

Isabella Hager

Statistik für die Sprachwissenschaft

Isabella Hager

Statistik für die Sprachwissenschaft

Teil 1: Grundbegriffe und wichtigste Testverfahren

1. Auflage

Isabella Hager

hat sich im Rahmen ihres Soziologiestudiums auf Statistik spezialisiert und ist heute Forschende, Beratende und Lehrende, unter anderem am Institut für Sprachwissenschaft der Universität Wien, der FH Campus Wien, bei prospect – Research & Solutions und beim Wiener Projektzentrum, Institut für Sozialforschung, Bildung, Integration und Kunst.

Mein Dank ergeht an

Eva Hellsberg, die die duftenden Korinthenkuchen aus dem Backrohr holt,
für ihr unermüdliches und unbarmherziges Wurmlesen

Maria Hellsberg, die mit ihrem analytischen, scharfen Blick alle Fehler findet

Elisabeth Brousek, die mich immer erinnert, was Wissenschaftlichkeit ausmacht

Philipp Weinfurter, für seine Kreativität und Geduld

Anselm Eder, der mir den Weg gewiesen hat

© 2021 Isabella Hager

1. Auflage

Isabella Hager

Umschlaggestaltung, Illustration: Philipp Weinfurter

Lektorat, Korrektorat: Eva Hellsberg, Maria Hellsberg, Elisabeth Brousek

Verlag: myMorawa von Dataform Media GmbH, Wien; www.mymorawa.com

ISBN: 978-3-99125-444-7 (Paperback)

ISBN: 978-3-99125-445-4 (Hardcover)

ISBN: 978-3-99125-471-3 (e-Book)



Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlages und der Autorin unzulässig. Dies gilt insbesondere für die elektronische oder sonstige Vervielfältigung, Übersetzung, Verbreitung und öffentliche Zugänglichmachung.

In diesem Buch wird die geschlechtsneutrale Sprache mittels * verwendet.

...dieses Buch...

...nimmt die Herausforderung an, die statistische Denkweise und die dazugehörigen Methoden verständlich zu erklären. Alles was Sie dabei tun müssen ist: interessiert sein. In diesem Buch werden Sie die wichtigsten Grundbegriffe und Verfahren der empirischen Sprachwissenschaft am Beispiel von echten Daten sowie Musterbeispielen kennen und anwenden lernen.

Ob in der Angewandten Linguistik, der Allgemeinen Linguistik, der quantitativen Linguistik oder der Korpuslinguistik, ob Experiment, Beobachtung oder Befragung – Statistik ist ein Werkzeug, das Ihnen dazu dient, Ihre Forschungsfragen zu beantworten. Das vorliegende Buch dient als Gebrauchsanweisung für die Konzeption und Auswertung Ihrer Forschungsprojekte.

Isabella Hager, Jänner 2021

Kennzeichnungen

Wichtiger Merksatz

Fragestellung

Beispiel

Aufgabe

Interpretation eines Ergebnisses

Technisch-methodische Bemerkung

Literatur: Ich empfehle die folgende Literatur zusätzlich zur Vertiefung.

Bortz, J. & Döring, N. (jeweils neueste Auflage): Statistik für Sozialwissenschaftler. Heidelberg: Springer-Verlag.	die „Bibel“ - das „Alte Testament“
Bortz, J. & Döring, N. (jeweils neueste Auflage): Forschungsmethoden und Evaluation. Heidelberg: Springer-Verlag.	die „Bibel“ - das „Neue Testament“
Field, A. (jeweils neueste Auflage): Discovering Statistics using IBM SPSS. London, California, New Dehli, Singapore. SAGE Publications Ltd. and sex and drugs and rock´n roll !!
Schobinski, P. (1996): Empirische Sprachwissenschaft. Westdeutscher Verlag GmbH, Opladen.	alt und bewährt
Meindl, C. (2011): Methodik für Linguisten. Narr Francke Attempo Verlag GmbH, Tübingen.	anspruchsvoll!
Hatzinger R., Nagel H. (2013): Statistik mit SPSS, Pearson-Studium.	gute SPSS-Anleitungen
Porst, R. (2011) Fragebogen: VS Verlag für Sozialwissenschaften, Springer Fachmedien Wiesbaden	
Bühner, M. (2004): Einführung in die Test- und Fragebogenkonstruktion. München: Pearson-Studium.	
Jede Art von Handbuch für SPSS ist fürs praktische Arbeiten geeignet!	

Inhalt

1	Fragestellung und Operationalisierung	9
1.1	Was ist eigentlich empirische Sozialforschung?	9
1.2	Das Formulieren einer Fragestellung	10
1.3	Konzeption und Operationalisierung	12
1.4	Hypothese und Hypothesenmodell	15

2	Grundgesamtheit und Stichprobe	16
2.1	Repräsentativität – oder: Der Zufall ist der „Master of the Universe“	18
2.2	Konfidenzintervall und Schwankungsbreiten	22
2.3	Wie groß soll/muss eine Stichprobe sein?	22
2.4	Stichprobenarten	24

3	Fragebogenkonstruktion	27
3.1	Gütekriterien von Messinstrumenten	27
3.2	Sozialwissenschaftliche Befragungen	28
3.3	Der Aufbau des Fragebogens	29
3.4	Die Elf Gebote zur Fragebogenerstellung	30

4	Datenniveaus – kategorisieren und messen	31
4.1	Nominalskala (kategorial)	32
4.2	Ordinalskala oder Rangskala (kategorial)	32
4.3	Intervallskala (metrisch)	33
4.4	Rationalskala/Verhältnisskala/Absolutskala (metrisch)	33
4.5	Das Stufenmodell der Datenniveaus	34
4.6	Kriterien einer Antwortskala	35
4.7	Arten von Antwortskalen	36
4.8	Zur Verwendung einer „neutralen“ Mittelkategorie	36
4.9	Textantworten	37
4.10	Messungen in der Korpuslinguistik	38
4.11	Die Erhebung: Vom Messinstrument zur Datenmatrix	39

5	Beschreibung von Häufigkeiten	41
5.1	Häufigkeitstabelle	41
5.2	Mehrfachantworttabelle	43
5.3	Berechnen eines Zähl-Index aus einer Mehrfachantwort	44
5.4	Kreuztabelle	45

6	Statistische Kennwerte	48
6.1	Berechnung der statistischen Kennwerte	49
6.2	Anwendung und Interpretation statistischer Kennwerte	54
6.2.1	Kennwerte bei nominalem Messniveau	54
6.2.2	Kennwerte bei ordinalem Messniveau	55
6.2.3	Kennzahlen bei metrischem Messniveau	56
6.2.4	Feinberechnung von Perzentilen	58
6.2.5	Kategorisierung bei metrischem Messniveau	59
6.2.6	Vergleich von Verteilungen bei metrischem Messniveau	61
6.2.7	Boxplot – grafische Darstellung von Verteilungen	63
6.3	Verteilungsformen	65

7	Die Normalverteilung	67
7.1	Die empirische Normalverteilung	67
7.2	Wann ist eine empirische Variable normalverteilt?	69
7.3	Die theoretische Normalverteilung	71
7.4	Standardnormen normalverteilter Testwerte	72
7.5	Berechnung von Konfidenzintervallen	74

8	Prüfmaß und Wahrscheinlichkeit – die Logik von Signifikanztests	77
8.1	Die Vorgangsweise bei einem Signifikanztest	79
8.2	Die Durchführung eines Signifikanztests am Beispiel Chi-Quadrat-Test	80
8.2.1	Übersicht zum Chi-Quadrat-Test – oder: „Baba“ vertuscht sich.....	85
8.2.2	Der Log-Likelihood-Test in der Korpuslinguistik – oder: „Friede der Sterne“	87
8.3	Hypothesen testen – eine Übersicht über die Signifikanztests	89
8.4	Anwendungsregeln – Übersicht	90
8.5	Entscheidungsbaum – Check-List	91

9	Parametrische Testverfahren	92
9.1	t-Test für unabhängige Stichproben: Testen eines Mittelwertunterschiedes zwischen zwei Gruppen	92
9.1.1	Überprüfen der Normalverteilung mit dem Kolmogorov-Smirnov-Test	94
9.2	Einfaktorielle Varianzanalyse: Testen eines Mittelwertunterschiedes zwischen mehreren Gruppen	96
9.3	Mehrfaktorielle Varianzanalyse: Testen eines Mittelwertunterschiedes anhand mehrerer Einflussfaktoren	100
9.4	t-Test für abhängige Messungen: Testen eines Mittelwertunterschiedes zwischen zwei Messzeitpunkten	102
9.5	Einfaktorielle Varianzanalyse mit Messwiederholung: Testen eines Mittelwertunterschiedes zwischen mehreren Messzeitpunkten	107
9.6	Mehrfaktorielle Varianzanalyse bei Messwiederholung: Testen einer Mittelwertveränderung anhand mehrerer Einflussfaktoren	111

10	Parameterfreie Testverfahren	115
10.1	Mann-Whitney-U-Test: Testen eines Verteilungsunterschiedes zwischen zwei Gruppen	115
10.2	Kruskal-Wallis-Test: Testen eines Verteilungsunterschiedes zwischen mehreren Gruppen	118
10.3	Wilcoxon-Test: Testen einer Veränderung zwischen zwei Messzeitpunkten	121
10.3.1	Exkurs: Die Signifikanz in Abhängigkeit von der Fallzahl	124
10.4	Friedman-Test: Testen einer Veränderung zwischen mehreren Messzeitpunkten	125

11	Korrelation – Stärke eines Zusammenhanges von zwei Merkmalen	128
11.1	Stärke des Zusammenhanges bei Kreuztabellen (=Assoziationsmaße)	129
11.2	Pearson-Korrelation	131
11.3	Reliabilitätsanalyse: Berechnen eines Summen-Index mithilfe der Korrelation	139
11.4	Spearman-Rang-Korrelation	144

12	Zufall oder Nicht-Zufall? – Das ist hier die Frage!	
	Überlegungen zum Zufall und zur Signifikanz	146
12.1	„The Art of Fehler“ oder: Des widerspenstigen Zufalls Zähmung	149
12.2	Stichprobengröße und Signifikanz	150
12.3	Effektstärke	152
12.4	Teststärke und Poweranalyse	155
12.5	Kleiner Ausblick auf multivariate Verfahren	157

13	Stichwortverzeichnis	158
14	Literatur	159

1 Fragestellung und Operationalisierung

Zu Beginn müssen wir einige Grundbegriffe klären, um uns das „quantitative Denken“ anzueignen. Um eine quantitative Studie erfolgreich durchzuführen, sind bei der Vorbereitung der Durchführung folgende Arbeitsschritte notwendig:

- Formulieren der **Fragestellung**
- Definition und **Operationalisierung** der darin enthaltenen Begriffe und Messdimensionen
- Formulieren von **Hypothesen** und Erstellen eines Hypothesenmodells
- Definition der **Grundgesamtheit** und Auswahl der **Stichprobe**

1.1 Was ist eigentlich empirische Sozialforschung?

Wenn wir empirische Sozialforschung in einem **Kontinuum von Strukturiertheit und Vorwissen** sehen, dann steht auf einer Seite der qualitative und auf der anderen Seite der quantitative Forschungsansatz. Je nach Fragestellung wird die dafür geeignete Methode ausgewählt, oder es erfolgt eine Kombination beider Methoden, wie bei der triangulativen Forschung.

Beim qualitativen Forschen geht es darum, die Logik eines Einzelfalls zu verstehen und daraus allgemeine Schlüsse zu ziehen. Das quantitative Paradigma benötigt eine ausreichende Menge an untersuchten Objekten, um daraus Gesetzmäßigkeiten abzuleiten.

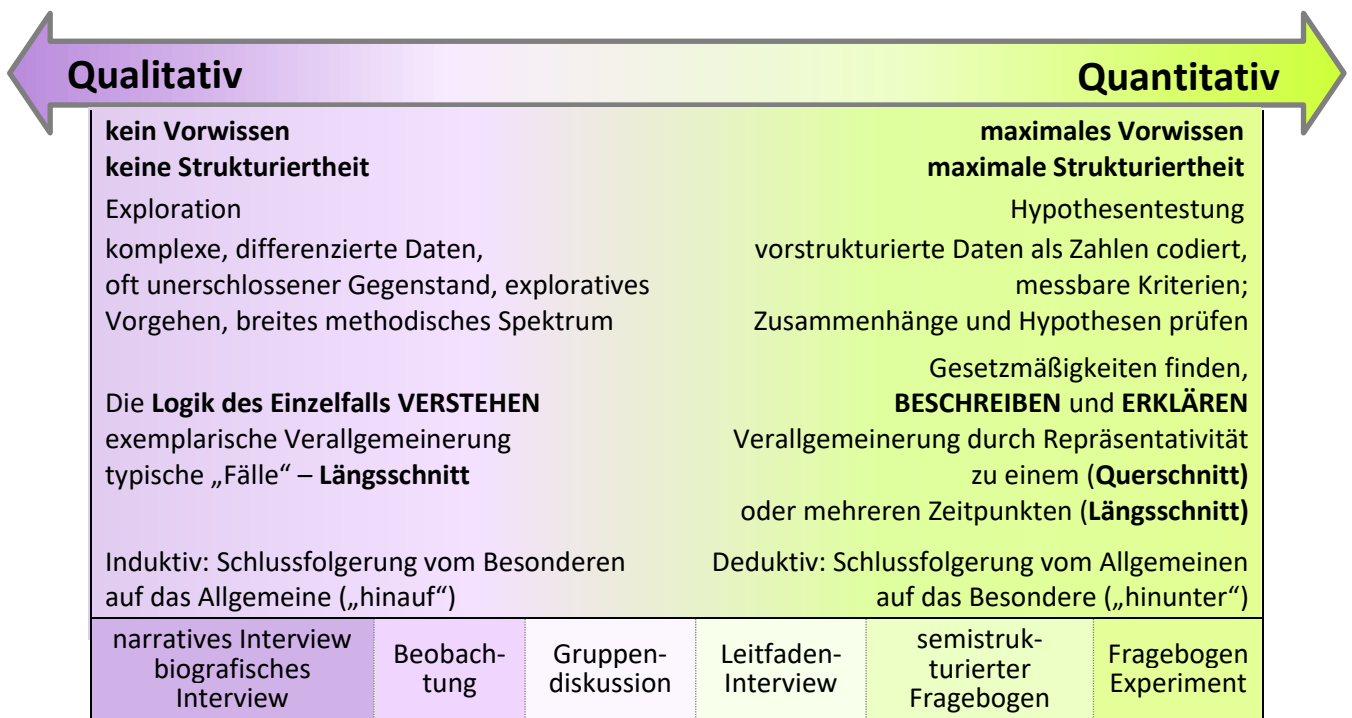


Abbildung 1: Überblick qualitative und quantitative Sozialforschung – Das Kontinuum der Datenerhebung

Die quantitative Forschung zeichnet sich dadurch aus, dass maximales Vorwissen bei maximaler Strukturiertheit benötigt wird. Antworten oder Eigenschaften von Menschen werden auf Messwerte oder Kategorien („Schubladen“) reduziert, die nur dann Sinn ergeben, wenn wir wissen, was wir da eigentlich messen. Das bedeutet: Beim quantitativen Forschen findet die meiste und wichtigste Arbeit **vor** der Erhebung statt: nämlich die Operationalisierung der Kriterien, die gemessen werden sollen und die Konstruktion des Messinstruments. Beim qualitativen Forschen kommt die meiste und wichtigste Arbeit erst **nach** der Erhebung: das qualitative Auswerten!

Der Durchführung einer quantitativen Studie muss daher eine umfangreiche Planung vorausgehen. Der erste und wichtigste Arbeitsschritt ist das Formulieren einer Fragestellung.

1.2 Das Formulieren einer Fragestellung

Am Anfang einer empirischen Forschungsarbeit steht die Entwicklung einer **Fragestellung**.

Eine wissenschaftliche Fragestellung muss systematisch formuliert und empirisch überprüfbar sein.

- „**Empirisch**“ heißt: „*erfahrungsgemäß*“: beobachtbar bzw. messbar anhand theoretisch formulierter Annahmen, definiert in Begriffen und Messdimensionen. Diese Annahmen (Hypothesen) werden anhand spezifischer sozialer Wirklichkeiten¹ überprüft.
- „**Systematisch**“ heißt: die Überprüfung erfolgt nach festgelegten Regeln (nach einer „Methode“) und ist daher intersubjektiv **nachvollziehbar**, im Unterschied zur Alltagsbeobachtung, die immer subjektiv bleibt. Die Systematik bezieht sich auf drei Grundkriterien der Wissenschaftlichkeit: Sie ist **strukturiert** nach einem geplanten Ablauf und einem festgelegten Design, sie ist **allgemein**, da sie auf einer Stichprobenauswahl beruht, die bestimmte Kriterien erfüllt (z.B. Repräsentativität) und sie ist **kontrolliert selektiv**, das heißt, es gibt festgelegte Kriterien darüber, was beobachtet bzw. gemessen wird.

Eine quantitative Fragestellung muss so formuliert sein, dass sie mit quantitativen Methoden beantwortbar ist. Das klingt trivial, ist aber oft ein schwieriges Unterfangen. Dabei müssen wir uns folgende Fragen stellen:

1.) WAS?	Was will ich eigentlich wissen?	→ Gefragt ist: O-RIGINALITÄT
2.) WARUM?	Wozu will ich das wissen? Wem nützen die Ergebnisse?	→ Gefragt ist: RE-LEVANZ
3.) WIE?	Was muss ich tun, um das herauszufinden? (Zugang, Erhebung, Design...)	→ Gefragt ist: SO-RGFALT

Abbildung 2: Die drei Fragen zur Entwicklung einer Fragestellung nach dem „ORESOSchema“²

Die erste Frage ist die schwerste und erfordert oftmals viel Geduld, bis sie beantwortet werden kann. Bei der zweiten Frage stellt sich heraus, ob die Studie realisiert wird – finden sich Geldgeber*innen, die für die Ergebnisse zahlen? – wenn ja, dann wird sie – sobald wir bei der dritten Frage angelangt sind – immer konkreter und auf bestimmte Gruppen, einen bestimmten Zeitraum, eine bestimmte Region eingegrenzt, um beantwortbar zu werden. Meist muss die erste Frage erneut angepasst werden, um die Fragestellung beantworten zu können. Dieser Kreislauf setzt sich so lange fort, bis die **Fragestellung systematisch formuliert und empirisch überprüfbar** ist.

Im **Entdeckungszusammenhang** wird das „Problem“ benannt, die Fragestellung formuliert und zeitlich, sachlich und räumlich eingegrenzt. Diese Eingrenzung ist essenziell, da immer nur Ausschnitte der sozialen Wirklichkeit empirisch erfasst werden können. Im Rahmen des **Begründungszusammenhangs** wird das Design geplant, das Messinstrument konstruiert und die Erhebung mit anschließender Analyse der gewonnenen Daten durchgeführt. Schließlich werden im **Verwertungszusammenhang** die Ergebnisse anwendungsorientiert dargestellt und interpretiert.

Gute Forschungsarbeit zeichnet sich dadurch aus, dass neue Fragen entstehen, an denen andere Forscher*innen weiter ansetzen können und eine ständige, reflexive Weiterentwicklung der Erkenntnisgewinnung stattfindet!

¹ Was ist **soziale Wirklichkeit**? Darunter verstehen wir Institutionen (z.B. Bildungseinrichtungen), Kommunikation, Interaktion, sozio-ökonomische Merkmale (z.B. Geschlecht, Alter, Migrationshintergrund, Erstsprache, Zweitsprache, Wortschatz, Sprachkenntnisse, ...), mitgeteilte Einstellungen, Empfindungen, Erfahrungen, Emotionen, (z.B. Wichtigkeit von bestimmten Sachverhalten, Fehlertoleranz...), beobachtetes oder mitgeteiltes Verhalten (z.B. Gewaltbereitschaft, Umgangsformen, Kontaktverhalten, Lernverhalten...) und vieles mehr. Soziale Wirklichkeit ist also (wie der Wiener Kabarettist Gunkl es ausdrücken würde: "eh alles"), was wir Menschen geschaffen haben, miteinander teilen und das von unserem Handeln abhängig ist.

² Das „ORESOSchema“ wurde nach Anregung von Anselm Eder, meinem Statistikprofessor, entwickelt.

Daraus folgt eine **Anleitung zum Formulieren einer quantitativen Fragestellung**:

- 1.) Erfasse dein Erkenntnisinteresse in einem Satz! Was ist das genaue Problem, das du bearbeiten willst?
- 2.) Formuliere die Problemstellung – dabei wird dir die Vielschichtigkeit des Themas klar...
- 3.) Aus der/den Problemstellungen heraus kannst du deine konkrete Forschungsfrage entwickeln.
 - Stelle konkrete Fragen zum Problem! Teile dein Problem in einzelne Teilaspekte auf!
 - Mehrere einzelne Fragen sind zulässig! Aber: Finde dazu eine übergeordnete Fragestellung!
 - Die Fragestellung bestimmt den weiteren Weg der Forschung: den Ansatz, das Design, die Methoden, den Datenerhebungsmodus und die Vorgangsweise bei der Auswertung.

Woran du erkennst, dass deine Fragestellung wissenschaftlich ist – im Unterschied zur Alltagsfrage:

- Sie ist theoretisch und ihre Beantwortung leistet einen Beitrag zur Entwicklung des Themas
- Sie ist relevant, d.h. sie besitzt praktischen bzw. wissenschaftlichen Nutzen
...ist sie es nicht? Frag dich, warum!

Stolpersteine bei der Formulierung einer Fragestellung: Eine quantitative Fragestellung...

- ...muss einen wissenschaftlichen **Erkenntniswert** haben.
Wissenschaftliche Forschungsfragen sind nicht zu einfach, sondern haben eine gewisse Komplexität und müssen an einen Erkenntnisaspekt anschließen!

⊗ Was ist der häufigste Grund, warum Menschen Sprachwissenschaft studieren?	Unterscheiden sich die Motivationen Sprachwissenschaft zu studieren zwischen älteren und jüngeren Menschen? ✓
---	---

- ...muss im Rahmen einer Studie **realistisch beantwortbar** sein.

⊗ Stirbt das österreichische Deutsch aus?	Inwiefern unterscheidet sich der Sprachgebrauch von österreichischem Deutsch zwischen Lehrer*innen und Schüler*innen an österreichischen Schulen? ✓
---	---

- ...muss mit **quantitativen Methoden beantwortbar** sein (messen und zählen!)

⊗ Sind Mundartgedichte beliebt?	Unterscheidet sich das Ausmaß der Beliebtheit von Mundartgedichten zwischen Menschen aus ländlichen und städtischen Regionen? ✓
---------------------------------	---

- ...muss als Ziel die **Beschreibung** eines Phänomens haben – und nicht das Verstehen!
Quantitative Forschungsfragen zielen auf das Auffinden von Gesetzmäßigkeiten ab, wodurch die beobachteten Umstände erklärt werden können.

⊗ Warum sprechen manche Menschen Dialekt und andere nicht?	Inwiefern hängt die Verwendung von Dialekt versus Standardsprache mit sozioökonomischen Merkmalen zusammen? ✓
--	---

- ...muss **begrenzt und präzise** sein! Was hilft? Eingrenzen!

⊗ Sind Medikamentenbeipacktexte verständlicher, wenn sie einfacher formuliert sind?	Wie wirkt sich die Formulierung der Texte auf die Verständlichkeit von Medikamentenbeipacktesten aus? ✓
---	---

- ...darf **nur eine (Haupt-)Frage** beinhalten (Eingrenzen und notfalls die Fragestellung aufsplitten)!

⊗ In welchen Bereichen möchten Studierende der Sprachwissenschaft nach Ende des Studiums arbeiten, und in welchem Zusammenhang steht dies mit den gesetzten Schwerpunkten, dem Studienfortschritt und der durchschnittlichen Leistungsbeurteilung?	In welchem Zusammenhang stehen die gesetzten Schwerpunkte während des Studiums mit den subjektiv eingeschätzten Berufsaussichten? ✓
--	---

Aufgabe: Formulieren Sie eine **quantitative sprachwissenschaftliche Fragestellung!**

Beschreiben Sie kurz den Bereich, in der/in dem die Forschungsarbeit zur Beantwortung der Fragestellung durchgeführt werden soll. Warum ist Ihre Fragestellung relevant?

Welche*r Geldgeber*in würde für die Ergebnisse bezahlen? Wem nützen die Ergebnisse?

1.3 Konzeption und Operationalisierung

Der nächste Arbeitsschritt ist die Operationalisierung. Wir erinnern uns: Beim quantitativen Forschen müssen wir vor der Erhebung bereits ganz genau wissen, was wir messen wollen:

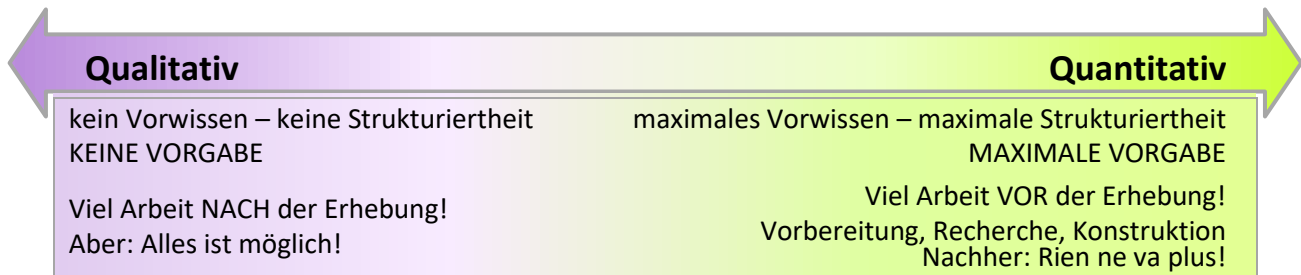


Abbildung 3: Arbeitsschwerpunkte bei qualitativer versus quantitativer Forschung

Diese Vorbereitung, das „Messbar-Machen“, sind konzeptionelle Definition und Operationalisierung der in der Fragestellung enthaltenen **Begriffe** und **Messdimensionen**. Dabei muss präzise definiert werden, was genau auf welche Weise gemessen werden soll.

Alle Begriffe/Messdimensionen in der Fragestellung müssen erst definiert und dann operationalisiert werden. Die Definition und die Operationalisierung der Inhalte einer Fragestellung bestimmen das Forschungsdesign und die Ergebnisse!

- **Konzeptionelle Definition:** Was bedeutet der Begriff / die Messdimension?
Welche theoretischen Aspekte / Konstrukte umfassen den Begriff / die Messdimension?
- **Operationalisierung:** Wie kann der Begriff / die Messdimension messbar gemacht werden?

Die Operationalisierung ist das UM-UND-AUF einer quantitativen Studie – sie bestimmt die Ergebnisse! Alles, was unpassend operationalisiert oder übersehen wurde, geht verloren! Um die relevanten Inhalte messbar zu machen, sind eine **theoretische Basis**, umfangreiche **Literaturrecherche** (Hat schon jemand an diesem Thema geforscht?), **explorative Interviews** mit Zielgruppenpersonen bzw. Expert*innen notwendig. Die Operationalisierung kann nur gelingen, wenn die Fragestellung und das Ziel präzise formuliert und die verwendeten Konzepte exakt definiert sind. Nur auf diese Weise kann sichergestellt werden, dass relevante Inhalte sinnvoll operationalisiert werden und nichts Wichtiges übersehen wird.

Wichtigste Kriterien für die Operationalisierung sind die Zweckmäßigkeit zur Beantwortung der Fragestellung sowie die Sicherstellung der Vergleichbarkeit mit anderen Studienergebnissen.

Beispiel Sprachentwicklung: Konzeptionelle Definition: „Der Entwicklungsprozess des Erwerbs einer Sprache in natürlicher Umgebung durch alltägliche soziale Kontakte.“ Operationalisiert wird die Sprachentwicklung anhand der Testergebnisse verschiedener Tests (Wortschatztest, Pluraltest, etc.) zu mehreren Zeitpunkten.

Beispiel Anregungsqualität: Konzeptionelle Definition: „Institutionelle oder familiäre Förderung kindlicher Lernprozesse hinsichtlich der Sprachentwicklung.“ Operationalisiert wird die Sprachentwicklung mittels...

- **Orientierungsaspekten:** Erziehungsstile, Erziehungsziele, Wertvorstellungen der der Betreuungs-/Bezugspersonen (z.B. elterliche Selbstwirksamkeitserwartung), bildungsbezogene Einstellungen...
- **Strukturasperkten:** Räumlichkeiten, Vorhandensein von Materialien, Ausbildungsgrad der Betreuungspersonen, Anzahl der Betreuungs- und Bezugspersonen, Anzahl der Klassenkamerad*innen/Geschwister, finanzielle Ressourcen/Fördergelder...
- **Prozessasperkten:** Anzahl und Ablauf von Interaktionen und Aktivitäten, Verhalten der Betreuungs-/Bezugspersonen, Konzepte der Betreuung/der Angebote...

Beispiel Armut: Um im Rahmen einer länderübergreifenden Umfrage die latente Dimension „**Armut**“ messbar zu machen, so benötigen wir zuerst eine **konzeptionelle Definition von Armut:** „*Armut ist die mangelnde Befriedigung der Grundbedürfnisse nach Kleidung, Nahrung, Wohnung und Erhaltung des Lebens.*“

Hierzu können mehrere **Indikatoren** herangezogen werden:

1. Einkommen/Haushaltseinkommen/Geld, das monatlich zur Verfügung steht
2. Auskommen mit dem Einkommen/das Vorhandensein von Schulden
3. Auskommen müssen ohne ausreichend Nahrung
4. Auskommen müssen ohne entsprechende Kleidung
5. Wohnsituation, Wohnungsmängel
6. Betroffenheit von Arbeitslosigkeit
7. Betroffenheit von Analphabetismus
8. Soziale Absicherung im Falle von Arbeitslosigkeit, Krankheit und Alter
9. usw. usw. usw.

„**Armut**“ kann noch viel mehr sein! Für das, wie ich „**Armut**“ messe (= operationalisiere), gibt es keine „richtigen“ oder „falschen“ Kriterien, vielmehr werden inhaltliche Entscheidungen getroffen, was unter „**Armut**“ verstanden werden **soll**. Diese Entscheidung bestimmt alle späteren Behauptungen und Aussagen über die von Armut betroffenen oder nicht betroffenen Personen. Das heißt auch: Keine Feststellung, die sich auf konstruierte Wirklichkeit bezieht, ist „objektiv“, sondern einfach nur „definiert“.

Worauf zielt die Fragestellung ab? Liegt der Fokus auf.....

- ...der **Kaufkraft**, benötige ich das Einkommen und das Auskommen mit dem Einkommen (ad 1. und 2.).
- ...**Lebensstandard** oder **Lebensqualität**, fokussiere ich das Auskommen-Müssen ohne ausreichend Nahrung bzw. entsprechende Kleidung, Wohnsituation (ad 3., 4. und 5.).
- ...**Exklusion aus sozialen Systemen versus Inklusion in soziale Systeme**, brauche ich die Betroffenheit von Arbeitslosigkeit bzw. Analphabetismus und die soziale Sicherheit (ad 6., 7. und 8.).
- ...**Gesundheit**, verwende ich das Auskommen müssen ohne ausreichend Nahrung oder entsprechender Kleidung, die Wohnsituation und Wohnungsmängel und soziale Absicherung (ad 3., 4., 5., 8.).
- ...**xxx** brauche ich..... (ad 9. und so weiter).

Große Umfrageinstrumente, die auf Ländervergleiche abzielen, beinhalten etliche vielschichtige Fragestellungen, für die meist standardisierte Indikatoren herangezogen werden, wie beispielsweise Systeme sozialer Indikatoren.

In der Forschungspraxis werden Begriffe bzw. latente Dimensionen wie „**Sprachentwicklung**“, „**Anregungsqualität**“, „**Armut**“, „**Lebenszufriedenheit**“ etc. als Messdimension verstanden, die exakt definiert und mithilfe mehrerer Indikatoren erhoben wird. Was ist eine Messdimension?

Eine **Messdimension** besteht aus mehreren Einzelaspekten (Items), welche alle „das Gleiche“ messen.

Wichtig ist die inhaltliche und empirische **Eindimensionalität** einer Messdimension.

Inhaltlich: Die erhobenen Aspekte passen alle unter **eine** Überschrift (z.B. „**Selbstwirksamkeitserwartung**“).

Empirisch: Die Aspekte hängen wechselseitig miteinander zusammen, was empirisch ermittelt wird.

Beispiel für eine Messdimension im Rahmen einer Untersuchung über die Gesundheit von Langzeitarbeitslosen: Die aus der Psychologie stammende Messdimension „**Selbstwirksamkeitserwartung**“ ergibt sich aus dem Grad der Zustimmung zu folgenden Aussagen:

		stimmt völlig					stimmt überhaupt nicht	
a	Für jedes Problem habe ich eine Lösung.	①	②	③	④	⑤	⑥	
b	Die meisten Probleme kann ich aus eigener Kraft gut meistern.	①	②	③	④	⑤	⑥	
c	Auch komplizierte Aufgaben kann ich in der Regel gut lösen.	①	②	③	④	⑤	⑥	
d	Misserfolg spornt mich zu noch größeren Anstrengungen an.	①	②	③	④	⑤	⑥	
e	Was auch immer passiert, ich werde schon klarkommen.	①	②	③	④	⑤	⑥	
f	①	②	③	④	⑤	⑥	

Quelle: Operationalisierung der Messdimension „Generalisierte Selbstwirksamkeitserwartung“ - eine Auswahl von 41 Items (Collani & Schyns (2002)).

Bei komplexeren Fragestellungen werden die in der Fragestellung verwendeten Konzepte in mehrere (latente) Messdimensionen aufgesplittet, die wiederum in Subdimensionen zerlegt werden und anhand verschiedener (manifeste) Indikatoren gemessen werden können. Anhand des Beispiels hinsichtlich der „Institutionellen Anregungsqualität zur Förderung der Sprachentwicklung“ könnte das wie folgt aussehen:

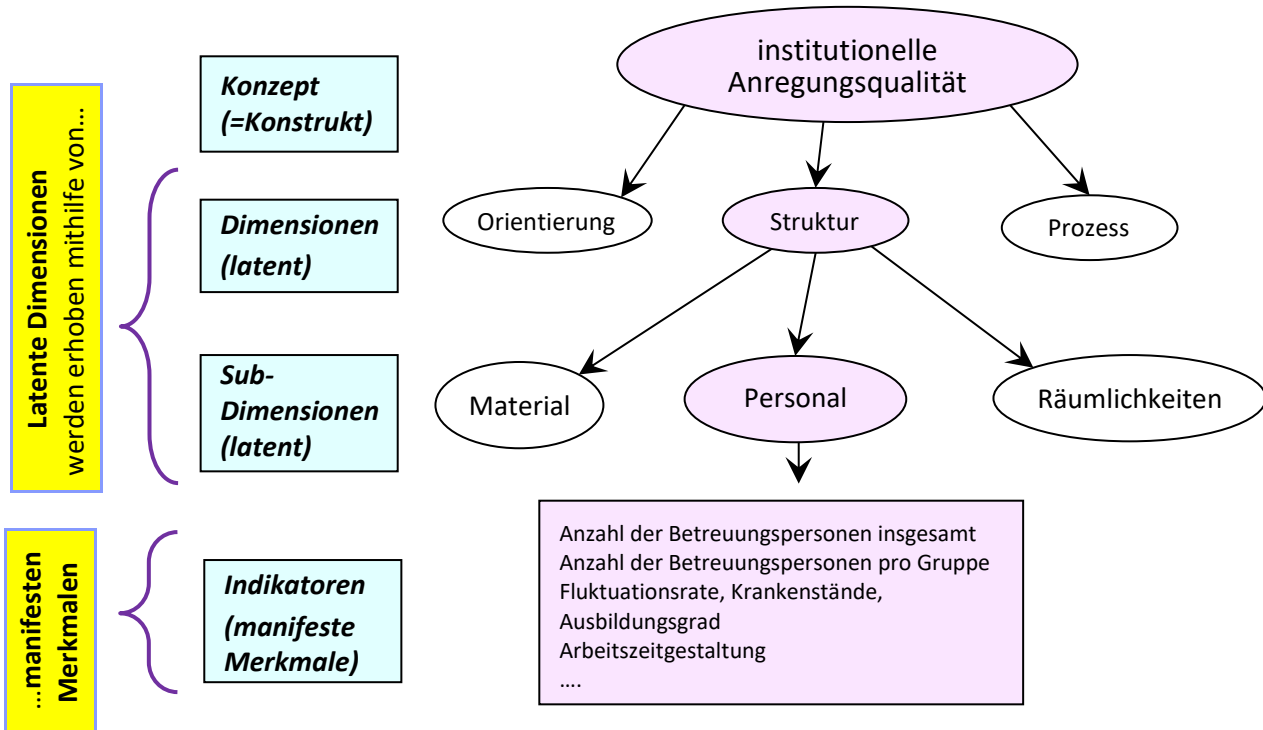


Abbildung 4: Beispiel einer Dimensionierung des Konzepts „Anregungsqualität“

Auch hier gilt: Es gibt keine „richtige“ oder „falsche“ Zusammenstellung der Items, vielmehr ist es eine auf inhaltlichen und methodischen Aspekten basierende Entscheidung, welche Items verwendet werden sollen.

Wir fragen also **nicht direkt** nach der latenten Dimension (etwa so: „Bitte schätzen Sie Ihre Selbstsicherheit auf der folgenden Skala ein.“), sondern: Als **Indikatoren** für die von uns gewählten Messdimensionen **bestimmen wir Teilaspekte in Form von formulierten Aussagen/Fragen**. Zu diesen Aussagen/Fragen erheben wir den Grad der Zustimmung bzw. die jeweilige Antwortmöglichkeit auf die Frage.

In diesem Zusammenhang wird auch von **Validität** (Gültigkeit) und **Reliabilität** (Zuverlässigkeit) von Messungen oder Messinstrumenten gesprochen. Die „**Konstruktvalidität**“ sagt aus, ob die ausgewählten Indikatoren das latente/theoretische Konstrukt (z.B. „Selbstwirksamkeit“) vollständig und passend messen (siehe auch Kapitel 3.1. auf Seite 27).

Um aus einer noch nicht getesteten „Fragebatterie“ (= Liste aus Items) die „besten“ (passendsten) auszuwählen, dienen Skalierungsverfahren:

- **Likert-Skala:** Sicherstellung möglichst großer Bandbreite der Antworten über die gesamte Antwortskala
- **Polaritätsprofil:** Gegensatzpaare von Begrifflichkeiten zur Messung von semantischen Bedeutungen, das sogenannte „semantisches Differenzial“
- **Guttman-Skala:** Konstruktion einer Skala mit steigender Zustimmung in einer theoretisch festgelegten und begründbaren Reihenfolge: Lässt sich die theoretische Abstufung auch empirisch beobachten?
- **Faktoranalyse:** Auffinden von latenten Subdimensionen innerhalb einer Batterie von Einzelaspekten
- **Reliabilitätsanalyse:** Testung der Güte der Eindimensionalität, d.h. alle Aspekte messen dasselbe.

Aufgabe: Definieren und operationalisieren Sie alle Begriffe und Messdimensionen in Ihrer sprachwissenschaftlichen Fragestellung!

1.4 Hypothese und Hypothesenmodell

Im besten Fall können die in der Fragestellung operationalisierten Begriffe und Messdimensionen in ein **Hypothesenmodell** integriert werden. Aber was ist eine Hypothese?

Eine **Hypothese** ist eine Vermutung / Behauptung über einen Sachverhalt, die überprüfbar ist.

Die Überprüfbarkeit ist der Unterschied zu Alltagshypothesen. Deshalb wird eine quantitative Hypothese niemals mit „**Wie ist...**“ beginnen, sondern ist in Form einer „**Je-Desto**“-Aussage formuliert.

Alltagshypothese	Wissenschaftliche Hypothese
„Sind Beipacktexte für Medikamente verständlich?“	„Die Verständlichkeit von Beipacktexten ist abhängig von der Gestaltung und Formulierung der Texte.“
„Die meisten Mitarbeiter*innen sind nicht an Weiterbildungen interessiert.“	„Je höher die Qualifikation der Mitarbeiter*innen, desto häufiger besuchen diese Weiterbildungsveranstaltungen.“
„Leuten, die lange arbeitslos sind, geht es schlecht.“	„Je länger die Dauer der Arbeitslosigkeit, desto negativer die Erfolgserwartung, einen Arbeitsplatz zu bekommen.“

Abbildung 5: Unterschied zwischen Alltagshypothesen und wissenschaftlichen Hypothesen

Hypothesen können verschiedene Sachverhalte betreffen, hier ein paar Beispiele: Die Fragestellung lautet:

Inwiefern ist der Spracherwerb bei Kindergartenkindern von der familialen Anregungsqualität abhängig?

➤ **Unterschied:**

Kinder aus Familien mit hoher Anregungsqualität erwerben die deutsche Sprache schneller als Kinder aus Familien mit niedriger Anregungsqualität.

➤ **Veränderung:**

Nach der Teilnahme an einer Sprachförderungsgruppe erfolgt der deutsche Spracherwerb schneller.

➤ **Zusammenhang:**

Je öfter einem Kind aus Büchern vorgelesen wird, desto schneller erfolgt der deutsche Spracherwerb.

Eine Fragestellung besteht zumeist aus mehreren Einzelhypothesen, die im Optimalfall in einem **Hypothesenmodell** dargestellt werden können. Das Modell postuliert den theoretischen Hintergrund mit den wichtigsten Hypothesen der Fragestellung.

Wie könnte so ein Hypothesenmodell aussehen? Zum Beispiel so.....

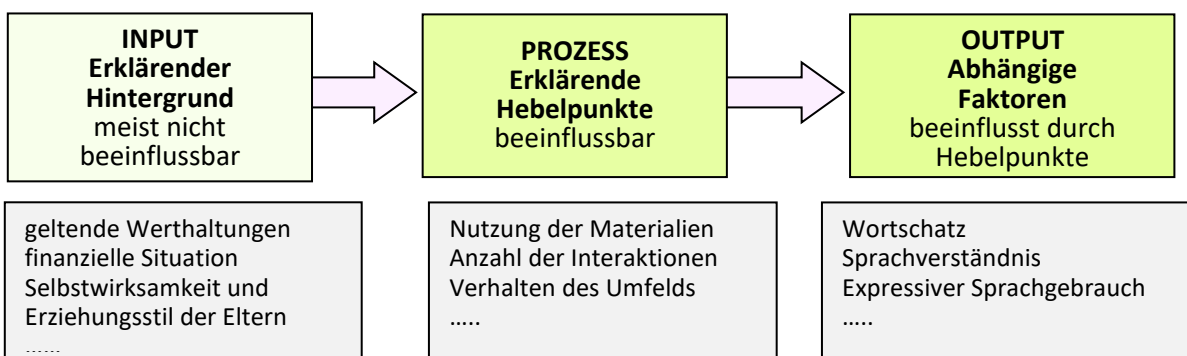


Abbildung 6: Schema eines möglichen Hypothesenmodells

Aus einem Hypothesenmodell ergeben sich zahlreiche Hypothesen. Die **Fragestellung** besteht somit aus einer **Zusammenstellung von inhaltlich aufeinander abgestimmten Einzelhypothesen**. Die Ergebnisse sind von den Indikatoren und Kategorien, die wir festlegen, abhängig.

Aufgabe: Formulieren Sie drei zu Ihrer Fragestellung passende Hypothesen.

Welches Ergebnis vermuten Sie? Begründen Sie Ihre Vermutung und beschreiben Sie das fiktive Ergebnis.

Erstellen Sie dazu ein (kleines und einfaches) Hypothesenmodell zu Ihrer Fragestellung. Probieren Sie es!

2 Grundgesamtheit und Stichprobe

Nachdem eine Fragestellung formuliert und die darin enthaltenen Begriffe und Messdimensionen definiert und operationalisiert sind, wird im nächsten Schritt die **Grundgesamtheit** (GG) bestimmt. Welche Personen (oder Elemente) betrifft meine Fragestellung?

Die **Grundgesamtheit** stellt die Menge aller Elemente (Personen, Texte, Interaktionen, Gerichtsakten, Regionen u.a.) dar, für welche die Fragestellung beantwortet werden bzw. für die die Ergebnisse gelten sollen. Die Grundgesamtheit muss **explizit definiert** und **eingegrenzt** werden.

Warum ist auch hier wieder eine genaue Definition erforderlich? Weil wir ganz genau wissen müssen, auf wen unsere Ergebnisse anwendbar sind und auf wen nicht.

Beispiel: Wenn Kinder eines Kindergartens hinsichtlich ihrer Sprachentwicklung untersucht werden, dann sind die Grundgesamtheit **alle Kinder, die zu diesem Zeitpunkt den betreffenden Kindergarten besuchen**.

Beispiel: Wenn die Studierenden einer Fakultät befragt werden, dann sind **alle Studierenden der betreffenden Fakultät in einem bestimmten Zeitraum oder zu einem bestimmten Zeitpunkt** die Grundgesamtheit.

Wenn wir untersuchen wollen, *ob sich Sprachverständnis und Sprachgebrauch von „österreichischem“, „deutschem“ und „schweizerischem“ Deutsch bei Schüler*innen und Lehrer*innen unterscheiden*, dann ist die Eingrenzung der Erhebung ausschlaggebend:
Räumlich: In ganz Österreich? Nur in Wien? Nur Niederösterreich? Welche Schulen?
Sachlich: Alle Schüler*innen? Bestimmte Altersgruppen? Ausschluss von Schüler*innen mit nicht-deutscher Erstsprache?
Zeitlich: Einmalig? In welchem Zeitraum? Oder Panelstudie (immer dieselben Personen) zu mehreren Zeitpunkten?
 Mögliche Grundgesamtheit: **Alle Schüler*innen und Lehrer*innen in öffentlichen Schulen in Wien im Jahr xxxx.**

Oft lässt sich eine sehr „ehrgeizig“ definierte Grundgesamtheit als Grundlage für eine Stichprobenziehung nicht realisieren, dazu müssten alle in Frage kommenden Einrichtungen motiviert sein, bei der Studie mitzuwirken. Meist sind aber einige dazu nicht bereit und Grundgesamtheit sowie Fragestellung müssen dementsprechend angepasst werden. Bei wissenschaftlichen Studien wird das Design an die herrschenden Gegebenheiten angepasst (Budget, Zeitrahmen, Zugang, Reichweite, Teilnahmebereitschaft, Realisierbarkeit...). Die Grundgesamtheit ergibt sich daher oft aus rein praktischen Gesichtspunkten, nämlich dem Zugang bzw. den faktischen Gegebenheiten. Die Definition der Grundgesamtheit erfordert also eine konkret formulierte (eingegrenzte) Fragestellung, die Abklärung des Untersuchungszeitpunktes sowie die Definitionen der verwendeten Begriffe. Meist erfolgen diese Schritte (Formulierung der Fragestellung, Operationalisierung, Eingrenzung und Festlegung der Grundgesamtheit) in einem reflexiven Abstimmungsprozess.

Sobald die Grundgesamtheit präzise definiert ist, kann überlegt werden, wer von der Grundgesamtheit untersucht werden soll. Alle? Eine Teilmenge?

Bei einer **Vollerhebung = Totalerhebung** werden alle Elemente der Grundgesamtheit untersucht.
 Bei einer **Stichprobe** wird eine Teilmenge aus der Grundgesamtheit untersucht.

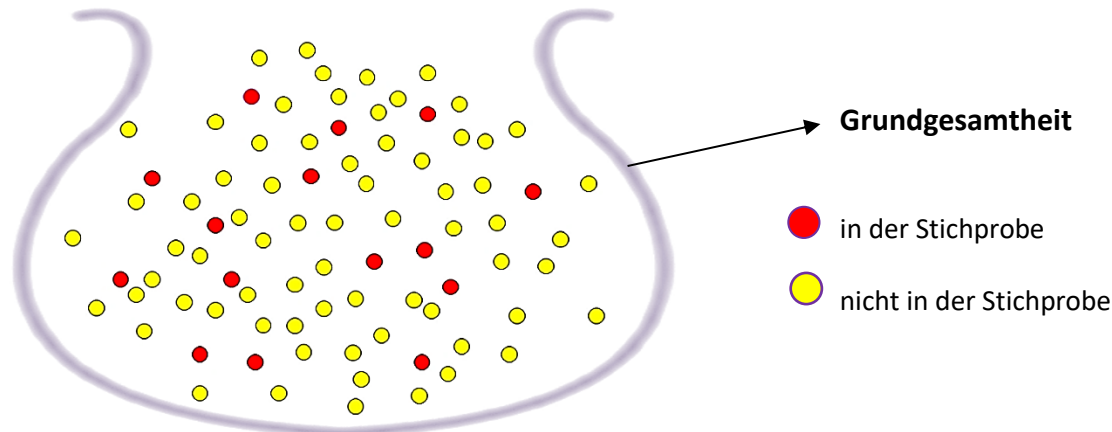


Abbildung 7: Grundgesamtheit und Stichprobe

Die folgende Grafik veranschaulicht die Überlegungen bei der **Stichprobenziehung**.

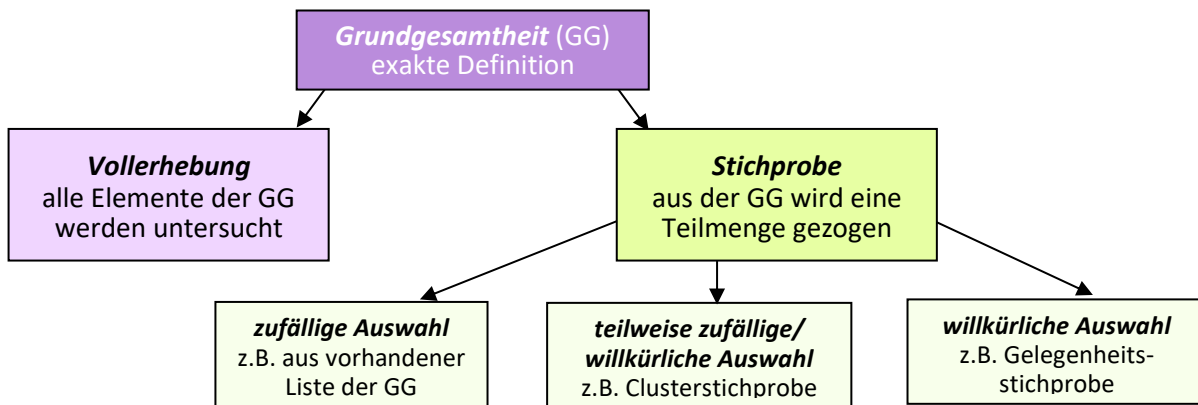


Abbildung 8: Grundgesamtheit und Stichprobe – Arten von Stichproben

Grundsätzlich gelten folgende Prioritäten: Eine **Totalerhebung** ist der Optimalfall, sie ist aber meist zu teuer oder nicht möglich. Wenn alle Elemente der Grundgesamtheit bekannt und in einer Liste vorhanden sind, dann ist eine **Zufallsstichprobe** (jede*r Zweite, jede*r Dritte, jede*r Zehnte auf der Liste – je nach Fallzahl und Budget) anzustreben. Wenn sowohl eine Totalerhebung als auch eine Zufallsauswahl nicht möglich sind, dann wird eine **willkürliche (=bewusste)** oder **teilweise willkürliche Auswahl** der untersuchten Einheiten/Personen getroffen.

Wir unterscheiden zwei Arten von Stichprobendesigns:

- Beim **repräsentativen Stichprobendesign** liegt der Schwerpunkt auf der **Beschreibung**. Hier ist es besonders wichtig, dass die Stichprobe die Grundgesamtheit in allen relevanten Merkmalen widerspiegelt, daher haben wir es meist mit einer sehr „bunten“ (heterogenen) Personengruppe zu tun und wollen (grob gesagt) wissen, wie es um diese Zielgruppe bestellt ist.

Ein typisches Beispiel für eine **repräsentative Studie** ist...

....die Befragung von Studierenden und Lehrenden zu einem bestimmten Zeitpunkt auf einer bestimmten Universität zum Gebrauch von österreichischem Dialekt in der universitären Lehre³.

- Das **experimentelle Design** hingegen zielt auf eine **Erklärung** ab. Im Fokus stehen hier die Bedingungen, unter denen der Versuch/die Testreihe stattfindet. Die untersuchten Gruppen sind meist homogen, um auf diese Weise „störende“ Einflüsse möglichst auszuschalten. Denn: Der gemessene, erzielte Effekt soll ausschließlich auf der Bedingung (z.B. Behandlungsmethode) beruhen und nicht durch andere („störende“) Umstände (z.B. unterschiedliches Alter der Proband*innen) (mit-)bedingt sein.

Ein Beispiel für ein **experimentelles Design** ist, ...

...wenn im Rahmen einer klinischen Studie die Schmerzdarstellung in Krankheitserzählungen untersucht wird, und hierbei die gesprochene Sprache von Ärzt*innen und Patient*innen auf der Schmerzambulanz in einem Krankenhaus in einem bestimmten Zeitraum aufgezeichnet und analysiert wird⁴.

³ Besuchen Sie Lisa Krammer <https://www.univie.ac.at/germanistik/lisa-krammer> und den Linguistischer Wissenschaftspodcast mundART <https://www.mundartpodcast.at/>

⁴ Menz (2013): Zum Vergleich von ärztlichen Konsultationen zu Gedanken bei gedolmetschten und nicht gedolmetschten Beziehungen, in: Menz (Hrsg): Migration und medizinische Kommunikation;

Die folgende Abbildung stellt diese beiden Designs anhand ihrer wesentlichsten Kennzeichen gegenüber.

Repräsentativ – deskriptiv verwendet bei Umfragen		Experimentell – kausal verwendet bei Interventionsstudien	
Stichprobe repräsentiert Grundgesamtheit (in allen „relevanten“ Merkmalen) Üblicherweise: Geschlecht, Alter, Region, Bildung, Einkommen, Erwerbstätigkeit – möglichst viele! Begründen, warum ein Merkmal nicht relevant ist! Die Ergebnisse sind auf die Grundgesamtheit generalisierbar – Bedingung: Zufallsauswahl (Liste, Urne – alle Elemente mit gleicher Chance)		Stichprobenauswahl durch Variation des zu untersuchenden Kriteriums oder Zeitpunkts Die Varianz des Kriteriums unterstellt Kausalität. „Bedingungen“ (z.B. Behandlung, Film, Unterrichtsmethode...) Gruppen unter verschiedenen Bedingungen: Bedingung A – nicht A (Versuchs- und Kontrollgruppe) oder Bedingung A – B – C...	
Beschreiben: Status quo in der Grundgesamtheit		Erklärung: Was bringt die „Bedingung“?	
Heterogene Stichprobe (möglichst viele Eigenschaften wie in der Grundgesamtheit) Rückschluss auf größere Personengruppe möglich		Homogene Gruppen (möglichst viele Eigenschaften gleich) Einfluss der Bedingung soll „sauber“ bleiben, soll kontrolliert werden „experimentell“ = Randomisierung: Gruppenzuteilung per Zufall „quasi-experimentell“ = Parallelisierung: homogene Gruppe natürliche, vorgegebene Gruppeneinteilung	
Repräsentativität als Kontinuum Je größer und heterogener die Stichprobe, je „zufälliger“ die Auswahl, desto repräsentativer die Ergebnisse...		eindeutige, kontrollierte Bedingungen mit wenig störenden Einflüssen (intervenierende bzw. „störende“ Variable) Interne Validität – hohe Kontrolle (<i>Laborexperiment</i>) Externe Validität – hohe Generalisierbarkeit (<i>Feldexperiment</i>)	

Abbildung 9: Repräsentatives und experimentelles Stichprobendesign im Vergleich

2.1 Repräsentativität – oder: Der Zufall ist der „Master of the Universe“

Repräsentativ ist eine Stichprobe dann, wenn sie die Grundgesamtheit repräsentiert. Was heißt das?

Repräsentativ heißt, die Stichprobe ist ein **Miniaturabbild der Grundgesamtheit**, sie entspricht in allen relevanten Merkmalen der Grundgesamtheit. (Was „relevant“ ist, definiert der*die Forscher*in!) Rückschlüsse auf die Grundgesamtheit sind nur dann zulässig, wenn die Stichprobe ein Abbild der Grundgesamtheit ist.

Bildlich können wir uns das „Miniaturabbild“ der Grundgesamtheit für die österreichische Bevölkerung ab 16 Jahren zu einem bestimmten Zeitpunkt so vorstellen:

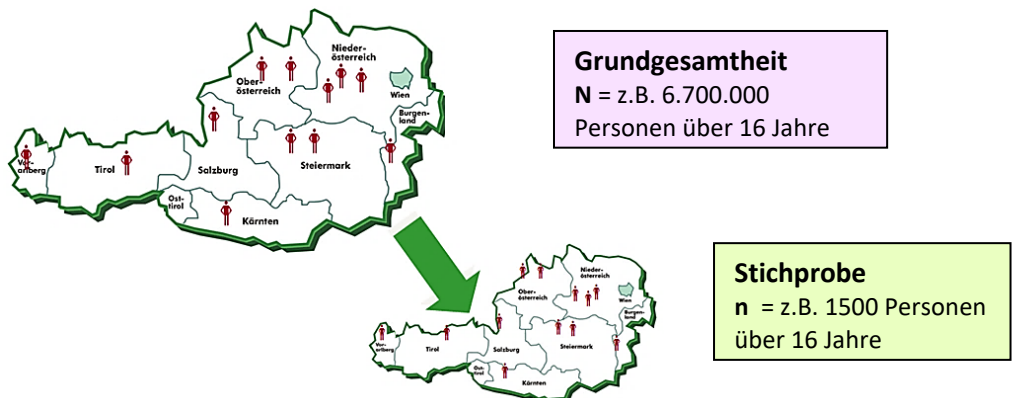


Abbildung 10: Die Stichprobe als Miniaturabbild der Grundgesamtheit am Beispiel Österreichische Bevölkerung
 Beachte die Schreibweise: Anzahl der Personen in der Grundgesamtheit = Großbuchstabe N und Anzahl der Personen in der Stichprobe = Kleinbuchstabe n. (Bildquelle: Fessel-GfK)

Gleich nach der Vollerhebung ist die reine Zufallsstichprobe die „beste“ Stichprobe: Eine reine Zufallsauswahl garantiert eine repräsentative Stichprobe! Das Prinzip: aus einer Urne werden zufällig Lose gezogen.

Wenn die untersuchten Personen oder Fälle nach einem reinen Zufallsprinzip ausgewählt werden („Lose ziehen“ oder „jede*r Zehnte von der Liste“), dann ist sichergestellt, dass alle Eigenschaften genauso vertreten sind, wie in der Grundgesamtheit: Der Anteil an Männern und Frauen, der Altersgruppen, der Bildungsabschlüsse, der Alkoholkonsument*innen, der Raucher*innen etc... entspricht **immer** dem Anteil der jeweiligen Gruppe in der Grundgesamtheit – und zwar gänzlich unabhängig davon, aus welchen konkreten Personen die Stichprobe real zusammengesetzt ist.⁵ Dies kann aufgrund der Gesetzmäßigkeiten des Zufalls garantiert werden, denn: **Der Zufall ist berechenbar!** Wenn Sie denken: „Was Zufall ist, ist unberechenbar, sonst wäre es ja nicht zufällig!“, dann werden Sie sehen: Wir werden den Zufall zähmen und für unsere Zwecke nutzen, ob er will oder nicht!⁶

Bei einer **Zufallsauswahl** hat jedes Element in der Grundgesamtheit die gleiche Chance, in die Stichprobe zu kommen. Bei einer **willkürlichen (= bewussten) Auswahl** entscheiden die Forscher*innen anhand inhaltlicher Kriterien, welche Personengruppen in die Stichprobe aufgenommen werden.

► Willkürliche oder bewusste Auswahl: Es gibt keine gleichen Chancen!

In Wien gibt es eine „Flüchtlingssiedlung“ namens Macondo, wo auf ehemaligem Kasernengelände bis zu 3.000 anerkannte Flüchtlinge aus etwa 30 verschiedenen Nationen leben.



Wir möchten untersuchen, auf welche Weise sich die Bewohner*innen untereinander verständigen.

Die Grundgesamtheit sind beispielsweise alle Bewohner*innen von Macondo über 16 Jahre im Jahr 2020.

Wir gehen an einem Sommernachmittag durch die Siedlung und befragen alle jugendlichen bzw. erwachsenen Personen, die sich auf öffentlichen Plätzen aufhalten.

Dies ist eine *Gelegenheitsstichprobe*: Wir wählen bewusst jene Personen, die am leichtesten zugänglich sind.

Repräsentativ ist diese Stichprobe nicht, denn es gibt etliche Bewohner*innen, die keine Chance haben, befragt zu werden: Jene, die an diesem Tag nicht draußen sind, jene, die arbeiten, die in Schule oder Ausbildung sind, die woanders unterwegs sind, jene, die nicht ausreichend mobil sind, jene, die in der Wohnung geblieben sind, weil sie Gäste haben, weil sie krank sind, weil... weil... weil....

Abbildung 11: „Macondo“ in Wien Simmering und die dort lebenden Nationen

Bildquelle: https://www.prearq.at/wp-content/uploads/2010/08/prearq_macondo014.jpg

► Zufallsstichprobe heißt: Gleiche Chancen für alle!

Wir erhalten einen Auszug aus dem Melderegister, wo die Adressen aller Bewohner*innen aus der Siedlung aufgelistet sind. Nun wissen wir genau über die Anzahl der Personen Bescheid und können eine Zufallsauswahl treffen. Wir können dann mittels per Post versandtem Papier-Fragebogen oder face-to-face-Kontakt an die zufällig ausgewählten Personen herantreten. Wer fehlt uns jetzt noch? Bewohner*innen, die zwar kontaktiert wurden, die aber bei der Befragung nicht mitmachen wollen – dies ist eine Frage des Rücklaufs. Aber: Es **ist** eine Zufallsstichprobe, denn: Alle, auf die unsere Einschlusskriterien (über 16 Jahre und in der Siedlung gemeldet) zutreffen, haben die gleiche Chance, in die Stichprobe zu kommen.

Kriterium für eine repräsentative Stichprobe ist also: Jede*r hat die gleiche Chance, reinzukommen.⁷

⁵ Ob in der Stichprobe Samira aus Syrien, Ayden aus Afghanistan, Leila aus dem Sudan, Abdi aus Somalia oder Rainer Zufall aus Dumpling-City drin sind, ist unerheblich – es kommt immer das gleiche Ergebnis raus. Mal ist zufällig die eine dabei, dann wieder zufällig der andere.

⁶ Erfahrungsgemäß hat der Zufall gar keinen eigenen Willen.

⁷ Dabei wird bei größeren Studien aus Kostengründen das Zufallsprinzip meist „ein bisschen“ umgangen, indem bereits vorher die Anteile an Personen eines bestimmten Geschlechts, Alters, bestimmter Bildung und Region festgelegt werden, und dann solange befragt wird, bis diese Quoten erfüllt sind. Eine solche "pragmatische" Stichprobe heißt „Quotenstichprobe“, sie entspricht der Grundgesamtheit in allen relevanten Merkmalen und gilt daher meist als repräsentativ.

Kriterien einer Zufallsstichprobe:

- **Jede Person/jedes Element der Grundgesamtheit, hat die gleiche bzw. eine angebbare Wahrscheinlichkeit** in die Stichprobe zu kommen („gezogen zu werden“).
- **Jede mögliche Stichprobe hat die gleiche Wahrscheinlichkeit** zustande zu kommen („gezogen zu werden“), das heißt: Jede mögliche Stichprobe bringt gleiche Ergebnisse! (siehe Fußnote 5 auf der vorigen Seite)
- Innerhalb bestimmter Grenzen (Schwankungsbreiten, sogenannter Konfidenzintervalle) kann der „wahre“ Anteil eines **Merkmals in der Grundgesamtheit geschätzt** werden.

Es geht also um die gleichen Chancen. Aber was bedeutet das mit der „**angebbaren Wahrscheinlichkeit**“? Es gibt auch Studiendesigns, in denen unterschiedliche Gruppen unterschiedliche – aber untereinander wieder gleiche – Chancen haben, in die Stichprobe zu gelangen.

Beispiel für angebbare Wahrscheinlichkeiten:

Nehmen wir an, wir untersuchen im Rahmen einer Feldstudie (Datenerhebung in der Wohnumgebung) lediglich die Art der Verständigung zwischen Flüchtlingen aus Syrien und Flüchtlingen aus Afghanistan und brauchen 50 Personen von jeder Gruppe. Nachdem jedoch viel mehr afghanische Flüchtlinge in der Siedlung leben als syrische, haben wir auf der Liste mit dem Herkunftsland Afghanistan vielleicht 500 Personen und auf der Liste mit Herkunftsland Syrien beispielsweise 100 Personen. Um jeweils 50 Personen zu erhalten, brauchen wir aus der 500er Liste jede zehnte und aus der 100er-Liste jede zweite Person. Die Personen auf der Afghanistan-Liste haben also eine 10%ige Chance, in die Stichprobe zu kommen (50/500), und jene auf der Syrien-Liste haben eine 50%ige Chance, ausgewählt zu werden (50/100).

Wir können diese Wahrscheinlichkeiten genau angeben, daher handelt es sich um eine Zufallsstichprobe.

Bei einer Zufallsstichprobe hat jede Person aus der Grundgesamtheit theoretisch die gleiche oder eine exakt angebbare Chance, teilzunehmen. Praktisch aber gibt es in jeder Stichprobe eine **systematische Verzerrung** (= „**Bias**“), die nicht zu verhindern ist. Diese ergibt sich aus der Erreichbarkeit und der Teilnahmebereitschaft.

Typische Verzerrungen sind:

- Die **Überrepräsentation** von Frauen, von Älteren, von „Ärmeren“: Diese Personengruppen sind erfahrungsgemäß leichter zu erreichen und bereitwilliger, an Untersuchungen teilzunehmen.
- Die **Unterrepräsentation** von Unzufriedenen: Bei Untersuchungen zur Zufriedenheit liegen von den Unzufriedeneren meist geringere Rücklaufquoten⁸ als von Zufriedeneren (Bewohner*innen, Mitarbeiter*innen...).

► **Gewichtung**

Nach Vorliegen der Stichprobe zeigen sich also meistens kleinere Abweichungen zur Grundgesamtheit. Im Fall, dass Informationen über die relevanten Merkmale in der Grundgesamtheit vorliegen, können diese Abweichungen „*zurecht-gewichtet*“ werden: Überrepräsentierte Gruppen können „*hinunter-gewichtet*“ und unterrepräsentierte Gruppen können „*hinauf-gewichtet*“ werden.⁹

Die Frage, ob die Stichprobe nach der Gewichtung noch repräsentativ ist, ist in Fachkreisen umstritten. Aus meiner Sicht ist eine Gewichtung für die Repräsentativität unerheblich, wenn die dabei zugewiesenen „Gewichte“ nicht stark von 1 abweichen, also nicht größer als 2 bzw. kleiner als 0,5 sind.

Auch das Durchführen von Signifikanztests mit gewichteten Umfragedaten ist nicht unproblematisch. Wenn die Gewichtungsfaktoren nicht stark von 1 abweichen, werden Signifikanztests in der Praxis meist durchgeführt. Sollte der Anteil der Fälle mit Gewichtungsfaktoren über 2 bzw. unter 0,5 jedoch beträchtlich sein, dann empfiehlt es sich, Signifikanztests mit ungewichteten Daten durchzuführen. Denn: Bei der Signifikanztestung kommt es sehr stark auf die Fallzahl an, die ja durch die Gewichtung verändert wird!

⁸ Die Rücklaufquote ist der Anteil der Personen, die an der Studie teilgenommen haben versus jener, die zwar für die Stichprobe ausgewählt wurden, aber nicht erreicht wurden oder nicht teilnehmen wollten.

⁹ Dabei können Sie sich ruhig eine Modellierung vorstellen: Hier kommt ein bisschen was weg, dort kommt ein bisschen mehr dazu, so lange, bis das gesamte Kunstwerk den gewünschten Proportionen entspricht. Das ist "iteratives Gewichten" – dabei wird schrittweise untersucht, wie gut die Stichprobe passt, dann wieder angepasst, dann wieder untersucht, dann wieder angepasst... usw.