос: Какой метод регрессии наиболее подходит для анализа крайних значений распределения?

А) Линейная регрессия

Б) Квантильная регрессия

В) Логистическая регрессия

Г) Полиномиальная регрессия

Правильный ответ:

Комбинированные задания (повышенный уровень сложности)

Задание 31

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Пропущенные значения всегда следует удалять из выборки.
2. Множественная импутация снижает риск искажения результатов анализа.
3. Импутация с помощью деревьев решений применима только для количественных переменных.
4. Взвешивание не влияет на результаты анализа.
5. Типы скоса не имеют значения при применении бутстраппинга.

Правильные ответы:

Эталонный пример обоснования:

Задание 32

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Дендрограмма — результат иерархического кластерного анализа.
2. Метод k-средних нельзя применять для категоризованных переменных.
3. Кластерный анализ предназначен только для больших выборок.
4. Определение числа кластеров может базироваться на "локте" кривой внутригрупповой дисперсии.
5. Кластеры формируются по критерию одинакового объёма подгрупп.

Правильные ответы:

Эталонный пример обоснования:

Задание 33

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Деревья решений помогают объяснить поведенческие паттерны за счёт логики разветвления.
2. Байесовские сети — это разновидность линейной регрессии.
3. Алгоритм CART используется при построении дерева решений.
4. Деревья решений не применяются к категориальным переменным.
5. Байесовские сети всегда имеют циклическую структуру.

Правильные ответы:

Эталонный пример обоснования:

Задание 34

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. Квантильная регрессия даёт возможность моделировать не только среднее значение отклика.
2. GAM предполагает строгую линейную связь между предикторами и зависимой переменной.
3. Квантильная регрессия чувствительна к выбросам меньше, чем OLS.
4. GAM модели используются для предсказания категориальных переменных.
5. В квантильной регрессии коэффициенты постоянны для всех квантилей.

Правильные ответы:

Эталонный пример обоснования:

Задание 35

Инструкция: Выберите два правильных утверждения и обоснуйте выбор:

1. В методе Difference-in-Difference необходимо наличие контрольной и экспериментальной групп.
2. Метод Discontinuity Regression применим при наличии случайной выборки.
3. Для анализа медиатора достаточно оценить только прямой эффект.
4. Метод Discontinuity предполагает наличие резкого порога в предикторе.
5. Difference-in-Difference не может применяться к панельным данным.

Правильные ответы:

Эталонный пример обоснования:

Комбинированные задания (высокий уровень сложности)

Задание 36

Инструкция: Выберите два правильных утверждения и обоснуйте выбор, используя результаты R-кода:

```
library(quantreg)
model <- rq(income ~ age + education, tau = 0.25, data = survey_data)
summary(model)
```

1. Эта модель оценивает медианный доход.
2. Коэффициенты модели описывают влияние предикторов на 25-й перцентиль распределения дохода.
3. Модель устойчива к выбросам в переменной income.
4. Классическая линейная регрессия лучше подходит для оценки влияния на нижний хвост распределения.
5. Переменная age здесь включена как категориальная.

Правильные ответы:

Эталонный пример обоснования:

Задание 37

Инструкция: Выберите два правильных утверждения и обоснуйте выбор, анализируя результат кластеризации:

```
library(cluster)
data <- scale(dataset)
clust <- pam(data, k = 3)
silhouette_score <- silhouette(clust)
plot(silhouette_score)
```

1. Метод РАМ создаёт кластеры на основе медиан, а не центроидов.
2. Silhouette-график помогает оценить качество кластеризации.
3. Кластеры сформированы с использованием дерева решений.
4. Высокие значения силуэта говорят о слабой кластеризации.
5. Метод требует, чтобы все переменные были категориальными.

Правильные ответы:

Эталонный пример обоснования:

Задание 38

Инструкция: Выберите два правильных утверждения и обоснуйте выбор на основе результатов модели:

```
library(mgcv)
model <- gam(outcome ~ s(age) + gender + s(income), data = dataset)
summary(model)
```

1. Модель предполагает линейную связь между возрастом и outcome.
2. Компонент $s(\text{age})$ позволяет учесть нелинейность зависимости.
3. GAM используется только для бинарных переменных.
4. Предположения модели аналогичны классической OLS-регрессии.
5. $s(\text{income})$ указывает на использование сглаженной функции.

Правильные ответы:

Эталонный пример обоснования:

Задание 39

Инструкция: Выберите два правильных утверждения и обоснуйте выбор, анализируя модель:

```
library(mediation)
model.m <- lm(mediator ~ treatment + age, data = data)
model.y <- lm(outcome ~ treatment + mediator + age, data = data)
med <- mediate(model.m, model.y, treat = "treatment", mediator = "mediator",
boot = TRUE)
summary(med)
```

1. Этот анализ позволяет оценить косвенный эффект переменной treatment через mediator.
2. Анализ медиатора не применим к непрерывным переменным.
3. Пакет mediation использует бутстраппинг для оценки доверительных интервалов.
4. Возраст (age) влияет только на медиатор, а не на outcome.
5. mediate() применяется только для дискретных переменных.

Правильные ответы:

Эталонный пример обоснования:

Задание 40

Инструкция: Выберите два правильных утверждения и обоснуйте выбор, интерпретируя Discontinuity Regression:

```
library(rdd)
rdd_data <- RDestimate(y ~ score, data = exam_data, cutpoint = 50)
summary(rdd_data)
```

1. Модель оценивает влияние прохождения порога score = 50 на результат y.
2. Метод Discontinuity требует, чтобы score был бинарной переменной.
3. Модель применима, если значения score можно интерпретировать как непрерывные.
4. cutpoint = 50 задаёт случайный уровень воздействия.
5. Метод неприменим к экспериментальным данным.

Правильные ответы:

Эталонный пример обоснования:

Открытые задания (повышенный уровень сложности)

Задание 41

Инструкция: Объясните, как интерпретировать отрицательный коэффициент при предикторе income в квантильной регрессии на уровне $\tau = 0.9$.

Ответ:

Задание 42

Инструкция: В модели `gam(outcome ~ s(age), data = df)` график функции `s(age)` имеет форму вогнутой кривой. Объясните, как это влияет на интерпретацию.

Ответ:

Задание 43

Инструкция: В диагностике модели регрессии наблюдается сильная гетероскедастичность. Объясните, почему это может быть проблемой.

Ответ:

Задание 44

Инструкция: При построении дерева решений используется `rpart()`, и дерево оказывается чрезмерно глубоким. Объясните, как это может повлиять на модель.

Ответ:

Задание 45

Инструкция: Что означает высокий показатель `sil_width` при кластеризации?

Ответ:

Задание 46

Инструкция: В выводе функции `summary(lm())` стандартная ошибка одного коэффициента значительно выше остальных. Объясните возможное следствие.

Ответ:

Задание 47

Инструкция: Объясните, зачем использовать множественную импутацию вместо удаления наблюдений с пропущенными значениями.

Ответ:

Задание 48

Инструкция: Вычислен `AIC()` для двух моделей, и модель 1 имеет $AIC = 482$, а модель 2 — $AIC = 489$. Какая модель предпочтительнее и почему?

Ответ:

Задание 49

Инструкция: В графике остатков наблюдается U-образная форма. Объясните, что это означает.

Эталонный ответ:

Задание 50

Инструкция: При анализе Difference-in-Difference обнаружено, что доинтервенционные тренды различаются между группами. Объясните, почему это критично.

Ответ:

Открытые задания (высокий уровень сложности)

Задание 51

Инструкция: В результате регрессии в R:

```
summary(lm(support ~ age + education + income, data = data))
```

коэффициент `income` статистически незначим. Объясните, можно ли утверждать, что доход не влияет на поддержку.

Ответ:

Задание 52

Инструкция: При диагностике модели `gam(outcome ~ s(age), data = data)` график сплайна показывает резкие скачки в начале и конце диапазона. Объясните, что это может означать и как исправить.

Ответ:

Задание 53

Инструкция: При кластеризации методом `kmeans()` визуализирована структура кластеров, но наблюдаются значительные перекрытия между группами. Объясните, как это повлияет на интерпретацию.

Ответ:

Задание 54

Инструкция: В модели Discontinuity Regression наблюдаются значительные различия между группами до точки порога. Объясните, как это влияет на достоверность оценки эффекта.

Ответ:

Задание 55

Инструкция: В результате анализа медиатора с помощью `mediate()` обнаружено, что прямой эффект значим, а опосредованный — нет. Объясните, как интерпретировать результат.

Ответ:

Задание 56

Инструкция: Вы строите модель квантильной регрессии в R:

```
rq_model <- rq(support ~ income + education, tau = 0.75, data = data)
summary(rq_model)
```

При этом коэффициент при `education` положителен, но в модели с `tau = 0.25` он близок к нулю. Объясните, как это можно интерпретировать.

Ответ:

Задание 57

Инструкция: При анализе иерархического кластерного анализа построена дендрограмма, и наблюдается, что несколько объектов соединяются на очень большой высоте. Объясните, как это влияет на интерпретацию и возможное число кластеров.

Ответ:

Задание 58

Инструкция: После применения бутстраппинга вы получили широкий интервал для оценки коэффициента (например, от -0.9 до 1.4). Объясните, что это говорит о модели и её выводах.

Ответ:

Задание 59

Инструкция: Вы провели анализ Difference-in-Difference и получили положительный

коэффициент при переменной взаимодействия, но $p\text{-value} = 0.15$. Объясните, можно ли считать эффект значимым, и какие шаги целесообразны.

Ответ:

Задание 60

Инструкция: На графике остатков регрессионной модели выявлены систематические участки с последовательным положительным и отрицательным отклонением. Объясните, что это означает и какие меры следует предпринять.

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, дискуссия, обсуждение литературы, решение и анализ практических заданий, 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
Обсуждение литературы	Магистрант в ходе подготовки и участия в обсуждениях литературы 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
Дискуссия	Магистрант должен быть готовым в ходе подготовки и участия в дискуссии по темам дисциплины, выполнять следующие действия:

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
	<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>