

Программа учебной дисциплины «Углубленные методы анализа данных в социальных науках: многоуровневый регрессионный анализ и структурное моделирование»

Утверждена

Академическим советом ООП

Протокол № 01 от «28» августа 2018 г.

Автор	Кузьмина Ю.В.
Число кредитов	4
Контактная работа (час.)	86
Самостоятельная работа (час.)	56
Курс	2 курс
Формат изучения дисциплины	Без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Углубленные методы анализа данных в социальных науках: многоуровневый регрессионный анализ и структурное моделирование» являются овладение студентами основными понятиями и умениями в области продвинутых методов анализа данных, а именно многоуровневого регрессионного анализа и моделирования структурными уравнениями.

В результате освоения дисциплины студент должен:

знать:

- Понятие множественной регрессии и логистической регрессии
- Основные требования к проведению множественного регрессионного анализа;
- Способы проверки соблюдения основных требований множественного регрессионного анализа и возможности действий в случае нарушения основных требований;
- Способы оценки соответствия модели данным
- Необходимость и преимущества многоуровневого регрессионного анализа

- Разные типы и виды многоуровневых структур
- Типы и виды многоуровневых моделей, различия между random и fixed коэффициентами
- Виды и способы центрирования и стандартизации переменных для включения в многоуровневые модели
- Особенности проведения многоуровневого моделирования для бинарных зависимых переменных
- Возможности структурного моделирования
- Отличия эксплораторного и конфирматорного факторного анализа
- Основные принципы проведения конфирматорного факторного анализа и путевого анализа
- Различие между экзогенными и эндогенными переменными
- Основные параметры оценки соответствия модели данным в моделировании структурными уравнениями
- Принципы проверки измерительной инвариантности латентных конструктов
- Различия между прямым и непрямым эффектом, медиатором и модератором.

уметь:

- Выбирать адекватный метод анализа данных в зависимости от исследовательских вопросов и задач и специфики данных
- Проверять способы соблюдения основных требований регрессионного анализа
- Включать в регрессионную модель переменные взаимодействия и интерпретировать коэффициенты переменных взаимодействия
- Выбирать релевантную многоуровневую модель в зависимости от исследовательского вопроса
- Рассчитывать коэффициент интер- и интраклассовой корреляции (unconditional и conditional) и интерпретировать его значения для двух- и трех-уровневых моделей
- Строить разные типы многоуровневых моделей и интерпретировать их результаты
- Использовать разные типы трансформации переменных и понимать особенности интерпретации полученных коэффициентов
- Проводить конфирматорный факторный анализ
- Строить путевые диаграммы и интерпретировать их

- Оценивать прямой и непрямой эффект переменных в путевых моделях
- Использовать интервальные и бинарные переменные как зависимые в многоуровневом регрессионном анализе и моделировании структурными уравнениями
- Проверять степень соответствия модели данным
- Интерпретировать результаты множественного регрессионного анализа, многоуровневого анализа, конфирматорного факторного анализа и путевого анализа

владеть:

- навыками работы в программной оболочке для анализа данных Stata и Mplus
- навыками практического применения изучаемых методов к существующим базам данных (PISA, PIRLS, проект «Анализ трудовых и образовательных траекторий»)
- навыками реализации самостоятельных проектов по анализу данных
- навыками написания академических текстов с применением результатов анализа данных
- навыками работы с исследовательскими текстами и анализа выбранной авторами стратегии анализа

Изучение дисциплины «Углубленные методы анализа данных в социальных науках: многоуровневый регрессионный анализ и структурное моделирование» базируется на следующих дисциплинах «Базовые методы анализа данных и работа со статистическими пакетами» на 1 курсе, «Качественные и количественные методы исследований в психологии», «Принципы измерения в психологии и образовании»

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

- Умение работать с одним из пакетов по анализу данных (SPSS, Stata, R)
- Знание основ регрессионного анализа
- Знание основных принципов факторного анализа

Основные положения дисциплины должны быть использованы в дальнейшем при подготовке и защите ВКР..

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Общее представление о многоуровневом и структурном анализе. Множественная регрессия для интервальных переменных

Многоуровневая регрессия – регрессия «в особых условиях». Моделирование структурными уравнениями и многоуровневая регрессия как расширение возможностей регрессионного анализа. Понятие множественной регрессии. Включение в анализ переменных взаимодействия. Интерпретация коэффициентов переменных взаимодействия. Использование регрессионного анализа для контроля условий. Оценка соответствия модели полученным данным.

История применения многоуровневой регрессии в исследованиях образования. Виды исследований и выборок, в которых используется многоуровневые методы анализа. Необходимость и преимущества многоуровневого анализа. Сравнение простого и многоуровневого регрессионного анализа. Виды структур: 2,3 и 4 уровневые структуры. Иерархические и неиерархические структуры. Структуры с пересекающейся классификацией. Структуры с множественным членством. Различия между уровнями и переменными, фиксированными и случайными классификациями. Размер и построение выборки для многоуровневого анализа.

Тема 2. Многоуровневая регрессия. Виды 2-уровневых моделей

Базовая 2-уровневая модель. Отклонения 1 и 2 уровня. Способы оценки модели и гипотез. Виды моделей: intercept-only model, модель с независимыми переменными 1 уровня с неизменяющимся коэффициентом регрессии в группах (fixed slope and random intercept), модель с независимыми переменными 1 уровня с изменяющимся коэффициентом регрессии в группах (random slope and random intercept), модель с независимыми переменными 2 уровня, модель с взаимодействием переменных 1 и 2 уровней. Коэффициент интраклассовой корреляции, безусловный и условный. Расчет процента объясненной дисперсии для 1 и 2 уровня. Использование центрированных и стандартизованных переменных. Виды центрирования. Интерпретация моделей. Анализ соответствия моделей полученным данным. Исследовательские вопросы для каждой модели.

Тема 3. Логистическая многоуровневая регрессия

Обобщенная линейная модель. Логит- и пробит анализ. Интерпретация логистических коэффициентов. Виды моделей многоуровневого анализа для бинарной переменной. Квазиравдоподобие.

Тема 4. Анализ лонгитюдных данных в рамках многоуровневого анализа

Способы работы с лонгитюдными данными. Виды анализа. Модели с фиксированными и изменяющимися эффектами. Выбор переменной «время». Линейные и нелинейные изменения. Включение переменных взаимодействия. Оценка разных типов траекторий изменений.

Тема 5. Введение в структурное моделирование (Structural Equation Modeling). Конфирматорный факторный анализ.

Основные характеристики SEM. Понятие латентной переменной. Экзогенные и эндогенные конструкты. Измерительные и структурные модели. Базовые принципы построения моделей.

Конфирматорный факторный анализ. Отличия от эксплораторного факторного анализа. Валидность и надежность коэффициентов КФА, ошибка измерения. Индексы соответствия модели данным.

Тема 6. КФА в Mplus. Измерительная инвариантность

Общее представление о программе и ее целях. Знакомство с интерфейсом. Подготовка файлов к работе в Mplus. Основные команды синтаксиса.

Измерительная инвариантность.

Конфирматорный факторный анализ с факторами второго порядка и бифакторные модели

Практические занятия:

Подготовка файлов к работе

Работа с меню, работы с синтаксисом. Примеры. Конфирматорный факторный анализ с двумя факторами. Работа с outputs. Интерпретация индексов соответствия модели данным, коэффициентов, modification indices. Мульти-групповой КФА (Multiple group CFA).

Тема 7. Путевые диаграммы. Прямой и непрямой эффект, медиаторы и модераторы.

Принципы построения путевых диаграмм. Идентификация и оценка моделей. Интерпретация результатов. Медиаторы. Модели с полным и частичным медиаторным эффектом. Модерация с латентными переменными.

III. ОЦЕНИВАНИЕ

Оценка за курс складывается из оценок за контрольные и самостоятельные работы, презентации ($O_{\text{текущий}}$) и итоговый тест ($O_{\text{экзамен}}$)

$$O_{\text{финальная}} = 0,7 * O_{\text{текущий}} + 0,3 * O_{\text{экзамен}}$$

где $O_{\text{текущий}}$ рассчитывается как взвешенная сумма всех форм текущего контроля, предусмотренных в РУП

$$O_{\text{текущий}} = 0,2 * O_{\text{презент}} + 0,4 * O_{\text{к/р}} + 0,4 * O_{\text{с.р}}$$

Выполняя тестовые и самостоятельные работы, студент должен продемонстрировать умение пользоваться компьютерными технологиями для ответа на исследовательские вопросы, находить релевантные вопросу методы анализа данных, интерпретировать полученные результаты, оформить и представить результаты своей деятельности в виде отчета по научной и практической работе.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

Итоговая зачетная работа: тест

Для выполнения итогового теста по курсу студентам предлагается несколько баз данных реальных исследований. В тесте сформулированы вопросы, для ответа на которые необходимо проанализировать одну из имеющихся баз и применить изучаемые методы для их анализа.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Оценочные средства для текущего контроля студента

Примерные вопросы для контрольных работ:

Какие из этих переменных можно использовать в качестве предикторов для результирующей переменной «достижения в чтении» без изменений:

- Место расположения школы (1-деревня, 2-маленький город (до 15 тысяч), 3 – средний город (до 100 тысяч), 4 – большой город (до 1 млн), 5 – мегаполис)
- Размер школы (количество учеников, учащихся в школе)
- Пол ребенка (1-мужской, 2 – женский)
- Уровень социально-экономического статуса (сводный индекс)
- Образование матери (1 – среднее и ниже, 2 – среднее профессиональное, 3-высшее и выше)

Какие вспомогательные переменные (dummy variables) могут быть образованы?

Напишите для каждой переменной, которую надо преобразовать

Какая модель больше подходит для ответа на следующий исследовательский вопрос: Изменяются ли результаты ЕГЭ по русскому языку в школах с разным количеством иммигрантов с учетом индивидуального уровня СЭС? (Опрошено i учеников в j школах):

- Балл ЕГЭ = $G00 + G10 * SES_{ij} + u_{0j} + e_{ij}$
- Балл ЕГЭ = $G00 + G01 * IMMIG_j + u_{0j} + e_{ij}$
- Балл ЕГЭ = $G00 + G10 * SES_{ij} + u_{0j} + u1_j * SES_{ij} + e_{ij}$
- Балл ЕГЭ = $G00 + G10 * SES_{ij} + G01 * IMMIG_j + u_{0j} + u1_j * SES_{ij} + e_{ij}$
- Балл ЕГЭ = $G00 + G10 * SES_{ij} + G11 * SES_{ij} * IMMIG_{ij} + G01 * IMMIG_j + u_{0j} + u1_j * SES_{ij} + e_{ij}$
- Балл ЕГЭ = $G00 + G10 * SES_{ij} + G01 * IMMIG_j + u_{0j} + e_{ij}$

В таблице ниже представлены оценки регрессионных параметров, полученные в результате многоуровневого анализа. Зависимая переменная – достижения в естественных науках. Все независимые интервальные переменные центрированы по общему среднему. Посчитайте R^2 для каждой модели на групповом уровне, на индивидуальном уровне, коэффициент интраклассовой корреляции. Напишите уравнение для каждой модели. Напишите, на какой вопрос можно ответить с помощью каждой модели.

	Null-model	Model 1	Model2	Model 3
Intercept	474,3 (3,8)	478,8 (3,6)	479,6 (3,4)	479,8 (3,4)

SES		36,8 (3)	36,6 (3)	37,2 (3,1)
Материальное благополучие		-21,17 (2,9)	-21,27 (2,9)	-21,4 (2,9)
Школьные ресурсы			11,16 (3,8)	10,7 (4)
U_{0j}	42,38	38,25	37,1	36,5
e_{ij}	77,78	75,29	75,29	74,4
U_{0j}^2	1796	1463	1373	1332
E_{ij}^2	6049	5669	5669	5540
u_{1j}				15,2
U_{2j}				11,7

Примеры вопросов для самостоятельной работы:

Используя базу данных PISA 2009 ответьте на следующий исследовательский вопрос (ответ должен включать описание переменных, описание модели, результаты и интерпретацию данных) – для каждого студента свой вопрос:

1. Как связаны результаты по чтению с материальным благополучием и СЭС учеников с учетом того, расположена школа в деревне или городе?
2. Отличаются ли результаты по естественным наукам в маленьких и больших школах с учетом уровня СЭС учеников?
3. Показывают ли ученики-иммигранты худшие результаты в чтении? Имеет ли значение количество детей иммигрантов в школе для результатов по чтению с учетом индивидуального иммигрантского статуса?
4. Изменяется ли эффект СЭС для результатов по чтению в городских и деревенских школах?
5. Изменяется ли эффект СЭС на результаты по математике в школах с разными ресурсами?

Примеры тем для эссе и презентаций:

- 1) Лонгитюдная инвариантность конструкторов. Для чего и как ее проверять. Какие следствия нарушения инвариантности в лонгитюдных исследованиях?
- 2) Как можно сравнивать бифакторные модели и модели второго порядка? В каких случаях бифакторные модели могут лучше подходить данным?

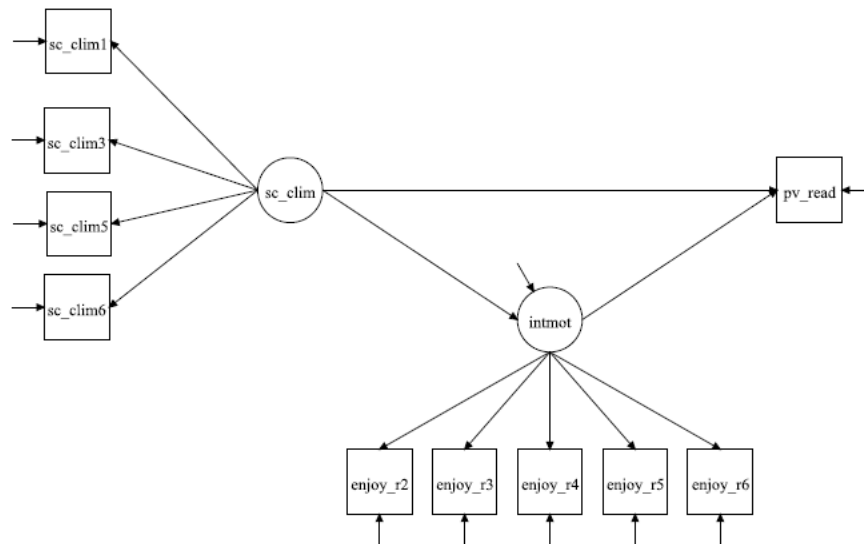
- 3) Как могут изменяться индексы соответствия модели данным на небольших выборках?
- 4) Возможности сравнения моделей при использовании MLR и WLSVM
- 5) Возможности и способы КФА для категориальных и дихотомических переменных

Примеры вопросов для итогового теста:

- 1. Рассчитайте коэффициент интраклассовой корреляции для достижений в математике в 8 классе (TIMSS (переменные `bsmmat01-bssmat05`, нужно сделать одну переменную – среднее для этих пяти)) и индекса СЭС (`escs`, не забудьте также про возможные миссинги). Выберите вариант ответа, который наиболее близок к полученным вами результатам.**
 - a. ИСС для СЭС выше, что говорит, о том, что сегрегация школ по СЭС выше, чем по академическим достижениям
 - b. ИСС для СЭС ниже, что говорит, о том, что сегрегация школ по СЭС выше, чем по академическим достижениям
 - c. ИСС для СЭС ниже, что говорит, о том, что сегрегация школ по СЭС ниже, чем по академическим достижениям
 - d. ИСС не отличаются для этих двух показателей
- 2. Исследователь захотел проверить гипотезу о том, что разрыв между школьниками из городских и деревенских школ (переменная `rural`, 0 – городские школы, 1 – деревенские) в баллах по математике будет выше в 9 классе (баллы PISA – переменная `PVMATH`, не забудьте проверить миссинги), по сравнению с 8-м классом (баллы TIMSS), при учете социально-экономического статуса родителей (`escs`, надо стандартизировать) и пола (переменная `girl`). Проверьте эту гипотезу (используйте модель `fixed slope`). Выберите все верные варианты ответа:**
 - a. Разрыв между городскими и деревенскими школьниками от 8-го к 9 классу значимо увеличился
 - b. Разрыв между городскими и деревенскими школьниками от 8-го к 9 классу значимо уменьшился
 - c. При контроле местоположения школы и СЭС в 9 классе мальчики показывают достижения выше
 - d. В 8 классе эффект СЭС значимо выше, чем в 9-м классе

- e. Разрыв между городскими и деревенскими школьниками от 8-го к 9 классу не изменился
- f. В 9 классе предсказанные значения для мальчиков из городских школ с СЭС = 0 равны 497.3 балла

3. Определите указанную ниже модель и впишите путевые коэффициенты (стандартизированные с 95% доверительным интервалом)



V. РЕСУРСЫ

5.1 Основная литература

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Hox J. (2002). *Multilevel Analysis: Techniques and Applications*

Kline, R. B. (2004). *Principles and Practice of Structural Equation Modeling*, (Methodology In The Social Sciences).

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

5.2 Дополнительная литература

- Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!*. Guilford Press.
- Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, 5(4), 362-387.
- Burchinal, M. R., Nelson, L., & Poe, M. (2006). IV. Growth curve analysis: An introduction to various methods for analyzing longitudinal data. *Monographs of the Society for Research in Child Development*, 71(3), 65-87
- De Leeuw, J., Meijer, E., & Goldstein, H. (2008). *Handbook of multilevel analysis*. New York: Springer.
- Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. *Psychological methods*, 12(1), 1.
- Hox, J. J., & Roberts, J. K. (2011). Multilevel analysis: Where we were and where we are. *Handbook of advanced multilevel analysis*, 1-11.
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and personality psychology compass*, 2(1), 302-317.
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92.
- Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology & Community Health*, 59(6), 443-449.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate behavioral research*, 42(1), 185-227.
- Rabe-Hesketh, S., & Skrondal, A. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Chapman and Hall/CRC.
- Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual review of psychology*, 52(1), 501-525.
- Snijders T., Bosker R. (1999). *Multilevel Analyses. An Introduction to Basic and Advanced Multilevel Modeling*.

Snijders, T. A. (2005). Power and sample size in multilevel modeling. *Encyclopedia of statistics in behavioral science*, 3, 1570-1573.

Tymms, P. (2004). Effect sizes in multilevel models. National Foundation for Educational Research.2865-2873.

Willms, J. D. (2006). *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems*. UNESCO Institute for Statistics

Zvoch, K., & Stevens, J. J. (2006). Longitudinal effects of school context and practice on middle school mathematics achievement. *The Journal of Educational Research*, 99(6), 347-357.

1. Программное обеспечение

№ п/п	Наименование	Условия доступа
1.	Microsoft Windows 10	<i>Из внутренней сети университета (договор)</i>
2.	Stata 13.0	<i>Лицензионное соглашение</i>
3.	Mplus	<i>Лицензионное соглашение</i>

5.3 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.	Базы данных PIRLS и TIMSS	https://timssandpirls.bc.edu/

2.	Базы данных PISA	http://www.oecd.org/pisa/
<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>		
1.	Сайт программы Mplus	URL: http://www.statmodel.com/
2.	Сайт программы Stata	https://www.stata.com/support/
3	Поддержка программы Stata, форум	https://www.statalist.org/forums/
4.	Материалы для изучения многоуровневой регрессии	http://www.bristol.ac.uk/cmm/learning/online-course/

5.4 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных и практических 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.