

Программа учебной дисциплины «Анализ данных в социальных науках»

Утверждена

Академическим советом ООП

Протокол № 01 от «30 » августа 2019 г.

Автор	Т.А. Чиркина, А.В. Капуза
Число кредитов	4
Контактная работа (час.)	52
Самостоятельная работа (час.)	100
Курс	1
Формат изучения дисциплины	без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Анализ в социальных науках» являются:

- формирование представления о многоуровневых методах анализа данных и структурном моделировании в исследованиях образования и социальных науках;
- формирование навыков выбора релевантного способа анализа данных;
- развитие навыков структурного моделирования;
- развитие навыков работы с данными, имеющими иерархическую, кластерную структуру для решения социально-экономических задач и анализа в сфере образования.

В результате освоения дисциплины студент должен:

знать:

- основные принципы и методы многоуровневого регрессионного анализа;
- основные принципы и методы структурного моделирования

уметь:

- применять изученные методы анализа данных для самостоятельного построения моделей
- анализировать и представлять результаты анализа данных в различных формах (отчет, презентация)

владеть:

- навыками анализа в R

Изучение дисциплины «Анализ данных в социальных науках» базируется на следующих дисциплинах:

- эконометрика.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- знать линейную и логистическую регрессию
- желательно: владение STATA

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Основные понятия многоуровневого регрессионного анализа.

Определение и области применения многоуровневого анализа. Достоинства и ограничения метода. Виды иерархических структур, константа (intercept), угол наклона (slope),

фиктивная переменная (dummy variable), стандартизация, центрирование, показатели ответственности модели, гомоскедастичность.

Тема 2. Многоуровневый линейный анализ. Базовая модель, модели с варьирующим интерсептом, наклоном, межуровневое взаимодействие

Базовая (нулевая) модель, составные части уравнения. Коэффициент интраклассовой корреляции. Модель с переменными первого уровня с неизменяющимся коэффициентом регрессии (fixed slope and random intercept). Модель с переменными первого и второго уровня (fixed slope). Модель с переменными первого уровня с изменяющимся коэффициентом (random slope). Модель с переменными первого и второго уровней с изменяющимся коэффициентом (random slope). Модель с переменными взаимодействия.

Тема 3. Многоуровневая логистическая регрессия

Области применения многоуровневой логистической регрессии. Виды преобразований - логит и пробит модели, логарифмический шанс и отношение шансов. Оценка R и R-квадрата. Виды моделей и их интерпретация.

Тема 4. Тестирование и спецификация многоуровневых моделей

От чего зависит выбор лучшей модели. Сравнение вложенных и невложенных моделей. Devian статистика. Contrast и constrains.

Тема 5. Основные концепты в анализе латентных переменных

Понятие латентных переменных. Понятие конструкторов. Определение модели КФА (конфирматорный факторный анализ): гипотезы, параметры модели, обозначения. Различия между КФА и РФА (разведывательный факторный анализ).

Тема 6. Конфирматорный факторный анализ

Идентификация модели. Идентифицированные, неидентифицированные, точно идентифицированные модели КФА. Виды ограничений параметров, используемых для идентификации модели. Показатели качества (Chi-Square, RMSEA, CFI): относительные преимущества и недостатки. Оценка моделей CFA с использованием R-пакета lavaan. Интерпретация параметров модели. Модификация и сравнение моделей (chi-square difference test). КФА второго порядка. Проблемы пропущенных данных.

Тема 7. Полные структурные модели

Основные понятия структурного моделирования (SEM). Экзогенные и эндогенные переменные. Путевые модели. Модели с латентными переменными и несколькими индикаторами. Причинность и эквивалентные модели. Пример путевых моделей. Примеры моделей SEM с эндогенными латентными переменными и наблюдаемыми экзогенными переменными.

Тема 8. Модерация и медиация в структурном моделировании

Понятия модерации и медиации. Прямой, непрямой, общий эффект. Модель с несколькими латентными переменными.

III. ОЦЕНИВАНИЕ

Домашние задания и проектная работа должны отражать способность использовать современные средства ИКТ для поиска и обработки информации, работы с базами данных профессиональной информации; быть оформлено в виде отчета по научной и практической работе и презентации. Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

Итоговая экзаменационная работа: исследовательский проект. Для выполнения исследовательского проекта по курсу студентам предлагается на выбор несколько баз дан-

ных реальных исследований. Выбрав одну из них, необходимо сформулировать постановку содержательной цели исследования, определить задачи, сформулировать гипотезы, подобрать соответствующие методы анализа данных, отобрать необходимые данные, проанализировать и проинтерпретировать результаты анализа. В работе обязательно должны быть задействованы все методы анализа данных, входящие в программу курса.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Оценочные средства для текущего контроля студента

1. Каковы условия применения многоуровневой регрессии?
2. В чем основные отличия моделей random и Fixed slope. В каких случаях какие модели используются?
3. Что такое агрегированные и дисагрегированные переменные?
4. Что такое латентная переменная? Что такое конструктор?
5. Факторная модель: обозначения, ключевые понятия (такие как загрузка факторов, общность, уникальность, собственное значение и коэффициент)
6. Конфирматорный факторный анализ (CFA). Различия между КФА и разведывательным факторным анализом.
7. Ключевые параметры КФА: факторные нагрузки, остатки, ковариации
8. Идентификация модели. Не идентифицированные, едва идентифицированные и идентифицированные модели КФА.
9. Различные типы ограничений параметров, используемых для идентификации модели в CFA.
10. Показатели качества моделей (Chi-Square, RMSEA, CFI, SRMR / WRMR)
11. Модификация модели и сравнение моделей (Chi-Square тест разности, AIC / BIC).
12. Чем медиация отличается от модерации? Приведите пример. Как моделируются модерация и медиация переменных?
13. Что означает низкое качество модели? Какие могут быть возможные причины низкого качества модели?

V. РЕСУРСЫ

5.1 Основная литература

1. Raudenbush S. W., Bryk A. S. Hierarchical linear models: Applications and data analysis methods. – Sage, 2002. – Т. 1.

5.2 Дополнительная литература

1. Chen F. F. Sensitivity of goodness of fit indexes to lack of measurement invariance //Structural equation modeling. – 2007. – Т. 14. – №. 3. – С. 464-504.
2. Sommet N., Morselli D. Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS //International Review of Social Psychology. – 2017. – Т. 30. – №. 1.
3. Austin P. C., Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis //Statistics in medicine. – 2017. – Т. 36. – №. 20. – С. 3257-3277.

5.3 Программное обеспечение

№ п/п	Наименование	Условия доступа

1.	Microsoft Windows 7 Professional RUS	<i>Из внутренней сети университета (договор)</i>
2.	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>
3.	IrfanView	<i>Свободное лицензионное соглашение</i>
4.	R и RStudio	<i>Свободное лицензионное соглашение</i>

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

№ п/п	Наименование	Условия доступа
<i>Профессиональные базы данных, информационно-справочные системы</i>		
1.	Scopus, Web of Sciences	<i>Из внутренней сети университета (договор)</i>

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 лабораторных и самостоятельных занятий по дисциплине оснащены ноутбуками с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.