

Markus Tausendpfund  
Simone Abendschön

# Quantitative Analyseverfahren. Eine Einführung

Redaktion und Überarbeitung:  
Felicitas Sander, Davin P. Akko 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. Wir weisen darauf hin, dass die vorgenannten Verwertungsalternativen je nach Ausgestaltung der Nutzungsbedingungen bereits durch Einstellen in Cloud-Systeme verwirklicht sein können. Die FernUniversität bedient sich im Falle der Kenntnis von Urheberrechtsverletzungen sowohl zivil- als auch strafrechtlicher Instrumente, um ihre Rechte geltend zu machen.

Der Inhalt dieses Studienbriefs wird gedruckt auf Recyclingpapier (80 g/m<sup>2</sup>, 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:<sup>1</sup>

- 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 demokratischen Werten im jungen Kindesalter. Simone Abendschön war bereits an den Universitäten Mannheim, Bamberg und Frankfurt tätig und hat seit 2015 die Professur für

---

<sup>1</sup> 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.

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 Felicitas Sander, M.A., Davin P. Akko, M.Sc. 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!

Felicitas Sander, Davin P. Akko 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](mailto:Markus.Tausendpfund@fernuni-hagen.de). Vielen Dank.

Hagen, im Januar 2020

Markus Tausendpfund und Simone Abendschön

## Inhaltsverzeichnis

Vorwort der Modulbetreuung.....	3
Vorwort der Autor*innen .....	5
Inhaltsverzeichnis .....	6
Abbildungsverzeichnis.....	9
Tabellenverzeichnis .....	10
1 Einführung.....	12
1.1 Einordnung im Forschungsprozess .....	12
1.2 Grundgesamtheit und Stichprobe .....	15
1.3 Klassifikationen von Variablen.....	17
2 Univariate Datenanalyse .....	20
2.1 Häufigkeitstabelle.....	20
2.2 Lagemaße .....	24
2.2.1 Modus.....	24
2.2.2 Median .....	26
2.2.3 Arithmetisches Mittel .....	28
2.3 Streuungsmaße .....	30
2.3.1 Varianz .....	31
2.3.2 Standardabweichung .....	34
2.4 Formmaße .....	35
2.4.1 Schiefe.....	36
2.4.2 Wölbung .....	39
2.5 Variablen standardisieren (z-Transformation) .....	40
2.6 Grafische Darstellungen.....	42
2.6.1 Säulen- und Balkendiagramm .....	42
2.6.2 Kreisdiagramm.....	43
2.6.3 Histogramm.....	44
2.6.4 Boxplot.....	45
3 Bivariate Datenanalyse.....	47
3.1 Kreuztabellen .....	48
3.2 Zusammenhangsmaße für nominale Merkmale .....	56
3.3 Zusammenhangsmaße für ordinale Merkmale .....	62
3.4 Zusammenhangsmaße für metrische Merkmale.....	67
3.5 Eta-Quadrat für metrische und nominale Merkmale.....	76

---

3.6	Zusammenfassung.....	80
4	Multivariate Datenanalyse.....	82
4.1	Einführung.....	82
4.2	Lineare Regression.....	84
4.2.1	Bivariate Regression.....	85
4.2.2	Multiple Regression.....	91
4.3	Logistische Regression.....	102
4.3.1	Bivariate Regression.....	104
4.3.2	Multiple Regression.....	107
5	Inferenzstatistik.....	113
5.1	Was ist das Problem?.....	113
5.2	Zentrale Konzepte der Inferenzstatistik.....	118
5.2.1	Zentraler Grenzwertsatz und Normalverteilung.....	118
5.2.2	Standardfehler.....	121
5.3	Schätzungsarten.....	127
5.3.1	Punktschätzung.....	127
5.3.1.1	Kriterien einer „guten“ Schätzung.....	127
5.3.1.2	Schätzer für Mittelwerte.....	129
5.3.1.3	Schätzer für Anteilswerte.....	129
5.3.1.4	Schätzer für Varianzen.....	130
5.3.2	Intervallschätzung.....	130
5.3.2.1	Konfidenzintervalle für Mittelwerte.....	133
5.3.2.2	Konfidenzintervalle für Anteilswerte.....	134
5.3.2.3	Interpretation von Konfidenzintervallen.....	136
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.....	145
5.4.2	Alpha- und Beta-Fehler.....	148
5.4.3	t-Test.....	149
5.4.3.1	t-Test für unabhängige Stichproben mit homogenen Varianzen.....	153
5.4.3.2	t-Test für unabhängige Stichproben mit heterogenen Varianzen.....	159
5.4.3.3	t-Test für abhängige Stichproben.....	161
5.4.3.4	Voraussetzungen für einen t-Test.....	163

6	Literatur.....	165
---	----------------	-----



## Abbildungsverzeichnis

Abbildung 1: Idealtypischer Ablauf eines quantitativen Forschungsprojekts .....	13
Abbildung 2: Grundgesamtheit und Stichprobe.....	16
Abbildung 3: Normalverteilung.....	36
Abbildung 4: Schiefe .....	37
Abbildung 5: Empirische Verteilungen mit unterschiedlicher Schiefe .....	38
Abbildung 6: Wölbung.....	39
Abbildung 7: Säulendiagramm des Interesses an Politik (in Prozent, n = 3490) .....	42
Abbildung 8: Balkendiagramm des Interesses an Politik (absolute Häufigkeiten, n = 3490).....	43
Abbildung 9: Zweitstimmen bei der Bundestagswahl 2017 (in Prozent).....	44
Abbildung 10: Histogramm des Alters (absolute Häufigkeiten, n = 3486).....	45
Abbildung 11: Elemente eines Boxplots.....	45
Abbildung 12: Boxplot der Interviewdauer (n = 3479) .....	46
Abbildung 13: IQ und Testergebnis beim räumlichen Denken – Streudiagramm .....	68
Abbildung 14: Weitere Arten des Zusammenhangs von zwei Merkmalen.....	69
Abbildung 15: Nettoeinkommen und Lebenszufriedenheit – Streudiagramm.....	74
Abbildung 16: Streudiagramm.....	87
Abbildung 17: Streudiagramm mit OLS-Regressionsgerade .....	89
Abbildung 18: Schematische Darstellung der vermuteten multivariaten Einflusstruktur.....	94
Abbildung 19: Streudiagramm mit Regressionskurve .....	106
Abbildung 20: Grundgesamtheit und Stichprobe.....	113
Abbildung 21: Rückschluss von der Stichprobe auf die Grundgesamtheit .....	114
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) .....	131
Abbildung 27: 95-Prozent-Konfidenzintervall .....	132
Abbildung 28: 99-Prozent-Konfidenzintervall .....	133
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 .....	144
Abbildung 32: t-Verteilung und Normalverteilung .....	151
Abbildung 33: Verschiedene t-Verteilungen .....	152
Abbildung 34:Varianten des t-Tests .....	152
Abbildung 35: Einseitiger und zweiseitiger t-Test.....	154

## Tabellenverzeichnis

Tabelle 1: Zulässige Rechenoperationen in Abhängigkeit vom Skalenniveau .....	18
Tabelle 2: Interesse an Politik .....	20
Tabelle 3: Subjektive Schichteinstufung .....	23
Tabelle 4: Lagemaße und Skalenniveau .....	24
Tabelle 5: Berechnung des Modus .....	25
Tabelle 6: Geschlecht .....	25
Tabelle 7: Berechnung des Medians (ungerade Fallzahl) .....	26
Tabelle 8: Berechnung des Medians (gerade Fallzahl) .....	26
Tabelle 9: Interesse an Politik .....	27
Tabelle 10: Berechnung des arithmetischen Mittels bei kleinen Fallzahlen .....	28
Tabelle 11: Berechnung des arithmetischen Mittels bei großen Fallzahlen .....	29
Tabelle 12: Mittelwerte und Ausreißer .....	30
Tabelle 13: Lebenszufriedenheit von zwei Gruppen .....	31
Tabelle 14: Arbeitstabelle für die Berechnung der Varianz (kleine Fallzahl) .....	33
Tabelle 15: Arbeitstabelle für die Berechnung der Varianz (große Fallzahl) .....	34
Tabelle 16: Variablen standardisieren .....	40
Tabelle 17: Bivariate Zusammenhangsmaße in Abhängigkeit vom Skalenniveau .....	48
Tabelle 18: Urliste – Abendliche Bibliotheksnutzung und Studiengang (n = 9) .....	49
Tabelle 19: Kreuztabelle – Abendliche Bibliotheksnutzung und Studiengang (n = 9) .....	50
Tabelle 20: Abendliche Bibliotheksnutzung und Studiengang – Zeilenprozente (n = 100) .....	51
Tabelle 21: Abendliche Bibliotheksnutzung und Studiengang – Spaltenprozente (n = 100) .....	51
Tabelle 22: Abendliche Bibliotheksnutzung und Studiengang – Gesamtprozente (n = 100) .....	52
Tabelle 23: Politisches Interesse und Geschlecht (Spaltenprozente) .....	53
Tabelle 24: Schulabschluss und elterlicher Bildungshintergrund (Spaltenprozente) .....	55
Tabelle 25: Schulabschluss und elterlicher Bildungshintergrund (Zeilenprozente) .....	56
Tabelle 26: Politisches Interesse und Geschlecht (beobachtete Häufigkeiten) – Kontingenztafel .....	57
Tabelle 27: Berechnung der erwarteten Häufigkeiten .....	58
Tabelle 28: Politisches Interesse und Geschlecht (erwartete Häufigkeiten) – Indifferenztafel .....	58
Tabelle 29: Arbeitstabelle zur Berechnung von Chi-Quadrat .....	59
Tabelle 30: Interpretation von Cramer's V .....	61
Tabelle 31: Interpretation von Spearman's Rho .....	63
Tabelle 32: Soziale Schicht und Gesundheitszustand .....	65
Tabelle 33: Arbeitstabelle zur Berechnung von Spearman's Rho .....	65
Tabelle 34: IQ und Testergebnis beim räumlichen Denken – Urliste .....	67
Tabelle 35: Arbeitstabelle zur Berechnung der Kovarianz .....	70
Tabelle 36: Interpretation von Pearson's r .....	71
Tabelle 37: Arbeitstabelle zur Berechnung von Pearson's r .....	71
Tabelle 38: Nettoeinkommen und Lebenszufriedenheit – Urliste .....	73
Tabelle 39: Arbeitstabelle zur Berechnung von Pearson's r .....	74
Tabelle 40: Zwischenergebnisse zur Berechnung von Pearson's r .....	75
Tabelle 41: Interpretation von Eta-Quadrat .....	78
Tabelle 42: Migrationshintergrund und politisches Wissen .....	78
Tabelle 43: Arbeitstabelle Migrationshintergrund und politisches Wissen .....	79

---

Tabelle 44: Arbeitstabelle Migrationshintergrund (Nein) und politisches Wissen .....	79
Tabelle 45: Arbeitstabelle Migrationshintergrund (Ja) und politisches Wissen.....	80
Tabelle 46: Unterschiedliche Bezeichnungen für Variablen der Regressionsanalyse.....	83
Tabelle 47: Bivariate lineare Regression mit Lebenszufriedenheit und Einkommen .....	86
Tabelle 48: Dummy-Kodierung für Familienstand .....	95
Tabelle 49: Bestimmungsfaktoren der Lebenszufriedenheit (Teil 1) .....	97
Tabelle 50: Bestimmungsfaktoren der Lebenszufriedenheit (Teil 2) .....	101
Tabelle 51: Bivariate logistische Regression mit Wahlbeteiligung und Alter .....	104
Tabelle 52: Bestimmungsfaktoren der Wahlbeteiligung .....	109
Tabelle 53: Mittelwerte in Zufallsstichproben (Stichprobengröße jeweils 1000 Personen) .....	115
Tabelle 54: Mittelwerte von Zufallsstichproben (Stichprobengröße jeweils 1000 Personen) ....	117
Tabelle 55: Vergleich zwischen Standardfehler und Standardabweichung .....	124
Tabelle 56: Mittelwerte von Zufallsstichproben.....	128
Tabelle 57: Erforderliche Stichprobengröße .....	140
Tabelle 58: Fehlerarten beim Hypothesentest .....	148
Tabelle 59: Lebenszufriedenheit von Frauen und Männern.....	155
Tabelle 60: Kritische Werte der t-Verteilung .....	157
Tabelle 61: Lebenszufriedenheit von West- und Ostdeutschen .....	159
Tabelle 62: Zufriedenheit mit der Demokratie.....	160
Tabelle 63: Beispieldaten für die Berechnung eines t-Tests bei abhängigen Stichproben.....	162

# 1 Einführung

Markus Tausendpfund

Vorschau



Dieses Kapitel macht Sie mit den Grundlagen der quantitativen Datenanalyse vertraut. Nach der Einordnung der Phase „Datenanalyse“ innerhalb des Forschungsprozesses werden die Begriffe „Grundgesamtheit“ und „Stichprobe“ erläutert. Bei einer empirischen Studie werden meist Aussagen über größere Gruppen angestrebt (z. B. die wahlberechtigte Bevölkerung in Deutschland). Allerdings liegen in den meisten Studien keine Informationen über alle Elemente dieser Gruppe vor, sondern nur über eine (zufällige) Auswahl dieser Gruppe. Die Gruppe, über die eine Aussage gemacht werden soll, wird als Grundgesamtheit oder Population bezeichnet. Die Gruppe, über die empirische Informationen vorliegen, wird als Stichprobe bezeichnet. Diese Begriffe werden knapp erläutert und es werden die Voraussetzungen skizziert, unter denen Befunde einer Stichprobe auf die zugehörige Grundgesamtheit übertragen werden können. Abschließend werden typische Klassifikationen von Variablen vorgestellt. Dabei liegt der Fokus auf dem Skalenniveau von Variablen, da das Skalenniveau eine wichtige Voraussetzung für die Anwendung bestimmter Analyseverfahren ist.

## 1.1 Einordnung im Forschungsprozess

Die quantitativen Analyseverfahren werden häufig mit dem quantitativen Forschungsprozess gleichgesetzt. Quantitativ arbeitende Sozial- und Bildungswissenschaftler\*innen nutzen statistische Analyseverfahren, um die theoretisch formulierten Hypothesen empirisch zu überprüfen. Sicherlich ist die Anwendung statistischer Analyseverfahren ein zentrales Merkmal des quantitativen Forschungsprozesses, aber die quantitative Datenanalyse sollte nicht isoliert betrachtet werden.

Vor der Datenanalyse bzw. Anwendung quantitativer Analyseverfahren müssen empirische Sozial- und Bildungsforscher\*innen wichtige vorgelagerte Entscheidungen treffen, die unmittelbare Auswirkungen auf die empirischen Befunde haben. Wie Abbildung 1 zeigt, stehen die Festlegung eines Forschungsthemas und die Entwicklung einer geeigneten Forschungsfrage am Beginn eines Forschungsprojekts. Auf dieser Grundlage werden die zentralen Konzepte identifiziert und theoretisch geklärt, ehe gehaltvolle Hypothesen formuliert und valide Operationalisierungen dieser Konzepte entwickelt werden. Diese Phasen in einem Forschungsprozess erfolgen in intensiver Auseinandersetzung mit dem existierenden Forschungsstand. Nur wer den Forschungsstand zu seinem\*ihrem Forschungsthema kennt, kann eine gehaltvolle Forschungsfrage entwickeln. Die Auseinandersetzung mit der Fachliteratur ist aber auch für die Konzeptspezifikation und die Entwicklung von Hypothesen erforderlich. Schließlich ist auch bei der „Übersetzung“ theoretischer Konzepte in empirische Indikatoren ein Überblick über existierende Operationalisierungen notwendig.

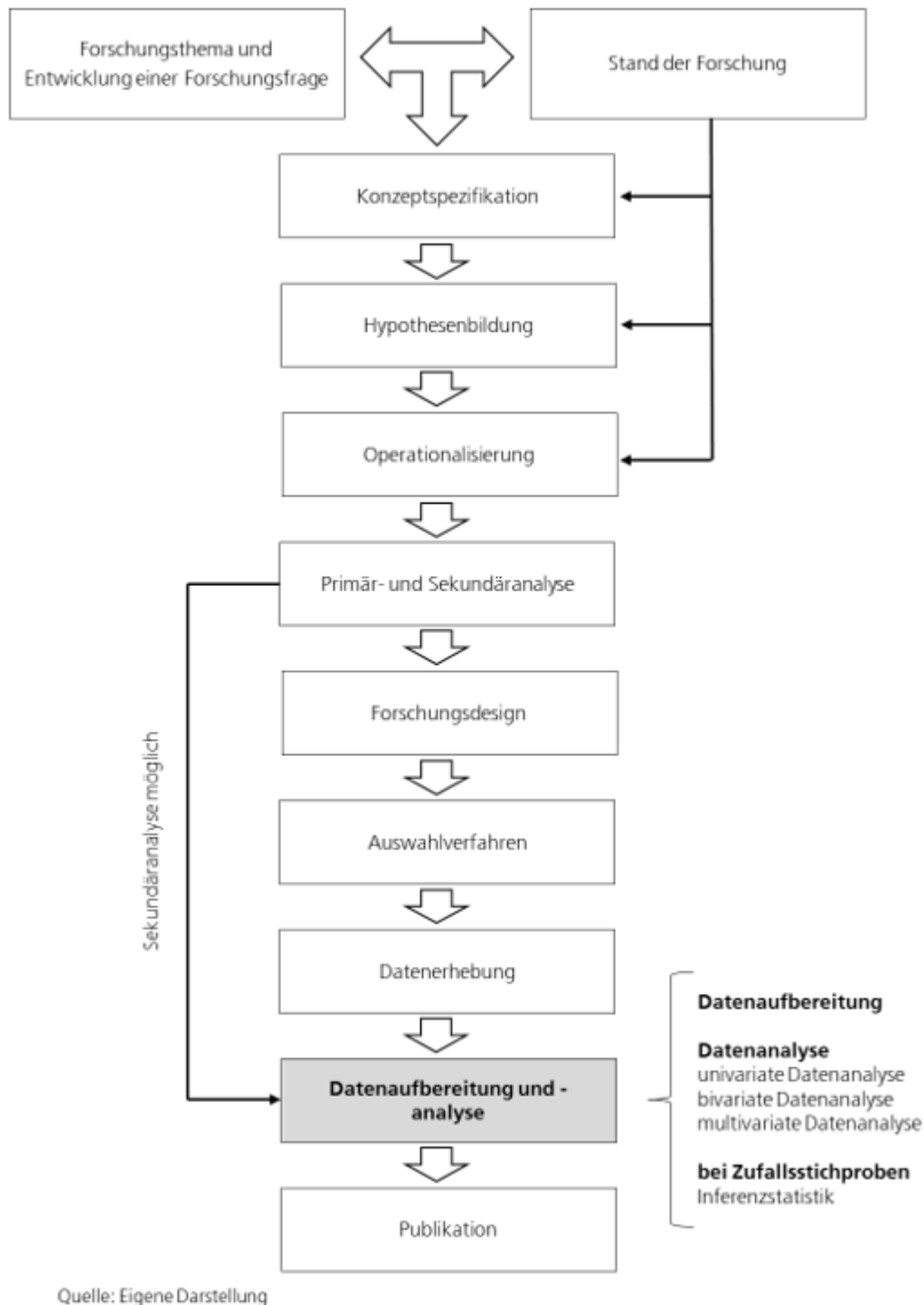


Abbildung 1: Idealtypischer Ablauf eines quantitativen Forschungsprojekts

Kein Analyseverfahren kann die intensive Auseinandersetzung mit dem existierenden Forschungsstand ersetzen. Ungeeignete Konzeptspezifikationen, schwammige Hypothesen oder auch ungültige Operationalisierungen führen zwangsläufig zu schlechten Daten und kein Analyseverfahren der Welt kann aus schlechten Daten valide empirische Befunde machen. Deshalb: Die Anwendung bzw. Durchführung quantitativer Analyseverfahren kann nur dann zu belastbaren empirischen Befunden führen, wenn die vorgelagerten Phasen erfolgreich bearbeitet wurden.

