

Markus Tausendpfund et al.

Fortgeschrittene Analyseverfahren in den Sozialwissenschaften

Fakultät für
**Kultur- und
Sozialwissen-
schaften**

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Der Inhalt dieses Studienbriefs wird gedruckt auf Recyclingpapier (80 g/m², weiß), hergestellt aus 100 % Altpapier.

Vorwort

Dieser Kurs enthält überblicksartige Darstellungen weiterführender Analyseverfahren bzw. Forschungsmethoden, um eine sozialwissenschaftliche Forschungsfrage zu bearbeiten. Jedes Verfahren bzw. jede Methode wird in ihren Grundlagen vorgestellt und an einem Anwendungsbeispiel aus der Forschung erläutert. Die Diskussion der Möglichkeiten und Grenzen eines Verfahrens gibt wertvolle Orientierungen für die (eigene) praktische Anwendung.

Folgende Verfahren bzw. Methoden wurden in dem Kurs berücksichtigt: Lineare Regression, logistische Regression, Aggregatdatenanalyse, Mehrebenenanalyse, Fehlende Werte, Faktorenanalyse, Qualitative Comparative Analysis, Teilstandardisierte Experteninterviews, Inhaltsanalyse, Experimente und Big Data. Die vorgestellten Verfahren sind zwar nur eine Auswahl der existierenden Analyseverfahren bzw. Forschungsmethoden, aber sie stellen sicherlich die am häufigsten eingesetzten Verfahren in den Sozialwissenschaften dar.

Ich bin den Kolleginnen und Kollegen, die zu diesem Kurs beigetragen haben, zu großem Dank verpflichtet. Sie alle sind der Einladung, einen Beitrag zu verfassen, nicht nur mit Freude und Sorgfalt nachgekommen, sondern haben auch mein Drängen, den gemeinsamen Vorgaben zu folgen, mit viel Geduld und Offenheit ertragen. Neben den Autoren bin ich insbesondere Juliane Döschner und Daniel Saar zu Dank verpflichtet, die an der Schlussredaktion der einzelnen Beiträge beteiligt waren.

Aktuelle Ergänzungen sowie eine mögliche Errata-Liste zu diesem Kurs finden Sie in der Moodle-Lernumgebung des Moduls MB 2 „Forschungsmethoden in den Sozialwissenschaften“. Dort werden auch Übungsaufgaben und Tests veröffentlicht, die die Auseinandersetzung mit den Inhalten dieses Kurses vertiefen und die praktische Anwendung einzelner Verfahren trainieren. Die aktuelle pdf-Version des Kurses ist über den Virtuellen Studienplatz (VU) der FernUniversität zugänglich.

Über Hinweise auf Fehler, Kommentare und Verbesserungsvorschläge freue ich mich. Sie erreichen mich unter der E-Mail-Adresse Markus.Tausendpfund@FernUni-Hagen.de.

Hagen, im Mai 2020

Markus Tausendpfund

Inhaltsverzeichnis

Die Autorinnen und Autoren	5
Einführung (Markus Tausendpfund)	7
Lineare Regression (Gerrit Bauer)	11
Logistische Regression (Felix Weiss)	42
Aggregatdatenanalyse (Sebastian Kuhn)	66
Mehrebenenanalyse (Markus Tausendpfund)	102
Fehlende Werte (Michael Bergmann und Fabio Franzese)	141
Explorative Faktorenanalyse und Skalenkonstruktion (Christian Schnaudt)	175
Qualitative Comparative Analysis (Felix Hörisch und Matthias Heiken)	205
Teilstandardisierte Experteninterviews (Benjamin von dem Berge)	232
Inhaltsanalyse (Daniela Braun)	254
Experimente (Sascha Huber)	284
Big Data (Dominic Nyhuis)	318

Die Autorinnen und Autoren

Dr. **Gerrit Bauer** ist Akademischer Rat an der Ludwig-Maximilians-Universität (LMU) München. Seine Forschungsschwerpunkte umfassen familiensoziologische (Fertilität, Scheidung, Arbeitsteilung) sowie ungleichheitsrelevante Fragestellungen (z.B. zum Zusammenhang zwischen räumlicher und sozialer Mobilität). Derzeit beschäftigt er sich zudem mit Veränderungen der Lebenszufriedenheit im Lebensverlauf (kurzfristige Effekte der Geburt von Kindern sowie langfristige Effekte der sozialen Herkunft).

Dr. **Michael Bergmann** ist wissenschaftlicher Mitarbeiter am Munich Center for the Economics of Aging (MEA) des Max-Planck-Instituts für Sozialrecht und Sozialpolitik in München. Er leitet dort die Abteilung „Survey Methodology“ des Survey of Health, Ageing and Retirement in Europe (SHARE). Seine Forschungsinteressen liegen im Bereich der methodischen Umfrageforschung, insbesondere in der Untersuchung von Panel Conditioning, Panel Attrition und Intervieweffekten sowie -fälschungen in Längsschnittstudien.

Dr. **Daniela Braun** ist wissenschaftliche Assistentin am Lehrstuhl für Vergleichende Politikwissenschaft am Geschwister-Scholl-Institut (GSI) der Ludwig-Maximilians-Universität in München (LMU). Ihre Arbeitsbereiche umfassen die politische Einstellungs- und Verhaltensforschung sowie die Untersuchung der Wahlen zum Europäischen Parlament.

Fabio Franzese ist wissenschaftlicher Mitarbeiter am Munich Center for the Economics of Aging (MEA) des Max-Planck-Instituts für Sozialrecht und Sozialpolitik in München. Er arbeitet dort im Database Management für den Survey of Health, Ageing and Retirement in Europe (SHARE) und forscht zu den Themen Gesundheits- und Familiensoziologie.

Matthias Heiken ist Masterstudent der Politischen Wissenschaft an der Ruprecht-Karls-Universität Heidelberg. Seine Arbeitsschwerpunkte sind Arbeitsmarkt-, Sozial- und Asylpolitik in Deutschland und im internationalen Vergleich.

PD Dr. **Felix Hörisch** ist derzeit Vertretungsprofessor am Institut für Politische Wissenschaft der Ruprecht-Karls-Universität Heidelberg. In seiner Forschung befasst er sich unter anderem mit Arbeitsmarkt-, Sozial-, Fiskal- und Asylpolitik im nationalen und internationalen Vergleich sowie der Methode Qualitative Comparative Analysis. Zudem ist er Mitglied des Advisory Boards von COM-PASSS (COMPARative Methods for Systematic cross-caSe analySis), dem QCA-Netzwerk.

Prof. Dr. **Sascha Huber** ist Professor am Institut für Politikwissenschaft der Universität Mainz. In seiner Forschung befasst er sich unter anderem mit der politischen Urteils- und Entscheidungsbildung in Demokratien und der Anwendung experimenteller Methoden.

Sebastian Kuhn hat Politikwissenschaft, Soziologie und Germanistik an der Universität Stuttgart studiert und als wissenschaftlicher Mitarbeiter an der Friedrich-Schiller-Universität Jena und an der FernUniversität in Hagen gearbeitet. Seit 2016 ist er bei der AOK Baden-Württemberg beschäftigt und leitet dort das Referat „Datenmanagement“. Seine Forschungsinteressen liegen hauptsächlich im Bereich quantitativer Methoden der Datenanalyse.

Dr. **Dominic Nyhuis** ist Akademischer Rat an der Leibniz Universität Hannover, Arbeitsbereich für Vergleichende Regierungslehre und das Politische System Deutschlands. Seine Tätigkeitsschwerpunkte liegen im Bereich von Parteien, Parlamenten und der vergleichenden Kommunalpolitikforschung. Methodisch liegt sein Fokus auf der Auswertung neuartiger Datenquellen für die sozialwissenschaftliche Forschung und neueren Entwicklungen der quantitativen Methodik.

Dr. **Christian Schnaudt** ist wissenschaftlicher Mitarbeiter (Post-Doc) bei GESIS – Leibniz-Institut für Sozialwissenschaften in Mannheim und dort als Survey Manager für die deutsche Teilstudie des European Social Survey (ESS) tätig. Seine Forschungsinteressen liegen in den Bereichen der politischen Einstellungs- und Verhaltensforschung, quantitativer Forschungsmethoden sowie der Umfragemethodik.

Dr. **Markus Tausendpfund** ist wissenschaftlicher Mitarbeiter an der Fakultät für Kultur- und Sozialwissenschaften, Arbeitsstelle Quantitative Methoden, an der FernUniversität in Hagen. Seine Arbeitsgebiete umfassen Methoden der empirischen Sozialforschung, Einstellungs- und Verhaltensforschung sowie lokale Politikforschung.

Dr. **Benjamin von dem Berge** ist Forschungsreferent und Wissenschaftler an der Zeppelin Universität in Friedrichshafen. Davor war er Wissenschaftlicher Mitarbeiter am Lehrstuhl „Vergleich Politischer Systeme“ (Prof. Dr. Thomas Poguntke) an der Heinrich-Heine-Universität Düsseldorf sowie am Mannheimer Zentrum für Europäische Sozialforschung (MZES) der Universität Mannheim (Prof. Dr. Jan W. van Deth). Seine wissenschaftlichen Arbeitsgebiete umfassen politische Parteien und Parteiensysteme, Demokratisierung, Europäische Integration und Vergleichende Politikwissenschaft im Allgemeinen.

Dr. **Felix Weiss** ist Assistant Professor für Bildungssoziologie an der Universität Aarhus, Dänemark. Seine Forschungsschwerpunkte sind Soziale Stratifizierung und Mobilität, Bildungsungleichheit, Arbeitsmärkte und Lebenslaufsoziologie.

Einführung

Markus Tausendpfund

1 Zielsetzung des Kurses

Regression, Faktorenanalyse, Qualitative Comparative Analysis oder auch Inhaltsanalyse – in den Sozialwissenschaften gibt es zahlreiche Analyseverfahren, um eine Forschungsfrage zu bearbeiten. Dabei gibt es nicht „das“ Analyseverfahren oder „die“ Methode, die sich für alle denkbaren sozialwissenschaftlichen Fragestellungen eignet. Im Gegenteil: Die Wahl eines bestimmten Analyseverfahrens oder einer Forschungsmethode ist in erster Linie von der konkreten Forschungsfrage abhängig.¹

Der vorliegende Kurs möchte mit wichtigen Analyseverfahren bzw. Forschungsmethoden der Sozialwissenschaften vertraut machen. Trotz der knapp 340 Seiten deckt der Kurs zum einen nicht alle existierenden Analyseverfahren/Forschungsmethoden ab und bietet zum anderen nur eine erste Einführung in die jeweiligen Methoden bzw. Analyseverfahren. Bei der Auswahl der Verfahren wurden insbesondere solche berücksichtigt, die in den Sozialwissenschaften besonders häufig verwendet werden.

Die einzelnen Beiträge stellen jeweils ein spezifisches Analyseverfahren bzw. eine Methode vor und diskutieren Voraussetzungen sowie Vor- und Nachteile. Dabei werden die einzelnen Verfahren nicht abstrakt dargestellt, sondern jeweils an konkreten sozialwissenschaftlichen Fragestellungen illustriert. Die Auseinandersetzung mit den unterschiedlichen Verfahren soll einerseits die passive Methodenkompetenz, also das grundlegende Verständnis des jeweiligen Verfahrens, fördern und andererseits das Interesse wecken, sich mit ausgewählten Verfahren intensiver zu beschäftigen.

Im Rahmen von Haus- und Abschlussarbeiten können (und sollen) einzelne Analyseverfahren bzw. Methoden genutzt werden, um konkrete Forschungsfragen zu bearbeiten. Zu diesem Zweck werden in der Moodle-Lernumgebung Übungsaufgaben zur Verfügung gestellt, um die Vertrautheit mit spezifischen Verfahren zu stärken und die eigene Forschungskompetenz zu fördern.

2 Beiträge im Überblick

Dieser Kurs umfasst insgesamt elf inhaltliche Kapitel und gibt einen Überblick über wichtige Analyseverfahren bzw. Forschungsmethoden in den Sozialwissenschaften.

¹ Ausschließlich aus Gründen der besseren Lesbarkeit wird in diesem Kurs nicht durchgängig eine geschlechtsneutrale Sprache verwendet. Männliche, weibliche und genderneutrale Formen wechseln sich in diesem Kurs zufallsverteilt ab. Mit den Bezeichnungen sind jeweils alle Geschlechter gemeint.

Lineare Regression

Die Regressionsanalyse ist sicherlich das wichtigste und bekannteste Analyseverfahren in den Sozialwissenschaften. Insbesondere bei großen Fallzahlen sind regressionsanalytische Verfahren ein beliebtes Instrument, mit dessen Hilfe der Effekt einer oder mehrerer unabhängiger Variablen (uVs) auf eine abhängige Variable (aV) geschätzt wird. Die Bezeichnung Regressionsanalyse ist dabei ein Oberbegriff für eine Vielzahl an regressionsanalytischen Verfahren. Das einfachste regressionsanalytische Verfahren ist die lineare Regression (OLS), die unter anderem eine (pseudo-)metrische abhängige Variable voraussetzt. Der Beitrag von **Gerit Bauer** bietet eine Auffrischung bereits erworbener, aber möglicherweise vergessener Kenntnisse. Das wesentliche Ziel dieses Beitrags ist es, in zwei wichtige Erweiterungen des OLS-Grundmodells einzuführen: in Interaktionseffekte und in die Modellierung nicht-linearer Zusammenhänge.

Logistische Regression

In den Sozialwissenschaften werden viele Fragestellungen untersucht, die in Variablen mit zwei Ausprägungen (sogenannte dichotome oder binäre Variablen) abgebildet werden können. Typische Themen sind etwa die Wahlbeteiligung, der Kauf eines bestimmten Produkts oder auch der Tabakkonsum. Aufbauend auf der Darstellung von Kreuztabellen stellt **Felix Weiss** das lineare Wahrscheinlichkeitsmodell (linear probability model) sowie die logistische Regression vor. Dabei werden unterschiedliche Interpretationsmöglichkeiten der logistischen Regression vorgestellt und ihre Vor- und Nachteile diskutiert.

Aggregatdatenanalyse

Unter der Bezeichnung Aggregatdatenanalyse wird in diesem Kurs ein Analyseverfahren verstanden, in dessen Rahmen sich die verwendeten Daten auf die Aggregat- bzw. Kollektivebene beziehen und diese mithilfe statistischer Verfahren ausgewertet werden. Während sich bei der linearen und logistischen Regression die uVs und die aV meist auf der Mikroebene befinden, liegen die uVs und die aV bei der Aggregatdatenanalyse auf der Makroebene. Der Beitrag von **Sebastian Kuhn** gibt einen Überblick über die Grundlagen der Aggregatdatenanalyse und diskutiert Stärken und Schwächen der Methode.

Mehrebenenanalyse

Die zentrale Annahme der Mehrebenenanalyse (Kontextanalyse) ist, dass individuelle Einstellungen und Verhaltensweisen nicht nur eine Folge von individuellen Merkmalen sind, sondern auch das Resultat der Einwirkungen von kontextuellen Faktoren. Bei der Mehrebenenanalyse handelt es sich um ein regressionsanalytisches Verfahren, bei dem die aV auf der Mikroebene, die uVs auf der Mikro- und Makroebene angesiedelt sind. Die Grundlage einer Mehrebenenanalyse ist eine hierarchische Datenstruktur, deren Elemente der unteren Ebene jeweils genau einem Element der höheren Ebene zugeordnet sind (z.B. Personen in Ländern). Der Beitrag von **Markus Tausendpfund** bietet eine Einführung in die Logik und Vorgehensweise der Mehrebenenanalyse. Dabei werden die Voraussetzungen der Mehrebenenanalyse (z.B. Fallzahl) sowie die typische Analysestrategie vorgestellt. Der Beitrag gibt auch einen knappen Überblick über Weiterentwicklungen (z.B. Drei-Ebenen-Modell).

Fehlende Werte

Die meisten Daten, die in der empirischen Sozialforschung verwendet werden, stammen aus Befragungen. Ein Problem bei Befragungen ist es, dass die Personen nicht alle Fragen beantworten. Insbesondere bei sensiblen Fragen (z.B. zum Einkommen) verweigern die Befragten ihre Antwort. Eine hohe Verweigerung kann allerdings zu systematischen Verzerrungen bei den empirischen Ergebnissen führen. Deshalb zielt der Beitrag von **Michael Bergmann** und **Fabio Franzese** darauf ab, ein Bewusstsein für die Problematik von fehlenden

Werten zu schaffen und stellt Möglichkeiten vor, wie bei der Analyse von Befragungsdaten mit Item Nonresponse umgegangen werden kann. Sie stellen die Ursachen für Item Nonresponse und zentrale Ausfallmechanismen vor und diskutieren die gängigsten Verfahren für den Umgang mit fehlenden Werten.

In der empirischen Sozialforschung sind wir häufig mit komplexen Konzepten wie Vertrauen oder Legitimität konfrontiert, die nicht mit einem Indikator operationalisiert werden können. Die Messung eines komplexen Konzepts mittels mehrerer Items ist zwar im Allgemeinen einer Messung über lediglich ein einzelnes Item vorzuziehen, da sich hierdurch die mit der Messung behafteten Fehler im Durchschnitt reduzieren. Allerdings ergibt sich aus der Verwendung mehrerer Items oder Variablen zwangsläufig die Frage, ob diese überhaupt zur Messung ein und desselben theoretischen Konzepts geeignet sind und folglich zu einer gemeinsamen Skala zusammengefasst werden können. Zur Prüfung, ob mehrere verschiedene Frageitems oder Variablen ein und dasselbe theoretische Konzept erfassen und somit zu einer gemeinsamen Skala zusammengefasst werden können, greifen Sozialwissenschaftlerinnen auf die sogenannte Faktorenanalyse zurück. **Christian Schnaudt** stellt die generellen Grundlagen der Faktorenanalyse vor und differenziert verschiedene faktorenanalytische Verfahren, die in der sozialwissenschaftlichen Forschung häufig eingesetzt werden.

Faktorenanalyse

Die Qualitative Comparative Analysis (QCA) ist ein vor rund 30 Jahren von Charles Ragin entwickeltes Verfahren, das sich in den letzten Jahren in den Sozialwissenschaften zunehmend etabliert hat. Zentrales Ziel der Methode ist die Identifikation von notwendigen und hinreichenden Bedingungen für eine bestimmte abhängige Variable, in der Terminologie der Methode eines bestimmten Outcomes. Während Sozialwissenschaftler bei großen Fallzahlen insbesondere auf regressionsanalytische Verfahren zurückgreifen, nutzen sie insbesondere bei mittleren Fallzahlen die QCA. Der Beitrag von **Felix Hörisch** und **Matthias Heiken** macht mit der Terminologie und den Grundbegriffen der QCA vertraut. Darauf aufbauend werden zwei Arten der QCA erläutert: crisp-set und fuzzy-set QCA.

Qualitative Comparative Analysis

Der Beitrag von **Benjamin von dem Berge** macht mit teilstandardisierten Experteninterviews vertraut. Dabei handelt es sich um eine Methode der Datenerhebung, nicht der Datenanalyse. Das Kapitel bietet eine Anleitung zur Durchführung von Experteninterviews. Wer im Rahmen teilstandardisierter Experteninterviews befragt wird, hängt vom Forschungsgegenstand der jeweiligen Studie ab. Werden beispielsweise Sachverhalte in politischen Parteien untersucht, dann können Parteifunktionäre befragt werden, bei Sachverhalten in kommunalen Verwaltungen könnten dies Mitarbeitende der Verwaltungen sein. Neben der Auswahl der Experten behandelt der Beitrag auch Fragen zur Datenqualität und gibt praktische Hinweise zur Durchführung von Experteninterviews.

Teilstandardisierte Experteninterviews

In der Methodenliteratur wird die Inhaltsanalyse zwar traditionell als ein Datenerhebungsverfahren eingeordnet, allerdings handelt es sich auch um ein Verfahren der Datenanalyse. Der Beitrag von **Daniela Braun** deckt beide Aspekte der Inhaltsanalyse ab. Zunächst wird der Schritt der Datenerhebung anhand des Euromanifesto-Projekts dargestellt. Im Rahmen des Euromanifesto-Projekts werden die Wahlprogramme der Parteien, die zu den Europawahlen antreten, gesammelt und anhand eines Codierschemas systematisch codiert. Darauf

Inhaltsanalyse

aufbauend wird auf die Möglichkeit eingegangen, die Inhaltsanalyse als Analyseverfahren zu verwenden. Dabei wird exemplarisch untersucht, wie sich die Salienz europäischer Polices im Europawahlkampf über die Zeit verändert hat.

| Experimente

Während regressionsanalytische Verfahren streng genommen nur Beziehungen (Korrelationen) zwischen Variablen untersuchen können, bieten Experimente die Möglichkeit, kausale Beziehungen zu entdecken. Das Experiment gilt zwar allgemein als Idealtyp empirischer Forschung, um kausale Zusammenhänge entdecken bzw. prüfen zu können, allerdings wurde lange Zeit der Gebrauch dieser Methode in der Politikwissenschaft oder Soziologie als nicht anwendbar betrachtet. Erst in jüngerer Zeit ist wieder eine stärkere Aufmerksamkeit gegenüber experimentellen Untersuchungsformen festzustellen. Der Beitrag von **Sascha Huber** bietet eine Einführung in sozialwissenschaftliche Experimente und gibt einen Überblick über Ziele und Formen von Experimenten. Daran anschließend arbeitet er heraus, unter welchen Voraussetzungen sozialwissenschaftliche Experimente eine lohnenswerte Forschungsstrategie darstellen.

| Big Data

„Big Data“ ist derzeit in aller Munde. Durch die Digitalisierung sind heute mehr Daten verfügbar als jemals zuvor. Zum Abschluss und als Ausblick auf aktuelle und künftige Entwicklungen behandelt **Dominic Nyhuis** in diesem Kurs das Thema Big Data, welches für die Sozialwissenschaften zunehmend an Bedeutung gewinnt. Er behandelt einerseits die Grundlagen der automatischen Webdatensammlung und stellt andererseits Überlegungen zur Auswertung großer, wenig strukturierter Daten vor. Dabei macht Nyhuis auch mit den technischen Grundlagen der webbasierten Datensammlung vertraut und stellt Beispiele aus der sozialwissenschaftlichen Big Data-Forschung vor.