

Markus Tausendpfund
Simone Abendschön

Quantitative Analyseverfahren. Eine Einführung

Redaktion und Überarbeitung:
Davin Akko, Anne-Kathrin Bestgen, Felicitas Kempf und Julia Schütz

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 der Modulbetreuung

Dieser Kurs bietet den Studierenden der Bildungswissenschaft eine Einführung in quantitative Analyseverfahren und einen verständlichen Einstieg, wie Daten in den Sozial- und Bildungswissenschaften quantitativ ausgewertet werden können. Die Bildungswissenschaft ist Teil der Sozialwissenschaften und beide greifen auf nahezu identische Forschungsmethoden zurück. Die empirische Bildungsforschung ist interdisziplinär zwischen Soziologie, Psychologie und auch Ökonomie aufgestellt, zentral jedoch erziehungs- und bildungswissenschaftlich verankert und untersucht mittels Methoden der empirischen Sozialforschung Bildungs-, Lern- und Reflexionsprozesse.

Mit der Bearbeitung dieses Kurses sind folgende Lernziele verbunden:¹

- Sie können verschiedene quantitative Analyseverfahren der empirischen Bildungsforschung benennen, beschreiben und anwenden.
- Sie können Verteilungen anhand statistischer Kennwerte beschreiben sowie deskriptive Ergebnisse interpretieren.
- Sie kennen verschiedene grafische Darstellungsformen und können geeignete Formen für die Darstellung statistischer Informationen auswählen.
- Sie können Strategien Methoden und Abläufe der empirischen Datenauswertung beschreiben, erklären und anwenden.
- Sie können verschiedene quantitative Analyseverfahren der empirischen Bildungsforschung unterscheiden und für ihre Forschungsfrage geeignete Verfahren auswählen.
- Sie können die Ergebnisse uni-, bi- und multivariater Analysen auswerten und interpretieren.
- Sie sind in der Lage Analyseverfahren im Hinblick auf ihre Angemessenheit für bestimmte Fragestellungen sowie Anwendungsvoraussetzungen zu beurteilen.

Den Autor Markus Tausendpfund und die Autorin Simone Abendschön möchten wir Ihnen gerne kurz vorstellen:

Dr. Markus Tausendpfund studierte Sozialwissenschaften mit den Schwerpunkten Soziologie, Sozialpsychologie, Methoden der empirischen Sozialforschung, Politische Soziologie und Arbeits- und Organisationspsychologie an der Universität Mannheim. 2012 schloss er seine Promotion zum Thema „Individuelle und kontextuelle Faktoren der politischen Unterstützung der Europäischen Union“ ab und leitet seit 2014 die Arbeitsstelle Quantitative Methoden an der FernUniversität in Hagen.

Prof.'in Dr.'in Simone Abendschön studierte Geschichte, Soziologie, Politikwissenschaft und Anglistik in Freiburg und Mannheim und promovierte 2010 zur Sozialisation von politischen und

¹ Die Lernziele orientieren sich an der Lernzieltaxonomie nach Bloom, B. et al. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York, Toronto: Longmans, Green.

demokratischen Werten im jungen Kindesalter. Simone Abendschön war bereits an den Universitäten Mannheim, Bamberg und Frankfurt tätig und hat seit 2015 die Professur für Politikwissenschaft mit dem Schwerpunkt Methoden der Politikwissenschaft an der Justus-Liebig-Universität Gießen inne.

An dieser Stelle möchten wir uns insbesondere bei Markus Tausendpfund für die angenehme Kooperation und den stets interessanten Austausch bedanken.

Der Studienbrief wurde von Davin Akko, M.Sc., Dr.'in Anne-Kathrin Bestgen, Felicitas Kempf, M.A. und Prof.'in Dr.'in Julia Schütz am Lehrgebiet Empirische Bildungsforschung redaktionell überarbeitet. Dabei wurden keine inhaltlichen Änderungen vorgenommen, sondern Änderungen aufgrund eines inklusiven Sprachgebrauchs eingefügt sowie eine barrierefreie Gestaltung beachtet. Zudem werden Bildungswissenschaftler*innen explizit als Zielgruppe angesprochen. In der Moodle-Lernumgebung des Moduls werden Lehrvideos und Übungsaufgaben veröffentlicht, die die Auseinandersetzung mit den Inhalten des Kurses fördern sollen.

Wir wünschen Ihnen viel Erfolg bei der Bearbeitung und eine anregende Lektüre!

Davin Akko, Anne-Kathrin Bestgen, Felicitas Kempf und Julia Schütz

Vorwort der Autor*innen

In der quantitativen Sozial- und Bildungsforschung wird zur Beschreibung von Daten und zur empirischen Überprüfung von Hypothesen auf statistische Verfahren zurückgegriffen. Wer eine (quantitative) Studie verstehen und kritisch bewerten möchte, der muss die grundlegenden Prinzipien, Anwendungsvoraussetzungen und auch Probleme der verwendeten statistischen Verfahren kennen. Für Sozial- und Bildungswissenschaftler*innen sind deshalb elementare Kenntnisse dieser quantitativen Analyseverfahren unverzichtbar.

Für die Sozial- und Bildungswissenschaften stellt die Statistik eine zentrale Hilfswissenschaft dar. Während sich Statistiker*innen – allgemeiner: Mathematiker*innen – häufig mit der Beweisführung und der Weiterentwicklung mathematischer Algorithmen beschäftigen, stehen für Studierende der Bildungswissenschaft das Kennenlernen und die praktische Anwendung statistischer Verfahren im Vordergrund. Im Mittelpunkt des Kurses steht das Verständnis quantitativer Analyseverfahren, mit denen Sozial- und Bildungswissenschaftler*innen bei der Auseinandersetzung mit quantitativen Studien konfrontiert werden.

Der vorliegende Kurs behandelt vier Themenbereiche: Univariate, bivariate und multivariate Datenanalyse sowie Grundlagen der Inferenzstatistik. Das Kapitel zur univariaten Datenanalyse behandelt die Häufigkeitsverteilung einzelner Merkmale. Dabei werden Lage- und Streuungsmaße sowie Formmaße vorgestellt. Die bivariate Datenanalyse untersucht Zusammenhänge zwischen zwei Merkmalen und Unterschiede zwischen zwei Merkmalen (Mittelwertvergleiche). Dabei werden Kreuztabellen sowie wichtige Zusammenhangsmaße behandelt. Bei der multivariaten Datenanalyse werden mit der linearen und logistischen Regression zwei zentrale Analyseverfahren der Sozialwissenschaften vorgestellt, die den Einfluss mehrerer unabhängiger Variablen auf eine abhängige Variable schätzen können. Aus zeitlichen, finanziellen und forschungspraktischen Gründen dominieren in den Sozial- und Bildungswissenschaften Stichproben. Deshalb behandelt der vierte Teil des Kurses die Grundlagen der Inferenzstatistik, die Instrumente zur Verfügung stellt, um zu entscheiden, ob und wie empirische Befunde aus Zufallsstichproben auf zugehörige Grundgesamtheiten übertragen werden dürfen.

In der Moodle-Lernumgebung des Moduls werden Videos und Übungsaufgaben veröffentlicht, die die Auseinandersetzung mit den Inhalten des Kurses fördern sollen. Für die kritische Durchsicht des Kurses sind wir Christian Cleve und Daniel Saar sehr dankbar.

Über Hinweise auf Fehler, Kommentare und Verbesserungsvorschläge freuen wir uns. Senden Sie Ihre Kommentare bitte an Markus.Tausendpfund@fernuni-hagen.de. Vielen Dank.

Hagen, im Januar 2020

Markus Tausendpfund und Simone Abendschön

Inhaltsverzeichnis

Vorwort der Modulbetreuung.....	III
Vorwort der Autor*innen	V
Inhaltsverzeichnis	VI
Abbildungsverzeichnis	VIII
Tabellenverzeichnis	IX
1 Einführung.....	11
1.1 Einordnung im Forschungsprozess	11
1.2 Grundgesamtheit und Stichprobe	14
1.3 Klassifikationen von Variablen	16
2 Univariate Datenanalyse.....	19
2.1 Häufigkeitstabelle	19
2.2 Lagemaße	23
2.2.1 Modus	23
2.2.2 Median	24
2.2.3 Arithmetisches Mittel	26
2.3 Streuungsmaße	29
2.3.1 Varianz	29
2.3.2 Standardabweichung	33
2.4 Formmaße	33
2.4.1 Schiefe	34
2.4.2 Wölbung	37
2.5 Variablen standardisieren (z-Transformation)	38
2.6 Grafische Darstellungen	40
2.6.1 Säulen- und Balkendiagramm	40
2.6.2 Kreisdiagramm	41
2.6.3 Histogramm	42
2.6.4 Boxplot	44
3 Bivariate Datenanalyse.....	46
3.1 Kreuztabellen	47
3.2 Zusammenhangsmaße für nominale Merkmale	55
3.3 Zusammenhangsmaße für ordinale Merkmale	61
3.4 Zusammenhangsmaße für metrische Merkmale	65
3.5 Eta-Quadrat für metrische und nominale Merkmale	74

3.6	Zusammenfassung	79
4	Multivariate Datenanalyse	80
4.1	Einführung	80
4.2	Lineare Regression	82
4.2.1	Bivariate Regression	83
4.2.2	Multiple Regression	90
4.3	Logistische Regression	101
4.3.1	Bivariate Regression	102
4.3.2	Multiple Regression	106
5	Inferenzstatistik	112
5.1	Was ist das Problem?	112
5.2	Zentrale Konzepte der Inferenzstatistik	117
5.2.1	Zentraler Grenzwertsatz und Normalverteilung	117
5.2.2	Standardfehler	121
5.3	Schätzungsarten	127
5.3.1	Punktschätzung	127
5.3.2	Intervallschätzung	130
5.3.3	Berechnung der benötigten Fallzahl	138
5.3.4	Anwendungsprobleme in der Praxis	140
5.4	Statistisches Testen	142
5.4.1	Allgemeine Vorgehensweise bei einem Signifikanztest	144
5.4.2	Alpha- und Beta-Fehler	147
5.4.3	t-Test	149
6	Literatur	163

Abbildungsverzeichnis

Abbildung 1: Idealtypischer Ablauf eines quantitativen Forschungsprojekts	12
Abbildung 2: Grundgesamtheit und Stichprobe.....	15
Abbildung 3: Normalverteilung	34
Abbildung 4: Schiefe	35
Abbildung 5: Empirische Verteilungen mit unterschiedlicher Schiefe	36
Abbildung 6: Wölbung.....	37
Abbildung 7: Säulendiagramm des Interesses an Politik (in Prozent, n = 3490)	41
Abbildung 8: Balkendiagramm des Interesses an Politik (absolute Häufigkeiten, n = 3490)	41
Abbildung 9: Zweitstimmen bei der Bundestagswahl 2017 (in Prozent)	42
Abbildung 10: Histogramm des Alters (absolute Häufigkeiten, n = 3486)	43
Abbildung 11: Elemente eines Boxplots	44
Abbildung 12: Boxplot der Interviewdauer (n = 3479)	45
Abbildung 13: IQ und Testergebnis beim räumlichen Denken – Streudiagramm	66
Abbildung 14: Weitere Arten des Zusammenhangs von zwei Merkmalen.....	67
Abbildung 15: Nettoeinkommen und Lebenszufriedenheit – Streudiagramm.....	72
Abbildung 16: Streudiagramm.....	85
Abbildung 17: Streudiagramm mit OLS-Regressionsgerade	87
Abbildung 18: Schematische Darstellung der vermuteten multivariaten Einflusstruktur	92
Abbildung 19: Streudiagramm mit Regressionskurve	105
Abbildung 20: Grundgesamtheit und Stichprobe.....	112
Abbildung 21: Rückschluss von der Stichprobe auf die Grundgesamtheit	113
Abbildung 22: Wiederholte Ziehung von Zufallsstichproben	119
Abbildung 23: Normalverteilung	121
Abbildung 24: Abweichungen einzelner Stichprobenmittelwerte vom wahren Mittelwert	122
Abbildung 25: Stichprobenverteilungen bei unterschiedlicher Fallzahl	123
Abbildung 26: Ergebnisse des Politbarometers zu zwei Zeitpunkten (in Prozent)	130
Abbildung 27: 95-Prozent-Konfidenzintervall	131
Abbildung 28: 99-Prozent-Konfidenzintervall	132
Abbildung 29: Fiktive Befragung zur Wahlentscheidung von 1000 Personen (in Prozent)	135
Abbildung 30: 95-Prozent-Konfidenzintervalle (Stichprobengröße jeweils 1000 Personen)	137
Abbildung 31: Schätzen und Testen im Vergleich	143
Abbildung 32: t-Verteilung und Normalverteilung	150
Abbildung 33: Verschiedene t-Verteilungen	151
Abbildung 34: Varianten des t-Tests	151
Abbildung 35: Einseitiger und zweiseitiger t-Test.....	153

Tabellenverzeichnis

Tabelle 1: Zulässige Rechenoperationen in Abhängigkeit vom Skalenniveau	17
Tabelle 2: Interesse an Politik	19
Tabelle 3: Subjektive Schichteinstufung.....	22
Tabelle 4: Lagemaße und Skalenniveau	23
Tabelle 5: Berechnung des Modus	23
Tabelle 6: Geschlecht	24
Tabelle 7: Berechnung des Medians (ungerade Fallzahl).....	25
Tabelle 8: Berechnung des Medians (gerade Fallzahl).....	25
Tabelle 9: Interesse an Politik	26
Tabelle 10: Berechnung des arithmetischen Mittels bei kleinen Fallzahlen.....	27
Tabelle 11: Berechnung des arithmetischen Mittels bei großen Fallzahlen.....	28
Tabelle 12: Mittelwerte und Ausreißer	28
Tabelle 13: Lebenszufriedenheit von zwei Gruppen	29
Tabelle 14: Arbeitstabelle für die Berechnung der Varianz (kleine Fallzahl).....	31
Tabelle 15: Arbeitstabelle für die Berechnung der Varianz (große Fallzahl).....	32
Tabelle 16: Variablen standardisieren	39
Tabelle 17: Bivariate Zusammenhangsmaße in Abhängigkeit vom Skalenniveau	46
Tabelle 18: Urliste – Abendliche Bibliotheksnutzung und Studiengang (n = 9)	48
Tabelle 19: Kreuztabelle – Abendliche Bibliotheksnutzung und Studiengang (n = 9)	48
Tabelle 20: Abendliche Bibliotheksnutzung und Studiengang – Zeilenprozente (n = 100).....	49
Tabelle 21: Abendliche Bibliotheksnutzung und Studiengang – Spaltenprozente (n = 100)	50
Tabelle 22: Abendliche Bibliotheksnutzung und Studiengang – Gesamtprozente (n = 100).....	50
Tabelle 23: Politisches Interesse und Geschlecht (Spaltenprozente).....	51
Tabelle 24: Schulabschluss und elterlicher Bildungshintergrund (Spaltenprozente)	54
Tabelle 25: Schulabschluss und elterlicher Bildungshintergrund (Zeilenprozente).....	55
Tabelle 26: Politisches Interesse und Geschlecht (beobachtete Häufigkeiten) – Kontingenztafel	56
Tabelle 27: Berechnung der erwarteten Häufigkeiten	56
Tabelle 28: Politisches Interesse und Geschlecht (erwartete Häufigkeiten) – Indifferenztafel ..	57
Tabelle 29: Arbeitstabelle zur Berechnung von Chi-Quadrat.....	58
Tabelle 30: Interpretation von Cramer's V	60
Tabelle 31: Interpretation von Spearman's Rho.....	62
Tabelle 32: Soziale Schicht und Gesundheitszustand.....	63
Tabelle 33: Arbeitstabelle zur Berechnung von Spearman's Rho	64
Tabelle 34: IQ und Testergebnis beim räumlichen Denken – Urliste	65
Tabelle 35: Arbeitstabelle zur Berechnung der Kovarianz	68
Tabelle 36: Interpretation von Pearson's r.....	69
Tabelle 37: Arbeitstabelle zur Berechnung von Pearson's r	70
Tabelle 38: Nettoeinkommen und Lebenszufriedenheit – Urliste.....	71
Tabelle 39: Arbeitstabelle zur Berechnung von Pearson's r	73
Tabelle 40: Zwischenergebnisse zur Berechnung von Pearson's r	73
Tabelle 41: Interpretation von Eta-Quadrat.....	76
Tabelle 42: Migrationshintergrund und politisches Wissen	76
Tabelle 43: Arbeitstabelle Migrationshintergrund und politisches Wissen.....	77

Tabelle 44: Arbeitstabelle Migrationshintergrund (Nein) und politisches Wissen.....	78
Tabelle 45: Arbeitstabelle Migrationshintergrund (Ja) und politisches Wissen.....	78
Tabelle 46: Unterschiedliche Bezeichnungen für Variablen der Regressionsanalyse.....	81
Tabelle 47: Bivariate lineare Regression mit Lebenszufriedenheit und Einkommen	84
Tabelle 48: Dummy-Kodierung für Familienstand	93
Tabelle 49: Bestimmungsfaktoren der Lebenszufriedenheit (Teil 1).....	95
Tabelle 50: Bestimmungsfaktoren der Lebenszufriedenheit (Teil 2).....	99
Tabelle 51: Bivariate logistische Regression mit Wahlbeteiligung und Alter	103
Tabelle 52: Bestimmungsfaktoren der Wahlbeteiligung	108
Tabelle 53: Mittelwerte in Zufallsstichproben (Stichprobengröße jeweils 1000 Personen)	114
Tabelle 54: Mittelwerte von Zufallsstichproben (Stichprobengröße jeweils 1000 Personen)	116
Tabelle 55: Vergleich zwischen Standardfehler und Standardabweichung	124
Tabelle 56: Mittelwerte von Zufallsstichproben.....	128
Tabelle 57: Erforderliche Stichprobengröße	139
Tabelle 58: Fehlerarten beim Hypothesentest	147
Tabelle 59: Lebenszufriedenheit von Frauen und Männern	154
Tabelle 60: Kritische Werte der t-Verteilung	156
Tabelle 61: Lebenszufriedenheit von West- und Ostdeutschen	157
Tabelle 62: Zufriedenheit mit der Demokratie.....	158
Tabelle 63: Beispieldaten für die Berechnung eines t-Tests bei abhängigen Stichproben.....	160