

Программа учебной дисциплины «Методы анализа неоднородных данных и паттерн-анализ»

Утверждена

Академическим советом ОП

Протокол № от __. __.2019

Разработчик	Сальникова Д. В., преподаватель, общеуниверситетская кафедра высшей математики
Число кредитов	6
Контактная работа (час.)	80
Самостоятельная работа (час.)	148
Курс, Образовательная программа	1 курс магистерской программы «Прикладная политология»
Формат изучения дисциплины	без использования онлайн курса

1. Цель, результаты освоения дисциплины и пререквизиты

Чаще всего в социальных науках, в частности, в политологии, исследователям приходится работать с неоднородными данными. Примером могут выступать данные с панельной структурой (наличие как пространственного, так и временного измерения: данные по каждой стране из выборки представлены за несколько лет), с иерархической структурой (к примеру, в массиве представлены данные на индивидуальном и страновом уровне, панельные данные также можно представить как данные с иерархической структурой). В таком случае предположение об одинаковых параметрах для всех единиц анализа является неправдоподобным. **Цель освоения дисциплины «Методы анализа неоднородных данных и паттерн-анализ» (1 курс магистратуры)** – овладение статистическими методами, позволяющих учесть отсутствие независимости наблюдений в подгруппах выборки, для решения политологических и социально-экономических задач.

В результате освоения дисциплины студент должен:

- Знать основы теории оценивания регрессионных моделей с фиксированными, случайными эффектами, моделей со смешанными эффектами

- Критически оценивать исследовательские стратегии применительно к анализу неоднородных данных, обоснованно выбирать подходящий метод к анализу неоднородных данных
- Приобрести опыт реализации изученных методов посредством Python, R
- Приобрести навык визуализации данных с иерархической структурой и результатов моделирования со смешанными эффектами (в частности, многоуровневого моделирования)
- Корректно интерпретировать результаты оценивания моделей со смешанными эффектами (в частности, многоуровневых моделей)

Пререквизиты изучения дисциплины: студенты должны знать основы статистики и регрессионный анализ в объеме следующих курсов: «Методология и методы политических исследований» (1 курс магистерской программы «Прикладная политология»), лекции 1 – 5 «Methodology and research methods of political science» (1 курс магистерской программы «Прикладная политология»).

2. Содержание учебной дисциплины

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	Лекции		
	Семинары		
	Сам. Работа		
Тема 1	4	Знает возможные стратегии работы с неоднородными данными и знает их область применения, корректно применяет и интерпретирует оценки моделей с переменными взаимодействия	Проверочная работа №1 (30 минут) Домашнее задание №1
	8		
	28		
Тема 2	6	Корректно интерпретирует результаты оценивания модели с фиксированными эффектами, знает область применения модели с фиксированными эффектами	Проверочная работа №2 (30 минут) Домашнее задание №2
	10		
	40		

Тема 3	6	Корректно интерпретирует оценки коэффициентов в моделях со смешанными эффектами, понимает различие между фиксированными и случайными эффектами в контексте модели со смешанными эффектами	Проверочная работа №3 (30 минут) Домашнее задание №3
	12		
	30		
Тема 4	4	Умеет выбрать и обосновать спецификацию модели со смешанными эффектами для анализа панельных данных, корректно интерпретирует оценки коэффициентов в моделях со смешанными эффектами	Проверочная работа №3 (30 минут) Домашнее задание №3
	12		
	30		
Тема 5	4	Корректно интерпретирует оценки логистических моделей со смешанными эффектами, знает их область применения	Домашнее задание №4
	6 + 8 (на презентацию эссе)		
	20		
Часов по видам учебных занятий:	24		
	56		
	148		
Итого часов:	228		

Тема 1. Введение: основы работы с гетерогенной выборкой

Типы структуры данных. Стратегии работы с неоднородным массивом данных: обсуждение преимуществ и ограничений. Pooled model + скорректированные стандартные ошибки. Разделение на подвыборки. Включение в модель переменных взаимодействия. Разумно ли ограничиваться линейным эффектом переменных взаимодействия? Обсуждение нелинейных эффектов. Инструменты визуализации.

Тема 2. Модели с фиксированными эффектами: возможности и ограничения

Модель с фиксированными эффектами. Применение к кросс-секционным вложенным данным и панельным данным. Выбор адекватной модели: модель с фиксированными эффектами VS модель со случайными эффектами. Уточнение спецификации модели с фиксированными эффектами. Обобщенный метод наименьших квадратов (GLS). Реализуемый обобщенный метод наименьших квадратов. Панельно-скорректированные стандартные ошибки. Пространственная корреляция: суть, условия возникновения, последствия. Инструменты работы с пространственной корреляцией. Автокорреляция:

суть, условия возникновения, последствия. Моделирование автокорреляции.

Тема 3. Модели со смешанными эффектами. Сравнение альтернативных подходов к работе с гетерогенной выборкой

Модель со смешанными эффектами. Многоуровневая модель как частный случай модели со смешанными эффектами. Преимущества модели со смешанными эффектами по сравнению с альтернативными методами при работе с данными, имеющими иерархическую структуру. Содержательные задачи, решаемые посредством моделей со смешанными эффектами: применение в социальных науках.

Обсуждение основных идей байесовского подхода. Почему многоуровневую модель можно назвать своеобразным компромиссом между классической статистикой и байесовским подходом? Что является источником априорной информации для получения распределения случайных эффектов? Формирование "веса" для правдоподобия выборки. Анализ кросс-секционных вложенных данных посредством моделей со смешанными эффектами. Сравнение со стратегией деления на подвыборки: обсуждение в контексте сравнения с многоуровневыми моделями эффекта сжатия, BLUP. Сравнение с моделями с фиксированными эффектами.

Тема 4. Анализ панельных данных при помощи моделей со смешанными эффектами

Представление панельных данных как данных с иерархической структурой. Структура ковариационной матрицы случайных эффектов. Различие между фиксированными и случайными эффектами, их интерпретация. Спецификация модели. Методы оценивания. Визуализация результатов.

Тема 5. Логистические модели со смешанными эффектами

3. Оценивание

Окончательная оценка по дисциплине рассчитывается в соответствии со следующей формулой:

$$\text{Итог.} = 0.2 \cdot \text{Одз.} + 0.15 \cdot \text{Осам.} + 0.25 \cdot \text{Оэссе} + 0.1 \cdot \text{Опрезент.} + 0.3 \cdot \text{Оэкз.},$$

где **Одз.** – **оценка за домашние задания.** Оценка рассчитывается как среднее арифметическое оценок за текущие домашние задания. Округление данного компонента итоговой оценки арифметическое. Оценка за каждое домашнее задание выставляется по 10-балльной шкале. Пересдача домашних заданий не предусмотрена. При сдаче домашнего задания с опозданием штраф составляет 1 балл за каждый день просрочки. После того, как домашние задания проверены и оценки за них выставлены, выполненные задания от опоздавших студентов могут быть приняты только для получения комментариев и диагностики ошибок: при опоздании более чем на 2 дня (48 часов после дедлайна) за данное домашнее задание выставляется оценка «0». Студентам предоставляется возможность повысить оценку за домашние задания посредством корректного выполнения бонусных домашних заданий.

Осам. – **оценка за самостоятельные работы.** Оценка рассчитывается как среднее арифметическое оценок за проверочные работы, выполняемые студентами на занятиях. Выполнение проверочных работ осуществляется в режиме closed-book. Оценка за каждую проверочную работу выставляется по 10-балльной шкале. Проверочные работы не переписываются, на их написание не выделяется дополнительное время вне зависимости от причины пропуска проверочной работы студентом. В случае пропуска студентом проверочной работы за нее выставляется оценка «0». Студентам предоставляется повысить данный компонент оценки посредством активной работы на семинарах (корректное решение задач у доски).

Оэссе – **оценка за эссе.** Оценка выставляется по 10-балльной шкале. При сдаче эссе с опозданием в пределах одного дня (24 часов) оценка снижается на 1 балл. При сдаче эссе с опозданием от 24 до 48 часов оценка снижается на 2 балла. Эссе, сданные позже, не принимаются. Пересдача и переписывание эссе с целью повышения оценки не предусмотрены. Требования к эссе представлены в следующем разделе программы курса.

Опрезент. – **оценка за презентацию своего мини-исследования (эссе).** Студент должен отразить в презентации исследовательский вопрос, как осуществлялся выбор методов, какие альтернативные стратегии работы с данными могут быть применены для ответа на поставленный исследовательский вопрос, как соотносится теоретическая и эмпирическая часть эссе.

Оэкс. – оценка за экзаменационную работу по курсу. Оценка выставляется по 10-балльной шкале. Экзаменационная работа выполняется в режиме closed-book. Примеры заданий представлены в следующем разделе программы курса. Первая и вторая пересдачи экзамена предусматривают выполнение письменной работы, по заданиям схожим с экзаменационной работой по курсу. На пересдаче окончательная оценка по дисциплине выставляется по той же самой формуле, что и до пересдач.

4. Примеры оценочных средств

Демонстрационный вариант экзаменационной работы доступен по ссылке: https://yadi.sk/i/0cH7_L3oUuey2Q

Блокирующие элементы не предусмотрены.

Эссе – самостоятельная (индивидуальная) письменная работа в формате статьи по результатам поиска и изучения соответствующей литературы и анализа массива политологических и социально-экономических данных с использованием изученных в рамках данного курса методов многомерной статистики.

Рекомендуемый объем эссе – 20 – 30 тыс. символов с пробелами.

Структура домашнего задания в обязательном порядке должна включать в себя следующие разделы:

1. Аннотация (от 500 до 800 знаков с пробелами)
2. Введение (постановка задачи, обзор литературы и краткая «приманка» для читателя в виде формулировки основных полученных в работе результатов)
3. Гипотезы (формулируются и обосновываются гипотезы работы)
4. Данные и используемые методы (описываются имеющиеся данные, объясняется выбор источников данных; описываются используемые методы, приводится спецификация моделей, обосновывается выбор методов)
5. Результаты (в табличной форме, соответствующей зарубежным публикационным стандартам, представляются полученные результаты, а также предлагается их интерпретация)
6. Обсуждение полученных результатов (описывается, в какой мере полученные результаты соответствуют существующей литературе; обсуждаются возможные причины

расхождений; обсуждается устойчивость полученных результатов к изменению массива данных и изменению методов анализа)

7. Заключение (формулируются выводы работы, обсуждаются теоретические следствия из полученных выводов, указываются возможные направления дальнейшего уточнения результатов)

8. Список использованной литературы

Оцениваются адекватность формулировки задачи/проблемы, соответствие заявленных для проверки гипотез сформулированной задаче, корректность подбора методов и интерпретации результатов, критический взгляд на полученные результаты, проверка результатов на устойчивость, обоснованность выводов. Обращайте внимание на грамотность русского языка и корректность использования терминов, аккуратность оформления, в т.ч. библиографии.

В случае обнаружения плагиата в работе студента, за эссе выставляется оценка «0».

Задание выполняется в Python или RStudio.

Оценка за эссе выставляется по 10-балльной шкале.

Преподавателю для проверки сдаются:

- 1) текст эссе
- 2) массив данных
- 3) **script** для возможности репликации результатов

5. Ресурсы

5.1 Рекомендуемая основная литература

1. Mostly harmless econometrics: an empiricist's companion / J. D. Angrist, J.- S. Pischke. – Princeton; Oxford: Princeton University Press, 2009. – 373 с. – На англ. яз.
2. Basic econometrics / D. N. Gujarati, D. C. Porter. – 5th. – Boston [etc.]: McGraw-Hill, 2009. – 921 с. – (Economics series) . – На англ. яз.
3. Econometric analysis of cross section and panel data / J. M. Wooldridge. – Cambridge: The MIT Press, 2002. – 752 p.
4. Steenbergen, M., Jones, B. (2002). Modeling Multilevel Data Structures. American Journal of Political Science. Vol. 46. No. 1, pp. 218 – 237.

5. Анализ панельных данных и данных о длительности состояний: учеб. пособие / Т. А. Ратникова, К. К. Фурманов. – М.: Изд. дом Высшей школы экономики, 2014. – 373 с.

5.2 Рекомендуемая дополнительная литература

1. Green, D. P., Kim, S. Y., Yoon, D. H. (2001). Dirty Pool. International Organization. Vol. 55, No.2, pp. 441 – 468.
2. Hainmueller, J., Mummolo, J., Xu, Y. (2018). How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice. Political Analysis, Vol. 26. No. 2. Pp. 163 – 192. DOI: 10.1017/pan.2018.46
3. Путеводитель по современной эконометрике: учеб. пособие для вузов / М. Вербик; Пер. с англ. В. А. Банникова; Науч. ред., предисл. С. А. Айвазяна. – М.: Научная книга, 2008. – 615 с. – (Б-ка Солев) . – НП.
4. Эконометрика: начальный курс: учебник для вузов / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий . – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 575 с.

5.3 Программное обеспечение

№п/п	Наименование	Условия доступа/скачивания
	Python	Свободное лицензионное соглашение
	R	Свободное лицензионное соглашение

5.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№п/п	Наименование	Условия доступа/скачивания
1	Открытое образование	URL: https://openedu.ru/

5.5 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для практических занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

6 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) предлагаются следующие варианты восприятия учебной информации, в том числе с применением электронного обучения и дистанционных технологий:

- 6.1 для лиц с нарушениями зрения:* в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные задания и консультации.
- 6.2 для лиц с нарушениями слуха:* в печатной форме; в форме электронного документа; индивидуальные задания и консультации.
- 6.3 для лиц с нарушениями опорно-двигательного аппарата:* в печатной форме; в форме электронного документа; индивидуальные задания и консультации.