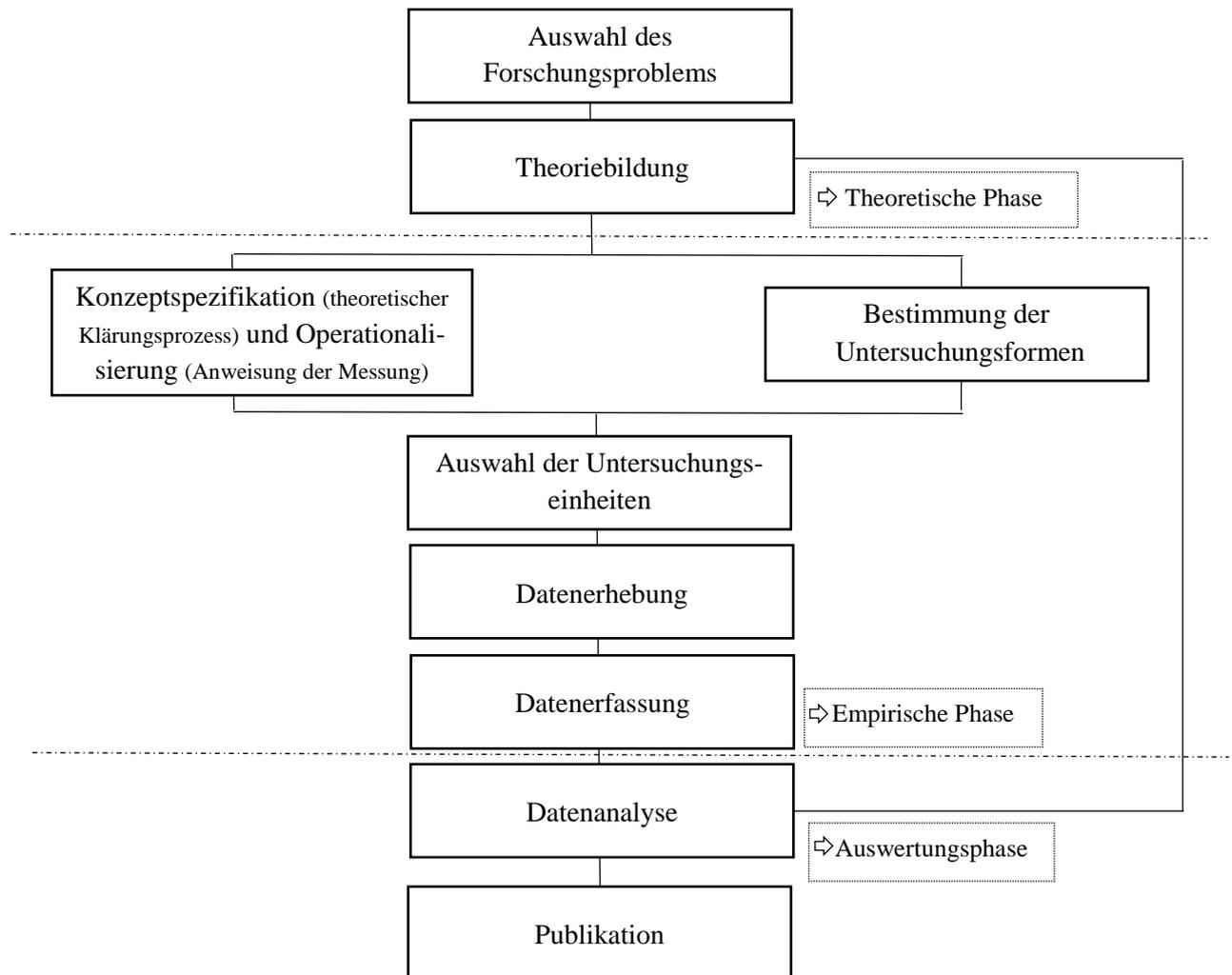


Skript – Grundlagen der quantitativen Sozialforschung

Bei der Durchführung eines empirischen Forschungsprojektes lassen sich einige grundlegende Arbeitsschritte unterscheiden:



Quelle: vgl. Schnell et al. 2008: 8. Eigene Darstellung.

I. Grundbegriffe der Statistik

- Deskriptive Statistik (beschreibende Statistik): tabellarische und graphische Darstellung der Untersuchungsergebnisse (univariate Analyse, bivariate Analyse, multivariate Analyse)
- Variable: Merkmal, mit mehr als einer Eigenschaft (Ausprägung)
- Im Falle eines standardisierten, quantitativen Fragebogens sind Variablen zumeist gleichgesetzt mit „Items“ bzw. Fragen
- Ausprägungen (Werte/ Kategorien) der Variable müssen bestimmten Regeln folgen:
 1. Ausprägungen sollten nicht überlappend (disjunkt) sein
und
 2. Ausprägungen sollten erschöpfend sein.

→ Beispiel: Wie viel wiegen Sie? 20-50 kg 50-80 kg 80-110 kg

→ Antwortkategorien sind *nicht disjunkt* und auch *nicht erschöpfend!*

Unterscheidung der Arten von Variablen nach bestimmten Eigenschaften:

1. Variablen nach der Art der Ausprägung	Beispielvariablen
Kontinuierlich/stetig	Geschwindigkeit
Diskret dichotom polytom	Besitz des Führerscheins, Schwangerschaft Staatsangehörigkeit
2. Variablen nach der Art der empirischen Verwertbarkeit	Beispielvariablen
Latent (Variablen, die nicht direkt messbar bzw. beobachtbar sind, sie werden mithilfe von Indikatoren erfasst)	Umweltbewusstsein, sozialer Status, Religiosität
Manifest (Variablen, die direkt messbar sind; wichtig Rolle als Indikatoren für latente Variablen)	Gewicht, Körpergröße, Einkommen
3. Variablen nach der Position in einer Hypothese	Beispiel zur Hypothese: Mütter, die alleinerziehend sind, beziehen häufiger Sozialleistungen vom Staat.
Abhängige Variable (zu erklärender Sachverhalt, Wirkung)	Bezug von Sozialleistungen: <input type="checkbox"/> Ja <input type="checkbox"/> Nein
Unabhängige Variable (vermuteter Erklärungsfaktor eines Sachverhalts, Ursache/Prädiktor)	Alleinerziehend: <input type="checkbox"/> Ja <input type="checkbox"/> Nein Geschlecht: <input type="checkbox"/> Frau <input type="checkbox"/> Mann
4. Variablen nach der Art des Skalenniveaus	Beispielvariablen
Nominales Skalenniveau (qualitative Variable)	Geschlecht, Familienstand
Ordinales Skalenniveau (qualitative Variable)	Sozialer Status, Schulnoten, Zufriedenheit
Intervallskalenniveau (quantitative Variable)	metrisch Temperatur in Grad Celsius
Ratioskalkenniveau (quantitative Variable)	

Skalenniveaus:

- **Nominalskalierte Daten** haben den niedrigsten Informationsgehalt. Es handelt sich meist um Kategorien, die zur Auswertung numerisch codiert werden. Sie werden daher manchmal auch "Kategorialskalen" genannt. Beispiele hierfür sind „Geschlecht" (1 für „männlich“, 2 für „weiblich“) oder „Religionszugehörigkeit" (1 für „evangelisch“, 2 für „katholisch“ etc.). Bei nur zwei Merkmalsausprägungen spricht man auch von einer dichotomen oder binären Variable. Bei mehr als zwei Merkmalsausprägungen spricht man von polytomen Variablen. Der Zahlenwert ist beliebig und ist nur für die Unterscheidung der Objekte bedeutsam. Mit nominalskalierten Daten lassen sich keine mathematischen Rechenoperationen (wie Addition und Subtraktion) durchführen, sondern nur Aussagen über relative und absolute Häufigkeiten machen. Als Lageparameter dient der Modus/ Modalwert.
- **Ordinalskalierte Variablen** folgen einer Rangreihe. Es lassen sich keine Aussagen über die absoluten Abstände zwischen den Werten machen. Ein Beispiel für ordinalskalierte Daten sind „Schulnoten“. Auch mit ordinalskalierten Variablen lassen sich keine mathematischen Rechenoperationen durchführen. Als Lageparameter dient der Median.

- Für **intervallskalierte Variablen** ist konstitutiv, dass die den Eigenschaften zuzuordnenden Zahlen nicht nur eine "namensgebende" und "ordnende" Funktion haben, sondern auch Aussagen über die - empirisch sinnvollen - Abstände zwischen Skalenwerten erlauben. Diese Eigenschaftsausprägungen ermöglichen es, die mathematischen Rechenoperationen von Addition, Subtraktion, Multiplikation und Division anzuwenden. Als Lageparameter dient der Arithmetische Mittelwert. Ein typisches Beispiel für eine intervallskalierte Variable ist „Temperatur in Grad Celsius“.
- **Ratioskalen** oder auch **Verhältnisskalen** stehen auf der höchsten Stufe der Skalenhierarchie. Sie weisen alle Eigenschaften der vorhergehenden Skalentypen auf und haben darüber hinaus einen absoluten Nullpunkt. Zu den bei Intervallskalen zulässigen mathematischen Rechenoperationen kommt hier die Möglichkeit hinzu Aussagen über Quotienten je zweier Zahlen zu treffen. Ein typisches Beispiel für eine ratioskalierte Variable ist „Lebensalter“.

Unbestritten ist, dass die vier genannten Skalentypen eine **kumulative Hierarchie** bilden - bei der die Ratioskala zugleich sämtliche Eigenschaften von Intervall-, Ordinal- und Nominalskalen aufweist; die Intervallskala jene von Ordinal- und Nominalskalen sowie die Ordinalskala jene der Nominalskalen. Es besteht auch die Möglichkeit Variablen zu transformieren oder auch polytome Variablen zu dichotomisieren. *In der statistischen Praxis werden (abhängige) ordinalskalierte Variablen häufig, wenn auch strittig, als metrischskalierte Variablen behandelt. Unter datenanalytischen Gesichtspunkten spricht wenig dagegen, Rangskalen mit vergleichsweise vielen Merkmalsausprägungen (wie beispielsweise Likertskalen mit fünf und mehr Ausprägungen) wie metrische Variablen zu behandeln, sofern ihre Verteilung als annähernd normalverteilt gelten kann* (vgl. Wittenberg et al. 2014: 150).

Hypothesenbildung und -prüfung:

Die durch eine Untersuchung zu überprüfenden Hypothesen lassen sich prinzipiell in Forschungs-, operationale und statistische Hypothesen unterscheiden:

- Für **Forschungshypothesen** ist charakteristisch, dass sie allgemein formuliert sind: Sie drücken ein generelles Gesetz aus, das für die Grundgesamtheit einer Untersuchungspopulation gilt. Ihre Aussagen müsse sich prinzipiell in der Form von „Wenn-dann-Sätze“ oder „Je-desto-Sätze“ formulieren lassen. Die Hypothesen müssen nicht in dieser Form gestaltet sein, jedoch sollten sie zumindest leicht in diese Form umformulierbar sein.
- Dahingehend beziehen sich **operationale Hypothesen** nicht auf die Grundgesamtheit, sondern ausschließlich auf die Untersuchungsgesamtheit. Sie stellen zumeist eine Prognose dar, wie die Ergebnisse einer Untersuchung ausfallen werden. Man nutzt sie auch, um vor Untersuchungsbeginn Ergebnisse zu definieren, die die Forschungshypothesen verifizieren oder falsifizieren sollen.
- **Statistische Hypothesen** werden schließlich verwendet um durch formale mathematische Überprüfung von zwei sich ausschließenden statistischen Entscheidungen die Zuverlässigkeit einer Aussage über die Grundgesamtheit zu ermitteln, die die Forschungshypothesen verifizieren oder falsifizieren sollen.

H₁ (Alternativhypothese): Entspricht inhaltlich der Forschungshypothese oder operationalen Hypothese. Es besteht eine wie auch immer geartete Beziehung zwischen den untersuchten Variablen.

H₀ (Nullhypothese): Entspricht inhaltlich nicht der Forschungshypothese oder operationalen Hypothese. Es besteht keine Beziehung zwischen den untersuchten Variablen.

Des Weiteren lassen sich die drei Hypothesengrundformen in Zusammenhangs-, Unterschieds- und Veränderungshypothesen kategorisieren:

- **Zusammenhangshypothesen** behaupten einen Zusammenhang zwischen zwei oder mehr Variablen. Zur Überprüfung einer solchen Behauptung werden statistische Signifikanztests durchgeführt.
- **Unterschiedshypothesen** behaupten, dass zwei oder mehr Teilgruppen, die die Ausprägung von einer oder mehreren unabhängigen Variablen repräsentieren, sich in Hinblick auf die abhängige Variable unterscheiden. Die Überprüfung von Unterschiedshypothesen erfolgt mit statistischen Signifikanztests des Unterschiedes von Mittelwerten und Varianzen.

- **Veränderungshypothesen** berücksichtigen den Zeitverlauf und behaupten, dass die über die Zeit variierende Ausprägung einer oder mehrerer unabhängiger Variablen die Werte der abhängigen Variablen einer oder mehrerer Populationen verändern. Die statistische Überprüfung ist häufig komplex und wird mit den Verfahren der Zeitreihenanalyse oder mit der Panelanalyse durchgeführt.

Art der Messung:

Bei **abhängigen Messungen** sind die untersuchten Variablen zeitverschoben jeweils bei derselben Untersuchungseinheit ermittelt worden. Dies ist beispielsweise bei Paneluntersuchungen und Experimenten der Fall. In drei Situationen kann von einer abhängigen Messung gesprochen werden:

- **Messwiederholung:** Die Messwerte stammen von der gleichen Person, zum Beispiel wird für jede Person das Gewicht bei der Krankenhausaufnahme (Gewicht [t0]) und bei der Entlassung (Gewicht [t1]) gemessen.
- **Natürliche Paare:** Die Messwerte stammen von verschiedenen Personen, diese gehören aber zusammen, zum Beispiel Ehefrau und Ehemann.
- **Matching:** Die Messwerte stammen von verschiedenen Personen, die aber einander zugeordnet wurden.

Bei **unabhängigen Messungen** werden Untersuchungseinheiten nur ein einziges Mal in die Stichprobe einbezogen.

II. Univariante Datenanalyse

Methoden der univariaten Datenanalyse haben nur eine Variable zum Gegenstand. Typisch sind Häufigkeitsanalysen sowie die Ermittlung von Lage- und Streuungsmaßen, wie der Median, das arithmetische Mittel und die Standardabweichung.

III. Bivariate Datenanalyse

Das Ziel der bivariaten Datenanalyse ist die Untersuchung der Beziehung zwischen zwei verschiedenen Variablen. Diese können in Form von Zusammenhängen oder in Form von Unterschieden untersucht werden. Dabei hängt die Wahl des geeigneten Analyseinstrumentes von den theoretischen Vorüberlegungen ab, aber auch von den mathematisch-statistischen Voraussetzungen der jeweils involvierten Variablen. Die gebräuchlichsten **Signifikanztests** werden in der nachfolgenden Abbildung dargestellt:

Anzahl der Kategorien der unabhängigen Variablen	Art der Messung	Messniveau der abhängigen Variablen		
		<i>nominal</i>	<i>ordinal (oder metrisch, aber nicht normalverteilt)</i>	<i>metrisch und normalverteilt</i>
<i>zwei</i>	<i>abhängige</i>	McNemar-Test	Wilcoxon-Test	T-Test
<i>mehrere</i>	<i>abhängige</i>	Cochran`s Q-Test	Friedman`s Rangvarianzanalyse; Kendall`s Konkordanz-Koeffizient W	Varianzanalyse
<i>zwei</i>	<i>unabhängige</i>	Fisher`s exakter Test; Chi-Quadrat-Test	Median-Test; Mann-Whitney-U-Test; Kolmogorov-Smirnov-Test; Wald-Wolfowitz-Test	T-Test
<i>mehrere</i>	<i>unabhängige</i>	Chi-Quadrat-Test	Median-Test; Kruskal-Wallis` ; Rangvarianzanalyse	Varianzanalyse

Quelle: vgl. Wittenberg/ Cramer/ Vicari (2014): 196. Eigene Darstellung.

Neben der reinen Deskription bivariater Zusammenhänge zwischen Variablen sind folgende Fragen konstitutiv:

1. Ist ein hypothetisch formulierter Zusammenhang zwischen zwei Variablen tatsächlich statistisch signifikant?
2. Wenn ja, wie stark ist dieser Zusammenhang?
3. Welche Richtung weist der Zusammenhang auf? (bei mindestens ordinalskalierten Variablen)
4. Welche Drittvariablen korrelieren ggf. mit der abhängigen Variable?

Tests zur Überprüfung der Stärke des Zusammenhangs:

Der Assoziationskoeffizient gibt die Stärke eines bivariaten Zusammenhangs an. Die Wahl des Zusammenhangsmaßes wird durch das Messniveau der Variablen bestimmt. Dabei bestimmt die Variable mit dem niedrigeren Messniveau das zu wählende Zusammenhangsmaß. Je näher der Wert bei 1 liegt, desto stärker ist auch der Zusammenhang. Maßzahlen für mindestens ordinalskalierte Variablen zeigen zusätzlich zur Stärke des Zusammenhangs auch die Richtung des Zusammenhangs an (-1 negativer Zusammenhang; 0 kein Zusammenhang; + 1 positiver Zusammenhang).

Skalenniveau	Zusammenhangsmaße
nominalskaliert	Phi Φ (in 2x2 Feldertabellen); Cramers V
mindestens ordinalskaliert	Spearman`s Korrelationskoeffizient Rho, Kendalls Tau
metrisch	Pearson`s Korrelationskoeffizient R
Zur sprachlichen Beschreibung der Stärke von Zusammenhängen	
$0,00 < Kor \leq 0,20$	sehr geringer Zusammenhang
$0,20 < Kor \leq 0,50$	geringer Zusammenhang
$0,50 < Kor \leq 0,70$	mittlerer Zusammenhang
$0,70 < Kor \leq 0,90$	starker Zusammenhang
$0,90 < Kor \leq 1,00$	sehr starker Zusammenhang

Quelle: Wittenberg/ Cramer/ Vicari (2014): 210. Eigene Darstellung.

IV. Multivariate Datenanalyse

Multivariate Verfahren der Datenanalyse geben die Möglichkeit mindestens drei Variablen zu betrachten. Multivariate Verfahren unterscheiden sich in strukturprüfende Verfahren (Dependanzanalysen), wie lineare Regression, logistische Regression und Diskriminanzanalyse und in strukturentdeckende Verfahren (Interdependanzanalyse), wie Faktoren- und Clusteranalyse.

Quellenangaben:

AKREMI, LEILA/ BAUR, NINA/ FROMM, SABINE (2011): Datenanalyse mit SPSS für Fortgeschrittene 1. Datenaufbereitung und uni- und bivariate Statistik. 3., überarbeitete und erweiterte Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.

BORTZ, JÜRGEN/ DÖRING, Nicola (2005): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer Medizin Verlag.

FROMM, SABINE (2012): Datenanalyse mit SPSS für Fortgeschrittene 2. Multivariate Verfahren für Querschnittsdaten. Wiesbaden: VS Verlag für Sozialwissenschaften.

SCHNELL, RAINER/ HILL, PAUL B./ ESSER, ELKE (2008): Methoden der empirischen Sozialforschung. 8. Auflage. München: Oldenbourg Verlag.

WITTENBERG, REINHARD/ CRAMER, HANS/ VICARI, BASHA (2014): Datenanalyse mit IBM SPSS Statistics. Eine syntaxorientierte Einführung. Konstanz und München: UVK Verlagsgesellschaft mbH.