

# Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (6. Auflage)

## Lösungen der Lernquiz

### Kapitel 1: Empirische Sozialforschung im Überblick

---

**Aufgabe 1.1:** Welche Wege der Produktion und Begründung von Wissen existieren außerhalb der wissenschaftlichen Forschung?

**Aufgabe 1.1:** Wissen durch Berufung auf

1. Autoritätspersonen,
2. Religion,
3. Tradition,
4. gesunden Menschenverstand,
5. Intuition,
6. anekdotische Evidenz und
7. Logik.

**Aufgabe 1.2:** Was versteht man unter Wissenschaftsbildung und Methodenkompetenz?

**Aufgabe 1.2:** Wissenschaftsbildung meint ein grundlegendes Verständnis des Wissenschaftssystems sowie des wissenschaftlichen Forschungsprozesses. Die Wissenschaftsbildung der Bevölkerung gewinnt an Bedeutung, weil in der heutigen Wissenschaftsgesellschaft in vielen gesellschaftlichen Bereichen auf wissenschaftliche Erkenntnisse Bezug genommen wird. Methodenkompetenz ist enger gefasst und bezeichnet die Kenntnisse und Fähigkeit, die notwendig sind, um am Wissenschaftssystem teilzunehmen, also wissenschaftliche Originalquellen zu verstehen und selbst Forschungsprojekte zu realisieren.

**Aufgabe 1.3:** Warum spricht man von „empirischer“ Sozialforschung? Wie ordnen sich Sozialwissenschaften in das System wissenschaftlicher Disziplinen ein?

**Aufgabe 1.3:** Unter „empirischer“ Sozialforschung versteht man die Untersuchung sozialer Sachverhalte auf der Basis einer systematischen Sammlung und Analyse von Daten über die Erfahrungswirklichkeit. Die Sozialwissenschaften werden den empirischen bzw. Erfahrungswissenschaften zugeordnet. Sie stehen dabei zusammen mit den Natur- sowie den Technikwissenschaften den nicht-empirischen Wissenschaften gegenüber, nämlich den Geistes- und Formalwissenschaften.

**Aufgabe 1.4:** Was sind Unterschiede und Gemeinsamkeiten quantitativer und qualitativer Sozialforschung?

**Aufgabe 1.4:** Die quantitative Sozialforschung arbeitet mit quantitativen Daten (d. h. mit numerischen Daten bzw. Messwerten), die statistisch auszuwerten sind. Die qualitative Sozialforschung arbeitet mit qualitativen Daten (d. h. mit verbalen, visuellen oder audiovisuellen bzw. narrativen Materialien), die interpretativ auszuwerten sind. Abgesehen von der Art des empirischen Datenmaterials und den damit verbundenen quantitativen und qualitativen Methoden der Datenerhebung und Datenanalyse unterscheiden sich beide Ansätze auch in ihren Methodologien und ihren wissenschaftstheoretischen

Grundlagen. Sowohl in der quantitativen als auch in der qualitativen Sozialforschung geht es darum, auf der Basis von empirischen Daten im Rahmen etablierter wissenschaftlicher Methodologien und Paradigmen Kenntnisse über soziale Sachverhalte zu gewinnen, die Kriterien der Wissenschaftlichkeit entsprechen.

**Aufgabe 1.5:** Was ist mit der SQ3R-Methode gemeint und wofür braucht man sie?

**Aufgabe 1.5:** Die SQ3R-Methode (Survey, Question, Read, Recite, Review) eine Methode, die dazu dient, sich einen wissenschaftlichen Text systematisch zu erschließen.

**Aufgabe 1.6:** Was versteht man unter interdisziplinärer Forschung? Welche Voraussetzungen müssen dafür gegeben sein?

**Aufgabe 1.6:** Bei interdisziplinärer Forschung werden Fragestellungen, Theorien, Methoden und/oder Ergebnisse unterschiedlicher wissenschaftlicher Disziplinen oder Subdisziplinen miteinander verknüpft, um eine Forschungsfrage umfassender zu bearbeiten. Dies wird möglich durch Forschende, die in mehreren Disziplinen ausgebildet sind und/oder durch Forschungsteams, die aus Vertretern unterschiedlicher Fächer zusammengesetzt sind.

**Aufgabe 1.7:** Was ist mit Mixed-Methods-Research im Kontext der empirischen Sozialforschung gemeint?

**Aufgabe 1.7:** Im Rahmen des Mixed-Methods-Ansatzes werden quantitative und qualitative Methoden und Methodologien innerhalb einer Studie in unterschiedlich aufwendiger Weise miteinander verknüpft. Gleichzeitig wird im Mixed-Methods-Ansatz auch die Vereinbarkeit der quantitativen und qualitativen Wissenschaftsverständnisse bzw. Paradigmen diskutiert.

**Aufgabe 1.8:** Grenzen Sie Grundlagen- und Anwendungsforschung hinsichtlich ihrer Fragestellungen, Methoden und Ergebnisse voneinander ab!

**Aufgabe 1.8:** In der Grundlagenforschung können sich Forschende ihre Themen relativ frei wählen, in der Anwendungsforschung dominiert Auftragsforschung, d. h. das Forschungsproblem wird vom Auftraggeber vorgegeben. In der Grundlagenforschung wird möglichst große methodische Strenge angestrebt, in der Anwendungsforschung sind unter den Randbedingungen der Praxis oft Einschränkungen der methodischen Strenge in Kauf zu nehmen. Während die Ergebnisse der Grundlagenforschung primär zum wissenschaftlichen Erkenntnisfortschritt beitragen sollen, wird von Ergebnissen der Anwendungsforschung vor allem hohe praktische Relevanz für den Auftraggeber verlangt.

**Aufgabe 1.9:** Was bedeutet es, wenn ein Zeitschriftenartikel sehr oft zitiert wird? Wie kommt dies üblicherweise zustande?

**Aufgabe 1.9:** Wenn ein wissenschaftlicher Fachartikel häufig zitiert wird, präsentiert er offenbar Theorien, empirische Befunde und/oder Methoden, die in vielen anderen wissenschaftlichen Studien aufgegriffen werden. Dies ist vor allem der Fall bei Beiträgen zu Themen, zu denen eine große Forschungscommunity existiert und bei Beiträgen zu Methoden, die in zahlreichen verschiedenen Forschungsfeldern zum Einsatz kommen können. Zudem muss ein entsprechend relevanter Beitrag bereits etwas älter sein, um mit der Zeit viele Zitationen zu sammeln.

**Aufgabe 1.10:** Wie lässt sich der Forschungsprozess in der quantitativen Sozialforschung beschreiben? Worin unterscheidet er sich vom qualitativen Forschungsprozess?

**Aufgabe 1.10:** Der quantitative Forschungsprozess lässt sich in neun Phasen einteilen, die sequenziell abgearbeitet werden:

1. Wahl von Forschungsthema und Forschungsproblem (inklusive Forschungsfragen bzw. Forschungshypothesen),
2. Erarbeitung von Forschungsstand und theoretischem Hintergrund,
3. Festlegung des Untersuchungsdesigns,
4. Operationalisierung aller interessierenden Merkmale der Erfahrungswirklichkeit,
5. Stichprobenziehung,
6. Datenerhebung,
7. Datenaufbereitung,
8. Datenanalyse,
9. Ergebnispräsentation.

Besonders arbeitsaufwendig sind bei quantitativen Studien in der Regel die Theoriearbeit (Phase 2) sowie die Operationalisierung (Phase 4).

Im Unterschied zum sequenziell strukturierten quantitativen Forschungsprozess ist der qualitative Forschungsprozess zumindest teilweise zirkulär organisiert, d. h. Zwischenergebnisse der Datenauswertung steuern oft die weitere Stichprobenziehung oder tragen zur Verfeinerung des Analyseinstruments bei. Zudem liegt die Hauptarbeit im qualitativen Forschungsprozess meist in der Phase der Datenanalyse und der damit einhergehenden Phase der Hypothesen- und Theoriebildung – also zeitlich nach der Datenerhebung und nicht wie im quantitativen Forschungsprozess vor der Datenerhebung.

## **Kapitel 2: Wissenschaftstheoretische Grundlagen der empirischen Sozialforschung**

---

**Aufgabe 2.1:** Welche Aussagen stimmen?

- a) Fragen nach dem Wesen wissenschaftlicher Untersuchungsgegenstände fallen in den Bereich der Ontologie.
- b) Die Erkennbarkeit der Wirklichkeit ist eine epistemologische Frage.
- c) Die Axiologie ist kein Teilbereich der Wissenschaftstheorie.
- d) Die Wissenschaftstheorie ist normativ ausgerichtet.
- e) Wenn die Wissenschaftspraxis von den Vorgaben der Wissenschaftstheorie abweicht, muss die Wissenschaftstheorie verworfen werden.

**Aufgabe 2.1:** Richtig sind: a, b, d

**Aufgabe 2.2:** Erläutern Sie das Falsifikationsprinzip des Kritischen Rationalismus!

**Aufgabe 2.2:** Gemäß Kritischem Rationalismus entsteht Erkenntnisfortschritt dadurch, dass falsche Theorien durch Falsifikation (Widerlegung) ausgesondert werden. Dazu werden aus der zu prüfenden Theorie empirisch prüfbare Hypothesen abgeleitet und mit aussagekräftigen Daten konfrontiert. Die Entscheidung über die Hypothese wird im quantitativen Paradigma in der Regel anhand eines statistischen Signifikanztests (oder eines anderen inferenzstatistischen Verfahrens) getroffen. Der Rückschluss vom Ergebnis des Hypothesentests (Widerlegung oder vorläufige Bestätigung der Forschungshypothese) auf die Theorie und deren Bewährungsgrad bedarf einer Diskussion der Hilfs- und Instrumententheorien der Hypothesenprüfung. Auch die heuristische Erklärungskraft einer

Theorie im Vergleich mit konkurrierenden inhaltlichen Theorien muss abgewogen werden, bevor eine falsifizierte Hypothese zum Anlass genommen wird, die zugrunde liegende Theorie zu verwerfen (raffinierter methodologischer Falsifikationismus).

**Aufgabe 2.3:** Welche Rolle spielt die Wissenschaftsgemeinschaft im Kritischen Rationalismus, vor allem vor dem Hintergrund, dass sowohl ein naiver Realismus als auch ein naiver Falsifikationismus erkenntnistheoretisch abgelehnt werden?

**Aufgabe 2.3:** Laut Kritischem Rationalismus soll die Wissenschaft sich der Wahrheitssuche widmen, indem sie Theorien bzw. Hypothesen mit Daten konfrontiert, um a) falsche Theorien auszusondern und um b) Theorien, die offenbar mit der empirischen Erfahrungswirklichkeit vereinbar sind, vorläufig beizubehalten. Dabei wird jedoch weder ein naiver Realismus, demgemäß Forschende direkt die Wirklichkeit erkennen, noch ein naiver Falsifikationismus, dem gemäß theoriekonträre Daten direkt die „Falschheit“ einer Theorie beweisen können, zugrunde gelegt. Vielmehr wird vor dem Hintergrund eines kritischen Realismus und eines raffinierten methodologischen Falsifikationismus ausdrücklich darauf hingewiesen, dass sowohl die Daten selbst theoriegeladen sind als auch die Bewertung von Theorien und Hypothesen im Lichte von Daten von Interpretationen abhängt. Als Regulativ fungiert die Wissenschaftsgemeinschaft (Scientific Community), vor der die mit dem jeweiligen methodischen Vorgehen verbundenen Hilfs- und Instrumententheorien einer Studie offengelegt, begründet und Theoriebewertungen im Vergleich zu konkurrierenden Theorien verteidigt werden müssen. Wissenschaftliche Forschung wird im quantitativen Paradigma als aktiver und kritisch reflektierter Produktionsprozess von Theorien und Daten durch die Wissenschaftsgemeinschaft und nicht als positivistisches Auffinden einer objektiven Wahrheit verstanden.

**Aufgabe 2.4:** Welche Kriterien werden an Theorien gemäß Kritischem Rationalismus angelegt?

**Aufgabe 2.4:**

1. Innere Widerspruchsfreiheit
2. Äußere Widerspruchsfreiheit
3. Falsifizierbarkeit
4. Möglichst hoher Grad an Falsifizierbarkeit
5. Möglichst große heuristische Erklärungskraft
6. Größtmögliche Sparsamkeit
7. Möglichst hoher empirischer Bewährungsgrad

**Aufgabe 2.5:** Welche Positionen vertritt der Kritische Rationalismus?

- a) Wissenschaftliche Erkenntnis entsteht durch die Sammlung von Daten über die Realität.
- b) Wenn Daten einer wissenschaftlichen Theorie widersprechen, muss diese verworfen werden.
- c) Forschende sollten gegenüber sozialen Sachverhalten eine wertfreie, neutrale Haltung einnehmen.
- d) Für den wissenschaftlichen Erkenntnisfortschritt sind die statistisch signifikanten Ergebnisse ausschlaggebend.
- e) Eine wissenschaftliche Theorie sollte die Komplexität der Wirklichkeit abbilden und möglichst alle Ursachenfaktoren eines Phänomens einbeziehen.

**Aufgabe 2.5:** Richtig: – (keine)

**Aufgabe 2.6:** Nennen Sie die fünf zentralen wissenschaftstheoretischen Grundprinzipien qualitativer Sozialforschung!

**Aufgabe 2.6:**

1. Ganzheitliche und rekonstruktive Untersuchung lebensweltlicher Phänomene
2. Reflektierte theoretische Offenheit zwecks Bildung neuer Theorien
3. Zirkularität und Flexibilität des Forschungsprozesses zwecks Annäherung an den Gegenstand
4. Forschung als Kommunikation und Kooperation zwischen Forschenden und Beforschten
5. Selbstreflexion der Subjektivität und Perspektivität der Forschenden

**Aufgabe 2.7:** Erläutern Sie das Prinzip der Zirkularität und Flexibilität des methodischen Vorgehens im qualitativen Paradigma! Wo stößt es an Grenzen?

**Aufgabe 2.7:** Qualitative Forschung zielt oft darauf ab, den jeweiligen Forschungsgegenstand möglichst genau aus Sicht der Beteiligten zu rekonstruieren, also ihre Sicht- und Handlungsweisen zu verstehen. Der Prozess des Fremdverstehens wird als schrittweise Annäherung konzeptualisiert, am Anfang bestehende Wissenslücken, Missverständnisse oder Vorurteile der Forschenden werden im Zuge des Forschungsprozesses idealerweise durch wachsendes Verständnis ersetzt. Dies soll durch reflektierte theoretische Offenheit, vor allem aber auch durch methodische Flexibilität und Zirkularität erreicht werden: Neue Erkenntnisse sollen im Verlauf eines Forschungsprojektes direkt das weitere methodische Vorgehen verbessern (z. B. zielgerichtete Suche nach besonders aussagekräftigen Fällen, die noch in die Stichprobe aufzunehmen sind; Modifikation des Interview-Leitfadens; Überarbeitung der Codes für die Datenanalyse), sodass man sich spiralförmig in mehreren Zyklen dem Gegenstand annähert. Das Prinzip der Flexibilität stößt an Grenzen, wenn eine genaue Vorplanung des Forschungsprojektes erforderlich ist (z. B. präzise Zeit-, Finanz-, Personalplanung zur Beantragung von Drittmittel-Projekten oder fristgerechten Durchführung von Qualifikationsarbeiten).

**Aufgabe 2.8:** Was trifft nicht zu?

- a) Im DN-Modell entspricht die abhängige Variable dem Explanandum.
- b) Durch Exhaustion steigt der Bewährungsgrad und sinkt der Informationsgehalt einer Theorie.
- c) Während im quantitativen Paradigma die theoretisch interessierenden Merkmale der Erfahrungswirklichkeit vor der Datenerhebung präzise zu definieren und zu operationalisieren sind, werden im qualitativen Paradigma die für das Verstehen der interessierenden Phänomene notwendigen theoretischen Konzepte erst im Zuge der Datenerhebung und Datenanalyse ausgearbeitet.
- d) In der qualitativen Forschung werden die subjektiven Sichtweisen der untersuchten Personen durch Kommunikation ergründet, in der quantitativen Forschung wird dagegen nur das Verhalten objektiv erfasst.
- e) Im qualitativen wie im quantitativen Paradigma wird die Forschungstätigkeit als aktiver Konstruktionsprozess verstanden.
- f) Die qualitative Sozialforschung konzentriert sich auf die Untersuchung lebensweltlicher Phänomene. Die quantitative Sozialforschung untersucht diese ebenfalls und analysiert darüber hinaus auch Sachverhalte auf der Mikroebene (z. B. Hirnaktivitäten auf der Basis physiologischer Messungen) sowie auf der Makroebene (z. B. Unterschiede in der Lebenserwartung unterschiedlicher Bevölkerungsgruppen im historischen Wandel oder interkulturellen Vergleich auf der Basis bevölkerungsstatistischer Erhebungen).

**Aufgabe 2.8:** Lösung: d

**Aufgabe 2.9:** Welche Positionen werden hinsichtlich der wissenschaftstheoretischen Vereinbarkeit quantitativer und qualitativer Forschungsstrategien zu einer Mixed-Methods-Forschungsstrategie vertreten?

**Aufgabe 2.9:**

1. Inkommensurabilitäts-These
2. Aparadigmatischer Standpunkt
3. Komplementaritäts-These
4. Dialektischer Standpunkt
5. Standpunkt eines neuen Mixed-Methods-Paradigmas

**Aufgabe 2.10:** Welche Position(en) vertritt die philosophisch-pragmatische Wissenschaftstheorie?

- a) Anwendungswissen ist immer interessenbezogen, Grundlagenwissen dagegen nicht.
- b) Forschungsaktivitäten in den Sozialwissenschaften sollten sich gemäß pragmatischem Wissenschaftsverständnis bevorzugt der Lösung sozialer Probleme widmen, die von der Allgemeinheit als vordringlich angesehen werden.
- c) Gemäß transaktionalem Mensch-Umwelt-Verhältnis geht der Pragmatismus davon aus, dass sich das Erleben und Verhalten von Menschen am besten durch das Zusammenspiel objektiver sozialer Gesetzmäßigkeiten und subjektiver Sinnkonstruktionen beschreiben, erklären und vorhersagen lässt.
- d) Wissen im Allgemeinen und wissenschaftliche Erkenntnisse im Besonderen sind als Hilfsmittel zur praktischen Problemlösung zu betrachten.
- e) Wenn eine Theorie nützliche Beiträge zur Lösung praktischer Probleme liefert, ist sie als gültig anzuerkennen.

**Aufgabe 2.10:** Richtig sind: b, c, d, e

**Kapitel 3: Qualitätskriterien in der empirischen Sozialforschung**

---

**Aufgabe 3.1:** Grenzen Sie verschiedene Formen der Pseudowissenschaft voneinander ab.

**Aufgabe 3.1:**

- a) Unabsichtliche Pseudowissenschaft
- b) Vorsätzliche Pseudowissenschaft
- c) Ideologische Pseudowissenschaft

**Aufgabe 3.2:** Nennen Sie vier Standards der Wissenschaftlichkeit und die zugehörigen Kriterien wissenschaftlicher Qualität!

**Aufgabe 3.2:**

1. Wissenschaftliches Forschungsproblem und inhaltliche Relevanz
2. Wissenschaftlicher Forschungsprozess und methodische Strenge
3. Wissenschafts- und Forschungsethik und ethische Strenge
4. Dokumentation des Forschungsprojekts und Präsentationsqualität

**Aufgabe 3.3:** Welche Aussagen stimmen?

- a) Eine zu geringe Stichprobengröße gefährdet die statistische Validität einer Studie.
- b) Durch eine repräsentative Stichprobenauswahl wird die interne Validität einer Studie gesteigert.
- c) Forschungsethische Qualitätskriterien sind vor allem bei der Datenerhebung und Ergebnispublikation zu beachten.
- d) In den meisten Phasen des empirisch-quantitativen Forschungsprozesses ist die methodische Strenge das zentrale Qualitätskriterium.

e) Im quantitativen und qualitativen Paradigma werden hinsichtlich methodischer Strenge dieselben Gütekriterien angelegt.

**Aufgabe 3.3:** Richtig sind: a, c und d.

**Aufgabe 3.4:** Welche Formen der Validität werden in der quantitativen Forschung in der Campbell-Tradition unterschieden und auf welche Phasen des Forschungsprozesses bzw. auf welche Aspekte einer Studie beziehen sie sich hauptsächlich?

**Aufgabe 3.4:**

1. Konstruktvalidität (Qualität der Theoriearbeit sowie der Operationalisierung)
2. Interne Validität (Qualität des Untersuchungsdesigns und dessen Umsetzung)
3. Externe Validität (Qualität des Untersuchungsdesigns und der Stichprobenziehung)
4. Statistische Validität (Qualität der statistischen Datenanalyse sowie zur Sicherung der Teststärke auch die Qualität des Untersuchungsdesigns, der Operationalisierung und der Stichprobenziehung)

**Aufgabe 3.5:** Welche Typen der Validität einer Studie werden durch geringe Reliabilität von Messinstrumenten gefährdet?

**Aufgabe 3.5:** Messinstrumente, die fehlerbehaftete Messwerte erzeugen (geringe Reliabilität), beschränken die Konstruktvalidität, die interne, die externe sowie die statistische Validität.

**Aufgabe 3.6:** Welche Aussagen sind falsch?

- a) Randomisierte Kontrollgruppenstudien garantieren maximale externe Validität.
- b) Testübung bedroht vor allem die interne Validität.
- c) Zu kleine Stichproben bedrohen die statistische Validität.
- d) Die Konzeptspezifikation ist die theoretische Basis der Konstruktvalidität.
- e) Interne Validität ist die Voraussetzung für externe Validität.
- f) Statistische Validität ist die Voraussetzung für interne Validität.

**Aufgabe 3.6:** Falsch sind: a und e

**Aufgabe 3.7:** Welche Gütekriterien müssen qualitative Studien gemäß Lincoln und Guba (1985) erfüllen und was sind die jeweiligen Pendants der Kriterien im quantitativen Ansatz?

**Aufgabe 3.7:** Vier Kriterien der Glaubwürdigkeit:

1. Vertrauenswürdigkeit (interne Validität)
2. Übertragbarkeit (externe Validität)
3. Zuverlässigkeit (Reliabilität)
4. Bestätigbarkeit (Objektivität)

**Aufgabe 3.8:** Was ist gemäß den Kernkriterien zur Beurteilung qualitativer Forschung nach Steinke (1999) unter den Kriterien „Indikation“ und „Limitation“ zu verstehen?

**Aufgabe 3.8:** Eine qualitative Studie erfüllt das Kriterium der Indikation, wenn mit Blick auf das konkrete Forschungsproblem sowohl die Wahl eines qualitativen Forschungsansatzes als auch die einzelnen Methodenentscheidungen gut begründet sind. Die Studie erfüllt das Kriterium der Limitation, wenn die Verallgemeinerbarkeit der Ergebnisse bzw. die Grenzen der Verallgemeinerbarkeit ausdrücklich diskutiert und ggf. auch untersucht werden. Zumindest sollten die

Merkmale der untersuchten Personen und Kontexte so genau beschrieben werden, dass die Übertragbarkeit auf konkrete andere Personen und Kontexte einschätzbar ist.

**Aufgabe 3.9:** Was stimmt?

- a) Aufgrund des wissenschaftstheoretischen Hintergrundes qualitativer Studien werden intersubjektiv nachvollziehbare Gütekriterien generell abgelehnt.
- b) Es existieren nur sehr wenige Kriterienkataloge zur Beurteilung qualitativer Studien.
- c) Nur wenn jedes einzelne Kriterium aus einem Kriterienkatalog qualitativer Gütekriterien erfüllt ist, handelt es sich um eine hochwertige qualitative Studie.
- d) Aufgrund des Prinzips der Offenheit spielt methodische Strenge als Qualitätskriterium in der qualitativen Forschung eine untergeordnete Rolle.
- e) Das Kriterium der intersubjektiven Nachvollziehbarkeit gilt nur im qualitativen Forschungsansatz, nicht im quantitativen.
- f) Das Kriterium der reflektierten Subjektivität ist spezifisch für qualitative Forschung.

**Aufgabe 3.9:** Zutreffend ist f.

**Aufgabe 3.10:** Was versteht man im Kontext der Mixed-Methods-Forschung unter Inferenzqualität und in welche Unterkriterien teilt sie sich auf?

**Aufgabe 3.10:** Die Inferenzqualität bezeichnet den Grad der Tragfähigkeit der – ggf. kausal interpretierbaren – Ergebnisse einer Mixed-Methods-Studie. Sie entspricht dem Gütekriterium der internen Validität in der quantitativen Forschung bzw. dem Kriterium der Vertrauenswürdigkeit in der qualitativen Forschung (gemäß dem Ansatz von Guba & Lincoln, 1985). Hohe Inferenzqualität erfordert hohe Mixed-Methods-Designqualität (d. h. angemessene Verknüpfung qualitativer und quantitativer Teilstudien/Forschungsphasen im Mixed-Methods-Design) sowie hohe Mixed-Methods-Interpretationsqualität (überzeugende integrative Interpretation qualitativer und quantitativer Einzelergebnisse).

## Kapitel 4 - Forschungs- und Wissenschaftsethik

---

**Aufgabe 4.1:** Wie lassen sich Forschungsethik und Wissenschaftsethik voneinander abgrenzen?

Aufgabe 4.1: Die Forschungsethik konzentriert sich auf den Forschungsprozess im engeren Sinne, insbesondere den ethisch verantwortungsvollen Umgang der Forschenden mit den Untersuchungsteilnehmenden und ihren Daten. Demgegenüber befasst sich die Wissenschaftsethik mit ihren Regeln guter wissenschaftlicher Praxis vor allem mit dem redlichen Zustandekommen von wissenschaftlichen Befunden, mit ihrer Interpretation, Darstellung und Publikation.

**Aufgabe 4.2:** Welche Formen der Täuschung von Untersuchungspersonen im Rahmen von empirischen Studien werden differenziert und unter welchen Umständen sind sie ethisch legitim?

**Aufgabe 4.2:**

1. Man unterscheidet bei Täuschungsstudien die aktive und passive Täuschung der Untersuchungspersonen.



2. Täuschungsstudien sind nur ethisch vertretbar, wenn sie einen hohen Nutzen im Sinne wichtigen Erkenntnisgewinns haben, geringen Schaden im Sinne von Beeinträchtigung der Untersuchungsteilnehmenden verursachen und wenn methodische Alternativlosigkeit vorliegt.

**Aufgabe 4.3:** Was versteht man unter informierter Einwilligung und wann ist sie notwendig?

**Aufgabe 4.3:**

1. Informierte Einwilligung in der Forschung meint die gründliche Aufklärung der Untersuchungsteilnehmenden über die Studie und auf dieser Basis deren ausdrückliche mündliche oder schriftliche Einwilligung zur Teilnahme.
2. Informierte Einwilligung ist notwendig bei wissenschaftlichen Studien, die mehr als minimale Risiken für die Teilnehmenden bergen.

**Aufgabe 4.4:** Welche Beiträge zu einer wissenschaftlichen Publikation qualifizieren gemäß guter wissenschaftlicher Praxis in der Regel für die Mitautorschaft?

- a. Die extrem zeitaufwendige Transkription aller Interviews,
- b. die umfassende Analyse der Daten samt Ergebnisinterpretation,
- c. die Akquise der Drittmittel für das betreffende Forschungsprojekt und/oder
- d. die Idee zu Studienthema und Untersuchungsdesign sowie die Überarbeitung des Manuskripts.

**Aufgabe 4.4:** b, d

**Aufgabe 4.5:** Welche Art von Plagiaten lässt sich durch Software automatisch entdecken?

**Aufgabe 4.5:** Mit Plagiatserkennungs-Software lassen sich am ehesten direkte Copy-Paste-Plagiate aus gut zugänglichen, digitalen Quellen aufdecken.

**Aufgabe 4.6:** Welche zwei Varianten der Wissenschaftsfälschung werden unterschieden?

**Aufgabe 4.6:**

1. Wissenschaftsfälschung im Sinne einer Manipulation empirischer Daten zur Erzeugung eines Wunschergebnisses: „data falsification“.
2. Wissenschaftsfälschung im Sinne einer Erfindung von Daten: „data fabrication“.

**Aufgabe 4.7:** Im Rahmen einer empirischen Studie werden Fokusgruppen-Diskussionen durchgeführt und auf Video aufgezeichnet. Welches Risiko birgt diese Studie für die Teilnehmenden und warum:

- a. kein Risiko,
- b. minimales Risiko oder
- c. größer als minimales Risiko.

Was folgt aus der Risikobewertung für den ethischen Umgang mit den Fokusgruppen-Teilnehmenden?

**Aufgabe 4.7:** Die Teilnahme an einer Fokusgruppen-Diskussion birgt ein „größer als minimales Risiko“ (richtig: c), da Fokusgruppen aufgezeichnet werden und es sich bei den Audio-/Videoaufzeichnungen um identifizierbares Datenmaterial handelt. Gefordert ist deswegen informierte Einwilligung der Teilnehmenden: Sie müssen über die Zielsetzung der Studie und den Umgang mit den Aufzeichnungen detailliert aufgeklärt werden und sich mit diesen Konditionen ausdrücklich – in der Regel schriftlich – einverstanden erklären.

**Aufgabe 4.8:** Welche Ethik-Richtlinien sind für die psychologische Forschung a) im deutschsprachigen Raum und b) international besonders relevant?

**Aufgabe 4.8:**

- a) Im deutschsprachigen Raum sind für die Psychologie die Ethikrichtlinien von DGPs und SGP relevant.
- b) Im internationalen Raum sind für die Psychologie die Ethikrichtlinien von APA und IUPsyS relevant.

**Aufgabe 4.9:** Nennen Sie drei Datenerhebungsmethoden, die als risikolose Forschung einzustufen sind und somit in der Regel keine formale informierte Einwilligung erfordern.

**Aufgabe 4.9:** Risikolose Datenerhebungsmethoden, die in der Regel keine informierte Einwilligung verlangen:

1. Beobachtung an öffentlichen Orten
2. Dokumentenanalyse öffentlicher Dokumente
3. Anonyme schriftliche Befragung

**Aufgabe 4.10:** Im Rahmen eines Experimentes soll der Einfluss von Stimmungen auf die Gedächtnisleistung untersucht werden. Die Probanden werden zufällig in drei Gruppen eingeteilt, die sich unter dem Vorwand der Überbrückung von Wartezeit jeweils entweder einen lustigen, neutralen oder traurigen Filmclip anschauen, um positive, neutrale oder negative Stimmung zu erzeugen. Wie ist mit dieser Situation aus forschungsethischer Sicht umzugehen?

- a. Die Probanden müssen vorher ihr Einverständnis zu der Stimmungsbeeinflussung erklären.
- b. Da Stimmungsänderungen im Alltag normal sind, handelt es sich um risikolose Forschung, bei der gar keine besonderen ethischen Vorkehrungen zu treffen sind.
- c. Das experimentelle Setting beinhaltet aktive Täuschung, sodass die Probanden am Ende des Versuchs über die wahre Zielsetzung der Studie aufzuklären sind.
- d. Den Probanden der neutralen und negativen Stimmungsbedingung sollte am Ende des Versuchs ebenfalls der lustige Filmclip gezeigt werden, damit sie in positiver Stimmung entlassen werden.

**Aufgabe 4.10:** c, d

## Kapitel 5: Forschungsthema

---

**Aufgabe 5.1:** Grenzen Sie Forschungsthema und Forschungsproblem voneinander ab!

**Aufgabe 5.1:** Das Forschungsthema beinhaltet die Nennung eines bestimmten Untersuchungsgegenstandes oder Sachverhaltes. Das allgemeine Forschungsthema wird durch Konkretisierung in ein spezifisches Forschungsproblem überführt. Die inhaltliche Eingrenzung des Gegenstandes, Bezüge zu früheren empirischen Studien, die Wahl von Theorie(n) und Methode(n) ermöglichen die Zuspitzung des Forschungsproblems.

**Aufgabe 5.2:** Wann entscheidet man sich bei einer Studie für Forschungshypothesen und wann für Forschungsfragen?

**Aufgabe 5.2:** Forschungshypothesen, welche die Existenz, Richtung und Stärke von Effekten postulieren, sind nur zu formulieren, wenn sie sich eindeutig aus etablierten Theorien und/oder gesicherten empirischen Befunden ableiten lassen. Sie werden in der Regel mithilfe quantitativer

Studien einer statistischen Überprüfung unterzogen. Forschungsfragen werden in Form von offenen W-Fragen aufgestellt, wenn gesicherte Vorkenntnisse fehlen und neue Hypothesen bzw. Theorien generiert werden sollen. Meist werden Forschungsfragen durch qualitative oder explorative quantitative Studien beantwortet.

**Aufgabe 5.3:** Welche drei empirischen Studientypen lassen sich gemäß Erkenntnisinteresse unterscheiden?

**Aufgabe 5.3:**

- Explorative (gegenstandserkundende) Studien untersuchen ein neues oder ungenügend erforschtes Themenfeld mithilfe qualitativer oder auch quantitativer Methoden. Zielsetzungen sind die Gegenstandsbeschreibung sowie die Hypothesen- bzw. Theoriebildung.

- Explanative (hypothesenprüfende) Studien streben die empirische Prüfung von Hypothesen und Weiterentwicklung von Theorien mit quantitativen Methoden an.

Deskriptive (populationsbeschreibende) Studien widmen sich der Verbreitung und Ausprägung einzelner Merkmale in Grundgesamtheiten anhand repräsentativer Stichproben.

**Aufgabe 5.4:** Welche Vor- und Nachteile hat es, ein Thema wissenschaftlich zu untersuchen, zu dem man einen starken persönlichen Bezug hat?

**Aufgabe 5.4:** Die Insiderperspektive kann Zugang zum Untersuchungsfeld erschließen und – z. B. durch persönliche Kontakte zu Szenemitgliedern – zu Datenerhebungen und Interpretationen verhelfen, die Außenstehenden nicht möglich sind. Allerdings kann durch persönliche Betroffenheit auch kritische Distanz zum Forschungsthema verloren gehen, was sich z. B. in mangelnder Reflexion subjektiver Überzeugungen oder zu starker emotionaler Involvierung niederschlagen kann.

**Aufgabe 5.5:** Welche Inhalte sollte ein Exposé für eine wissenschaftliche Qualifikationsarbeit enthalten?

**Aufgabe 5.5:** Das Exposé benennt das Forschungsthema sowie das Forschungsproblem und ggf. erste konkrete Forschungshypothesen bzw. Forschungsfragen. Es führt in groben Zügen den Forschungsstand bzw. den theoretischen Hintergrund aus. Zudem sollte es das geplante methodische Vorgehen (Untersuchungsdesign, Stichprobe, Datenerhebungsmethode/n, Datenauswertungsmethode/n, Forschungsethik) beschreiben und damit auch Umfang und Anspruch der Arbeit abstecken. Ein vorläufiges Literaturverzeichnis mit deutsch- und englischsprachigen Quellen, ggf. ein Entwurf des Inhaltsverzeichnisses der Arbeit sowie ein wochengenaue Zeitplan sind beizufügen.

**Aufgabe 5.6:** Was ist gemeint, wenn in einer Publikation zu lesen ist, dass eine bestimmte Methodenentscheidung aus „forschungsökonomischen“ oder „forschungspragmatischen“ Gründen getroffen wurde? Bitte erläutern Sie den Fachbegriff und geben Sie ein realistisches Beispiel!

**Aufgabe 5.6:** Forschungsökonomische bzw. forschungspragmatische Gründe beziehen sich auf den Aufwand, der mit einer empirischen Studie verbunden ist. Es kann sich um zeitlichen, finanziellen und/oder personellen Aufwand handeln. Der Aufwand muss sich dabei an den vorhandenen Ressourcen orientieren und auch in vernünftigem Verhältnis zum Erkenntnisgewinn und seinem möglichen Nutzen stehen. Aus forschungsökonomischen Gründen kann man sich z. B. dafür entscheiden, eine einfache Gelegenheitsstichprobe anstelle einer sehr aufwendigen echten Zufallsstichprobe einzusetzen.

**Aufgabe 5.7:** Welche sieben Probleme können die empirische Untersuchbarkeit eines Forschungsthemas einschränken? Bitte die Probleme nennen und kurz erläutern!

**Aufgabe 5.7:** Folgende sieben Probleme können die empirische Untersuchbarkeit eines Forschungsthemas einschränken:

1. Ethische Grenzen: Die Forschungsethik verbietet alle Vorgehensweisen, durch die es für Untersuchungsteilnehmende zu Beeinträchtigungen kommen kann. Viele interessante Forschungsprobleme sind deswegen nicht oder nur eingeschränkt untersuchbar.
2. Politisch brisante Themen: Insbesondere anwendungswissenschaftliche Studien, Auftrags- und Evaluationsforschung können für die jeweiligen Anspruchsgruppen (Stakeholder) große politische Brisanz besitzen und Einmischung, Protest etc. heraufbeschwören. Wissenschaftliche Studien mit Praxispartnern erfordern dementsprechend besonderes Geschick und Verantwortungsbewusstsein hinsichtlich der Integration der Ansprüche der Stakeholder, ohne dass dabei die Wissenschaftlichkeit geopfert wird. In manchen Fällen kann der politische Druck, bestimmte Resultate zu liefern, so stark sein, dass man besser auf eine Studie verzichten sollte.
3. Aufwand für die Untersuchungsteilnehmenden: Methoden der Datenerhebung unterscheiden sich sehr stark im Zeitaufwand für die Untersuchungsteilnehmenden. Welcher Aufwand noch als akzeptabel empfunden wird und ab wann es verstärkt zu Teilnahmeverweigerung und Untersuchungsabbruch kommt, hängt stark von der jeweiligen Zielgruppe, vom Thema sowie von der Datenerhebungsmethode ab und sollte im Vorfeld eruiert werden.
4. Aufwand für die Forschenden: Methoden der Datenerhebung und Datenauswertung unterscheiden sich beträchtlich im Zeitaufwand für die Forschenden. Zur Umsetzung einer Studie ist eine vorhergehende Schätzung des Zeitaufwandes notwendig. Oft muss der Umfang der Studie reduziert werden, damit der Aufwand beherrschbar bleibt.
5. Schwer erreichbare Zielgruppen bzw. schwer zugängliche Materialien: Sollen schlecht erreichbare Zielgruppen oder schwer zugängliche Materialien untersucht werden, so ist mit großem zeitlichen und finanziellen Aufwand bei der Rekrutierung bzw. Beschaffung zu rechnen. Für den Fall der Erfolglosigkeit ist vorab eine Alternative zu planen.
6. Abhängigkeit von Dritten bei der Datenerhebung: Als problematisch kann es bewertet werden, wenn die Möglichkeit zur Datenerhebung von der Kooperationsbereitschaft Dritter abhängt. Erfahrungsgemäß braucht man daher Zeitpuffer und einen „Plan B“.
7. Verfügbarkeit technischer Hilfsmittel: Werden für die Studie spezielle Laborräume, Messgeräte etc. benötigt, müssen diese rechtzeitig reserviert werden. Auch müssen Funktionsfähigkeit und regelmäßige Wartung während des Untersuchungszeitraums gesichert sein.

**Aufgabe 5.8:** Was kennzeichnet die praktische (inklusive gesamtgesellschaftliche) Relevanz eines Forschungsthemas und welche Grenzen ergeben sich bei der Bearbeitung?

**Aufgabe 5.8:** Praktisch bedeutsam sind Forschungsfragen, die sich auf die Gestaltung von Maßnahmen, das Füllen von Entscheidungen oder die Lösung von Problemen beziehen. Entsprechende Forschungsfragen spielen in der Anwendungs- und Auftragsforschung sowie der Evaluationsforschung eine zentrale Rolle. Praxisorientierte Arbeiten entstehen in der Regel in Kooperation mit Praxispartnern, was den Forschenden die Möglichkeit gibt, durch ihre Ergebnisse zu greifbaren Veränderungen beizutragen. Die Kooperation mit Auftraggebern und Praxispartnern kann jedoch zeitaufwendig und konfliktgeladen sein. Gesamtgesellschaftlich relevante Studien sollen durch ihren Erkenntnisgewinn dazu beitragen, größere gesellschaftlicher Probleme zu lösen. Themen aufzugreifen, die aktuell die breite Öffentlichkeit bewegen, ist im Sinne der gesellschaftlichen Verantwortung empirischer Sozialforschung sinnvoll und wünschenswert. Allerdings können diese Themen manchmal

schnell „aus der Mode“ kommen. Zudem kann der selbstgestellte Anspruch, mit einer eigenen Studie zur Verbesserung der Gesellschaft beizutragen, überfordern.

**Aufgabe 5.9:** „Einzelkinder sind sehr viel egoistischer als Geschwisterkinder.“ Charakterisieren Sie den Typ dieser Forschungshypothese mit einschlägigen Fachbegriffen und formulieren Sie die zum Hypothesenpaar gehörende Nullhypothese.

**Aufgabe 5.9:** Es handelt sich bei der Forschungshypothese um eine gerichtete Unterschiedshypothese, die einen großen Effekt postuliert. Die entsprechende Nullhypothese lautet: Einzelkinder sind nicht sehr viel egoistischer als Geschwisterkinder.

**Aufgabe 5.10:** Welche Vor- und Nachteile hat es, ein bislang stark erforschtes oder sehr wenig erforschtes Thema zu bearbeiten?

**Aufgabe 5.10:** Bei einem stark erforschten Thema kann man sich einer etablierten Forschungstradition (hinsichtlich Theorien, Methoden) anschließen, allerdings muss dafür umfassende Literaturarbeit geleistet werden. Bei einem wenig erforschten Thema kann gegenstandsbeschreibend und theoriebildend Pionierarbeit geleistet werden. Forschende mit wenig Erfahrung können aber Gefahr laufen, sich mangels einschlägiger Vorläuferstudien zu „verrennen“.

## Kapitel 6: Forschungsstand und theoretischer Hintergrund

---

**Aufgabe 6.1:** Wie stellt man die Suchbegriffe für eine wissenschaftliche Literaturrecherche zusammen?

**Aufgabe 6.1:** Die Suchbegriffe lassen sich in primäre Suchbegriffe (Oberbegriffe, z. B. „Computerspiel“) sowie sekundäre Suchbegriffe (Unterbegriffe, z. B. „Ego-Shooter“) einteilen. Die Schlagwörter sollten mit ihrer englischen Übersetzung vertreten sein, um entsprechende Quellen einbeziehen zu können. Eine erste Liste kann vor der Recherche erstellt werden und wird im Verlauf der Suche verfeinert und ergänzt (u. a. mittels Thesaurus). Um Suchen einzugrenzen sind häufig auch Suchbegriffskombinationen notwendig (z. B. Geschlecht UND Computerspiel).

**Aufgabe 6.2:** Auf welche Quellen stützt sich die Literaturrecherche bei wissenschaftlichen Arbeiten hauptsächlich?

**Aufgabe 6.2:** Hauptsächlich wird bei wissenschaftlichen Arbeiten auf andere wissenschaftliche Quellen zugegriffen, in den Human- und Sozialwissenschaften vorrangig auf Artikel in Fachzeitschriften sowie Fachbücher. Diese sind zu großen Teilen mittels wissenschaftlicher Literaturdatenbanken verfügbar. Sekundärquellen wie z. B. Presseartikel, Sachbücher, Lexika oder Websites sollten nur im Ausnahmefall zur Stützung einer wissenschaftlichen Arbeit herangezogen werden.

**Aufgabe 6.3:** Wie funktioniert eine Suche nach wissenschaftlicher Literatur gemäß dem Schneeballsystem?

**Aufgabe 6.3:** Beim Schneeballverfahren sichtet man die Literaturverzeichnisse aktueller Zeitschriften- oder Fachbuchartikel und übernimmt dort aufgeführte Quellen in die eigene Literatursammlung. Beschafft man die so gefundenen Artikel, so kann man wiederum in deren Literaturverzeichnissen nach weiteren Quellen Ausschau halten.

**Aufgabe 6.4:** Welche Typen von wissenschaftlichen Publikationen sind für die Literaturrecherche die nützlichsten?

**Aufgabe 6.4:** Die nützlichsten Quellen für die wissenschaftliche Literaturrecherche und -beschaffung sind folgende:

1. Übersichtsartikel (Research Reviews, Systematic Reviews, Scoping Reviews), die den Forschungsstand zum Thema zusammenfassen
2. Metaanalysen, die vorliegende statistische Befunde zu einem bestimmten Effekt aus dem Themenbereich statistisch aggregieren
3. aktuelle Monografien und Sammelbände mit großer Nähe zur Forschungsfrage
4. aktuelle theoretische und/oder empirische Zeitschriftenbeiträge mit großer Nähe zur Forschungsfrage, die teilweise gebündelt in Sonderheften zu finden sind
5. Pionierstudien bzw. „Klassiker“, die wegbereitende Funktion in einem Forschungsfeld haben

**Aufgabe 6.5:** Welche Bezugswege gibt es für wissenschaftliche Volltexte?

**Aufgabe 6.5:**

1. Wissenschaftliche Literaturdatenbanken
2. Repositorien
3. Internet
4. Lokaler Bibliotheksbestand
5. Elektronische Zeitschriftenbibliothek (EZB)
6. Fernleihe
7. Dokumentenlieferdienste
8. Buchhandel
9. Autorinnen und Autoren

**Aufgabe 6.6:** Welche fünf Fragen sollten bei der Darstellung des Forschungsstandes beantwortet werden?

**Aufgabe 6.6:**

1. Wie hat sich die Forschung zum Thema in den letzten Jahren bzw. Jahrzehnten entwickelt?
2. Welche zentralen Forschungslinien und Befunde zum Thema sind auszumachen?
3. Welche Anwendungsfelder hat die Forschung zu diesem Themengebiet?
4. Wie ist der Forschungsstand zum Thema insgesamt zu beurteilen?
5. Wie ordnet sich die eigene Studie in den Forschungsstand ein?

**Aufgabe 6.7:** Wie können Theorien gemäß der Reichweite ihrer Erklärungen eingeordnet werden?

**Aufgabe 6.7:** Man unterscheidet drei Gruppen von Theorien gemäß der Reichweite ihrer Erklärungen:

1. Theorien geringer Reichweite erklären konkrete Sachverhalte in ihrem Kontext (gegenstandsnahe Theorien).
2. Theorien mittlerer Reichweite erklären eine umschriebene Menge von Sachverhalten über unterschiedliche Kontexte hinweg.
3. Groß- bzw. Metatheorien sind übergeordnete Theorien, die auf eine Vielzahl von Sachverhalten in unterschiedlichen Kontexten anwendbar sind.

**Aufgabe 6.8:** Welches Ziel verfolgt die grafische Darstellung von Theorien?

**Aufgabe 6.8:** Grafische Darstellungen von Theorien sollen dabei helfen, die Struktur ihrer Aussagen sichtbar und vergleichbar zu machen sowie Inkonsistenzen und Lücken aufzudecken.

**Aufgabe 6.9:** Welche Computersimulationen werden in der Theorieentwicklung unterschieden und welche Ziele verfolgen sie?

**Aufgabe 6.9:** Eine Computersimulation ist ein lauffähiges Programm, das die von einer Theorie postulierten Prozesse simuliert. Dabei werden quantitative und qualitative Computermodelle unterschieden. Quantitative Computermodelle basieren auf einem System mathematischer Gleichungen und haben das Ziel, für unterschiedliche Anfangssituationen die entsprechenden, theoriekonformen Konsequenzen in Form von Parameterschätzungen zu berechnen. Sie können zur Theorieprüfung und Prognose verwendet werden. Bei qualitativen Computermodellen geht es nicht um korrekte Parameterschätzungen, sondern darum, ob die von einer Theorie beschriebenen Phänomene oder Effekte überhaupt nachgestellt werden können.

**Aufgabe 6.10:** Aus welchen unterschiedlichen Ausgangspunkten lassen sich Forschungsfragen und Forschungshypothesen ableiten?

**Aufgabe 6.10:**

1. Ableitung von Forschungsfragen aus dem Forschungsstand.
2. Ableitung von Forschungsfragen und/oder Forschungshypothesen aus einer oder aus mehreren wissenschaftlichen Theorien.
3. Ableitung von Forschungsfragen und/oder Forschungshypothesen aus Alltagstheorien.

## Kapitel 7: Untersuchungsdesign

---

**Aufgabe 7.1:** Kennzeichnen Sie Labor- und Feldstudie hinsichtlich ihrer Validität!

**Aufgabe 7.1:** Die Laborstudie hat eine höhere interne und geringere externe Validität, die Feldstudie eine höhere externe und geringere interne Validität.

**Aufgabe 7.2:** Grenzen Sie nicht-experimentelle, quasi-experimentelle und experimentelle Designs voneinander ab!

**Aufgabe 7.2:** Bei experimentellen Designs werden mindestens eine Experimental- und eine Kontrollgruppe durch Randomisierung gebildet, unterschiedlich behandelt (experimentelle Manipulation) und hinsichtlich der abhängigen Variable(n) miteinander verglichen. Bei einem quasi-experimentellen Design werden vorgefundene bzw. nicht-randomisiert gebildete Gruppen unterschiedlich behandelt und miteinander verglichen. Bei einer nicht-experimentellen Studie werden vorgefundene Gruppen – ohne Behandlung – miteinander verglichen. Die interne Validität der experimentellen Studie ist am höchsten, die der nicht-experimentellen Studie am geringsten. Das experimentelle Design erlaubt die zuverlässigsten Aussagen über Kausalhypothesen.

**Aufgabe 7.3:** Kennzeichnen Sie Studientypen nach der Anzahl der Untersuchungsteilnehmenden!

**Aufgabe 7.3:**

1. Einzelfallstudien mit einer Untersuchungseinheit;
2. Gruppenstudien mit mehreren Untersuchungseinheiten existieren in den Varianten

- a) Stichprobenuntersuchung oder
- b) Vollerhebung.

**Aufgabe 7.4:** Was versteht man unter Replikationsstudien, wozu dienen sie und welche Varianten werden unterschieden?

**Aufgabe 7.4:** Replikationsstudien sind Wiederholungsstudien, in denen die Ergebnisse früherer Studien überprüft werden; entweder durch exakte Wiederholung des Ursprungsdesigns (direkte/exakte Replikationsstudie) oder durch dessen systematische Variation (systematische Replikationsstudie).

**Aufgabe 7.5:** Was charakterisiert angewandte Forschung und welche Varianten gibt es?

**Aufgabe 7.5:** Im Unterschied zur Grundlagenforschung, die primär auf wissenschaftlichen Erkenntnisgewinn abzielt, geht es bei der Anwendungsforschung um praxisbezogene Fragestellungen. Man unterscheidet dabei unabhängige Studien von finanzierten Auftragsstudien.

**Aufgabe 7.6:** Handelt es sich bei der Metaanalyse um eine theoretische Studie oder eine empirische Studie? Begründen Sie!

**Aufgabe 7.6:** Die Metaanalyse kann als theoretische Studie eingeordnet werden, weil sie als Gegenstand nicht Methoden (Methodenstudie) oder Datensätze (empirische Studie) untersucht, sondern sich auf Forschungsliteratur stützt. Sie kann aber auch als empirische Studie aufgefasst werden, weil sie Theorien, Methoden und Ergebnisse anderer Studien nicht zusammenfasst (theoretische Studie im Sinne eines narrativen Reviews), sondern die statistischen Befunde einschlägiger vorliegender Studien mit Methoden der statistischen Datenanalyse zu einem neuen empirischen Gesamtbefund zusammenfasst.

**Aufgabe 7.7:** Was versteht man unter einem Mixed-Methods-Design?

**Aufgabe 7.7:** Eine systematische Kombination quantitativer und qualitativer Designs.

**Aufgabe 7.8:** Worin unterscheiden sich Trendstudie und Panelstudie?

**Aufgabe 7.8:** Bei einer Trendstudie werden in zeitlichem Abstand unterschiedliche Stichproben mit demselben Instrument untersucht (wiederholte Querschnittstudien), dabei werden gesellschaftliche Trends deutlich. Bei Panelstudien wird in zeitlichem Abstand dieselbe Stichprobe mit demselben Instrument untersucht (Längsschnittstudie), dabei werden individuelle Veränderungen deutlich, allerdings sind diese mit Epochen- und Kohorteneffekten konfundiert.

**Aufgabe 7.9:** Worin unterscheiden sich explorative, deskriptive und explanative Studien?

**Aufgabe 7.9:** In ihrem Erkenntnisinteresse: Explorative Studien dienen der Gegenstandserkundung und Theoriebildung. Deskriptive Studien dienen der Populationsbeschreibung, explanative Studien der Hypothesen- und Theorieprüfung.

**Aufgabe 7.10:** Was versteht man unter Störvariablen und welche Typen von Störvariablen sind voneinander abzugrenzen?



**Aufgabe 7.10:** Im Kontext explanativer experimenteller und quasi-experimenteller Studien spricht man von Störvariablen oder konfundierenden Variablen als Variablen, die neben der/den eigentlich interessierenden unabhängigen Variablen ebenfalls die abhängige(n) Variable(n) beeinflussen. Störvariablen verhindern, dass die Effekte in den abhängigen Variablen eindeutig auf die unabhängigen Variablen zurückgeführt werden können. Man unterscheidet umwelt- bzw. untersuchungsbedingte und personengebundene Störvariablen.

## Kapitel 8: Operationalisierung

---

**Aufgabe 8.1:** Was versteht man unter Nominaldefinition und Nominalskala?

**Aufgabe 8.1:**

- Bei einer Nominaldefinition wird der zu definierende Begriff (Definiendum) durch einen oder mehrere bereits bekannte Begriffe (Definiens) festgelegt. Voraussetzung für Nominaldefinitionen ist ein Bestand an anerkannten Begriffen, die als erklärt gelten und somit zur Präzisierung des zu definierenden Konzeptes verwendbar sind.
- Eine Nominalskala bezeichnet das niedrigste Skalenniveau. Sie ordnet den Objekten eines empirischen Relativs Zahlen zu, die so geartet sind, dass Objekte mit gleicher Merkmalsausprägung gleiche Zahlen und Objekte mit verschiedener Merkmalsausprägung verschiedene Zahlen erhalten.

**Aufgabe 8.2:** Was sind eine Ratingskala, eine Rangskala und eine psychometrische Skala?

**Aufgabe 8.2:**

- Eine Ratingskala ist eine abgestufte Antwortskala (z.B. stimmt gar nicht – stimmt wenig – stimmt teils-teils – stimmt ziemlich – stimmt völlig).
- Eine Rangskala bzw. Ordinalskala bezeichnet das zweitniedrigste Skalenniveau; sie ordnet den Objekten eines empirischen Relativs Zahlen zu, die so geartet sind, dass Objekte mit stärkerer Merkmalsausprägung jeweils auch größere Messwerte erhalten.
- Eine psychometrische Skala ist ein standardisiertes Messinstrument, das aus mehreren Items besteht und ein latentes Merkmal misst, wobei die Testgütekriterien des Item-Satzes (Objektivität, Reliabilität, Validität, Dimensionalität, Skalierbarkeit etc.) bekannt und mindestens zufriedenstellend sind.

**Aufgabe 8.3:** Wie wird in der qualitativen und quantitativen Forschung jeweils zwischen theoretischen Konstrukten einerseits und empirischen Daten andererseits eine Verknüpfung hergestellt?

**Aufgabe 8.3:**

- In der qualitativen Forschung erfolgt die Bildung theoretischer Konstrukte primär induktiv („bottom-up“) aus Daten, die in nicht-strukturierter Weise erhoben wurden. Die theoretischen Konstrukte stellen Verdichtungen und Abstraktionen des qualitativen Datenmaterials dar.
- In der quantitativen Forschung wird deduktiv („top-down“) anhand theoretischer Überlegungen und früherer Studien eine Konzeptspezifikation sowie eine Operationalisierung (Auswahl bzw. Konstruktion des Messinstrumentes) für das interessierende theoretische Konzept erarbeitet. Mithilfe dieses Messinstrumentes werden dann quantitative Daten zu dem betreffenden theoretischen Konzept erhoben.

**Aufgabe 8.4:** Was ist bei der Messung soziodemografischer Variablen zu beachten?

**Aufgabe 8.4:** Konventionalität, Sparsamkeit, Diskriminierungsfreiheit, theoretische Reflexion

**Aufgabe 8.5:** Charakterisieren Sie folgende Ratingskala: „stimmt gar nicht – stimmt wenig – stimmt teils-teils – stimmt ziemlich – stimmt völlig“.

**Aufgabe 8.5:**

- Intensitätsskala
- Unipolar
- Fünf Stufen
- Ungerade Stufenzahl (Mittelkategorie vorhanden)
- Verbale Etiketten

**Aufgabe 8.6:** Was unterscheidet die Operationalisierung von „Depression“ mittels Einzelindikator und eindimensionaler psychometrischer Skala?

**Aufgabe 8.6:**

- Einzelindikator: Das Konstrukt Depression wird mit nur einer einzigen Frage/Aussage plus Antwortformat sehr ökonomisch, aber auch relativ ungenau erfasst.
- Eindimensionale psychometrische Skala: Das Konstrukt Depression wird mit einem Satz an inhaltsähnlichen Items erhoben, deren psychometrische Gütekriterien bekannt und ausreichend sind. Gegenüber dem Einzelindikator ist die Messgenauigkeit erhöht. Dafür steigen auch die Untersuchungsdauer und damit die Belastung für die Teilnehmenden sowie die Kosten für die Studie.

**Aufgabe 8.7:** Eine Likert-Skala zur Messung der Studienmotivation enthält 12 Items, die auf 5-stufigen Ratingskalen (1 = stimmt gar nicht bis 5 = stimmt völlig) zu beantworten sind. In welchem Wertebereich bewegen sich Summen- und Durchschnittsscore der Skala?

**Aufgabe 8.7:**

- Summenscore: 12–60
- Durchschnittsscore: 1–5

**Aufgabe 8.8:** Eine Untersuchungsperson hat auf der Thurstone-Skala zur Kontaktsituation in Wohnsiedlungen (Abschn. 8.6.3; Exkurs: Menschliche Kontakte in Siedlungen) die Items 1, 2 und 4 bejaht und die anderen Items verneint. Welchen Skalenwert erhält die Person und wie ist dieser inhaltlich zu interpretieren?

**Aufgabe 8.8:** Der Skalenwert der Person ergibt sich bei einer Thurston-Skala aus der Summe der Skalenwerte der bejahten Items:  $(-2.0) + (-3.05) + (-0.53) = -5.58$ , d. h. nachbarschaftliche Kontakte sind bei dieser Befragungsperson laut Selbstausskunft kaum vorhanden.

**Aufgabe 8.9:** Eine Untersuchungsperson hat auf der Guttman-Skala zur vorehelichen Permissivität (Abschn. 8.6.2) das Item a bejaht und die anderen Items verneint. Welchen Skalenwert erhält die Person und wie ist dieser inhaltlich zu interpretieren?

**Aufgabe 8.9:** Der Skalenwert der Person ergibt sich bei einer Guttman-Skala als Summe der bejahten Items: hier also Skalenwert 1 für ein bejahtes Items, d. h. die sexuelle Freizügigkeit ist bei der Befragungsperson sehr gering ausgeprägt.

**Aufgabe 8.10:** Auf welchem Skalenniveau sollten Daten erhoben werden? Was ist zu tun, wenn Daten in einem ungeeigneten Skalenniveau vorliegen?

**Aufgabe 8.10:**

- Die Datenerhebung erfolgt normalerweise auf dem höchstmöglichen Skalenniveau, sofern die Untersuchungsteilnehmenden dadurch nicht überfordert werden und die Information im Rahmen der Studie sinnvoll nutzbar ist.
- Daten auf zu geringem Skalenniveau können unter bestimmten Voraussetzungen mit speziellen statistischen Techniken auf ein höheres Skalenniveau transformiert werden (z. B. Ordinaldaten aus direkter Rangordnung lassen sich mittels „Law of Categorical Judgement“ und Ordinaldaten aus indirekter Rangordnung mittels „Law of Comparative Judgement“ jeweils in Intervalldaten überführen).
- Daten auf zu hohem Skalenniveau können immer auf jedes beliebige niedrigere Skalenniveau heruntertransformiert werden, was allerdings mit Informationsverlust verbunden ist.

**Kapitel 9: Stichprobenziehung**

---

**Aufgabe 9.1: Was versteht man unter einer Vollerhebung und wann ist sie der Stichprobenziehung vorzuziehen?**

**Aufgabe 9.1:** Unter einer Vollerhebung versteht man die Untersuchung aller Objekte der Zielpopulation. Vollerhebungen sind möglich, wenn es sich um überschaubare und gut erreichbare Populationen handelt. Sie sind notwendig, wenn eine kleine Population sehr heterogen ist und Stichproben das Gesamtbild verfehlen würden.

**Aufgabe 9.2: Erklären Sie die Unterschiede zwischen Ziel-, Auswahl- und Interferenzpopulation!**

**Aufgabe 9.2:** Unter der Zielpopulation (angestrebte Grundgesamtheit) versteht man die Gesamtmenge aller Untersuchungseinheiten (z. B. Personen, Verhaltensweisen, Objekte), über die im Rahmen einer empirischen Studie Aussagen getroffen werden sollen. Die Auswahlpopulation meint dagegen die Gesamtheit aller Objekte, die sich im Auswahlrahmen befinden und somit prinzipiell eine Chance haben, in die Stichprobe zu gelangen. Angesichts von Stichprobenausfällen (z. B. Personen, die zur geplanten Stichprobe gehören, aber nicht erreichbar sind oder die Teilnahme verweigern) erlaubt eine konkrete Stichprobe meist keinen direkten Rückschluss auf die Auswahlpopulation, sondern nur auf die Inferenzpopulation, die Gesamtheit aller Objekte, aus der die Stichprobe tatsächlich stammt (also die Auswahlpopulation abzüglich der nicht-erreichbaren und nicht-teilnahmebereiten Personen).

**Aufgabe 9.3: In der Presse ist oft von „repräsentativen Studien“ die Rede. Warum ist diese Bezeichnung irreführend?**

**Aufgabe 9.3:** Repräsentativität ist ein Qualitätsmerkmal, das sich nur auf das Verhältnis zwischen Stichprobe und Population bezieht, nicht auf eine gesamte Studie. Eine quantitative Studie, die mit einer repräsentativen Stichprobe operiert, kann dennoch wissenschaftlich wertlos sein, wenn z. B. beim Untersuchungsdesign, bei der Operationalisierung oder bei der Datenanalyse Fehler gemacht werden, welche die Validität der Studie gefährden.

**Aufgabe 9.4: Was versteht man unter einer „Theoretischen Stichprobe“ und wann wird sie eingesetzt?**

**Aufgabe 9.4:** Die Theoretische Stichprobe („theoretical sample“) ist eine Auswahlmethode aus der Grounded-Theory-Methodologie, die in der qualitativen Forschung bei theoriebildenden Studien zum

Einsatz kommt. Die Fallauswahl wird ganz bewusst so gesteuert, dass ein maximaler theoretischer Erkenntniswert resultiert. Dabei erfolgt die Fallauswahl schrittweise im Zuge der Datenerhebung und Datenauswertung: Anhand der bisherigen Ergebnisse der Studie wird jeweils entschieden, welche **weiteren** Fälle (Personen, Gruppen, Institutionen, Ereignisse) **als nächstes** in das Sample aufzunehmen sind. Stichprobenauswahl, Datenerhebung und Datenanalyse werden nicht linear, sondern mehrfach zirkulär abgearbeitet. Es können verschiedene und wechselnde Kriterien zur bewussten Auswahl von Fällen herangezogen werden. Wird durch weitere Fälle kein neuer Informationsgehalt für die Theoriebildung generiert, so ist die Stichprobenbildung abgeschlossen (theoretische Sättigung).

#### **Aufgabe 9.5:** Wie ist die „einfache Zufallsstichprobe“ definiert?

**Aufgabe 9.5:** Man zieht eine einfache Zufallsstichprobe („simple random sample“), indem man aus einer möglichst vollständigen Liste aller  $N$  Objekte der Zielpopulation (Auswahlrahmen) nach einem „blinden“ statistischen Zufallsprinzip eine Anzahl  $n$  von Objekten auswählt, wobei die Auswahlwahrscheinlichkeiten aller Objekte gleich groß sein müssen.

#### **Aufgabe 9.6:** Ist die Gelegenheitsstichprobe eine Zufallsstichprobe? Begründen Sie Ihre Antwort!

**Aufgabe 9.6:** Gelegenheitsstichproben werden oft fälschlich als „Zufallsstichproben“ bezeichnet. Die Gelegenheitsstichprobe basiert aber auf einer willkürlichen Auswahl von Fällen, die gerade verfügbar sind. Die Gelegenheitsstichprobe weist die niedrigste, die Zufallsstichprobe dagegen die höchste Stichprobenqualität im Sinne globaler Repräsentativität auf.

#### **Aufgabe 9.7:** Was versteht man unter „probabilistischen Stichproben“, wozu werden sie eingesetzt und welche Arten von probabilistischen Stichproben sind zu unterscheiden?

**Aufgabe 9.7:** Probabilistische Stichproben basieren auf einem Auswahlrahmen, d. h. einer möglichst vollständigen Liste aller Objekte der Zielpopulation, und einem darauf angewendeten statistischen Zufallsverfahren. Sie sind aufwendig und teuer, aber für populationsbeschreibende quantitative Untersuchungen unerlässlich, in denen es auf globale Repräsentativität der Stichprobe ankommt. Es werden vier Arten von probabilistischen Stichproben unterschieden: die einfache Zufallsstichprobe, die geschichtete Zufallsstichprobe, die Klumpenstichprobe und die mehrstufige Zufallsstichprobe.

#### **Aufgabe 9.8:** Worin unterscheidet sich die Klumpenstichprobe von der geschichteten Stichprobe?

**Aufgabe 9.8:** Bei einer Klumpenstichprobe wird eine Zufallsauswahl von natürlichen Gruppen (= Klumpen) von Untersuchungseinheiten gezogen (z. B. Schulklassen, Abteilungen), die vollständig untersucht werden. Bei einer geschichteten Stichprobe wird die Population gemäß den Ausprägungen mindestens einer Schichtungsvariable (z. B. Einkommen: unterdurchschnittliches, durchschnittliches, überdurchschnittliches Einkommen) in Teilpopulationen (= Schichten) eingeteilt, aus denen dann jeweils einfache Zufallsstichproben gezogen werden.

**Aufgabe 9.9:** In der Zeitung lesen Sie unter der Überschrift „Haschisch macht müde und faul“ folgende Meldung: „Wie eine amerikanische Repräsentativstudie zeigt, haben 70 % aller Haschischkonsumenten unterdurchschnittliche Schulleistungen. Gleichzeitig schlafen sie überdurchschnittlich lange. Diese Befunde belegen eindrücklich, wie gefährlich eine liberale Drogenpolitik ist.“ In dieser Nachricht sind 5 Fehler versteckt. Welche?

**Aufgabe 9.9:** Die Zeitungsmeldung enthält die folgenden Ungenauigkeiten und Fehler:

1. Vom korrelativen Zusammenhang zwischen Müdigkeit und Schulleistung einerseits und Haschischkonsum andererseits kann nicht auf kausale Wirkungsrichtungen geschlossen werden (vielleicht motiviert Müdigkeit zum Haschischkonsum, oder es spielen noch ganz andere Faktoren eine Rolle).
2. Es werden Aussagen über die Population aller Haschischkonsumenten gemacht. Diese Population ist unbekannt, sodass keine probabilistischen („repräsentativen“) Stichproben gezogen werden können. Statt von einer „Repräsentativstudie“ zu sprechen, wäre hier eine kurze Beschreibung des Auswahlverfahrens und der resultierenden Größe und Zusammensetzung der Stichprobe sinnvoller gewesen.
3. Auf der Basis von Stichprobenergebnissen lassen sich Populationsparameter nur unter Unsicherheit schätzen. Die „70 %“-Angabe müsste also in irgendeiner Weise relativiert werden (etwa durch Angabe des Konfidenzintervalls).
4. „Überdurchschnittliches“ Schlafen und „unterdurchschnittliche“ Leistungen – damit ist allein die Richtung von Unterschieden angesprochen. Die entscheidende Frage lautet nun, wie groß diese Effekte sind. „Überdurchschnittlich“ kann z. B. bedeuten „3 Minuten länger“ oder „4 Stunden länger“ – die Effektgröße sollte in der Darstellung konkretisiert werden. Aussagen über Effekte ohne Angabe der Effektgröße sind weitgehend informationslos.
5. Die praktische Schlussfolgerung, eine „liberale Drogenpolitik“ sei „gefährlich“, steht mit dem Thema der Untersuchung in keinerlei Zusammenhang. Die Studie behandelt Merkmalsausprägungen bei Haschischkonsumenten und kann weder sagen, ob die Merkmalsausprägungen vom Haschischkonsum verursacht werden, noch ob eine „liberale Drogenpolitik“ zum Haschischkonsum führt (wie implizit angedeutet wird).

**Aufgabe 9.10:** Nennen Sie Vor- und Nachteile probabilistischer Stichprobenverfahren.

**Aufgabe 9.10:** Probabilistische Stichprobenverfahren sind die beste Möglichkeit, um aus einer definierten Zielpopulation eine global repräsentative Stichprobe zu ziehen. Sie sind deswegen für populationsbeschreibende quantitative Studien meist unverzichtbar. Nachteilig ist, dass probabilistische Stichproben mit hohem Aufwand und hohen Kosten verbunden sind. Oft können sie auch gar nicht angewendet werden, weil für viele Zielpopulationen kein passender Auswahlrahmen (d. h. keine Liste aller oder fast aller Populationselemente) zur Verfügung steht.

## Kapitel 10.1: Datenerhebung - Beobachtung

---

**Aufgabe 10.1.1:** Nennen Sie fünf Vorzüge der Beobachtung gegenüber der Befragung mit jeweils einem Beispiel.

**Aufgabe 10.1.1:** Fünf Vorzüge der Beobachtung gegenüber der Befragung:

1. beobachtbar sind Personen, die nicht befragt werden können (z. B. Säuglinge);
2. beobachtbar sind Phänomene, über die Personen keine genauen Selbstauskünfte geben können (z. B. Details der eigenen Mimik);
3. beobachtbar sind Phänomene, über die Personen keine Auskunft geben wollen (z. B. normabweichende Verhaltensweisen);
4. beobachtbar sind Phänomene im kontinuierlichen Zeitverlauf, während die Befragung summarische Aussagen erhebt (z. B. Schülerverhalten im Laufe einer ganzen Unterrichtsstunde) und
5. beobachtet werden kann nonreaktiv, d. h. ohne direkten Eingriff in das beobachtete Geschehen, während die Befragung auf die aktive und bewusste Teilnahme der Befragten angewiesen ist und entsprechenden Verzerrungen unterliegt (z. B. Beobachtung des hinterlassenen Hausmülls).

**Aufgabe 10.1.2:** Was ist unter Perspektivität, Selektivität und Konstruiertheit der Beobachtung zu verstehen?

**Aufgabe 10.1.2:**

- Perspektivität der Beobachtung: Mit welchem Ziel, aus welcher Perspektive, von welchem Standpunkt aus wird beobachtet?
- Selektivität der Beobachtung: Was geht in die Beobachtung ein, und was geht nicht ein?
- Konstruktivität/Konstruiertheit der Beobachtung: Wie werden die Beobachtungen registriert, dokumentiert und interpretiert? Mit welchen Begriffen und Konzepten wird auf das beobachtete Geschehen Bezug genommen?

**Aufgabe 10.1.3:** In welchem Zusammenhang fällt bei der qualitativen und quantitativen Beobachtung jeweils besonders viel Arbeit an?

**Aufgabe 10.1.3:** Die Beobachtungsmethode ist generell aufwendiger als die Datenerhebung per Befragung:

- Besonderer Zeitaufwand bei der qualitativen Beobachtung: Dokumentation und qualitative Auswertung der nicht- oder teil-standardisierten Beobachtungsprotokolle.
- Besonderer Zeitaufwand bei der quantitativen Beobachtung: Entwicklung und Überprüfung des standardisierten Beobachtungsinstrumentes.

**Aufgabe 10.1.4:** Wie und wozu berechnet man in quantitativen Beobachtungsstudien die Beobachterübereinstimmung?

**Aufgabe 10.1.4:** Man ermittelt die Beobachterübereinstimmung einer Kategorie/Variable im Beobachtungssystem, indem mindestens 2 Beobachter:innen dieselbe Stichprobe von Untersuchungseinheiten beobachten und die Messwerte für die interessierende Kategorie/Variable erfassen. Dann wird (je nach Skalenniveau der Variable) mit einem passenden Übereinstimmungskoeffizienten der Grad der Übereinstimmung quantifiziert. Die Beobachterübereinstimmung gilt als Maß für die Reliabilität (Messgenauigkeit) des standardisierten Beobachtungsinstrumentes bzw. seiner einzelnen Kategorien/Variablen.

**Aufgabe 10.1.5:** Welche Varianten der Stichprobenziehung sind für quantitative Beobachtungsstudien relevant und wann werden sie eingesetzt?

**Aufgabe 10.1.5:** Bei einer quantitativen Beobachtung kann generell zwischen Zeit- und Ereignisstichprobe gewählt werden. Die Zeitstichprobe ist bei zeitlich ausgedehnten Verhaltensweisen (z. B. Flirtverhalten, Essrituale) indiziert. Dabei erfassen die Beobachter in regelmäßigen Abständen die aktuelle Ausprägung des Verhaltens. Die Ereignisstichprobe ist indiziert, wenn Einzelereignisse oder kurze Verhaltenssequenzen erfasst werden sollen (z. B. passierende Fahrzeuge).

**Aufgabe 10.1.6:** Welche Aussage/-n stimmt/stimmen?

- a) Eine ethnografische Feldbeobachtung ist immer teilnehmend.
- b) Ohne die Zustimmung des Gatekeepers darf eine ethnografische Feldbeobachtung nicht stattfinden.
- c) Während einer ethnografischen Feldbeobachtung können Notizen, Fotos und Videos erstellt werden.
- d) Eine ethnografische Feldbeobachtung ist verdeckt durchzuführen.

### **Aufgabe 10.1.6:** c

**Aufgabe 10.1.7:** Was ist eine strukturierte Verhaltensbeobachtung, wie und wo wird sie durchgeführt?

**Aufgabe 10.1.7:** Bei der strukturierten Verhaltensbeobachtung als einer Variante der quantitativen Beobachtung geht es darum, zahlreiche und komplexe Merkmale und Verhaltensweisen in ihrer Häufigkeit, Dauer und/oder Intensität zu messen. Es kommen komplexe und hinsichtlich ihrer Gütekriterien überprüfte standardisierte Beobachtungsinstrumente zum Einsatz. Strukturierte Verhaltensbeobachtungen finden sowohl im Feld als auch im Labor statt.

**Aufgabe 10.1.8:** Was ist eine nonreaktive Beobachtung, wie und wo kommt sie zum Einsatz?

**Aufgabe 10.1.8:** Die nonreaktive Beobachtung ist eine Spezialform der nonreaktiven Datenerhebung und bezieht sich auf die Erfassung von quantifizierbaren Verhaltensspuren mithilfe menschlicher Sinnesorgane und/oder technischer Sensoren zum Zeitpunkt (oder im Zeitraum) ihres Auftretens. Typischerweise findet sie im Feld statt und zwar entweder offline oder online.

**Aufgabe 10.1.9:** In einem Schnellrestaurant wird von einem Sitzplatz in der Nähe des Verkaufstresens aus beobachtet, wie oft und von wem das Aktionsangebot der Woche bestellt wird und wie oft das Personal ausdrücklich auf das Angebot hinweist. Was trifft auf diese Studie zu?

- a) Es handelt sich um eine ethnografische Feldbeobachtung.
- b) Es handelt sich um eine nicht-teilnehmende Beobachtung.
- c) Es handelt sich um eine strukturierte Verhaltensbeobachtung.
- d) Es handelt sich um eine Zeitstichprobe.
- e) Es handelt sich um eine quantitative Beobachtung mit geringem Komplexitätsgrad.

**Aufgabe 10.1.9:** b, e

**Aufgabe 10.1.10:** Was versteht man unter der Lost-Letter-Technik?

**Aufgabe 10.1.10:** Bei der Lost-Letter-Technik werden frankierte und vermeintlich an verschiedene Organisationen adressierte Briefe im öffentlichen Raumausgelegt, sodass sie den Eindruck erwecken, verloren worden zu sein. Der Rücklauf der Briefe zum Forschungsteam wird als Indikator für die Einstellung gegenüber den genannten Organisationen betrachtet.

## **Kapitel 10.2: Datenerhebung – Interview**

---

**Aufgabe 10.2.1:** Nennen Sie drei Techniken der unstrukturierten mündlichen Befragung!

**Aufgabe 10.2.1:**

1. Narratives Interview
2. Ethnografisches Feldinterview
3. Methode des lauten Denkens

**Aufgabe 10.2.2:** Was stimmt?

- a) Bei einer offenen Frage werden keine Antwortalternativen vorgegeben.
- b) Qualitative Interviews arbeiten fast ausschließlich mit offenen Fragen.
- c) In quantitativen Interviews werden vereinzelt auch offene Fragen eingesetzt.

**Aufgabe 10.2.2:** Alle richtig

**Aufgabe 10.2.3:** Grenzen Sie die Datenerhebung mittels Interview von der Fragebogenmethode ab und zwar hinsichtlich ihrer zentralen Merkmale sowie ihrer Vor- und Nachteile!

**Aufgabe 10.2.3:** Beim Interview findet ein interaktiver Kontakt zwischen Interviewer:in und Befragungsperson statt. Die Äußerungen erfolgen im Modus der Mündlichkeit. Demgegenüber wird ein Fragebogen selbstständig ausgefüllt und schriftlich beantwortet. Interviews sind zu bevorzugen, wenn man offene Fragen stellen möchte (diese werden eher mündlich beantwortet als schriftlich ausformuliert) und dabei individuell auf die Befragungsperson eingehen will (z. B. Rückfragen, Ergänzungsfragen) und/oder wenn man viele geschlossene Fragen vorlegen möchte (hier motiviert der Interviewer zur Fortsetzung der Befragung, während bei langen Fragebögen eher abgebrochen wird). Die Interview-Methode hat durch den Einsatz der Interviewenden den Nachteil deutlich größeren Aufwandes und höherer Kosten als das Austeilen und Einsammeln von Fragebögen. Zudem kann die Interaktion mit den Interviewenden bei heiklen Fragen, sozial ängstlichen Befragungspersonen etc. die Auskunftswilligkeit einschränken, sodass hier die schriftliche Befragung zu bevorzugen ist.

**Aufgabe 10.2.4:** Worin bestehen die besonderen Herausforderungen bei Experteninterviews?

**Aufgabe 10.2.4:** Herausforderungen des Experteninterviews:

- a) Definition der Experten
- b) Erreichbarkeit der Experten
- c) Festlegung der Rolle des Interviewers als Laie oder Co-Experte

**Aufgabe 10.2.5:** Was ist zu beachten, wenn ein standardisierter Interview-Fragebogen entwickelt wird, der sowohl in mündlichen als auch telefonischen Interviews zum Einsatz kommen soll?

**Aufgabe 10.2.5:** Herausforderungen einer Mixed-Mode-Interview-Studie, bei der derselbe Interview-Fragebogen mündlich und fernmündlich eingesetzt werden soll:

- a) Die Fragebogen-Länge muss für mündliche und fernmündliche Interviewformen adäquat sein (bei Telefon-Interviews Obergrenze von ca. 20 Minuten)
- b) Die Fragebogen-Items müssen bei mündlichen und fernmündlichen Interviewformen gleichermaßen anwendbar sein (Verzicht auf visuelle Vorlagen sowie auf zu viele verschiedene Antwortvorgaben pro Item, da diese telefonisch nicht funktionieren)

**Aufgabe 10.2.6:** Welche Unterschiede und Gemeinsamkeiten haben das fokussierte Interview und die Fokusgruppen-Diskussion?

**Aufgabe 10.2.6:** Gemeinsamkeiten und Unterschiede zwischen fokussiertem Interview und Fokusgruppen-Diskussion:

- a) Bei beiden Verfahren werden die Befragungspersonen mit konkreten Anschauungsobjekten konfrontiert (z. B. Filmausschnitte, Produktproben), zu denen sie Auskunft geben sollen
- b) Bei beiden Verfahren können mehrere Personen gleichzeitig befragt werden
- c) Beim fokussierten Interview stehen die Reaktionen auf das fokussierte Objekt (meist Medienangebote wie z. B. Filmausschnitte) im Zentrum, während bei Fokusgruppen-Diskussionen oft auch neue Gestaltungsvorschläge erörtert werden (z. B. Verbesserung von Produkten)
- d) Während das fokussierte Interview als Einzel- oder Gruppeninterview stattfinden kann, ist eine Fokusgruppen-Diskussion nur mit mehreren Personen möglich.



**Aufgabe 10.2.7:** Wie gehen Sie vor, um ein bevölkerungsrepräsentatives Sample von  $n = 500$  Personen Ihres Bundeslandes telefonisch zu befragen (Interview-Fragebogen liegt vor)? Erläutern Sie Stichprobenziehung, Rekrutierung und Voraussetzungen der Interview-Durchführung!

**Aufgabe 10.2.7:** Um eine bevölkerungsrepräsentative Stichprobe von  $n = 500$  Personen eines Bundeslandes für eine Interview-Studie zusammenzustellen, sind aufgrund der Struktur des Rufnummernsystems in Deutschland prinzipiell das Gabler-Häder-Design sowie das ADM-Design nutzbar. Da das ADM-Design nur ADM-Mitgliedern zur Verfügung steht, ist auf das Gabler-Häder-Design zurückzugreifen.

- a) Beschaffung einer repräsentativen Stichprobe von regionalen Telefonnummern von Privathaushalten gemäß Gabler-Häder-Design über die Abteilung SDM der GESIS (die angeforderte Zahl der Telefonnummern – die Brutto-Stichprobe – muss wegen der hohen Zahl der zu erwartenden Ausfälle ca. doppelt so groß sein wie die angezielte Netto-Stichprobe),
- b) Rekrutierung und Schulung der Interviewer,
- c) Kontaktierung aller Telefonnummern der Stichprobenliste über ein Telefonstudio,
- d) bei telefonisch erreichten Haushalten Zufallsauswahl der Befragungsperson durch die geschulten Interviewer anhand Schwedenschlüssel oder Last/Next-Birthday-Methode,
- e) erneute Kontaktversuche, wenn die Befragungsperson nicht direkt erreichbar ist,
- f) Durchführung der Interviews.

**Aufgabe 10.2.8:** Wie gehen Sie hinsichtlich Stichprobenkonstruktion vor, um in einer bestimmten Stadt eine bevölkerungsrepräsentative persönlich-mündliche Interviewstudie mit  $n = 300$  Personen durchzuführen?

**Aufgabe 10.2.8:** Um eine bevölkerungsrepräsentative Stichprobe von  $n = 300$  Personen einer Stadt für eine Interview-Studie zusammenzustellen, sind prinzipiell die Verfahren der Register-Stichprobe, Random-Route und das ADM-Mastersample nutzbar. Da das ADM-Mastersample nur von ADM-Mitgliedern genutzt und das Random-Route-Verfahren sehr aufwendig ist, empfiehlt sich die Register-Stichprobe: Beim zuständigen Einwohnermeldeamt wird eine Zufallsauswahl von Adressen von Privatpersonen zusammen mit Basisinformationen (Geschlecht, Nationalität, akademischer Titel etc.) für wissenschaftliche Zwecke angefordert. Bei der Festlegung der angeforderten Stichprobengröße (Brutto-Stichprobe) werden zu erwartende Ausfälle (nicht erreichte Personen, Teilnahmeverweigerung) berücksichtigt, d. h. die angeforderte Brutto-Stichprobe ist entsprechend größer zu wählen (z. B.  $n = 600$ ) als die angestrebte Netto-Stichprobe.

**Aufgabe 10.2.9:** Welche Fehler beeinträchtigen gemäß Total-Survey-Error-Modell die Aussagekraft von quantitativen Interviewstudien?

**Aufgabe 10.2.9:** Das Total-Survey-Error-Modell benennt fünf Fehlerquellen bei quantitativen Umfragen:

1. Coverage Error
2. Nonresponse Error
3. Sampling Error
4. Measurement Error
5. Postsurvey Error

**Aufgabe 10.2.10:** Wie kann man Befragungsfehler durch die Interviewenden verhindern bzw. verringern?

**Aufgabe 10.2.10:** Verringerung bzw. Verhinderung von Interviewfehlern, die durch die Interviewenden verursacht werden:

- a) Auswahl geeigneter Interviewer:innen
- b) Schulung der Interviewer:innen
- c) Kontrolle der Interviewer:innen

### **Kapitel 10.3: Datenerhebung – Fragebogen**

---

**Aufgabe 10.3.1:** Welche Fehler enthält dieses Item? „Intensive Computernutzung beeinträchtigt Kinder nicht pauschal in ihrer Intelligenzentwicklung, vielmehr kommt es auf die Förderung ihrer Interessen im Elternhaus an.“

stimmt gar nicht – stimmt wenig – stimmt teils-teils – stimmt ziemlich – stimmt völlig

**Aufgabe 10.3.1:** Das Item enthält drei Fehler:

- Lange und komplizierte Aussage
- Vermischung mehrerer Inhalte
- Doppelte Verneinung zwischen Aussage und Antwortskala

**Aufgabe 10.3.2:** Nennen Sie zwei Vor- und zwei Nachteile selbstadministrierter Fragebögen im Vergleich zu intervieweradministrierten Befragungen!

**Aufgabe 10.3.2:**

- Vorteile selbstadministrierter Fragebögen im Vergleich zum Interview: höherer Grad an Diskretion und höhere Effizienz
- Nachteile selbstadministrierter Fragebögen im Vergleich zum Interview: höhere Anforderungen an die Befragten (Lese-/Schreibkompetenzen, Konzentration, Motivation) und geringere Gestaltbarkeit/Transparenz der Befragungs- bzw. Ausfüllsituation für die Forschenden

**Aufgabe 10.3.3:** Nennen Sie wichtige Klassifikationskriterien zur Differenzierung wissenschaftlicher Fragebogenstudien!

**Aufgabe 10.3.3:** Sechs wichtige Kriterien zur Klassifikation wissenschaftlicher Fragebogenstudien:

1. Grad der Standardisierung des Fragebogens (teilstandardisiert, vollstandardisiert)
2. Modus der schriftlichen Befragung (z. B. Paper-Pencil-Fragebogen, elektronischer Fragebogen)
3. Verbreitungsweg des Fragebogens (z. B. Austeilen-und-Einsammeln, postalisch)
4. Art der Befragungspersonen (Laien oder Experten)
5. Befragung zu Einzelpersonen oder zu Gruppen
6. Spezielles Abfrageformat (z. B. Tagebuch-Methode; Delphi-Methode)

**Aufgabe 10.3.4:** Was versteht man unter T-ACASI?

**Aufgabe 10.3.4:** Bei T-ACASI (Telephone Audio Computer Assisted Self-Interviewing) handelt es sich um eine selbstadministrierte Telefon-Befragung, die über ein Computersystem gesteuert wird. Das System liest den Fragebogen vor, die Befragungsperson antwortet per Tastendruck auf dem Telefon oder mündlich.

**Aufgabe 10.3.5:** Grenzen Sie die strukturierte Tagebuch-Methode und das Ambulante Assessment voneinander ab!

**Aufgabe 10.3.5:** Die strukturierte Tagebuch-Methode und das Ambulante Assessment dienen beide der fortlaufenden Datenerfassung im Feld über längere Zeitspannen hinweg. Dabei fragt die Tagebuch-Methode Selbstauskünfte ab, während beim Ambulanten Assessment u. a. physiologische Messungen durchgeführt werden.

**Aufgabe 10.3.6:** Mit welchen Maßnahmen lässt sich ein neu konstruierter standardisierter Fragebogen vor der Haupterhebung überprüfen und verbessern?

**Aufgabe 10.3.6:** Im Zuge der Konstruktion eines standardisierten Fragebogens, sollten drei Überprüfungen vorgenommen werden:

1. qualitativer Pretest
2. Fragebogen-Konferenz
3. quantitativer Pretest

**Aufgabe 10.3.7:** Nennen Sie fünf Typen von Funktions-Items! Welche davon sind Wegwerf-Fragen bzw. Wegwerf-Items?

**Aufgabe 10.3.7:**

- Fünf Typen von Funktionsitems: 1) Eisbrecher-, 2) Puffer- und 3) Distraktor- 4) Filter- und 5) Kontroll-Items
- davon Wegwerf-Items: Eisbrecher-, Puffer- und Distraktor-Items

**Aufgabe 10.3.8:** Was versteht man unter projektiven Fragetechniken?

**Aufgabe 10.3.8:** Bei projektiven Fragetechniken werden u. a. uneindeutige Reizvorlagen (z. B. Fotos, Geschichten- oder Satzanfänge) präsentiert, die von den Befragungspersonen zu beschreiben oder zu ergänzen sind. Dabei projizieren die Befragten innerpsychische Vorgänge auf die Reizvorlage, sodass unbewusste Einstellungen oder Motive hervorgehoben werden können.

**Aufgabe 10.3.9:** Differenzieren Sie zwischen Rücklaufquote, Rücklaufkurve und Rücklaufstatistik!

**Aufgabe 10.3.9:**

- Rücklaufquote: Anteil der beantworteten von den ausgesendeten Fragebögen
- Rücklaufkurve: Zeitverlauf des Eingangs der beantworteten Fragebögen
- Rücklaufstatistik: sozialstatistische Merkmale der Respondenten im Vergleich zu Nicht-Respondenten

**Aufgabe 10.3.10:** Grenzen Sie eine Mixed-Mode-Umfrage von einer Mixed-Methods-Studie ab!

**Aufgabe 10.3.10:** Mixed-Mode-Umfrage: Eine vollstrukturierte Umfrage, die mindestens zwei verschiedene Befragungs-Modi kombiniert, z. B. strukturierte Online-Befragung plus strukturierte postalische Befragung mit demselben Fragebogen. Mixed-Methods-Studie: Eine empirische Untersuchung, die mindestens eine quantitative und eine qualitative Datenerhebungsmethode systematisch kombiniert.

## Kapitel 10.4: Datenerhebung - Psychologischer Test

---

**Aufgabe 10.4.1:** Nennen Sie die Haupt- und Nebengütekriterien psychometrischer Tests!

**Aufgabe 10.4.1:**

- Drei Hauptgütekriterien: Objektivität, Reliabilität, Validität
- Sieben Nebengütekriterien: Skalierung, Normierung, Testökonomie, Nützlichkeit, Zumutbarkeit, Nicht-Verfälschbarkeit, Testfairness

**Aufgabe 10.4.2:** Welche Techniken zur Abschätzung der Reliabilität eines psychometrischen Tests werden verwendet?

**Aufgabe 10.4.2:** Vier Methoden der Reliabilitätsschätzung:

- Retest-Reliabilität
- Paralleltest-Reliabilität
- Testhalbierungs-Reliabilität
- Interne Konsistenz

**Aufgabe 10.4.3:** Was versteht man unter einer Item-Charakteristik und wie wird sie dargestellt?

**Aufgabe 10.4.3:** Die Item-Charakteristik beschreibt den Zusammenhang zwischen den latenten Merkmalsausprägungen von Personen und den Wahrscheinlichkeiten bestimmter Itembeantwortungen. Sie wird im Rahmen der PTT bzw. IRT als Funktion oder Kurve dargestellt.

**Aufgabe 10.4.4:** Bei der Korrelation eines Intelligenztests mit der aktuellen Schulleistung spricht man von: a) Vorhersagevalidität, b) Übereinstimmungsvalidität, c) Konstruktvalidität, d) Augenscheinvalidität?

**Aufgabe 10.4.4:** Richtig: b

**Aufgabe 10.4.5:** Die Korrelation zwischen zwei Testhälften ist eine a) Unterschätzung, b) exakte Schätzung, c) Überschätzung der Reliabilität des Tests.

**Aufgabe 10.4.5:** Richtig: a (Unterschätzung durch verminderte Testlänge bei der Testhalbierung)

**Aufgabe 10.4.6:** Welche Maßnahmen werden zur Kontrolle bzw. Verhinderung sozial erwünschten Antwortens in Persönlichkeitstests diskutiert und welche werden im Normalfall eingesetzt?

**Aufgabe 10.4.6:**

- Normalfall: Zusicherung von Anonymität und Aufforderung zu ehrlichem Antworten
- Spezialtechniken:
  1. Bogus-Pipeline-Methode
  2. Objektive Persönlichkeitstests
  3. Random-Response-Technik
  4. Ausbalancierte Antwortvorgaben
  5. Kontrollskalen

**Aufgabe 10.4.7:** Sie planen eine Untersuchung zur sozialen Kompetenz von Polizist:innen. In der Literatur finden Sie zwei Kompetenz-Tests. Der eine hat eine Reliabilität von .76 und eine

Kriteriumsvalidität von .48. Der andere weist eine Reliabilität von .41 und eine Kriteriumsvalidität von .75 auf. Welchen Test wählen Sie und warum?

**Aufgabe 10.4.7:** Den ersten Test, denn der zweite Test wird offensichtlich durch falsche Angaben gekennzeichnet: die Validität kann nicht größer sein als die Wurzel aus der Reliabilität

**Aufgabe 10.4.8:** Wie ist die Item-Trennschärfe definiert und in welcher Relation steht sie zur Reliabilität?

**Aufgabe 10.4.8:** Die Itemtrennschärfe ist definiert als Korrelation eines Items mit dem Testwert. Items mit hoher Trennschärfe steigern die Reliabilität eines Tests, Items mit geringer Trennschärfe reduzieren die Reliabilität eines Tests und können im Zuge einer Itemanalyse identifiziert und ggf. aus der Skala herausgenommen werden.

**Aufgabe 10.4.9:** Welche Klassifikationskriterien sind zur Einteilung psychometrischer Tests einschlägig und welche Gruppen von Tests werden durch sie gebildet?

**Aufgabe 10.4.9:**

1. Art des gemessenen Merkmals: Persönlichkeitstests mit Unterformen versus Leistungstests mit Unterformen
2. Testtheoretische Grundlagen: Tests auf der Basis der KTT versus Tests auf der Basis der PTT bzw. IRT

**Aufgabe 10.4.10:** Wozu dient das 3PL-Modell?

**Aufgabe 10.4.10:** Das 3PL-Modell (Drei-Parameter-Logistische-Modell) gehört zur Gruppe der dichotomen logistischen Latent-Trait-Modelle: Ein kontinuierliches latentes Merkmal wird mit einem Test gemessen, der aus dichotomen Items besteht. Die Lösungswahrscheinlichkeit der Items wird berechnet auf der Basis einer logistischen Funktion, wobei die Ausprägung des latenten Merkmals (Personenparameter) sowie zusätzlich drei weitere Parameter eingehen: Itemschwierigkeits-, Itemtrennschärfe- und Rateparameter.

## Kapitel 10.5: Datenerhebung - Physiologische Messungen

---

**Aufgabe 10.5.1:** Warum werden physiologische Messungen auch als „objektive“ Datenerhebungsmethoden bezeichnet?

**Aufgabe 10.5.1:** Die physiologischen Messverfahren, bei denen die Daten apparativ erhoben werden, sind als objektive Messmethoden von Datenerhebungsmethoden abzugrenzen, die subjektive Selbstaussagen erfassen (z. B. Interview, Fragebogen) und dadurch stärker von kognitiven und sozialen Verzerrungen und Verfälschungen betroffen sind.

**Aufgabe 10.5.2:** Welche Messprobleme treten bei der Erhebung physiologischer Daten auf?

**Aufgabe 10.5.2:**

1. Probleme mit physiologischen und technischen Artefakten
2. Spezifitätsproblematik
3. Ausgangswertproblematik

**Aufgabe 10.5.3:** Grenzen Sie EEG und fMRT zur Messung von Hirnaktivität hinsichtlich physiologischer Grundlagen, Kosten und Informationsgehalt voneinander ab!

**Aufgabe 10.5.3:**

1. EEG: elektrische Spannungsfelder generiert von Nervenzellverbänden des Gehirns; fMRT: Sauerstoffgehalt im Blut des Gehirns
2. fMRT ist um ein Vielfaches teurer als EEG
3. fMRT ermöglicht präzisere Lokalisation aktivierter Hirnareale sowie Untersuchung subkortikaler Prozesse
4. das EEG hat eine bessere zeitliche Auflösung

**Aufgabe 10.5.4:** In welchen Bereichen bewegt sich üblicherweise die Herzschlagfrequenz beim gesunden Erwachsenen im Ruhe- sowie im maximalen Belastungszustand?

**Aufgabe 10.5.4:**

1. Herzfrequenz bei gesunden Erwachsenen im Ruhezustand: ca. 60–70 bps/min<sup>-1</sup>
2. Herzfrequenz bei gesunden Erwachsenen im maximalen Belastungszustand ca. 150–180 bps/min<sup>-1</sup>

**Aufgabe 10.5.5:** Warum wird bei EP im EEG mit der „Mittelungstechnik“ gearbeitet?

**Aufgabe 10.5.5:** Bei EP (evozierten Potenzialen) wird mit der Mittelungs- oder Summationstechnik gearbeitet, um die kleinen Amplituden der EP sichtbar zu machen und von den Spontan-EEG-Amplituden abzugrenzen.

**Aufgabe 10.5.6:** Welche Parallele besteht zwischen Spontan-EEG und Ruhe-EKG?

**Aufgabe 10.5.6:** Sowohl beim Spontan-EEG als auch beim Ruhe-EKG handelt es sich um die Erfassung der tonischen/langfristigen – im Unterschied zur phasischen/ereignisbezogenen – Aktivität des betreffenden Organsystems Gehirn bzw. Herz-Kreislauf-System.

**Aufgabe 10.5.7:** Wie lässt sich ein Ansteigen der Hautleitfähigkeit psychologisch interpretieren?

**Aufgabe 10.5.7:** Das Ansteigen der Hautleitfähigkeit kommt durch vermehrte Schweißdrüsenaktivität zustande und signalisiert wachsende Aktiviertheit des Organismus, vor allem im Kontext negativer Emotionen wie z. B. Angst.

**Aufgabe 10.5.8:** Welche psychologischen Rückschlüsse lassen sich aus der Messung der Aktivität von Gesichtsmuskeln ziehen?

**Aufgabe 10.5.8:** Anhand der Messung der Gesichtsmuskelaktivität lassen sich vor allem Rückschlüsse auf Emotionen ziehen.

**Aufgabe 10.5.9:** Wie werden Blickbewegungen gemessen und in welchen sozialwissenschaftlichen Forschungsfeldern spielen sie eine wichtige Rolle?

**Aufgabe 10.5.9:**

1. Blickbewegungen werden meist mittels Blickbewegungskamera bzw. Eyetracker gemessen, das ist genauer als per Elektrookulografie.

2. Die Messung von Blickbewegungen spielt u. a. in der sozialwissenschaftlichen Forschung zum Leseverhalten, zur visuellen Informationsverarbeitung, zu psychischen Erkrankungen wie z. B. Schizophrenie, zu Werbewirkungen sowie zu Bildschirm-/Web-Design und Usability eine Rolle. Auch in der Methodenforschung kommen Blickbewegungsmessungen zum Einsatz (z. B. grafische Gestaltung von Fragebogen-Instrumenten).

**Aufgabe 10.5.10:** Intensivierte Genitaldurchblutung ist ein spezifischer physiologischer Indikator sexueller Erregung des Organismus. In welcher Beziehung steht er zu subjektiv erlebter sexueller Erregung?

**Aufgabe 10.5.10:** Genau wie bei anderen physiologischen Indikatoren kann nicht direkt auf das subjektive Erleben geschlossen werden.

## **Kapitel 10.6: Datenerhebung – Dokumentenanalyse**

---

**Aufgabe 10.6.1:** Wodurch kann die Qualität von vorgefundenen Dokumenten eingeschränkt sein?

**Aufgabe 10.6.1:** Mögliche Einschränkungen der Qualität bei vorgefundenen Dokumenten:

1. unklare Authentizität/Herkunft
2. zweifelhafte Glaubwürdigkeit
3. begrenzte Repräsentativität
4. beschränkte Interpretierbarkeit

**Aufgabe 10.6.2:** Welche Arbeitsschritte gehören zur zusammenfassenden qualitativen Inhaltsanalyse nach Mayring?

**Aufgabe 10.6.2:** Vier Arbeitsschritte der zusammenfassenden qualitativen Inhaltsanalyse:

1. Paraphrasierung
2. Generalisierung auf das Abstraktionsniveau
3. Erste Reduktion
4. Zweite Reduktion

**Aufgabe 10.6.3:** Auf welche Dokumententypen ist die Interpretative Phänomenologische Analyse (IPA) nicht anwendbar?

**Aufgabe 10.6.3:** Die IPA ist nicht anwendbar auf Dokumente, die keinen autobiografischen Charakter haben.

**Aufgabe 10.6.4:** Wie erfolgt eine Kodierung gemäß Grounded-Theory-Methodologie?

**Aufgabe 10.6.4:** Gemäß GTM erfolgt die Kodierung nach der Methode des permanenten Vergleichs, dies beinhaltet Offenes Kodieren, Axiales Kodieren und Selektives Kodieren.

**Aufgabe 10.6.5:** Welche Arbeitsschritte sind im Kontext der quantitativen Inhaltsanalyse notwendig für die Konstruktion eines standardisierten Kategoriensystems?

**Aufgabe 10.6.5:** Fünf Arbeitsschritte zur Konstruktion eines vollstandardisierten Kategoriensystems im Kontext der quantitativen Inhaltsanalyse:

1. Deduktiver Entwurf des Kategoriensystems (sowie des Codebuchs) anhand von Forschungsliteratur/Theorien
2. Induktive Überarbeitung des Kategoriensystems (sowie des Codebuchs) anhand der Inspektion von Dokumenten aus der Zielpopulation
3. Pretest und Revision des Kategoriensystems (sowie des Codebuchs)
4. Kodiererschulung
5. Reliabilitätsanalyse des Kategoriensystems mit geschulten Kodierern und Finalisierung des Kategoriensystems (sowie des Codebuchs)

**Aufgabe 10.6.6:** Die Darstellung von Wissenschaftler:innen in TV-Wissensmagazinen (z. B. „Galileo“, „Planet Wissen“, „Quarks“) soll untersucht werden. Nennen Sie

- a) eine sinnvolle ordinalskalierte formale Kategorie mit ihren Ausprägungen;
- b) eine sinnvolle kardinalskalierte formale Kategorie mit ihren Ausprägungen;
- c) eine sinnvolle nominalskalierte inhaltliche Kategorie mit ihren Ausprägungen und
- d) eine sinnvolle kardinalskalierte inhaltliche Kategorie mit ihren Ausprägungen!

**Aufgabe 10.6.6:** Kategorien für die quantitative Inhaltsanalyse von Wissensmagazinen im Fernsehen:

- a) Platzierung des Beitrags, in dem Wissenschaftler:innen auftreten, im Kontext der Sendung: 1., 2., 3., 4., 5. usw. Beitrag
- b) Dauer des Beitrags, in dem Wissenschaftler:innen auftreten: in Minuten und Sekunden
- c) Fachrichtung der Wissenschaftler:innen: Naturwissenschaften / Technikwissenschaften / Sozialwissenschaften / Geisteswissenschaften / Formalwissenschaften
- d) Rating der Darstellung fachlicher Kompetenz der Wissenschaftler:innen: sehr geringe – eher geringe – mittlere – eher hohe – sehr hohe fachliche Kompetenz (muss im Codebuch durch Erläuterung und Beispiele unterlegt werden)

**Aufgabe 10.6.7:** Wozu dient Cohens Kappa im Rahmen der Dokumentenanalyse und was zeichnet diesen Koeffizienten aus?

**Aufgabe 10.6.7:** Cohens Kappa-Koeffizient ist ein statistischer Reliabilitäts-Koeffizient für den Grad der Intra- oder Inter-Kodierer-Übereinstimmung einer nominalen Kategorie im Rahmen einer quantitativen Inhaltsanalyse. Er bereinigt die Kodiererübereinstimmung um Zufallsübereinstimmungen.

**Aufgabe 10.6.8:** Im Rahmen einer quantitativen Dokumentenanalyse werden 50 Zeitungsartikel von 2 Kodierern mit einem aus 30 dichotomen Kategorien bestehenden inhaltsanalytischen Kategoriensystem kodiert. Wie viele Reliabilitätskoeffizienten sind zur Beurteilung der Messgenauigkeit des Kategoriensystems zu bestimmen?

**Aufgabe 10.6.8:** Es sind 30 Reliabilitätskoeffizienten zu bestimmen (je ein Inter-Kodierer-Reliabilitätskoeffizient pro Kategorie)

**Aufgabe 10.6.9:** Wozu dient im Kontext der quantitativen Inhaltsanalyse die Intra-Klassen-Korrelation (ICC)?

**Aufgabe 10.6.9:** Mittels der Intra-Klassen-Korrelation ICC wird bei einer quantitativen Inhaltsanalyse die Inter- oder Intra-Kodierer-Übereinstimmung für kardinalskalierte Kategorien bestimmt.



**Aufgabe 10.6.10:** Was versteht man unter der Transitorik von Dokumenten und welche Bedeutung hat sie für die Online-Inhaltsanalyse?

**Aufgabe 10.6.10:** Transitorik meint die Flüchtigkeit von Online-Dokumenten, die fortwährend geändert werden können und nicht zentral archiviert werden. Die Transitorik erfordert im Kontext wissenschaftlicher Dokumentenanalysen eine umfangreiche Archivierung des auszuwertenden Rohdaten-Materials durch die Forschenden, wobei alle relevanten Merkmale der Online-Dokumente – z. B. Verlinkungen – zu erhalten sind.

## **Kapitel 11: Datenaufbereitung**

---

**Aufgabe 11.1:** Nennen Sie drei Funktionen, die eine Datenaufbereitung zwecks Steigerung der Datenqualität erfüllt!

**Aufgabe 11.1:** Drei Funktionen der Datenaufbereitung:

1. Verhinderung von Schwierigkeiten bei der Datenanalyse (z. B. wenn die Bedeutung von Messwerten wegen mangelnder Kommentierung des Datensatzes später nicht mehr rekonstruierbar ist)
2. Verhinderung fehlerhafter Ergebnisse der statistischen Analyse (z. B. verzerrte Mittelwerte, wenn sich im Datensatz unrealistische Ausreißerwerte oder Tippfehler befinden)
3. Verhinderung von ethischen Problemen (z.B. Datenschutzverletzung durch mangelnde Anonymisierung)

**Aufgabe 11.2:** Warum hat die Datenaufbereitung eine zunehmend größere Bedeutung in der sozialwissenschaftlichen Forschung?

**Aufgabe 11.2:** Vier Gründe für die wachsende Bedeutung systematischer Datenaufbereitung:

1. Die computergestützte qualitative und quantitative Datenanalyse erfordert eine digitale Aufbereitung des Datenmaterials.
2. Kooperative Forschungsprojekte erfordern eine sorgfältige Aufbereitung der Datensätze, damit Forschungspartner mit ihnen arbeiten können.
3. Sekundäranalysen und Datenarchive verlangen ebenfalls eine systematische Aufbereitung des Datenmaterials, damit externe Forschungsteams mit ihnen arbeiten können.
4. In der digitalisierten Gesellschaft fallen große Mengen an digitalen Daten an, die für die sozialwissenschaftliche Forschung nur durch entsprechende Datenaufbereitung erschlossen werden können.

**Aufgabe 11.3:** Wie können Fehler in Interview-Transkripten entdeckt werden?

**Aufgabe 11.3:** Aufdeckung von Fehlern in Transkripten:

1. Abgleich zwischen Transkript und Audio-/Videoaufzeichnung
2. Abgleich von zwei Transkripten (erstellt von verschiedenen Personen) derselben Audio- oder Videoaufzeichnung
3. Kontrolle und ggf. Kommentierung des Transkripts durch die interviewte Person

**Aufgabe 11.4:** Welche Probleme können bei der Anonymisierung von Transkripten auftreten?

**Aufgabe 11.4:** Mögliche Probleme bei der Anonymisierung von Transkripten:

1. Hoher Arbeitsaufwand.

2. Eliminierung identifizierender Details kann zu Informationsverlust führen.
3. Manchmal ist schwer abschätzbar, welche Details zur Identifizierbarkeit führen könnten.

**Aufgabe 11.5:** Wie sind Messwerte außerhalb des zulässigen Wertebereichs im Zuge der Datenaufbereitung zu verhindern bzw. zu erkennen?

**Aufgabe 11.5:**

1. Verhinderung von Werten außerhalb des gültigen Wertebereichs: Manuelle Dateneingabe über Dateneingabemasken (die keine ungültigen Werte zulassen) anstatt direkt in die Datenmatrix; Datenerhebung mittels elektronischer Fragebögen oder Formulare (die keine ungültigen Werte zulassen) anstelle von Paper-Pencil-Instrumenten.
2. Erkennung von Werten außerhalb des gültigen Wertebereichs: Systematische Betrachtung der Häufigkeitsverteilungen aller Variablen im Datensatz.

**Aufgabe 11.6:** Nennen Sie die sechs Kriterien für gute Datenqualität bei quantitativen Daten!

**Aufgabe 11.6:** Kriterien guter Qualität bei quantitativen Daten:

1. Vollständigkeit
2. Einheitlichkeit
3. Ausschluss von Dubletten
4. Behandlung fehlender Werte
5. Behandlung von Ausreißer-Werten
6. Plausibilität der Antwortmuster

**Aufgabe 11.7:** Was versteht man unter einem Masterdatensatz?

**Aufgabe 11.7:** Ein Masterdatensatz ist ein kommentierter/gelabelter, anonymisierter und bereinigter quantitativer Datensatz mit allen Variablen und Fällen im Ursprungszustand, der als Referenz archiviert wird. Die Datentransformation (z. B. Umkodierung und Zusammenfassung von Variablen) erfolgt anhand von Arbeitskopien.

**Aufgabe 11.8:** Nennen Sie drei Möglichkeiten, fehlende Werte in einem quantitativen Datensatz zu behandeln!

**Aufgabe 11.8:** Drei Verfahren zu Behandlung von Missings:

1. Ausschluss-Verfahren
2. Imputationsbasierte Verfahren
3. Modellbasierte Verfahren

**Aufgabe 11.9:** Was soll durch eine Umpolung von Variablen erreicht werden?

**Aufgabe 11.9:**

1. Durch Umpolung von Variablenausprägungen sollen hohe Werte starke Merkmalsausprägungen repräsentieren, um die intuitive Verständlichkeit der Ergebnisse zu erhöhen.
2. Variablen, die gemeinsam ein Konstrukt erfassen, müssen alle gleichgerichtet gepolt sein, um sinnvoll interpretierbare Skalen- oder Indexwerte zu berechnen. Dies kann eine Umpolung erforderlich machen.

**Aufgabe 11.10:** Erläutern Sie die Funktion von Gewichtungsverfahren und wann man sie einsetzt!

**Aufgabe 11.10:** Gewichtungungsverfahren zielen auf die Anpassung der Stichprobenstruktur an die Populationszusammensetzung durch Hochgewichten unterrepräsentierter und Heruntergewichten überrepräsentierter Fälle ab. Anwendungsgebiete sind vor allem probabilistische Stichproben im Rahmen populationsbeschreibender Studien.

## **Kapitel 12.1: Datenanalyse - Qualitative Datenanalyse**

---

**Aufgabe 12.1.1:** Nennen Sie vier Beispiele für Datenmaterial, das einer qualitativen Datenanalyse zu unterziehen ist!

**Aufgabe 12.1.1:** Qualitatives Datenmaterial:

1. Notizen und Bilddokumente ethnografischer Feldbeobachtungen
2. Transkripte von nicht- oder halbstrukturierten Interviews
3. Freitextantworten aus nicht- oder halbstrukturierten Fragebögen
4. vorgefundene Dokumente wie z. B. Zeitungsartikel, Online-Tagebücher, Social Media Beiträge

**Aufgabe 12.1.2:** Worin unterscheiden sich spezialisierte und allgemeine Verfahren der qualitativen Datenanalyse?

**Aufgabe 12.1.2:**

- a) Spezialisierte Verfahren der qualitativen Datenanalyse sind auf bestimmte Arten von Datenmaterial und/oder bestimmte inhaltliche Fragestellungen zugeschnitten.
- b) Allgemeine Verfahren der qualitativen Datenanalyse sind relativ breit für unterschiedliches Datenmaterial und für unterschiedliche inhaltliche Fragestellungen einsetzbar.

**Aufgabe 12.1.3:** Grenzen Sie die diskursive Analyse und die Kritische Diskursanalyse voneinander ab und nennen Sie eine Gemeinsamkeit!

**Aufgabe 12.1.3:** Die diskursive Analyse bzw. Gesprächs- oder Konversationsanalyse befasst sich mit der Auswertung von interpersonalen Gesprächen (z. B. Alltagsgespräche oder Fokusgruppendifkussionen). Die Kritische Diskursanalyse zielt auf die Analyse gesellschaftlicher Machtverhältnisse, die sich in öffentlichen Diskursen widerspiegeln und durch diese erzeugt werden, wobei als Analysematerial oft auf Beiträge aus den Massenmedien zurückgegriffen wird. Beide Methoden zählen zu den spezialisierten qualitativen Analysemethoden.

**Aufgabe 12.1.4:** Grenzen Sie Tiefenhermeneutik und Objektive Hermeneutik als qualitative Analysemethoden voneinander ab und nennen Sie eine Gemeinsamkeit!

**Aufgabe 12.1.4:** Die Tiefenhermeneutik ist eine spezialisierte qualitative Auswertungsmethode, die kulturelle Artefakte (z. B. Märchen, Kinofilme) psychoanalytisch deutet. Die Objektive Hermeneutik ist eine allgemeine qualitative Auswertungsmethode, die Texte bzw. Protokolle hinsichtlich ihrer algorithmischen Erzeugungsregeln sequenzanalytisch auswertet. Beide Methoden sind im deutschsprachigen Raum beheimatet.

**Aufgabe 12.1.5:** Schildern Sie die Arbeitsschritte bei der kategorienbildenden qualitativen Datenanalyse und benennen Sie die nicht-kategorienbildende Auswertungsstrategie!

**Aufgabe 12.1.5:**

- Bei der kategorienbildenden Auswertung wird der Text 1.) in sinnvolle Texteinheiten (Analyseeinheiten) segmentiert, 2.) den Textstellen werden Codes zugeordnet und 3.) inhaltsähnliche Codes werden zu übergeordneten Kategorien zusammengefasst.
- Neben der kategorienbildenden Auswertung existiert die sequenzanalytische Auswertung.

**Aufgabe 12.1.6:** Welche Auswertungsebenen werden bei der qualitativen Datenanalyse unterschieden?

**Aufgabe 12.1.6:** Ebenen der qualitativen Datenanalyse:

- a) fallbezogene Auswertung
- b) fallübergreifende Auswertung

**Aufgabe 12.1.7:** Nennen Sie drei Varianten der fallübergreifenden qualitativen Datenauswertung!

**Aufgabe 12.1.7:** Drei Varianten der fallübergreifenden qualitativen Datenanalyse:

1. Thematische Analyse
2. Typenbildung
3. Theoriebildung

**Aufgabe 12.1.8:** Mit welchen Hilfsmitteln kann eine qualitative Datenanalyse erfolgen?

**Aufgabe 12.1.8:** Hilfsmittel für die qualitative Datenanalyse:

- a) manuelle Auswertung mit Farbstiften, Schere, Klebstoff etc. zum Unterstreichen, Ausschneiden, Gruppieren des Materials
- b) computergestützte Analyse mit Bürosoftware (v. a. Microsoft Word, Excel)
- c) computergestützte Analyse mit QDA-Software

**Aufgabe 12.1.9:** Nennen Sie zwei QDA-Programme, die in Deutschland verbreitet sind!

**Aufgabe 12.1.9:** Atlas.ti und MAXQDA sind in Deutschland aktuell verbreitete QDA-Programme.

**Aufgabe 12.1.10:** Lässt sich qualitatives Datenmaterial auch statistisch auswerten? (Begründung)

**Aufgabe 12.1.10:** Qualitatives Datenmaterial lässt sich quantifizieren und dann statistisch auswerten. Sinnvoll ist das nur bei ausreichender Stichprobengröße.

## Kapitel 12.2: Datenanalyse - Quantitative Datenanalyse

---

**Aufgabe 12.2.1:** Wovon hängt die statistische Auswertungsstrategie bei quantitativen Studien ab?

**Aufgabe 12.2.1:** Die statistische Auswertungsstrategie hängt primär vom Erkenntnisinteresse einer quantitativen Studie ab:

1. Explorativ-hypothesenbildend
2. Deskriptiv-populationsbeschreibend
3. Explanativ-hypothesenprüfend

**Aufgabe 12.2.2:** Mit welchen Datenerhebungsmethoden wird quantitatives Datenmaterial in den Human- und Sozialwissenschaften üblicherweise gewonnen?

**Aufgabe 12.2.2:** Wege der Gewinnung quantitativer Daten (Messwerte) in der empirischen Sozialforschung:

1. Strukturierte Beobachtung
2. Strukturierte mündliche Befragung (Interview)
3. Strukturierte schriftliche Befragung
4. Psychometrische Tests
5. Physiologische Messungen
6. Quantitative Inhaltsanalyse

**Aufgabe 12.2.3:** Welche Ansätze der inferenzstatistischen Hypothesenprüfung sind zu unterscheiden?

**Aufgabe 12.2.3:** Verschiedene Ansätze der inferenzstatistischen Hypothesenprüfung:

1. Klassischer Signifikanztest
2. Minimum-Effektgrößen-Test
3. Strukturgleichungsmodelle
4. Resampling-Verfahren
5. Verfahren der Bayes-Statistik

**Aufgabe 12.2.4:** Die Stichprobenbeschreibung einer Studie sollte mit Tabellen und Grafiken veranschaulicht werden. Diskutieren Sie diese Aussage!

**Aufgabe 12.2.4:** Die Stichprobenbeschreibung ist eine Hintergrundinformation jeder quantitativen Studie, die in der Regel in einem Absatz im Fließtext behandelt wird. Tabellarische und grafische Darstellungen sind in der Regel den zentralen inhaltlichen Befunden einer Studie vorbehalten.

**Aufgabe 12.2.5:** Mit der inferenzstatistischen Datenanalyse ist die statistische Prüfung der Forschungshypothesen gemeint. Richtig oder falsch?

**Aufgabe 12.2.5:** Falsch, denn zur Inferenzstatistik gehören neben Methoden der Hypothesenprüfung auch Verfahren der Parameterschätzungen (Punkt- und Intervallschätzungen) sowie Tests zur Voraussetzungsprüfung statistischer Verfahren (z. B. Test auf Varianzhomogenität, Test auf Normalverteilung).

**Aufgabe 12.2.6:** Was stimmt?

- a) Der statistische Signifikanztest stellt die einzige Möglichkeit der Prüfung quantitativer Hypothesen dar.
- b) Zur Deskriptivstatistik zählen die Bestimmung von Stichprobenkennwerten sowie von Populationsparametern.
- c) Wenn ein Ergebnis statistisch nicht signifikant ist, dann gilt automatisch die Nullhypothese als bestätigt.
- d) Wenn ein Ergebnis statistisch signifikant ist, dann handelt es sich um einen praktisch bedeutsamen Effekt.
- e) Ein Ergebnis ist signifikant, wenn das Alpha-Fehler-Niveau unter 5 % liegt.

**Aufgabe 12.2.6:** Alle Antwortvorgaben sind falsch.

**Aufgabe 12.2.7:** Grenzen Sie den klassischen Signifikanztest von Resampling-Verfahren und Verfahren der Bayes-Statistik ab!

**Aufgabe 12.2.7:** Beim klassischen Signifikanztest wird als Prüfverteilung eine theoretische Verteilung verwendet, bei Resampling-Verfahren dagegen eine empirische (d. h. aus den Daten erzeugte) Verteilung. Bei Verfahren der Bayes-Statistik wird das Vorwissen der Forschenden (bzw. der aktuelle Forschungsstand) über die getesteten Effekte einbezogen.

**Aufgabe 12.2.8:** Nennen Sie mindestens drei allgemeine Statistik-Programmpakete, die in den Sozialwissenschaften verbreitet sind!

**Aufgabe 12.2.8:** In den Sozialwissenschaften verbreitete Statistik-Programmpakete:

- SPSS
- SAS
- R

**Aufgabe 12.2.9:** Deskriptivstatistische Methoden werden im Rahmen jeder quantitativen Studie eingesetzt. Wozu?

**Aufgabe 12.2.9:** Deskriptivstatistische Verfahren werden im Rahmen jeder quantitativ-empirischen Studie für die Stichprobenbeschreibung und die Dateninspektion genutzt.

**Aufgabe 12.2.10:** Das Ergebnis einer empirisch-quantitativen Studie zeigt sich im Output des verwendeten Statistik-Programms. Diskutieren Sie diese Aussage!

**Aufgabe 12.2.10:** Der Output eines Statistik-Programms ist nicht das Ergebnis einer Studie, ihm müssen vielmehr zunächst die für das Forschungsproblem relevanten statistischen Kennwerte entnommen und diese dann im Gesamtbild der Befunde inhaltlich interpretiert werden.

## **Kapitel 12.3: Datenanalyse - Quantitative Datenanalyse bei explorativen Studien**

---

**Aufgabe 12.3.1:** Was ist der Unterschied zwischen der explorativen Datenanalyse (EDA) und der konfirmativen Datenanalyse (CDA)?

**Aufgabe 12.3.1:** EDA dient dazu, Strukturen und Trends in einem quantitativen Datensatz zu entdecken, in der CDA wird die Analyse auf die laut Hypothesen interessierenden Effekte beschränkt.

**Aufgabe 12.3.2:** Welche Techniken werden im Rahmen der explorativen Datenanalyse eingesetzt?

**Aufgabe 12.3.2:** Im Rahmen der explorativen Datenanalyse werden vor allem drei Techniken eingesetzt:

1. Visualisierungstechniken
2. Explorative multivariate Verfahren
3. Techniken des Data Mining

**Aufgabe 12.3.3:** Welche Werte können Boxplots entnommen werden?

**Aufgabe 12.3.3:** Einem Box-Plot können folgende statistische Kennwerte entnommen werden:

- a) Medianwert
- b) Interquartilbereich
- c) Streuung
- d) Ausreißerwerte
- e) Extremwerte

**Aufgabe 12.3.4:** Grenzen Sie die Zielsetzungen einer explorativen Faktorenanalyse und einer Cluster-Analyse voneinander ab.

**Aufgabe 12.3.4:**

- Explorative Faktorenanalyse: Bündelung von Variablen zu Faktoren, also zu Variablen höherer Ordnung
- Clusteranalyse: Zusammenfassung von Untersuchungsobjekten zu Clustern, also zu Typen, deren Untersuchungsobjekte sich jeweils innerhalb eines Clusters stark ähneln und zwischen den Clustern unterscheiden

**Aufgabe 12.3.5:** Was unterscheidet einen „Signifikanztest auf Probe“ von einem „Pseudo-Signifikanztest“?

**Aufgabe 12.3.5:** Bei einem „Signifikanztest auf Probe“ wird ein induktiv in den Daten gefundener – d. h. nicht per theoretisch fundierter Hypothese vorausgesagter – Effekt probenhalber gegen den Zufall getestet, um auf dieser Basis Hypothesen für zukünftige Studien zu bilden. Beim Pseudo-Signifikanztest wird ein induktiv in den Daten gefundener Effekt gegen den Zufall getestet und das Ergebnis im Nachhinein als Hypothesentest ausgegeben.

**Aufgabe 12.3.6:** Welches Hauptmerkmal kennzeichnet eine A-priori-Hypothese?

**Aufgabe 12.3.6:** A-priori-Hypothesen sind dadurch gekennzeichnet, dass sie vor der Datenerhebung auf der Basis von Theorien und/oder früheren Studienergebnissen formuliert werden. Typischerweise handelt es sich um gerichtete Hypothesen, idealerweise auch mit Angabe zur Effektgröße.

**Aufgabe 12.3.7:** Handelt es sich bei Auswertungen mit dem Tool Google „Trends“ (<https://trends.google.com>) um „web usage mining“? (Begründung)

**Aufgabe 12.3.7:** Das Tool „Trends“ von Google bietet der Öffentlichkeit nur einfache Häufigkeitsanalysen für Suchbegriffe an, die nicht unter Data-Mining fallen. Die dahinter liegende Datenbank ist jedoch geeignet, um Web Usage Mining durchzuführen und Muster in den weltweiten Google-Suchanfragen zu erkunden.

**Aufgabe 12.3.8:** Was versteht man unter Big Data?

**Aufgabe 12.3.8:** Unter Big Data versteht man die Sammlung, Archivierung und Analyse sehr großer Datenmengen, die durch automatische Datenaufzeichnung entstehen.

**Aufgabe 12.3.9:** Wozu nutzt man die Multidimensionale Skalierung?

**Aufgabe 12.3.9:** Die Multidimensionale Skalierung dient dazu, die Ähnlichkeit bzw. Unähnlichkeit von Urteilsobjekten in einem 2- oder 3-dimensionalen Wahrnehmungsraum abzubilden.

**Aufgabe 12.3.10:** Welcher Statistiker hat die explorative Datenanalyse namhaft propagiert?

**Aufgabe 12.3.10:** Ein wichtiger Vertreter des EDA-Ansatzes war der Statistiker John Tukey.

## **Kapitel 12.4: Datenanalyse - Quantitative Datenanalyse bei deskriptiven Studien**

---

**Aufgabe 12.4.1:** Welche Qualitätskriterien sollen Punktschätzer erfüllen?

**Aufgabe 12.4.1:** Punktschätzer sollen fünf Qualitätskriterien erfüllen:

1. Erwartungstreue
2. Konsistenz
3. Effizienz
4. Suffizienz
5. Robustheit

**Aufgabe 12.4.2:** Bei der Körpergröße handelt es sich um eine normalverteilte Zufallsvariable. Wie groß ist die Wahrscheinlichkeit für das Auftreten einer ganz bestimmten Körpergröße bei einer untersuchten Person?

**Aufgabe 12.4.2:** Wie bei allen kontinuierlichen Zufallsvariablen, die mit theoretisch beliebiger Genauigkeit erfasst werden können, ist die Wahrscheinlichkeit für ein Einzelereignis Null, es kann nur die Wahrscheinlichkeit für einen Wertebereich angegeben werden.

**Aufgabe 12.4.3:** Welche Aussagen stimmen?

- a) Je höher der Konfidenzkoeffizient, umso breiter ist das Konfidenzintervall.
- b) Je größer die Populationsstreuung, umso breiter ist das Konfidenzintervall.
- c) Die Varianz einer einfachen Zufallsstichprobe ist ein erwartungstreuer Schätzer der Populationsvarianz.
- d) Der Mittelwert einer einfachen Zufallsstichprobe ist ein erwartungstreuer Schätzer des Populationsmittelwertes.

**Aufgabe 12.4.3:** Richtig sind: a, b und d

**Aufgabe 12.4.4:** Warum ist es erforderlich, Parameterschätzungen mit einem Konfidenzintervall zu versehen?

**Aufgabe 12.4.4:** Die Schätzung von Populationsparametern auf der Basis von Daten über nur einen (meist verhältnismäßig kleinen) Teil der Population (Stichprobenuntersuchung) ist immer mit Unsicherheit behaftet. Bei der Intervallschätzung mittels Konfidenzintervall kann diese Unsicherheit in Form eines Wahrscheinlichkeitswertes (Konfidenzkoeffizient) quantifiziert und kontrolliert werden.

**Aufgabe 12.4.5:** Welche Methoden der Punktschätzung werden unterschieden?

**Aufgabe 12.4.5:** Methoden der Punktschätzung von Populationsparametern:

1. Momenten-Methode
2. Kleinstquadrat-Methode
3. Maximum-Likelihood-Methode
4. Resampling-Verfahren



## 5. Bayes-Verfahren

**Aufgabe 12.4.6:** Welche Aussagen stimmen nicht?

- a) Die z-Transformation überführt jede Verteilung in eine Verteilung mit Mittelwert 1 und Streuung 0.
- b) Die z-Transformation überführt jede Verteilung in eine Verteilung mit Mittelwert 1 und Varianz 0.
- c) Die z-Transformation überführt jede Verteilung in eine Verteilung mit Mittelwert 0 und Varianz 1.
- d) Die z-Transformation überführt jede Verteilung in eine Standardnormalverteilung.

**Aufgabe 12.4.6:** Falsch sind: a, b und d

**Aufgabe 12.4.7:** Bei dem sog. Standardfehler handelt es sich um:

- a) die Streuung der Stichprobe,
- b) die Streuung der Population,
- c) die Streuung der Stichprobenkennwerteverteilung,
- d) die Differenz zwischen Schätzwert und Populationsparameter.

**Aufgabe 12.4.7:** Richtig ist: c

**Aufgabe 12.4.8:** Für die Berechnung des Konfidenzintervalls des Mittelwertes liegen verschiedene Formeln vor. Warum?

**Aufgabe 12.4.8:** Die Berechnung des Konfidenzintervalls des Mittelwertes erfolgt bei unbekannter Populationsstreuung und einem kleinen Stichprobenumfang ( $n < 30$ ) anhand kritischer Werte aus der  $t$ -Verteilung, bei bekannter Populationsstreuung sowie bei großem Stichprobenumfang ( $n \geq 30$ ) anhand kritischer Werte aus der Standardnormalverteilung.

**Aufgabe 12.4.9:** Um im Vorfeld einer deskriptiven Studie zu bestimmen, welcher optimale Stichprobenumfang für die Intervallschätzung des interessierenden Populationsparameters notwendig ist, müssen welche Entscheidungen getroffen werden?

**Aufgabe 12.4.9:** Um den optimalen Stichprobenumfang für die Intervallschätzung des Populationsparameters zu bestimmen, müssen a) der Konfidenzoeffizient (meist 95 % oder 99 %) und b) der zulässige Schätzfehler (in Prozent der Merkmalsstreuung) festgelegt werden.

**Aufgabe 12.4.10:** Eine Lehrerin interessiert sich für die Frage, wie viel Zeit 11-jährige Schulkinder täglich für ihre Hausaufgaben aufwenden. Für ihre Untersuchung nimmt sie in Kauf, dass die wahre durchschnittliche Hausaufgabenzeit um maximal 3 Minuten überschätzt wird. Das zu ermittelnde Konfidenzintervall soll mit einem Konfidenzoeffizienten von 99 % abgesichert werden. Die Streuung der Hausaufgabenzeit bei den Schulkindern der betrachteten Altersgruppe möge bei  $\sigma = 35$  min liegen. Mit welchem optimalen Stichprobenumfang  $n_{\text{opt}}$  sollte eine entsprechende empirische Studie angelegt sein, um – ohne Ressourcenvergeudung – ein aussagekräftiges Konfidenzintervall berechnen zu können?

**Aufgabe 12.4.10:** Schätzfehler:  $3/35 \approx 0.09 \times \sigma$ ;  $n_{\text{opt}} = 821$

## Kapitel 12.5.1: Datenanalyse - Quantitative Datenanalyse bei explanativen Studien

---

**Aufgabe 12.5.1.1:** Welche Aussage ist richtig?

- a) Wenig aussagekräftige Daten können vor der Hypothesenprüfung so aufbereitet werden, dass eine statistische Überprüfung sinnvoll stattfinden kann.
- b) Das Untersuchungsdesign entscheidet über den Informationsgehalt der Daten in Bezug auf die zu überprüfenden Hypothesen.
- c) Ungerichtete Hypothesen haben einen höheren Aussagegehalt als gerichtete Hypothesen.
- d) Hypothesen sollten nach der Betrachtung des empirischen Stichprobenergebnisses an die Datenlage angepasst werden.
- e) In wissenschaftlichen Publikationen werden immer die inhaltlichen und die statistischen Hypothesen genannt.

**Aufgabe 12.5.1.1:** Richtig: b

**Aufgabe 12.5.1.2:** Wovon hängt die Auswahl eines passenden Signifikanztestes für einen inhaltlich sinnvollen statistischen Hypothesentest ab?

**Aufgabe 12.5.1.2:**

1. Art der Hypothese (Zusammenhang, Unterschied, Veränderung, Einzelfall)
2. Anzahl der Variablen
3. Skalenniveau der Variablen
4. Stichprobenumfang und Verteilungseigenschaften

**Aufgabe 12.5.1.3:** Warum wird der klassische Signifikanztest auch als Nullhypothesen-Signifikanztest bezeichnet?

**Aufgabe 12.5.1.3:** Weil das empirische Ergebnis der jeweiligen Studie in der Logik des klassischen Signifikanztests anhand des  $H_0$ -Modell (Stichprobenkennwerteverteilung unter der Annahme, dass in der Population kein Effekt vorliegt) geprüft wird.

**Aufgabe 12.5.1.4:** Warum wird als Entscheidungsregel für die Signifikanz eines Ergebnisses einerseits  $[p(D|H_0) \leq \alpha]$  und andererseits  $[|\text{empirischer Wert}| \geq |\text{kritischer Wert}|]$  angegeben?

**Aufgabe 12.5.1.4:** Beide Aussagen sind inhaltlich gleichbedeutend, da der empirische Wert der Prüfverteilung genau die Irrtumswahrscheinlichkeit  $p(D|H_0)$  definiert und der kritische Wert das Alpha-Fehler-Niveau. In der Praxis wird die Signifikanzentscheidung beim Rechnen mit Statistik-Software direkt anhand der ausgegebenen Irrtumswahrscheinlichkeit getroffen, beim Rechnen per Hand anhand des empirischen Wertes.

**Aufgabe 12.5.1.5:** Im klassischen Nullhypothesen-Signifikanztest werden bei großem Stichprobenumfang auch kleine Effekte statistisch signifikant, denn die Streuung des  $H_0$ -Modells reduziert sich und der kritische Wert wird dem Betrag nach kleiner. Illustrieren Sie diese Aussage, indem Sie den kritischen  $t$ -Wert für die 5 %-Signifikanzschwelle zur einseitigen Prüfung der bivariaten Korrelation jeweils für eine Stichprobengröße von  $n = 10$ ,  $n = 20$ ,  $n = 30$  und  $n = 100$  gegenüberstellen!

**Aufgabe 12.5.1.5:** Der kritische  $t$ -Wert zur Prüfung der bivariaten Korrelation hat  $df = n - 1$  Freiheitsgrade. Für die genannten  $n$  ergeben sich somit folgende kritische  $t$ -Werte laut austabellierter  $t$ -Verteilung für das 5 %-Niveau:

- $df = 9$ ;  $t = 1.833$
- $df = 19$ ;  $t = 1.792$

- $df = 29$ ;  $t = 1.699$
- $df = 120$  (99 ist nicht austabelliert)  $\rightarrow t = 1.658$

**Aufgabe 12.5.1.6:** In einer Veröffentlichung lesen Sie, dass das Ergebnis eines Signifikanztests „höchst signifikant ( $p < .001$ ) ausgefallen“ sei. Diskutieren Sie diese Aussage!

**Aufgabe 12.5.1.6:** Das Signifikanz- bzw. Alpha-Fehler-Niveau wird vor der Hypothesenprüfung festgelegt und muss begründet werden. Das Ergebnis kann dann nur signifikant (wenn die vorher festgelegte Schwelle von 5 %, 1 % oder 0.1 % unterschritten wird) oder nicht signifikant ausfallen. „Höchst signifikant“ oder „hoch signifikant“ sind demnach statistisch unsinnige Steigerungsformen.

**Aufgabe 12.5.1.7:** „Bei einer Korrelation von  $r = -.14$  handelt es sich um einen kleinen und somit praktisch bedeutungslosen Effekt.“ Diskutieren Sie diese Aussage!

**Aufgabe 12.5.1.7:** Auch wenn ein Korrelationskoeffizient von  $r = -.14$  nominell als „klein“ klassifiziert wird, muss zur Einordnung der praktischen Bedeutsamkeit in Rechnung gestellt werden, um welche Inhalte es geht und mit welchen Effektgrößen in diesem Forschungsfeld normalerweise zu rechnen ist. Auch kleine Effekte können unter Umständen praktisch bedeutsam sein (z. B. wenn es um Gesundheit oder Menschenleben geht).

**Aufgabe 12.5.1.8:** Welche Aussagen sind falsch?

- Ein einseitiger Signifikanztest ist teststärker als ein zweiseitiger Signifikanztest.
- Bei einem einseitigen Signifikanztest wird das Signifikanzniveau auf 10 % festgelegt, weil der kritische Wert nur auf einer Seite der Prüfverteilung abgetragen wird.
- Ein statistisch signifikantes Ergebnis deutet auf praktische Bedeutsamkeit hin.
- Die Verwendung von optimalen Stichprobenumfängen garantiert, dass statistisch signifikante Ergebnisse erzielt werden.

**Aufgabe 12.5.1.8:** Falsch: b), c), d)

**Aufgabe 12.5.1.9:** Grenzen Sie Alpha- und Beta-Fehler-Wahrscheinlichkeit voneinander ab und geben Sie jeweils das korrekte Symbol an.

**Aufgabe 12.5.1.9:**

- Alpha-Fehler-Wahrscheinlichkeit:  $p(D|H_0)$ : Bedingte Wahrscheinlichkeit des empirischen Stichprobenergebnisses unter Annahme der Gültigkeit der  $H_0$ . Wenn bei geringer Alpha-Fehler- bzw. Irrtumswahrscheinlichkeit  $p(D|H_0) \leq \alpha$  die  $H_1$  angenommen wird, so entspricht die Irrtumswahrscheinlichkeit der Wahrscheinlichkeit, die  $H_1$  fälschlicherweise anzunehmen (denn auch wenn sich das Ergebnis nur sehr schlecht mit dem  $H_0$ -Modell vereinbaren lässt – z. B.  $p = 3\%$  – so ist es doch in 3 % der Fälle möglich, dass ein extremes Stichprobenergebnis zustande kommt, obwohl in der Population in Wirklichkeit die  $H_0$  gilt).

- Beta-Fehler-Wahrscheinlichkeit:  $p(D|H_1)$ : Bedingte Wahrscheinlichkeit des empirischen Stichprobenergebnisses unter Annahme der Gültigkeit der  $H_1$ . Wenn bei geringer Beta-Fehler-Wahrscheinlichkeit  $p(D|H_1) \leq \beta$  die  $H_0$  angenommen wird, so entspricht die Beta-Fehler-Wahrscheinlichkeit der Wahrscheinlichkeit, die  $H_0$  fälschlicherweise beizubehalten.

**Aufgabe 12.5.1.10:** Welches der folgenden Konfidenzintervalle (95 %) lässt keine Aussage über die Richtung des Effekts zu? Welches Konfidenzintervall schätzt den Populationseffekt am genauesten?

- $-.21 \leq \rho \leq -.48$

b)  $+0.15 \leq \rho \leq +0.23$

c)  $-0.01 \leq \rho \leq +0.34$

d)  $+0.31 \leq \rho \leq +0.42$

**Aufgabe 12.5.1.10:** keine Aussage über die Richtung des Effekts: c); genauester Schätzer: b)

## Kapitel 12.5.2: Datenanalyse - Quantitative Datenanalyse bei explanativen Studien

---

**Aufgabe 12.5.2.1:** Welche interne Validität weisen Interdependenzanalysen auf und wie ist diese im Hinblick auf kausale Interpretationen zu bewerten?

**Aufgabe 12.5.2.1:** Die interne Validität von Interdependenzanalysen ist in der Regel gering, vor allem, wenn es sich um Querschnittsstudien handelt. Denn der in einer Interdependenzanalyse gefundene Zusammenhang sagt zunächst nichts über Kausalbeziehungen der untersuchten Merkmale aus. Eine belastbare Kausalitätsaussage lässt sich nur auf Basis experimenteller Designs treffen. Die interne Validität kann durch zeitversetzte Messung von Prädiktor- und Kriteriumsvariablen (Zukünftiges kann Vergangenes nicht kausal bestimmen), Neutralisierung von Kontroll- und Störvariablen durch Partialkorrelation und Detailanalysen von Wirkungspfaden in komplexen Kausalmodellen (Pfadanalyse, Strukturgleichungsmodelle) erhöht werden.

**Aufgabe 12.5.2.2:** Lässt sich durch einen Korrelationskoeffizienten von  $r = .58$  auf einen starken Zusammenhang in der Population schließen?

**Aufgabe 12.5.2.2:** Nein, der Korrelationskoeffizient  $r$  quantifiziert nur Enge und Richtung des Zusammenhangs in der Stichprobe (deskriptiv-statistisches Ergebnis). Für den Rückschluss auf die Population wird ein inferenzstatistisches Ergebnis benötigt (Konfidenzintervall oder statistischer Signifikanztest).

**Aufgabe 12.5.2.3:** Die multiplen Regressionskoeffizienten (d. h. die Beta-Gewichte einer multiplen Regression) entsprechen den bivariaten Korrelationskoeffizienten zwischen Prädiktoren und Kriterium. Diskutieren Sie diese Aussage!

**Aufgabe 12.5.2.3:** Den Beta-Gewichten einer multiplen Regressionsanalyse ist zu entnehmen, ob und wie die einzelnen Prädiktoren unter Berücksichtigung aller anderen Prädiktoren mit dem Kriterium korrelieren. Korrelationskoeffizienten in der bivariaten Korrelation werden dagegen unabhängig von anderen Prädiktoren berechnet. Der multiple Regressionskoeffizient eines Prädiktors entspricht nur dann seiner bivariaten Korrelation mit dem Kriterium, wenn die betrachteten Prädiktoren untereinander nicht korrelieren.

**Aufgabe 12.5.2.4:** Was stimmt nicht?

- a) Das Ausschließen von Ausreißerwerten vor der Datenanalyse ist nicht zwingend notwendig, da Korrelationskoeffizienten durch diese kaum verzerrt werden.
- b) Eine kleine homogene Teilstichprobe aus der Population kann den wahren Zusammenhang unterschätzen.
- c) Die Extremgruppenselektion führt zu einer Unterschätzung des Zusammenhanges.
- d) Zur Prüfung von Zusammenhangshypothesen sollten Stichproben gezogen werden, die die Population möglichst gut repräsentieren.

**Aufgabe 12.5.2.4:** Falsch sind: a und c

**Aufgabe 12.5.2.5:** Grenzen Sie die Partialkorrelation von der Semipartialkorrelation ab!

**Aufgabe 12.5.2.5:** Bei der Partialkorrelation wird der bivariate Zusammenhang der Variablen X und Y um den Einfluss einer Variablen Z bereinigt, z. B. um zu überprüfen, ob zwischen den Variablen X und Y eine Scheinkorrelation vorliegt, die durch Z verursacht wird. Eine Semipartialkorrelation kommt zur Anwendung, wenn lediglich eine Variable (X oder Y) um den Einfluss der Variable Z bereinigt wird.

**Aufgabe 12.5.2.6:** Was trifft zu?

- a) Multiple Zusammenhangshypothesen beziehen sich auf Zusammenhänge zwischen zwei kardinalskalierten Variablen.
- b) Multiple Zusammenhangshypothesen spielen eine untergeordnete Rolle in der Hypothesenprüfung.
- c) Die Überprüfung multipler Zusammenhangshypothesen erfolgt in der Regel über die multiple Korrelations- und Regressionsanalyse.
- d) In der sozialwissenschaftlichen Forschung sind lediglich lineare Zusammenhänge von Interesse.
- e) Multiple Korrelation und kanonische Korrelation sind Synonyme.
- f) Multiple Zusammenhänge sollten mittels multivariater Verfahren überprüft werden und nicht über eine Mehrfachberechnung von bivariaten Verfahren.

**Aufgabe 12.5.2.6:** Richtig sind: c und f

**Aufgabe 12.5.2.7:** Wie ist ein Cross-Lagged-Panel-Design aufgebaut und wozu braucht man es?

**Aufgabe 12.5.2.7:** Ein Cross-Lagged-Panel-Design dient dazu, die Richtung einer theoretisch bzw. hypothetisch vorgegebenen Kausalbeziehung zwischen Variablen mit hoher interner Validität zu prüfen. Dabei macht man sich den Zeitfaktor zunutze, indem man die Variablen, zwischen denen eine Kausalrelation bestehen soll, mit zeitlichem Abstand (Lag) zweimal (oder auch mehrfach) erfasst (1. Messung:  $x_1, y_1$ ; 2. Messung:  $x_2, y_2$ ). Berechnet werden dann (im Fall von zwei Messungen) alle sechs bivariaten Korrelationen zwischen den vier Messwertreihen ( $x_1x_2, y_1y_2, x_1y_1, x_1y_2, x_2y_1, x_2y_2$ ). Anhand der Größenverhältnisse dieser Korrelationen kann die Kausalrichtung abgeschätzt werden.

**Aufgabe 12.5.2.8:** Was versteht man unter einer Pfadanalyse?

**Aufgabe 12.5.2.8:** Die Pfadanalyse nutzt die Techniken der Korrelations- und Regressionsanalyse (insbesondere die Partial- und Semipartialkorrelation), um ein a priori formuliertes komplexes Kausalmodell mit höherer interner Validität zu prüfen als es einfache bivariate (oder multiple) Korrelationen zwischen den beteiligten Variablen ermöglichen. Auch die Pfadanalyse kann jedoch Kausalhypothesen nur falsifizieren und nicht verifizieren.

**Aufgabe 12.5.2.9:** Grenzen Sie Moderator- und Mediatorvariable voneinander ab!

**Aufgabe 12.5.2.9:** Eine Moderatorvariable beeinflusst Enge und/oder Richtung des Zusammenhangs zwischen zwei Variablen, während eine Mediatorvariable kausal zur Entstehung des Effektes beiträgt.

**Aufgabe 12.5.2.10:** Welche Rolle spielen Metaanalysen bei der Überprüfung von kausalen Zusammenhangshypothesen?

**Aufgabe 12.5.2.10:** Wird eine Zusammenhangshypothese in einer Querschnittstudie einmalig überprüft, ist die Generalisierbarkeit dieses Effekts sehr eingeschränkt. Wird die gleiche Hypothese von mehreren Forschenden und unter verschiedenen Bedingungen geprüft, und wird ein entsprechender Zusammenhang immer wieder nachgewiesen, kann dieser Effekt überzeugender generalisiert werden. Metaanalysen können einen Überblick über statistische Ergebnisse vergleichbarer Studien geben. Neben einer Zusammenfassung zu einem Gesamtwert und einer Gesamtinterpretation können sie auch differenzieren, hinsichtlich welcher Subgruppen oder Untersuchungsbedingungen die Generalisierbarkeit eingeschränkt ist.

### **Kapitel 12.5.3: Datenanalyse - Quantitative Datenanalyse bei explanativen Studien**

---

**Aufgabe 12.5.3.1:** Ein experimenteller und ein quasi-experimenteller Viergruppenplan sollen jeweils statistisch ausgewertet werden. Welche der beiden Studien geht mit höherem Aufwand bei der statistischen Datenanalyse einher (Begründung)?

**Aufgabe 12.5.3.1:** Die experimentelle Studie ist einfacher auszuwerten, da durch Randomisierung die Störvariablen ausgeschaltet sind, während sie bei der quasi-experimentellen Studie möglichst umfassend statistisch kontrolliert werden müssen, was meist mit Zusatzaufwand verbunden ist (z. B. durch Kovarianzanalyse, Propensity-Score-Analyse).

**Aufgabe 12.5.3.2:** In einem Seniorenwohnheim wird ein Jahr lang eine wöchentliche Kochgruppe durchgeführt. Die 25 Mitglieder berichteten nach dem Ende der Kochgruppe auf einem standardisierten Ernährungsfragebogen (1–10 Punkte) eine überdurchschnittlich gesundheitsbewusste Ernährung ( $M = 7.69$ ). Eine Kontrollgruppe von 25 Senioren, die nicht an der Kochgruppe teilgenommen hatten, zeigte signifikant ( $t$ -Test für unabhängige Stichproben) weniger gesundheitsbewusstes Ernährungsverhalten ( $M = 3.45$ ). Ist damit belegt, dass die Kochgruppe gesundheitsbewusste Ernährung bei Senioren fördert? (Begründung) Unter welchen Bedingungen kann eine solche Studie Auskunft über die Wirksamkeit der Kochgruppe geben? Welches statistische Auswertungsverfahren ist indiziert?

**Aufgabe 12.5.3.2:** Der Zwei-Gruppen-Plan mit ausschließlicher Nachher-Messung gibt wenig Auskunft über die Wirksamkeit der Kochgruppe, da unklar ist, wie gesundheitsbewusst sich die Senioren von vorne herein ernährten. Beruhte die Teilnahme am Kochkurs auf reiner Selbstselektion, könnten sich verstärkt ernährungsbewusste Personen bei der Kochgruppe angemeldet haben. Wurde die Kochgruppen-Teilnahme aus einer Liste von Interessierten per Los entschieden und somit Treatment- und Kontrollgruppe randomisiert, wären weniger Verzerrungen zu erwarten als bei reiner Selbstselektion. Aufgrund der geringen Gruppengrößen von  $n = 25$  sind Verzerrungen aber auch bei Randomisierung nicht ausgeschlossen. Eine Möglichkeit, um die interne Validität zu steigern, bestünde in der Erhebung von Kontrollvariablen, die ebenfalls auf gesundheitsbewusste Ernährung Einfluss haben (z. B. Gesundheitszustand, Einstellungen etc.) und deren Einfluss aus der abhängigen Variablen herauspartialisiert werden könnte, auch eine Propensity-Score-Analyse wäre möglich. Beide Verfahren wären aber mit Mehraufwand für die Untersuchungspersonen verbunden (verlängerter Fragebogen) und würden Störeinflüsse nur insofern kontrollieren, als man sie vorab antizipieren kann. Eine aussagekräftige Studie müsste als Messwiederholungsdesign (Zwei-Gruppen-Prä-Post-Design) angelegt werden, um in beiden Gruppen mögliche Veränderungen des Ernährungsverhaltens zu erfassen, d. h. es müssten beim Untersuchungsdesign und bei der statistischen Auswertung Prinzipien zur Überprüfung von Veränderungshypothesen berücksichtigt werden. Die statistische Analyse könnte als Varianzanalyse mit Messwiederholung erfolgen, die Wirksamkeit der Maßnahme würde sich dann

in einer signifikanten Interaktion zwischen Messzeitpunkt und Behandlungsgruppe zeigen (d. h. die Kontrollgruppe verändert sich nicht, während in der Kochgruppe das ernährungsbewusste Verhalten zunimmt). Alternativ könnte auch eine Kovarianzanalyse durchgeführt werden, bei der die Posttest-Werte statistisch um den Einfluss der Prätest-Werte bereinigt werden, d. h. die Prätest-Werte werden als Kovariate behandelt. Geht man davon aus, dass die Pretest-Messung mit dem Ernährungsfragebogen bereits einen Einfluss ausübt, wäre auch ein Solomon-Vier-Gruppen-Plan mit entsprechender Auswertung möglich.

**Aufgabe 12.5.3.3:** Eine groß angelegte experimentelle Studie zum Vergleich zwischen herkömmlichen Frontalvorlesungen, Blended-Learning-Veranstaltungen (in denen Unterrichtsmaterialien begleitend auf einer E-Learning-Plattform bereitgestellt werden) und Online-Kursen (in denen alle Inhalte auf einer E-Learning-Plattform inklusive interaktiver Lernmaterialien bereitgestellt werden) habe ergeben, dass sich die Lernleistung signifikant zwischen den verschiedenen Unterrichtsformen unterscheidet. Welche Aussagen können Sie über die Effektivität der jeweiligen Lernformen treffen?

**Aufgabe 12.5.3.3:** Eine Aussage über die Effektivität der einzelnen Lernformen kann aufgrund dieser Information nicht gegeben werden. Wenn sich die Lernleistung signifikant zwischen den Gruppen unterscheidet (Overall-Effekt der einfaktoriellen Varianzanalyse), kann zunächst nur festgehalten werden, dass sich mindestens zwei Gruppen überzufällig unterscheiden. Für genauere Aussagen müssen die deskriptiven Statistiken und insbesondere Effektgrößen betrachtet werden. Möchte man angeben, welche der drei Gruppen sich jeweils signifikant voneinander unterscheiden, betrachtet man die Ergebnisse von Post-hoc-Einzelvergleichen oder (sofern entsprechende Hypothesen vorab formuliert wurden) von Kontrasten.

**Aufgabe 12.5.3.4:** Grenzen Sie mehrfaktorielle und multivariate Varianzanalyse voneinander ab.

**Aufgabe 12.5.3.4:** Eine mehrfaktorielle Varianzanalyse berücksichtigt mehr als eine unabhängige Variable (d. h. mehr als einen Faktor), eine multivariate Varianzanalyse mehr als eine abhängige Variable. Diese Varianten können auch miteinander kombiniert werden (mehrfaktorielle multivariate Varianzanalyse).

**Aufgabe 12.5.3.5:** Welche Aussage ist richtig?

- a) Wenn in einer zweifaktoriellen Varianzanalyse beide Haupteffekte signifikant werden, wird auch der Interaktionseffekt signifikant.
- b) Ein signifikanter Interaktionseffekt A x B besagt, dass beide Faktoren additiv zusammenwirken.
- c) Ordinale Interaktion und hybride Interaktion besagen das gleiche.
- d) Bei einer disordinalen Interaktion erster Ordnung kann keiner der beiden Haupteffekte global interpretiert werden.
- e) Interaktionseffekte sind in der Forschung kaum von Interesse.

**Aufgabe 12.5.3.5:** Richtig: d

**Aufgabe 12.5.3.6:** Welche Möglichkeiten der Kontrolle von personengebundenen Störvariablen werden bei der Prüfung von Unterschiedshypothesen eingesetzt?

**Aufgabe 12.5.3.6:**

- a) Designtechnische Kontrolle: Randomisierung im Experiment.
- b) Statistische Kontrolle: Berücksichtigung von Kontrollvariablen bei quasi- und nichtexperimentellen Studien: Störvariablen können als zusätzliche unabhängige Variablen (Kontrollfaktoren) in die

Berechnung eingehen (→ mehrfaktorielle Varianzanalyse); Störvariablen können als Kontrollvariablen aus den Werten der abhängigen Variablen herauspartialisiert werden (→ Kovarianzanalyse); Störvariablen können als Kontrollvariablen mit speziellen statistischen Verfahren wie der Propensity-Score-Analyse kontrolliert werden.

**Aufgabe 12.5.3.7:** Was ist eine Interaktion zweiter Ordnung und wie benennt man sie?

**Aufgabe 12.5.3.7:** Eine Interaktion zweiter Ordnung liegt vor, wenn die Art der Interaktion zwischen zwei Faktoren von den Ausprägungen eines dritten Faktors abhängt. Sie wird oft als  $A \times B \times C$ - oder Triple-Interaktion bezeichnet.

**Aufgabe 12.5.3.8:** „Hierarchische und teilhierarchische bzw. allgemein: unvollständige Pläne bieten die Möglichkeit, trotz fehlender Werte im Datensatz eine aussagekräftige statistische Analyse der Haupteffekte durchzuführen.“ Diskutieren Sie diese Aussage!

**Aufgabe 12.5.3.8:** Einzelne fehlende Werte im Datensatz (Missings) sind abzugrenzen von unvollständigen Plänen, bei denen einzelne Faktoren systematisch untereinander geschachtelt werden. Fehlende Werte sind im Zusammenhang mit der Datenaufbereitung zu behandeln. Hierarchische Pläne sind bereits bei der Untersuchungsplanung anzulegen. Bei unvollständigen Plänen sind Interaktionseffekte gar nicht oder nur teilweise überprüfbar.

**Aufgabe 12.5.3.9:** Was stimmt nicht?

- a) Einflussreiche Kontrollvariablen korrelieren mit der abhängigen Variablen.
- b) Durch den Einfluss von Störvariablen werden Gruppenunterschiede generell abgeschwächt.
- c) Nominalskalierte Kontrollvariablen können bei der varianzanalytischen Auswertung von Unterschiedshypothesen nicht berücksichtigt werden.
- d) Eine Kovarianzanalyse kann nur in univariaten Versuchsplänen eingesetzt werden.

**Aufgabe 12.5.3.9:** Falsch sind: b, c und d

**Aufgabe 12.5.3.10:** Im Rahmen eines Experiments soll überprüft werden, inwiefern die Publikumsbewertung einer Zeichnung von der Person des Künstlers abhängt. Dazu wurde dieselbe Zeichnung 4 Gruppen von Untersuchungspersonen zur Bewertung vorgelegt. Die Gruppen erhielten jeweils eine unterschiedliche Information dazu, von wem die Zeichnung stammt: a) von einem Kind, b) von einem Patienten einer Kunsttherapiegruppe, c) von einem Hobbymaler, d) von einem renommierten Künstler. Die Untersuchungsteilnehmenden wurden gebeten, jeweils auf Ratingskalen anzugeben, wie schön sie die Zeichnung finden und wie gut sie ihnen insgesamt gefällt. Was für ein Untersuchungsplan liegt vor und wie muss diese Studie statistisch ausgewertet werden?

**Aufgabe 12.5.3.10:** Es handelt sich um einen einfaktoriellen (eine UV: Person des Künstlers), multivariaten (zwei AV: Rating von „schön“, Rating von „gefallen“) Vier-Gruppen-Plan (vier Ausprägungen der UV). Die statistische Auswertung sollte über eine einfaktorielle multivariate Varianzanalyse oder eine Diskriminanzanalyse erfolgen.

## **Kapitel 12.5.4: Datenanalyse - Quantitative Datenanalyse bei explanativen Studien**

---

**Aufgabe 12.5.4.1:** Welches sind die typischen Anwendungsfälle der inferenzstatistischen Überprüfung von Veränderungshypothesen in den Sozialwissenschaften?



**Aufgabe 12.5.4.1:** Typische Anwendungsfälle der inferenzstatistischen Überprüfung von Veränderungshypothesen in den Sozialwissenschaften:

1. Prüfung der Wirkung eines Treatments im Rahmen einer experimentellen Gruppenstudie.
2. Prüfung der Wirkung eines Treatments im Rahmen einer quasi-experimentellen Gruppenstudie.
3. Prüfung der Wirkung eines Treatments im Rahmen einer Einzelfallstudie.
4. Prüfung des Einflusses von Lebensalter, Geburtsjahrgang/Kohorte und/oder Epoche auf Entwicklungen im Rahmen von Längsschnitt-, Querschnitt- und Zeitwandelstudien.
5. Prüfung von Effekten in Zeitreihen.

**Aufgabe 12.5.4.2:** Welche drei Hypothesenarten können mit der Box-Jenkins-Zeitreihenanalyse statistisch geprüft werden?

**Aufgabe 12.5.4.2:** Drei Hypothesenarten, die mit der Box-Jenkins-Zeitreihenanalyse prüfbar sind:

1. Vorhersagemodelle: Hypothesen über die zukünftige Entwicklung der Zeitreihe.
2. Interventionsmodelle: Hypothesen über den Einfluss einer Intervention auf die Zeitreihe.
3. Transferfunktionsmodelle: Hypothesen über den Einfluss einer anderen Zeitreihe auf die betrachtete Zeitreihe.

**Aufgabe 12.5.4.3:** Was sind Sequenzeffekte und wie kann man sie kontrollieren?

**Aufgabe 12.5.4.3:** Sequenzeffekte treten bei Messwiederholungen auf, wenn dieselben Untersuchungspersonen mehrere Aufgaben nacheinander bearbeiten und die Reihenfolge der Bearbeitung einen Einfluss auf die Beantwortung bzw. Reaktion hat. Sequenzeffekte können als untersuchungsbedingte Störvariablen aufgefasst werden und sind z. B. durch Konstanthalten auszuschalten (allerdings auf Kosten der externen Validität). Häufig lässt man die Untersuchungspersonen die Aufgaben auch in unterschiedlicher Reihenfolge bearbeiten und berücksichtigt die Reihenfolge als Kontrollfaktor im Untersuchungsdesign bzw. bei der varianzanalytischen Auswertung.

**Aufgabe 12.5.4.4:** Skizzieren Sie die Vorgehensweise bei der Regressions-Diskontinuitäts-Analyse!

**Aufgabe 12.5.4.4:** Beim Regressions-Diskontinuitäts-Plan wird zur Bildung einer Experimental- und einer Kontrollgruppe eine kontinuierliche Personenvariable herangezogen (Zuweisungsvariable). Die Gruppeneinteilung wird anhand eines Cutoff-Points (z. B. Medianwert: Mediansplit) vorgenommen. Vor der Intervention wird für beide Gruppen der Zusammenhang bzw. die Regression zwischen der Zuweisungsvariablen und der abhängigen Variablen berechnet, nach dem Treatment geschieht dasselbe. Eine Wirkung des Treatments ist daran ablesbar, dass die Regression in der Experimentalgruppe nach dem Treatment anders verläuft als vor dem Treatment („Knick“ bzw. Diskontinuität) und dass die Regression in der Experimentalgruppe anders verläuft als in der Kontrollgruppe.

**Aufgabe 12.5.4.5:** Diskutieren Sie Vor- und Nachteile von Querschnitt- und Längsschnittstudien und geben Sie an, welches statistische Auswertungsverfahren typisch ist!

**Aufgabe 12.5.4.5:** Mit untersuchungstechnisch sehr aufwendigen und teuren Längsschnittstudien können intraindividuelle Veränderungen (von Angehörigen einer oder mehrerer Geburtskohorten in einer bestimmten Epoche) verfolgt werden. Längsschnittstudien sind zur Prüfung von Entwicklungshypothesen besser geeignet als Querschnittuntersuchungen, weil letztere Alterseffekte

nur interpersonal erfassen (die unterschiedlichen Altersgruppen werden parallel untersucht und entstammen verschiedenen Geburtsjahrgängen). Querschnittstudien sind untersuchungstechnisch weniger aufwendig. Die Konfundierung von Alter, Generation/Kohorte und Epoche kann mit beiden Plänen nicht gänzlich aufgelöst werden und beide Untersuchungsformen sind mit Fehlerquellen behaftet: Selektive Ausfälle (bei Querschnittstudien in der Population, bei Längsschnittstudien in der Stichprobe), mangelnde Vergleichbarkeit der Messinstrumente und damit auch der Messergebnisse (für unterschiedliche Altersgruppen) oder Testübung (beim Längsschnitt). Die statistische Auswertung erfolgt typischerweise über Varianzanalysen mit und ohne Messwiederholungen auf einem bzw. mehreren Faktoren.

**Aufgabe 12.5.4.6:** Was versteht man im Kontext der Zeitreihenanalyse unter einer Autokorrelation?

**Aufgabe 12.5.4.6:** Im Kontext der Zeitreihenanalyse meint die Autokorrelation die Korrelation der Zeitreihe mit sich selbst bei Verschiebung der Messzeitpunkte.

**Aufgabe 12.5.4.7:** Im Rahmen einer Längsschnittstudie zur Veränderung des Selbstwertgefühls über die Lebensspanne wird eine Stichprobe 20-Jähriger untersucht. Vier Folgerhebungen finden jeweils im Abstand von 10 Jahren statt. Es liegen also Selbstwertdaten der Stichprobe im Alter von 20, 30, 40, 50 und 60 Jahren vor.

- a) Welche Hypothese wird mit einer einfaktoriellen Varianzanalyse geprüft?
- b) Lässt ein signifikanter Haupteffekt der Varianzanalyse den Schluss zu, dass sich der Selbstwert durch das Lebensalter verändert? Begründung!
- c) Mit welchen Verfahren ließe sich die Hypothese testen, dass der Selbstwert im Laufe des Lebens linear ansteigt?
- d) Wie ließe sich die Hypothese prüfen, dass sich v. a. der erste und der letzte Messzeitpunkt unterscheiden?

**Aufgabe 12.5.4.7:**

- a)  $H_1$ : Zu mindestens zwei Messzeitpunkten unterscheidet sich der Selbstwert in der Population.
- b) Ein signifikanter Haupteffekt lässt sich nicht eindeutig als Alterseffekt interpretieren, denn Unterschiede zwischen den Altersstufen können durch Konfundierung auch auf Einflüsse des Messzeitpunktes zurückgehen (z. B. könnte ein Absinken des Selbstwertes mit 50 Jahren nicht durch das Lebensalter, sondern durch eine zum Untersuchungszeitpunkt auftretende Wirtschaftskrise verursacht sein). Diese Konfundierung kann z. B. aufgehoben werden, wenn vergleichend weitere Längsschnitte (mit anderen Messzeitpunkten) herangezogen werden.
- c) 1.) Mithilfe der Varianzanalyse: Trendtests; 2.) mithilfe von Strukturgleichungsmodellen: Latente Wachstumsmodelle.
- d) Durch A-priori-Kontraste anstelle der Berechnung der Overall-Signifikanz der Varianzanalyse.

**Aufgabe 12.5.4.8:** Die monatlichen Umsatzzahlen des Online-Versandhauses Amazon.de liegen für März 2005, Februar 2006, Mai 2007, Februar 2008, Mai 2009, Juni 2010, Mai 2011, Juni 2012 und Juni 2013 vor. Handelt es sich um eine Zeitreihe? Begründung!

**Aufgabe 12.5.4.8:** Es handelt sich bei diesen Umsatzzahlen nicht um eine Zeitreihe, da bei einer Zeitreihe die Messwerte in exakt gleichen zeitlichen Abständen erhoben werden.

**Aufgabe 12.5.4.9:** Auf welche drei Einflussfaktoren wird die Ausprägung eines Messwertes in einer Zeitreihe im Rahmen einer ARIMA-Analyse zurückgeführt?

**Aufgabe 12.5.4.9:** Bei einer Zeitreihenanalyse gemäß ARIMA-Modell werden drei Komponenten als Einflussfaktoren auf die Messwerte der Zeitreihe berücksichtigt:

1. vorausgehender/zurückliegender Messwert (AR-Komponente)
2. spezifischer stochastischer Trend (I-Komponente)
3. vorausgehender/zurückliegender Zufallseinfluss (MA-Komponente)

**Aufgabe 12.5.4.10:** Wozu dient der Solomon-Viergruppenplan?

**Aufgabe 12.5.4.10:** Der Solomon-Viergruppenplan dient dazu, die Veränderung bei einem Vorher-Nachher-Gruppenvergleich zu prüfen und dabei den möglichen Einfluss der Vorher-Messung zu kontrollieren.

## Kapitel 12.5.5: Datenanalyse - Quantitative Datenanalyse bei explanativen Studien

---

**Aufgabe 12.5.5.1:** Was ist unter einer hypothesenprüfenden Einzelfalluntersuchung zu verstehen?

**Aufgabe 12.5.5.1:** In einer hypothesenprüfenden Einzelfalluntersuchung geht es darum, Einzelfallhypothesen – also Annahmen über Merkmale oder Verhaltensweisen einer einzelnen Person bzw. eines einzelnen Objektes – anhand eines systematisch erhobenen quantitativen Datensatzes inferenzstatistisch zu überprüfen. Für hypothesenprüfende Einzelfallanalysen werden Verhaltensstichproben derselben Person in verschiedenen Situationen, zu unterschiedlichen Zeitpunkten und/oder unter variierenden Aufgabenstellungen gezogen.

**Aufgabe 12.5.5.2:** Welche Vor- und Nachteile haben Einzelfalluntersuchungen gegenüber Stichprobenuntersuchungen?

**Aufgabe 12.5.5.2:** Der Hauptvorteil der quantitativen Einzelfallstudie gegenüber der Stichprobenuntersuchung besteht darin, dass die Besonderheiten des Einzelfalls gut überschaubar und besser kontrollierbar sind. Bei Interventionen, die passgenau auf eine Einzelperson zugeschnitten sind (z. B. individualisierte Therapie- und Trainingsprogramme), ist die Untersuchung der entsprechenden Intervention nur am Einzelfall möglich. Der Nachteil für die Untersuchungspersonen von hypothesenprüfenden quantitativen Einzelfallstudien besteht jedoch darin, dass sie sich jeweils sehr vielen wiederholten Messungen unterziehen müssen, was je nach Art der Datenerhebungsmethode mehr oder minder aufwendig sein kann.

**Aufgabe 12.5.5.3:** Eine Schülerin erreicht in einem Intelligenztest einen IQ von 96. Heißt dies, dass sie unterdurchschnittlich intelligent ist?

**Aufgabe 12.5.5.3:** Nein, denn der Testwert ist messfehlerbehaftet. Zur genaueren Einschätzung im Rahmen der Einzelfalldiagnostik sollte unter Berücksichtigung der Reliabilität des betreffenden Tests das Konfidenzintervall des Testwertes der Schülerin bestimmt werden, das möglicherweise den Durchschnitts-IQ von 100 enthält.

**Aufgabe 12.5.5.4:** Was unterscheidet den A-BC-B-BC-Plan von dem A-B-A-B-Plan?

**Aufgabe 12.5.5.4:** Im Unterschied zum A-B-A-B-Plan, der neben der Baselinephase (A) eine einzelne Intervention (B) untersucht, wird beim A-BC-B-BC-Plan die Intervention (B) in Kombination mit einer

zweiten Intervention (C) untersucht. Die Intervention B wird dazu allein sowie in Kombination mit der Intervention C betrachtet.

**Aufgabe 12.5.5.5:** Sind die Messfehler der Daten aus Einzelfallstudien voneinander unabhängig oder abhängig – und wie ist damit umzugehen?

**Aufgabe 12.5.5.5:** Bei Einzelfallstudien handelt es sich durch die Messwiederholung an ein und derselben Untersuchungseinheit um abhängige Messungen bzw. Messwerte mit abhängigen Messfehlern. Um das Problem der seriellen Abhängigkeit von Einzelmessungen in der Single-Subject-Forschung zu umgehen, kann man Einzelmessungen z. B. zu Phasenmittelwerten zusammenfassen, die bei mindestens 10 Messwerten pro Phase in der Regel nahezu unabhängig sind.

**Aufgabe 12.5.5.6:** In einem validen Test zur Messung von Sozialkompetenz (0–100 Punkte) erzielt ein Polizist einen Testwert von 56. Nach einem Kompetenztraining wird derselbe Test erneut durchgeführt und es resultiert ein Testwert von 67. Lässt sich aus diesen Daten schlussfolgern, dass das Training die Sozialkompetenz verbessert hat?

**Aufgabe 12.5.5.6:** Nein, denn es kann sich auch um eine reine Zufallsschwankung handeln, die umso größer ausfällt, je ungenauer ein Test ist. Unter Berücksichtigung der Reliabilität des Tests sollte im Rahmen der Einzelfalldiagnostik also geprüft werden, ob die gemessene Testwert-Verbesserung tatsächlich überzufällig ist. Zudem lässt die einfache Vorher-Nachher-Messung keinen sehr eindeutigen kausalen Rückschluss zu (geringe interne Validität), da auch viele andere mögliche Einflüsse einen geringeren Vorher- und/oder einen höheren Nachher-Wert erzeugt haben könnten. Überzeugender wäre z.B. ein A-B-A-B-Design.

**Aufgabe 12.5.5.7:** Auf welche Weise kann bei dem Randomisierungstest die Teststärke erhöht werden?

**Aufgabe 12.5.5.7:** Die Teststärke kann bei einem Randomisierungstest durch mehr Messzeitpunkte – und somit mehr Messwerte – im Untersuchungsplan erhöht werden oder durch Vergrößerung der Stichprobe über eine Untersuchungseinheit ( $n = 1$ ) hinaus. Bei sehr kleinen Stichproben mit sehr ähnlichen Untersuchungseinheiten wird immer noch von Single-Subject-Research gesprochen.

**Aufgabe 12.5.5.8:** Für die Überprüfung von Einzelfallhypothesen stehen meist exakte und asymptotische Varianten des gleichen Signifikanztests zur Verfügung. Wann wird die jeweilige Testvariante eingesetzt?

**Aufgabe 12.5.5.8:** Allgemein werden exakte Signifikanztests bei kleineren Messwertreihen eingesetzt. Asymptotische Signifikanztests dienen zur Überprüfung von Einzelfallhypothesen bei größeren Messwertreihen bzw. größerer Anzahl an Messungen. Die genaue Anzahl der Messungen, ab der ein asymptotischer Signifikanztest eingesetzt wird, variiert je nach Typ des Signifikanztests.

**Aufgabe 12.5.5.9:** In einem Artikel, der den Einfluss von Gartenarbeit auf die Depressivität eines Patienten mit einer Minor Depression untersucht, lesen Sie, dass Gartenarbeit zu keiner Verbesserung der Depressivität führt. Es wurde ein A-B-A-B-Plan eingesetzt und die Auswertung der Daten erfolgte mit dem Randomisierungstest. Ist die Beibehaltung der Nullhypothese korrekt? Bitte begründen Sie Ihre Antwort!

**Aufgabe 12.5.5.9:** Die Teststärke des Randomisierungstests ist bei einem 4-Phasen-Plan nicht ausreichend. Es werden bei  $n = 1$  insgesamt mindestens 8 Phasen benötigt, damit der Randomisierungstest eine ausreichende Teststärke aufweist. Da die Teststärke nicht ausreichend war, ist bei einem nicht-signifikanten Ergebnis nicht klar, ob wirklich kein Effekt vorliegt, oder ein vorliegender Effekt nicht aufgedeckt werden konnte.

**Aufgabe 12.5.5.10:** Wie ist im Rahmen der Einzelfalldiagnostik bei einem Vergleich von einem Untertestwert mit dem Gesamttestwert vorzugehen?

**Aufgabe 12.5.5.10:** Um zu bewerten, ob ein Untertestwert einer Testperson signifikant von ihrem Gesamttestwert abweicht, greift man auf die anhand von Eichstichproben ermittelten Testkennwerte zurück. Anhand der bekannten Korrelation des Untertests mit dem Gesamttest wird für den Gesamttestwert der Person der theoretisch erwartete zugehörige Untertestwert regressionsanalytisch bestimmt. Der empirische Untertestwert der Testperson wird mit diesem theoretischen erwarteten Wert verglichen, d. h. die Differenz wird auf Signifikanz geprüft.

## Kapitel 13: Ergebnispräsentation

---

**Aufgabe 13.1:** Was versteht man unter doppelblindem Peer-Review?

**Aufgabe 13.1:** Die Begutachtung (review) von wissenschaftlichen Fachartikeln, Vorträgen, Forschungsanträgen etc. durch Fachkollegen (peers). Dabei bleiben sowohl die Gutachtenden als auch die Begutachteten wechselseitig füreinander anonym (doppelblindeg Vorgehen).

**Aufgabe 13.2:** Worin unterscheidet sich der Peer-Review-Prozess bei Zeitschriftenartikeln von dem bei Konferenzvorträgen?

**Aufgabe 13.2:**

- Peer-Review-Prozess bei wissenschaftlichen Konferenzvorträgen: einstufiges Peer Review, keine Nachbesserung möglich, dafür relativ zügig binnen mehrerer Wochen abgeschlossen.
- Peer-Review-Prozess bei wissenschaftlichen Zeitschriftenartikeln: mehrstufiges Peer Review mit Möglichkeit zur ein- oder mehrmaligen Nachbesserung, kann mehrere Monate bis Jahre in Anspruch nehmen.

**Aufgabe 13.3:** Was ist der Impact-Factor und wann steigt er?

**Aufgabe 13.3:** Der Impact-Faktor ist eine bibliometrische Maßzahl, die die Relevanz einer wissenschaftlichen Fachzeitschrift beziffert. Je häufiger Artikel einer Zeitschrift in anderen Zeitschriften zitiert werden, umso höher wird ihr jährlich neu berechneter Impact-Faktor.

**Aufgabe 13.4:** Nennen Sie Unterschiede im Umgang mit Konferenzvorträgen in den Geistes-, Sozial- und Technikwissenschaften!

**Aufgabe 13.4:** Bedeutung von Konferenzvorträgen in unterschiedlichen Fachdisziplinen:

- Technikwissenschaften: Konferenzvorträge haben einen hohen Stellenwert und gelten als vollwertige Publikationen durch ihre Veröffentlichung in Konferenzbänden bzw. Proceedings.
- Sozialwissenschaften: Konferenzvorträge haben einen geringeren Stellenwert als Zeitschriftenartikel, werden frei vorgetragen.

- Geistes- und Kulturwissenschaften: Konferenzvorträge werden oft wörtlich ausformuliert und vorgelesen.

**Aufgabe 13.5:** Welches Format haben wissenschaftliche Poster üblicherweise?

**Aufgabe 13.5:** Wissenschaftliche Poster haben meist das Format DIN A0 (Hochformat).

**Aufgabe 13.6:** Welche Vortragsformen auf Konferenzen sind zu unterscheiden?

**Aufgabe 13.6:** Drei Formen von Konferenzvorträgen:

1. Forschungsreferat: Präsentation einer empirischen Studie
2. Positionsreferat: Forschungsüberblick zu einem Thema
3. Grundsatzreferat: eingeladener Beitrag von einer bekannten Forscherpersönlichkeit zu einem bestimmten Thema oder Forschungsgebiet

**Aufgabe 13.7:** Warum sind die Erfolgchancen beim Peer-Review bei wissenschaftlichen Postern meist höher als bei Konferenzvorträgen?

**Aufgabe 13.7:** Die Beurteilungskriterien sind bei Vorträgen strenger, u. a. weil die Zeitfenster für Vorträge auf einer Konferenz limitiert sind, während in Posterausstellungen organisatorisch relativ problemlos zusätzliche Poster aufgenommen werden können.

**Aufgabe 13.8:** Worin unterscheiden sich wissenschaftliche Fachzeitschriften und Transferzeitschriften?

**Aufgabe 13.8:** Wissenschaftliche Fachzeitschriften richten sich an ein wissenschaftliches Publikum, Transferzeitschriften dagegen vermitteln wissenschaftliche Ergebnisse an Fachleute aus der Praxis, vor allem aus der Wirtschaft.

**Aufgabe 13.9:** Welche Ergebnisse kann die Begutachtung eines bei einer Fachzeitschrift eingereichten Artikels haben?

**Aufgabe 13.9:** Mögliche Ergebnisse beim Peer Review einer Artikeleinreichung bei einer Fachzeitschrift:

- Ablehnung
- Einladung zu einer erneuten Einreichung und Begutachtung nach grundlegender Überarbeitung
- Annahme zur Publikation (mit ggf. kleinen Korrektur-/Überarbeitungsaufträgen)

**Aufgabe 13.10:** Welche Einreichungsformen für Vorträge fordern Calls wissenschaftlicher Fachkonferenzen?

**Aufgabe 13.10:** Calls von wissenschaftlichen Konferenzen verlangen für die Einreichung von Vorträgen üblicherweise

- Abstract,
- Extended Abstract oder
- Full Paper.

## Kapitel 14: Bestimmung von Teststärke, Effektgröße und optimalem Stichprobenumfang

---

**Aufgabe 14.1:** Was trifft zu? Ein Signifikanztest wird umso eher signifikant,

- a) je größer/kleiner der Effekt,
- b) je größer/kleiner der Stichprobenumfang,
- c) je größer/kleiner das Signifikanzniveau,
- d) je größer/kleiner die Teststärke.

**Aufgabe 14.1:**

- a) Je größer der Effekt
- b) Je größer der Stichprobenumfang
- c) Je größer (z. B. 5 % > 1 %) das Signifikanzniveau
- d) Je größer die Teststärke

**Aufgabe 14.2:** Was versteht man unter der „Effektgröße“?

**Aufgabe 14.2:** Die inhaltlich interessierenden Populationsparameter (z. B. Häufigkeiten, Mittelwertunterschiede, Variablenzusammenhänge; Häufigkeits- oder Mittelwertveränderungen) bezeichnet man als „Effekte“ und ihre Größe bzw. ihren Betrag als „Effektgröße“ oder „Effektstärke“. Die unstandardisierte Effektgröße (z. B. einfache Mittelwertdifferenz) ist nur bei verhältnisskalierten Variablen gut interpretierbar, meist werden deswegen standardisierte Effektgrößenmaße verwendet, die von der Messeinheit und vom Stichprobenumfang unabhängig sind. Standardisierte Effektgrößen werden gebildet, indem der Effekt an der Merkmalsstreuung relativiert wird (z. B. statt der einfachen Mittelwertdifferenz wird das standardisierte Effektgrößenmaß  $d$  von Cohen verwendet).

**Aufgabe 14.3:** Was ist mit der „Power“ im Kontext von Signifikanztests gemeint und wovon hängt sie ab?

**Aufgabe 14.3:** Die Power (Teststärke) ist die Wahrscheinlichkeit, mit der ein Signifikanztest zu einem signifikanten Ergebnis führt, wenn in der Population die  $H_1$  (es muss sich dabei um eine spezifische Hypothese handeln) gilt. Man sagt auch: Die Teststärke ist die Wahrscheinlichkeit, mit der ein Signifikanztest einen vorhandenen Populationseffekt bestimmter Größe durch ein signifikantes Ergebnis aufdecken kann. Die Teststärke ist umso größer, je größer der Effekt, je größer der Stichprobenumfang und je größer das Signifikanzniveau (z. B. 5 % > 1 %).

**Aufgabe 14.4:** Anhand von Urlauberbefragungen hat ein Tourismusunternehmen 30 zufällig ausgewählte Ferienorte am Meer mit einem Punktwert versehen, der die Zufriedenheit mit dem Urlaubsort ausdrückt. Dann wurde untersucht, welche Bedeutung folgende Faktoren für die Bewertung des Urlaubsortes haben: Tageshöchsttemperatur, Wassertemperatur, Seegang, Wasserverschmutzung, Verschmutzung des Strandes, Anzahl der Besucher am Strand sowie Anzahl der aktiven Wassersportler im Umkreis von 1 km vor dem Strand. Interessanterweise konnte kein einziger signifikanter Effekt nachgewiesen werden. Woran könnte das liegen?

**Aufgabe 14.4:** Der Stichprobenumfang der Studie und somit die Teststärke waren vermutlich zu klein, vor allem in Anbetracht der Menge der untersuchten Variablen. Es kann natürlich auch sein, dass tatsächlich kein Zusammenhang zwischen den untersuchten Merkmalen des Urlaubsortes und der Zufriedenheit besteht, z. B. weil sich die Zufriedenheit an anderen Merkmalen orientierte.

**Aufgabe 14.5:** Welches Problem tritt auf, wenn eine empirische Studie „underpowered“ ist?

**Aufgabe 14.5:** Wenn eine Studie „underpowered“ ist, hat sie eine ungenügende Teststärke, etwa weil sie auf einem – im Hinblick auf den untersuchten Effekt und dessen Größe – zu geringen Stichprobenumfang basiert, weil viele Störvariablen einwirken oder weil die unabhängige Variable zu schwach dosiert wurde und somit nur ein sehr kleiner Effekt untersuchbar war. Im Ergebnis liefert eine solche Studie nicht-signifikante Resultate, die hinsichtlich Hypothesenprüfung nicht aussagekräftig sind.

**Aufgabe 14.6:** Grenzen Sie Vorgehensweise und Zielsetzung der A-priori- und der Post-hoc-Teststärkeanalyse voneinander ab!

**Aufgabe 14.6:** Bei der A-priori-Teststärkeanalyse wird mit Blick auf die interessierende Effektgröße der optimale Stichprobenumfang für eine geplante Studie im Voraus bestimmt. Bei der Post-hoc-Teststärkeanalyse wird bei einer bereits durchgeführten Studie im Nachgang (vor allem im Falle eines nicht-signifikanten Ergebnisses) die erreichte Teststärke bestimmt, um die Aussagekraft des Ergebnisses einschätzen zu können. Denn ein nicht-signifikantes Resultat darf nur bei ausreichender Teststärke als Hinweis auf die Gültigkeit der Nullhypothese gewertet werden. Bei einem signifikanten Ergebnis erübrigt sich meist eine Post-hoc-Teststärkeanalyse: Denn wenn ein signifikantes Ergebnis vorliegt, hatte der Test offenbar genügend Power, um Signifikanz erreichen zu können. (Es sei denn, das signifikante Ergebnis resultiert aus einem Alpha-Fehler, worauf geringe Teststärke bei einem signifikanten Ergebnis hindeuten würde.) Anstelle bzw. ergänzend zu einer Post-hoc-Teststärkeanalyse wird bei einem signifikanten Ergebnis vor allem die Effektgröße betrachtet, um sicherzugehen, dass der Test nicht durch übermäßige Teststärke einen praktisch bedeutungslosen Effekt hat signifikant werden lassen.

**Aufgabe 14.7:** Warum gibt es viele verschiedene Effektgrößenmaße?

**Aufgabe 14.7:** Ein Effektgrößenmaß ist eine standardisierte Maßzahl für einen Effekt. Für unterschiedliche Arten von Effekten (z. B. Unterschied zwischen zwei oder mehr Gruppen, Zusammenhang zwischen zwei oder mehr Variablen, Veränderung zwischen zwei oder mehr Messzeitpunkten etc.) liegen jeweils unterschiedliche Effektgrößenmaße vor.

**Aufgabe 14.8:** Welche Typen von Effektgrößenmaßen werden unterschieden?

**Aufgabe 14.8:** Die unterschiedlichen Effektgrößenmaße lassen sich in vier Gruppen einteilen: Gruppendifferenz-, Zusammenhangs-, Varianzaufklärungs- und Risiko-Effektgrößenmaße. Andere Klassifikationen sind möglich, etwa die Einteilung in zwei übergeordnete Gruppen: Effektgrößenmaße der  $d$ -Familie (Gruppendifferenz- und Risiko-Effektgrößenmaße) sowie Effektgrößenmaße der  $r$ -Familie (Zusammenhangs- und Varianzaufklärungsmaße).

**Aufgabe 14.9:** Warum sollten die Ergebnisse einer quantitativen empirischen Studie immer auch in Form von Effektgrößenmaßen berichtet werden?

**Aufgabe 14.9:** Im Unterschied zur bloßen Signifikanzaussage (die wesentlich vom Stichprobenumfang der Studie abhängt), liefern Effektgrößenmaße Hinweise auf die theoretische und/oder praktische Bedeutsamkeit eines Effektes (und zwar unabhängig vom Stichprobenumfang der Studie). Darüber hinaus sind Angaben zur Effektgröße notwendig, um verschiedene Studien zur selben Fragestellung miteinander vergleichen oder im Rahmen einer Metaanalyse zu einer Schätzung des Gesamteffektes verrechnen zu können.



**Aufgabe 14.10:** Welche Typen von Teststärkeanalysen werden differenziert?

**Aufgabe 14.10:**

1. Post-hoc-Teststärkeanalyse
2. A-priori-Teststärkeanalyse
3. Sensitivitäts-Teststärkeanalyse
4. Kriteriums-Teststärkeanalyse
5. Kompromiss-Teststärkeanalyse

## Kapitel 15: Minimum-Effektgrößen-Tests

---

**Aufgabe 15.1:** Worin unterscheidet sich die traditionelle Nullhypothese des statistischen Signifikanztests von einer Minimum-Effekt-Nullhypothese?

**Aufgabe 15.1:** Die traditionelle Nullhypothese bzw. Nil-Nullhypothese ( $H_{00}$ ) postuliert einen exakten Nulleffekt in der Population, es handelt sich um eine Punkthypothese, die in der Praxis sehr unrealistisch ist. Demgegenüber postuliert eine Minimum-Effekt-Nullhypothese um Null herum einen ganzen Bereich vernachlässigbar kleiner Effekte (typischerweise Varianzaufklärungen bis zu 1 %:  $H_{01}$  bzw. bis zu 5 %:  $H_{05}$ ) und ist somit als Bereichshypothese auch realistischer.

**Aufgabe 15.2:** Was stimmt? Die Teststärke eines Signifikanztests ist umso höher,

- a) je größer der Stichprobenumfang,
- b) je liberaler das Signifikanzniveau,
- c) je größer der Populationseffekt und/oder
- d) je geringer die Beta-Fehler-Wahrscheinlichkeit.

**Aufgabe 15.2:** Richtig sind: a, b, c und d

**Aufgabe 15.3:** Empfiehlt es sich, in Zukunft nur noch  $H_{05}$ -Hypothesen zu überprüfen um sicherzustellen, dass keine praktisch zu vernachlässigenden Effekte postuliert werden? Begründen Sie!

**Aufgabe 15.3:** Nein. Die  $H_{05}$  sollte nur in gut begründeten Ausnahmefällen geprüft werden. Denn 5 % Varianzaufklärung sind in den Sozialwissenschaften oft nicht als vernachlässigbar geringe, sondern bereits als inhaltlich bedeutsame kleine bis mittlere Effekte einzuordnen. Stattdessen empfiehlt es sich eher, als Minimum-Effekt-Nullhypothese die  $H_{01}$  zu prüfen. Denn bis zu 1 % Varianzaufklärung ist vermutlich eine Größenordnung, die für die meisten Forschungsbereiche tatsächlich theoretisch und/oder praktisch zu vernachlässigen ist. Derartige Pauschalaussagen über die Bedeutung von Effektgrößen sind jedoch immer gegenstandsspezifisch zu relativieren.

**Aufgabe 15.4:** Welche Aussage trifft jeweils zu?

- a) Eine  $H_{05}$  erklärt eine bivariate Populationskorrelation von  $\geq .15$  für unbedeutend klein.
- b) Bei einem traditionellen Signifikanztest kann eine multiple Korrelation von  $R = .007$  in der Stichprobe statistisch signifikant werden.
- c) Beim Test einer  $H_{05}$  kann eine bivariate Korrelation von  $r = .20$  nicht signifikant werden.
- d) Wenn ein Test der  $H_{01}$  signifikant wird, ist von mangelhafter Teststärke auszugehen.
- e) Wenn ein Test gegen die  $H_{05}$  signifikant geworden ist, weiß man, dass auch ein Test gegen  $H_{01}$  und  $H_{00}$  statistisch signifikant ausfallen würde.

**Aufgabe 15.4:** Richtig sind: b, c und e

**Aufgabe 15.5:** Begründen Sie, warum Untersuchungen mit einer Teststärke unter 50 % nicht durchgeführt bzw. nicht publiziert werden sollten. Welche Teststärke ist zu fordern?

**Aufgabe 15.5:** Bei Studien mit einer Teststärke  $1 - \beta \leq .50$  ist im Falle eines nicht-signifikanten Ergebnisses unklar, ob tatsächlich kein Effekt in der Population vorliegt oder der Test nicht in der Lage war, den Effekt aufzudecken (z. B. aufgrund zu geringer Stichprobengröße oder zu großer Merkmalsstreuungen). Das Ergebnis der Studie muss deswegen als nicht interpretierbar gelten. Teststärken ab 80 % gelten als ausreichend.

**Aufgabe 15.6:** Was versteht man unter dem optimalen Stichprobenumfang und wann und wie wird er festgelegt?

**Aufgabe 15.6:** Der optimale Stichprobenumfang bezeichnet denjenigen Stichprobenumfang, der – je nach erwarteter Effektgröße und festgelegtem Signifikanzniveau – eine ausreichende Teststärke von 80 % sicherstellt. Optimale Stichprobenumfänge für unterschiedliche Tests sind austabelliert und können mit entsprechenden Software-Tools bestimmt werden (z. B. G\*Power). Der optimale Stichprobenumfang wird idealerweise vor der Untersuchungsdurchführung bestimmt (A-priori-Teststärkeanalyse).

**Aufgabe 15.7:** Was lässt sich über die Teststärken von Signifikanztests aussagen, bei denen eine Nil-Nullhypothese, eine 1 %-Minimum-Effekt-Nullhypothese oder eine 5 %-Minimum-Effekt-Nullhypothese verwendet wird?

**Aufgabe 15.7:** Unter sonst gleichen Bedingungen ist die Teststärke bei einem Signifikanztest mit Nil-Nullhypothese höher als bei 1 %-Minimum-Effekt-Nullhypothese. Bei einer 5 %-Minimum-Effekt-Nullhypothese ist die Teststärke am geringsten.

**Aufgabe 15.8:** Was ist zu beachten, wenn bei einem Signifikanztest die Nullhypothese (Nil-Nullhypothese oder Minimum-Effekt-Nullhypothese) die von den Forschenden inhaltlich begründete und zu prüfende Wunschhypothese ist?

**Aufgabe 15.8:** Wenn die Nullhypothese die Wunschhypothese ist, muss die Teststärke des Signifikanztests möglichst hoch sein. Dies wird erreicht durch

- a) großen bzw. optimalen Stichprobenumfang,
- b) liberalisiertes Signifikanzniveau ( $\alpha = 10\%$ ) sowie
- c) zweiseitigen Test.

**Aufgabe 15.9:** Eine Grundsatzkritik am klassischen Nil-Nullhypothesen-Signifikanztest lautet, er sei informationslos, weil letztlich jede empirische Studie ein signifikantes Ergebnis erzeugen könne, wenn nur eine entsprechend große Stichprobe gewählt wird. Wie ist mit dieser Kritik bei der statistischen Hypothesenprüfung umzugehen?

**Aufgabe 15.9:** Der klassische Signifikanztest ist nicht informationslos, wenn die statistische Signifikanz im Kontext der Stichprobengröße, Teststärke und Effektgröße interpretiert wird.

**Aufgabe 15.10:** Was versteht man im Kontext von Minimum-Effekt-Tests unter einem F-Äquivalent?

**Aufgabe 15.10:** Ein  $F$ -Äquivalent drückt einen Populationseffekt aus, bei dem der Signifikanztest eine definierte Teststärke von 50 % oder 80 % hat.

## Kapitel 16: Metaanalyse

---

**Aufgabe 16.1:** Grenzen Sie vier verschiedene Formen der wissenschaftlichen Forschungssynthese charakterisierend voneinander ab!

### **Aufgabe 16.1:**

1. Narratives Review: Überblick über ein Forschungsfeld oder eine Forschungsfrage, wobei Literatursammlung und Ergebnisbewertung keinen festen Regeln folgen.
2. Systematisches Review: Überblick über ein Forschungsfeld oder eine Forschungsfrage, wobei Literatursammlung und Ergebnisbewertung festen Regeln folgen (z. B. PRISMA-Framework) und oft auch eine (quantitative) Metaanalyse enthalten.
3. Quantitative Metaanalyse: Statistische Zusammenfassung eines einheitlichen Pools an quantitativen Studien zur Schätzung des Gesamteffekts (meist inklusive Moderatorvariablen-Analyse).
4. Qualitative Metaanalyse: Interpretative Zusammenfassung eines einheitlichen Pools an qualitativen Studien zur Beantwortung einer konkreten Forschungsfrage (in der Praxis bislang kaum etabliert).

**Aufgabe 16.2:** Was versteht man im Kontext der Metaanalyse unter abhängigen Untersuchungsergebnissen? Wie ist mit ihnen zu verfahren?

**Aufgabe 16.2:** Abhängige Untersuchungsergebnisse sind Teilergebnisse einer einzelnen Primärstudie, die an derselben Stichprobe gewonnen wurden. Teilergebnisse einer Primärstudie sollten nicht mehrfach in eine Metaanalyse eingebracht werden. Entweder man wählt ein Teilergebnis pro Primärstudie aus oder man aggregiert die Teilergebnisse einer Primärstudie zu einem Gesamtbefund, der einmalig mit dem Stichprobenumfang der Primärstudie in die Metaanalyse eingeht.

**Aufgabe 16.3:** Was ist das  $\Delta$ -Maß und wozu wird es benötigt?

**Aufgabe 16.3:** Das  $\Delta$ -Maß ist ein universelles Effektgrößenmaß, in das fast jedes testspezifische Effektgrößenmaß (z. B. Gruppendifferenz-, Zusammenhangs- oder Varianzaufklärungsmaß) überführt werden kann. Die  $\Delta$ -Maße verschiedener Tests bzw. Untersuchungen sind dann direkt vergleichbar bzw. statistisch im Rahmen einer Metaanalyse aggregierbar. Das  $\Delta$ -Maß entspricht der bivariaten Produkt-Moment-Korrelation, d. h., es ist ein Korrelationsäquivalent.

**Aufgabe 16.4:** Was sind kombinierte Signifikanztests?

**Aufgabe 16.4:** Wenn Primärstudien metaanalytisch zusammengefasst werden sollen, bei denen Angaben über Effektgrößen ganz oder teilweise fehlen, können mithilfe eines kombinierten Signifikanztests zumindest drei Typen von Aussagen getroffen werden:

- Aussagen über statistisch signifikante Ergebnisse:
  - Deskriptiv: Ob signifikante Ergebnisse häufiger als nicht-signifikante auftreten (Vote Counting).
  - Test: Ob überzufällig mehr Ergebnisse signifikant als nicht-signifikant sind (Binomialtest).
- Aussagen über positive (hypothesenkonforme) Ergebnisse:

- Test: Ob mehr positive (hypothesekonforme) als negative Ergebnisse (unabhängig von ihrer Signifikanz) vorliegen (Vorzeichenstest).
- Aussagen über die Irrtumswahrscheinlichkeiten:
  - Test: Ob die exakten Irrtumswahrscheinlichkeiten der Studien insgesamt die Signifikanzschwelle von 5 % unterschreiten (Stouffer-Test).
  - Test: Zusammenfassung exakter Irrtumswahrscheinlichkeiten und Stichprobenumfänge bei Zweigruppenvergleichen durch ein Korrelationsäquivalent ( $r_{\text{equivalent}}$ ) und Konfidenzintervallbestimmung.

**Aufgabe 16.5:** Was versteht man unter dem Publication-Bias? Wie geht man mit diesem Problem im Kontext der Metaanalyse um?

**Aufgabe 16.5:** Mit dem Publication-Bias ist gemeint, dass veröffentlichte Studien oftmals in der Richtung verzerrt sind, dass sie verstärkt hypothesekonforme und signifikante Ergebnisse berichten und somit die untersuchten Effekte systematisch überschätzen. Denn hypothesekonträre und nicht-signifikante Ergebnisse, die ebenfalls vorliegen und für die Gesamtbewertung des Effektes wichtig sind, bleiben oft unveröffentlicht.

Da der Publication-Bias zu einer systematischen Überschätzung des interessierenden Effektes führt, sollten im Kontext der Metaanalyse folgende Gegenmaßnahmen ergriffen werden:

1. Bei der Studienbeschaffung bemüht man sich, auch unveröffentlichte Studien (z. B. interne Projektberichte) zu finden und einzubeziehen, um einen weniger verzerrten Studienpool zu bilden (Maßnahmen hierzu sind jedoch nicht standardisiert, sodass von diesem Vorgehen teilweise auch abgeraten wird).
2. Man fertigt einen Funnel-Plot an und prüft, ob die Darstellung auffällige Asymmetrien aufweist. Ist der Funnel-Plot weitgehend symmetrisch, liegt kein nennenswerter Publication-Bias vor.
3. Man berechnet eine widerlegungssichere Untersuchungszahl, d. h. das „Fail-Safe- $N$ “ ( $N_{FS}$ ), und prüft, ob  $N_{FS} \geq 5 \cdot k + 10$  ist. In diesem Falle kann man davon ausgehen, dass ein signifikantes metaanalytisches Ergebnis der  $k$  Untersuchungen „widerlegungssicher“ ist.
4. Ist auf der Basis des Funnel-Plots und/oder der Fail-Safe- $N$ -Methode von einem deutlichen Publication-Bias auszugehen, so kann mittels Trim-and-Fill-Prozedur eine korrigierte Effektgrößenschätzung vorgenommen werden, die darauf basiert, dass der Einfluss unveröffentlicher „Schubladen-Studien“ abgeschätzt wird.
5. Verzerrte Ergebnisse im Studienpool können auch durch eine P-Curve- bzw. Z-Curve-Analyse identifiziert werden.

**Aufgabe 16.6:** Wozu dient eine A-priori-Teststärkeanalyse im Kontext der Metaanalyse, welche Probleme sind damit verbunden und wie ist mit ihnen umzugehen?

**Aufgabe 16.6:** Mittels A-priori-Teststärkeanalyse wird generell bestimmt, welcher Mindeststichprobenumfang („optimaler Stichprobenumfang“) notwendig ist, damit ein Signifikanztest anhand der verwendeten Datenbasis überhaupt die Chance hat, einen Populationseffekt bestimmter Größe durch ein signifikantes Ergebnis aufzudecken. Prinzipiell sind A-priori-Teststärkeanalysen bei Metaanalysen ebenso wie bei Primärstudien sinnvoll, damit keine Studien geplant und durchgeführt werden, die sich im Nachgang mangels Teststärke als gar nicht aussagekräftig erweisen („underpowered studies“).

Im Kontext der Metaanalyse durch A-priori-Teststärkeanalysen festzulegen, wie viele Primärstudien der zu aggregierende Studienpool enthalten muss, um a) die Homogenität der Einzeleffekte und b) die Signifikanz des Gesamteffekts mit ausreichender Teststärke prüfen zu können, ist vor allem deswegen problematisch, weil für die Teststärkebestimmung (vor allem des Homogenitätstests) im Vorfeld sehr

genaue Kenntnisse des zu sammelnden Studienpools notwendig sind (Anzahl der Studien, Fallzahl pro Studie, mögliche Moderatorvariablen, Anzahl der Subgruppen pro Moderatorvariable, Anzahl der Studien pro Subgruppe, Effektgrößen in den Subgruppen etc.). Entsprechende Vorgaben sind meist sehr spekulativ. Selbst wenn es plausibel möglich wäre, diese Randbedingungen realistisch abzuschätzen und auf dieser Basis einen optimalen Stichprobenumfang festzulegen, stellt sich dann die Frage, ob es im jeweiligen Forschungsfeld überhaupt möglich ist, eine entsprechend große Zahl an thematisch einschlägigen und methodisch hochwertigen Studien zu finden.

Angesichts dieser Probleme bleibt oft nur die Möglichkeit, die A-priori-Teststärkeanalyse zu überspringen und das Teststärkeproblem a posteriori zu behandeln, also nachträglich mögliche Teststärkeprobleme der Metaanalyse aufzudecken und in der Ergebnis-Diskussion anzusprechen. Es lässt sich argumentieren, dass eine metaanalytische Auswertung eines Studienpools auch bei geringer Studienzahl sinnvoll ist, da sie in eine systematische Zusammenfassung mündet, die methodenkritisch bewertet wird. Wird auf diese Form der systematischen Forschungssynthese verzichtet, versuchen Leser:innen oft intuitiv den Gesamteffekt abzuschätzen (z. B. durch einfaches Auszählen von Signifikanzen), was noch stärkeren Verzerrungen unterworfen ist.

**Aufgabe 16.7:** Was unterscheidet eine Metaanalyse von einer Primärstudie und was sind die Vorteile der Metaanalyse?

**Aufgabe 16.7:** Eine Metaanalyse fasst die Ergebnisse einzelner Primärstudien zusammen, die sich auf die gleiche Fragestellung oder Hypothese beziehen. In der Metaanalyse sind jeweils ganze Primärstudien die Untersuchungseinheiten, während in Primärstudien in der Regel einzelne Personen oder Objekte die Untersuchungseinheiten bilden. Während Primärstudien Einzeleffekte abschätzen, wird mittels Metaanalysen ein Gesamteffekt in der Population geschätzt. Durch die breitere Datenbasis und damit höhere Teststärke ist die Metaanalyse besser als eine Primärstudie in der Lage, einen wahren Populationseffekt aufzudecken und dessen Größe präzise abzuschätzen. Zusätzlich können im Zuge einer Metaanalyse mittels Moderatorvariablen-Analyse die Umstände untersucht werden, unter denen der Effekt stärker oder schwächer ausfällt. Zusammenfassend lässt sich festhalten, dass eine Metaanalyse eine Primärstudie hinsichtlich interner und statistischer Validität sowie externer und Konstrukt-Validität in der Regel übertrifft.

**Aufgabe 16.8:** Welche Aussagen sind richtig?

- Das Fixed-Effect-Modell geht davon aus, dass sich die zu aggregierenden Primärstudien systematisch unterscheiden.
- Das Random-Effects-Modell postuliert, dass in jeder Primärstudie eine andere Populationseffektgröße geschätzt wird.
- Ob ein Fixed-Effect- oder ein Random-Effects-Modell zum Einsatz kommen soll, ist vor der Durchführung der Metaanalyse inhaltlich-konzeptuell zu entscheiden.
- Im Fixed-Effect-Modell und im Random-Effects-Modell werden Studien mit geringem Stichprobenumfang geringer gewichtet.

**Aufgabe 16.8:** Richtig sind: b, c und d

**Aufgabe 16.9:** Welche Aussagen sind falsch?

- Die Teststärke einer Metaanalyse übertrifft durch den größeren Gesamtstichprobenumfang die Teststärke der Primärstudien.
- Werden die unabhängigen Variablen in den Primärstudien unterschiedlich operationalisiert, sollte die Art der Operationalisierung als Moderatorvariable in der Metaanalyse berücksichtigt werden.

- c. Eine Zusammenfassung von Effektgrößen im Fixed-Effect-Modell ist nur sinnvoll, wenn die untersuchungsspezifischen Effektgrößen zumindest schwach heterogen sind.
- d. Eine vergleichbare Operationalisierung der abhängigen Variablen ist bei der Metaanalyse im Random-Effects-Modell nicht notwendig.

**Aufgabe 16.9:** Falsch sind: c und d

**Aufgabe 16.10:** In welchen inhaltlichen Forschungsfeldern wurde die quantitative Metaanalyse ursprünglich entwickelt und wo kommt sie heute vermehrt zum Einsatz?

**Aufgabe 16.10:** Die quantitative Metaanalyse wurde in der Psychologie in den 1970er-Jahren entwickelt, vor allem im Zusammenhang mit der Erforschung von Psychotherapie-Wirkungen und interpersonalen Erwartungseffekten (Versuchsleiter-Effekt, Lehrererwartungs-Effekt). Heute spielt sie in der Psychologie, der Medizin und der Erziehungswissenschaft eine große Rolle. Sie dient vor allem dazu, den Forschungsstand zur Wirksamkeit von Maßnahmen (psychotherapeutische, medizinische, pädagogische Interventionen) zusammenfassend bewerten zu können. In der Medizin hat sich die Cochrane-Collaboration und in den Sozialwissenschaften die Campbell-Collaboration jeweils als Expertennetzwerk gebildet, das systematische Reviews und Metaanalysen nach strengen Qualitätskriterien erstellt und publiziert.

## Kapitel 17: Strukturgleichungsmodelle

---

**Aufgabe 17.1:** Welche Vorteile bietet die Analyse von Strukturgleichungsmodellen gegenüber Methoden auf der Ebene von beobachteten Variablen, wie z. B. Regressionsanalysen?

**Aufgabe 17.1:** Es gibt zumindest vier mögliche Vorteile.

1. **Reliabilität:** Beobachtete Variablen enthalten oftmals Messfehleranteile. Bei mehreren Messungen pro Konstrukt können zufällige Messfehler in Strukturgleichungsmodellen berücksichtigt werden, sodass Schätzungen der Beziehungen zwischen Konstrukten unbeeinträchtigt von Messfehlern sind.
2. **Validität:** Jedes Konstrukt kann durch mehrere unterschiedliche beobachtete Variablen operationalisiert werden (z. B. mehrere Tests oder Fragebögen). Wenn keine beobachtete Variable für sich alleine das Konstrukt optimal repräsentiert, werden durch die gleichzeitige Berücksichtigung mehrerer beobachteter Variablen validere Aussagen über das Konstrukt möglich.
3. **Komplexität:** In Strukturgleichungsmodellen können komplexe Beziehungen vieler Variablen simultan analysiert werden, für die sonst mehrere einzelne Analysen durchgeführt werden müssten.
4. **Konfirmatorische Analyse:** Mit Strukturgleichungsmodellen kann man konfirmatorisch prüfen, ob alle insgesamt bestehenden Beziehungen zwischen den beobachteten Variablen durch aufgestellte theoretische Annahmen zu erklären sind oder nicht.

**Aufgabe 17.2:** Was sind manifeste und latente Variablen?

**Aufgabe 17.2:** Manifeste Variablen sind beobachtete Variablen (Indikatorvariablen), die zur Messung von latenten Variablen verwendet werden. Latente Variablen entsprechen theoretischen Konstrukten und sind damit nicht direkt beobachtbar. Ein inhaltlich definiertes psychologisches Konstrukt kann als latente Variable eines Strukturgleichungsmodells aufgefasst werden, wenn es mit lokal stochastisch unabhängigen manifesten Indikatorvariablen gemessen werden kann.

**Aufgabe 17.3:** Was ist eine Mediatorvariable?

**Aufgabe 17.3:** Eine Mediatorvariable ist eine Variable, die ein Zwischenglied darstellt in einer angenommenen Beeinflussungskette von einer ersten Variable (Prädiktor) über eine zweite Variable (Mediator) auf eine dritte Variable (Kriterium). Die Mediatorvariable ist damit gleichzeitig abhängige Variable (sie wird durch die erste Variable vorhergesagt) wie auch unabhängige Variable (sie selbst sagt die dritte Variable vorher). Die Analyse solcher Mediatoreffekte ist mit Strukturgleichungsmodellen gut möglich, weil Variablen darin gleichzeitig Prädiktor- und Kriteriumsvariable sein können. Zur Abgrenzung: Bei Mediation verläuft die Beziehung zwischen Prädiktor- und Kriteriumsvariable vermittelt (mediert) über eine dritte Variable (Mediatorvariable) zwischen Prädiktor und Kriterium. Bei Moderation dagegen verändert sich die Stärke der Beziehung zwischen Prädiktor- und Kriteriumsvariable abhängig vom Wert einer dritten Variable (der Moderatorvariablen).

**Aufgabe 17.4:** Welche Teilmodelle kann man in Strukturgleichungsmodellen unterscheiden? Welche Beziehungen werden darin festgelegt?

**Aufgabe 17.4:** Man kann Messmodelle und Strukturmodelle unterscheiden. In Messmodellen wird dargestellt, wie Konstrukte gemessen werden, d. h., es werden die Beziehungen zwischen latenten und manifesten Variablen festgelegt. In Strukturmodellen werden Beziehungen zwischen Konstrukten dargestellt, d. h. Effekte zwischen den latenten Variablen untereinander. Dies können ungerichtete Beziehungen zwischen den latenten Variablen sein, gerichtete Effekte der exogenen Variablen auf endogene Variablen sowie gerichtete Effekte zwischen endogenen Variablen.

**Aufgabe 17.5:** Was ist ein Skalierer? Wofür wird er benötigt?

**Aufgabe 17.5:** Pro latenter Variable wird die Faktorladung einer zugeordneten manifesten Variablen auf den Wert 1 fixiert. Diese manifeste Variable, deren Faktorladung vorgegeben wird, dient als Skalierer, um die Varianz der latenten Variable festzulegen. Die nicht beobachtbare Varianz der latenten Variablen wird so an der beobachtbaren Varianz der zugeordneten manifesten Variablen orientiert.

**Aufgabe 17.6:** Was ist die modelltheoretische Kovarianzmatrix? Welche Einträge stehen darin?

**Aufgabe 17.6:** Die modelltheoretische Kovarianzmatrix enthält die im Modell reproduzierten Varianzen und Kovarianzen der manifesten Variablen, d. h. Werte für die Varianzen und Kovarianzen, wie sie sich aus den angenommenen Beziehungen und den dafür geschätzten Parametern ergeben.

**Aufgabe 17.7:** Welche Hypothese prüft der  $\chi^2$ -Test im Zusammenhang mit Strukturgleichungsmodellen? Soll der Test für ein Modell, das zu den Daten passt, signifikant sein oder nicht?

**Aufgabe 17.7:** Der  $\chi^2$ -Test prüft die Hypothese, dass modelltheoretische und empirische Kovarianzmatrix nicht voneinander abweichen, d. h., dass das Modell die empirischen Varianzen und Kovarianzen perfekt reproduzieren kann. Für ein passendes Modell sollte das Ergebnis nicht signifikant sein, also keine bzw. nur geringe Abweichungen anzeigen (hier ist somit die Nullhypothese die Wunschhypothese).

**Aufgabe 17.8:** Angenommen, Sie möchten die vermuteten Zusammenhänge zwischen 4 Konstrukten mit einer konfirmatorischen Faktorenanalyse untersuchen. Zu jedem Konstrukt haben Sie Daten von zwei beobachteten Variablen erhoben. Ist dies ausreichend, damit das Modell identifiziert ist?

**Aufgabe 17.8:** Vorausgesetzt, dass die Konstrukte wie vermutet zusammenhängen, ist das Modell identifiziert: Bei  $k = 8$  beobachteten Variablen erhalten Sie  $s = k(k + 1) / 2 = 36$  empirische Varianzen und Kovarianzen. Hieraus würden Sie  $t = 22$  Parameter schätzen: 8 Fehlervarianzen, 4 Faktorladungen (die 4 übrigen wären zur Skalierung auf den Wert 1 festgelegt), 4 Varianzen und 6 Kovarianzen der Konstrukte, entsprechend  $df = 36 - 22 = 14$ . Dies funktioniert allerdings nur, wenn die Konstrukte tatsächlich zusammenhängen, denn mit nur jeweils 2 beobachteten Variablen wären die Messmodelle jedes einzelnen Konstruktes für sich alleine nicht identifiziert. Es wäre also sicherer, mehr beobachtete Variablen pro Konstrukt zu erheben.

**Aufgabe 17.9:** Sie prüfen einen angenommenen Effekt von Arbeitsanforderungen auf Burnout mit einem Strukturgleichungsmodell. Die Analyse der Daten von  $N = 185$  Beschäftigten ergibt einen standardisierten Pfadkoeffizienten der Arbeitsanforderungen auf Burnout von  $.38$ ,  $p = .042$ . Als Modellgütekriterien erhalten Sie  $\chi^2 = 55.40$ ,  $df = 30$ ,  $p = .003$  sowie  $RMSEA = .18$  und  $CFI = .83$ . Wie interpretieren Sie das Ergebnis? Besteht der angenommene Effekt?

**Aufgabe 17.9:** Der Effekt ist zwar signifikant, aber eine inhaltliche Aussage nicht sinnvoll, da das Modell nicht zu den Daten passt. Der  $\chi^2$ -Test zeigt signifikante Abweichungen zwischen Modell (modelltheoretische Kovarianzmatrix) und Daten (empirische Kovarianzmatrix) an. Zwar kann ein signifikantes Testergebnis in großen Stichproben auch durch sehr geringe Abweichungen zwischen Modell und Daten zustande kommen, die womöglich praktisch unbedeutend wären. Die deskriptiven Kriterien  $RMSEA$  und  $CFI$  weisen jedoch auf erhebliche Abweichungen zwischen Modell und Daten hin. Bei einem Modell, welches nicht zu den Daten passt, können die Ergebnisse der Parameterschätzung sinnfrei sein und sollten daher nicht interpretiert werden.

**Aufgabe 17.10:** Sie untersuchen den Effekt von Burnout bei Mitarbeitern auf deren Kündigungsneigung mithilfe eines Strukturgleichungsmodells. Sie erhalten Daten von  $N = 112$  Beschäftigten. Die Analyse ergibt einen signifikanten Effekt von Burnout auf die Kündigungsneigung (standardisierter Pfadkoeffizient  $.32$ ,  $p = .011$ ). Das Modell passt offenbar gut zu den Daten ( $\chi^2 = 9.32$ ,  $df = 8$ ,  $p = .316$ ). Für das Item „Ich suche gezielt nach Stellenangeboten, die für mich infrage kommen“ erhalten Sie als standardisierte Parameterschätzungen eine Faktorladung von  $1.05$  und eine Fehlervarianz von  $-0.10$ . Wie interpretieren Sie dieses Ergebnis? Besteht der vermutete Effekt?

**Aufgabe 17.10:** Die negative Fehlervarianzschätzung des Items zusammen mit der standardisierten Faktorladung  $> 1$  weisen auf einen Heywood Case hin, d. h., die Parameterschätzung hat Werte geliefert, die zwar die dem Modell zugrunde liegenden Gleichungen erfüllen, die aber empirisch unmöglich sind: Varianzen können nicht negativ sein. Selbst wenn das Item „Ich suche gezielt nach Stellenangeboten, die für mich infrage kommen“ die Kündigungsneigung gut misst, wird es keine perfekte Messung sein, sondern in der Praxis eine **positive** Fehlervarianz haben. Schätzergebnisse, die einen Heywood Case darstellen, sollten nicht interpretiert werden, eine Aussage über den vermuteten Effekt ist daher so nicht möglich. Ein Grund könnte in der für ein Strukturgleichungsmodell hier recht kleinen Stichprobe liegen, sodass mehr Daten oder womöglich eine andere Schätzmethode Abhilfe schaffen könnten.



## Kapitel 18: Evaluationsforschung

---

### Aufgabe 18.1: Worin besteht der Unterschied zwischen formativer und summativer Evaluation?

**Aufgabe 18.1:** Die bilanzierend-summative Evaluation zielt auf eine abschließende Bewertung des Evaluationsgegenstandes, die gestaltend-formative Evaluation soll dagegen zu dessen schrittweiser Verbesserung beitragen.

### Aufgabe 18.2: Was kennzeichnet interne versus externe Evaluation und Selbst- versus Fremdevaluation?

**Aufgabe 18.2:** Rollenverteilung: Bei der Selbstevaluation wird die Evaluation einer Maßnahme von den Praktiker:innen selbst organisiert und durchgeführt, sie sind Evaluierte und Evaluierende in Personalunion. Bei der Fremdevaluation sind Evaluierte und Evaluierende dagegen unterschiedliche Personen. Institutionelle Zugehörigkeit: Bei der Fremdevaluation wird weiter differenziert, ob die Evaluierenden derselben Institution angehören wie diejenigen, die auch für die Maßnahme verantwortlich ist (interne Fremdevaluation), oder ob sie aus einer anderen Institution stammen (externe Fremdevaluation). Eine Selbstevaluation ist definitionsgemäß immer intern.

### Aufgabe 18.3: Grenzen Sie Interventions- und Evaluationsforschung voneinander ab!

**Aufgabe 18.3:** Im Rahmen der Interventionsforschung werden Produkte, Methoden oder Veränderungsmaßnahmen entwickelt, im Rahmen der Evaluationsforschung werden sie bewertet.

### Aufgabe 18.4: Was versteht man unter dem CIPP-Modell?

**Aufgabe 18.4:** Das CIPP-Modell ist eine Variante des systemischen Evaluationsansatzes, der darauf abzielt, den Evaluationsgegenstand möglichst ganzheitlich im Kontext sowie im zeitlichen Verlauf zu betrachten. Gemäß CIPP-Modell werden Evaluationsstudien konzipiert und durchgeführt, die aus vier Teil-Evaluationen bestehen (Context-, Input-, Process- und Product-Evaluation).

### Aufgabe 18.5: Was unterscheidet die nutzungsorientierte von der partizipativen Evaluationsforschung?

**Aufgabe 18.5:** Sowohl bei nutzungsorientierten als auch bei partizipativen Evaluationsansätzen spielt die Einbeziehung der Stakeholder eine herausgehobene Rolle. Dabei orientiert sich die nutzungsorientierte Evaluation gemäß einer pragmatischen Philosophie vor allem an den intendierten Nutzerinnen und Nutzern der empirischen Evaluationsergebnisse. Sie sollen darin unterstützt werden, die Evaluationsergebnisse praktisch fruchtbar zu machen (z. B. Auftraggeber, Entscheidungsträger in Wirtschaft und Politik). Demgegenüber verfolgt die partizipative Evaluation gemäß einer demokratischen Philosophie das Ziel, die Betroffenen einer Maßnahme (insbesondere unterprivilegierte Zielgruppen) möglichst umfassend in Konzeption, Durchführung, Auswertung und Interpretation der Evaluationsstudie einzubeziehen und dadurch auch einen Beitrag zu ihrem Empowerment zu leisten.

### Aufgabe 18.6: Welche Ziele im Kontext der Evaluationsforschung verfolgen Metaanalyse und Metaevaluation?

**Aufgabe 18.6:** Eine Metaanalyse fasst die Befunde mehrerer einzelner Evaluationsstudien zum selben Evaluationsgegenstand statistisch zusammen und liefert somit eine Gesamtbewertung des Evaluationsgegenstandes, etwa seiner Effektivität oder Effizienz. Eine Metaevaluation betrachtet Evaluationsstudien dagegen nicht im Hinblick auf ihre Ergebnisse, sondern auf ihre methodische Qualität; vor allem die Einhaltung von Evaluationsstandards wird geprüft.

**Aufgabe 18.7:** Gute grundlagenwissenschaftliche Studien müssen Kriterien der methodischen und ethischen Strenge genügen. Wo sind diese in den Evaluationsstandards abgebildet?

**Aufgabe 18.7:** Methodische Strenge: v. a. Genauigkeitsstandards; ethische Strenge: v. a. Fairnessstandards.

**Aufgabe 18.8:** Mit welchen Methoden wird die Effizienz eines Evaluationsgegenstandes im Kontext der ökonomischen Evaluation ermittelt?

**Aufgabe 18.8:** Kosten-Effektivitäts-Analyse, Kosten-Nutzwert-Analyse, Kosten-Nutzen-Analyse

**Aufgabe 18.9:** Angenommen, die zahnmedizinische Versorgung soll dadurch verbessert werden, dass bei schmerzhaften Behandlungen auf die Injektion von Schmerzmitteln verzichtet und stattdessen mit modernen Hypnosetechniken gearbeitet wird. Hypnosetherapie hätte den Vorteil, dass Medikamente eingespart werden (Kostensparnis, geringere Belastung des Organismus) und sich die Patienten während und nach der Behandlung evtl. besser fühlen. In einer Großstadt werden zufällig 5 Zahnarztpraxen, die mit Hypnose arbeiten, und 3 Praxen, die herkömmliche Methoden der Schmerzbehandlung einsetzen, ausgewählt. Während einer 4-wöchigen Untersuchungsphase wird direkt nach jeder Behandlung auf gesonderten Erhebungsbögen notiert, wie unangenehm die Behandlung für den Patienten war (gar nicht, wenig, teils teils, ziemlich, völlig) und ob der Patient eine bessere Schmerzversorgung wünscht (ja, nein). Zudem werden Alter (in Jahren), Geschlecht (weiblich/männlich; zur Operationalisierung von Geschlecht Abschn. 8.5.1) und Art der Behandlung (z. B. Wurzelbehandlung, Krone, Inlay) aufgezeichnet. Am folgenden Tag wird telefonisch nachgefragt, ob nach der Behandlung noch unangenehme Nachwirkungen auftraten (gar nicht, wenig, teils teils, ziemlich, völlig). Wie würden Sie diese Evaluationsstudie anhand der folgenden Merkmale charakterisieren?

1. Evaluationsfrage?
2. Unabhängige Variable (und Skalenniveau)?
3. Moderatorvariablen (und Skalenniveau)?
4. Abhängige Variablen (und Skalenniveau)?
5. Datenerhebungsmethode?
6. Untersuchungsdesign?
7. Verhältnis von Interventions- und Evaluationsstichprobe?
8. Erfolgskriterien?

**Aufgabe 18.9:**

1. Evaluationsfrage: Ist eine Hypnosebehandlung der herkömmlichen Schmerztherapie überlegen oder zumindest gleichwertig?
2. Unabhängige Variable: Form der Schmerztherapie (Hypnose oder Schmerzmittel), nominale bzw. dichotome Variable
3. Moderatorvariablen: Geschlecht und Art der Zahnbehandlung (beide nominalskaliert), Alter (kardinalskaliert)

4. Abhängige Variablen: Intensität negativer Empfindungen während und nach der Behandlung (jeweils fünf äquidistante Intensitätsstufen; kardinalskaliert); Zufriedenheit mit der Behandlung (bessere Versorgung gewünscht ja/nein, nominalskaliert)
5. Datenerhebungsmethode: Standardisierte Befragung der Patienten (mündlich und fernmündlich)
6. Untersuchungsdesign: Quasi-experimentelle Untersuchung mit einer Experimentalgruppe (Hypnose) und einer Kontrollgruppe (Schmerzmittel). (Randomisierung ist vermutlich nicht möglich, da viele Patienten selbst über die Behandlungsform entscheiden wollen)
7. Verhältnis von Interventions- und Evaluationsstichprobe: Da es sich nicht um ein gezielt initiiertes Interventionsprojekt handelt, ist die Interventionsstichprobe (Anzahl der Patienten, die unter Hypnose behandelt werden) unbekannt
8. Erfolgskriterien: Die Experimentalgruppe empfindet die Behandlung deutlich und statistisch signifikant weniger unangenehm und wünscht seltener eine bessere Versorgung als die Kontrollgruppe (gerichtete Alternativhypothese als „starkes“ Erfolgskriterium).

**Aufgabe 18.10:** Diskutieren Sie für das obige Beispiel die Problematik der internen und externen Validität. Angenommen, es stellt sich heraus, dass die Hypnosegruppe tatsächlich weniger Beschwerden berichtet als die Kontrollgruppe; welche Störeinflüsse könnten die interne Validität beeinträchtigt haben? Inwiefern sind Ergebnisse dieser Studie generalisierbar?

**Aufgabe 18.10:**

- Interne Validität: Wie könnte ein positives Resultat bei der Experimentalgruppe zustande kommen, obwohl „in Wirklichkeit“ die Hypnosebehandlung gar nicht wirksam ist?
  - Probandeneffekt: Wer sich für die Hypnosebehandlung entscheidet, ist weniger schmerzempfindlich (Problem der Selbstselektion bei dem vorliegenden quasi-experimentellen Design, Schmerzempfindlichkeit wäre als Kontrollvariable zu erheben).
  - Arzteeffekt: Ärzte, die Hypnose anbieten, arbeiten besonders gut und schmerzfrei.
  - Behandlungseffekt: Hypnose wird besonders bei „leichteren“, weniger schmerzintensiven Behandlungen eingesetzt (die Art der Behandlung wäre als Kontrollvariable bzw. Moderatorvariable zu berücksichtigen).
- Externe Validität: Zur Zielpopulation gehören prinzipiell alle Patient:innen. Bei der realisierten Stichprobe von Zahnarztpraxen ist jedoch zunächst nur auf die Gesamtheit der Patient:innen der lokalen Zahnarztpraxen der untersuchten Großstadt zu generalisieren.

## Kapitel 19: Computationale Methoden in den Sozial- und Humanwissenschaften

---

**Aufgabe 19.1:** Was versteht man in den Sozial- und Humanwissenschaften unter computationalen Methoden?

**Aufgabe 19.1:** Als Computationale Methoden werden in den Sozial- und Humanwissenschaften solche Methoden der Datenerhebung, Datenaufbereitung und Datenanalyse bezeichnet, die zur Lösung der Forschungsprobleme besonders große digitale Datensätze verarbeiten (Big Data) und dafür fortgeschrittene Algorithmen und Rechenkapazitäten benötigen.

**Aufgabe 19.2:** Nennen Sie drei Verfahren der Operationalisierung im Kontext automatischer Textanalysen!

**Aufgabe 19.2:** (1) Diktionärsbasierte Verfahren, (2) überwachtes Maschinenlernen, (3) unüberwachtes Maschinenlernen

**Aufgabe 19.3:** Welche Wege der Datenerhebung sind in der computationalen Forschung typisch, um aussagekräftige, große Mengen an digitalen Daten zu erheben?

**Aufgabe 19.3:** (1) Zusammenarbeit mit Unternehmen, (2) Kauf von Datensätzen, (3) API-Calls, (4) Web-Scraping, (5) Nutzer-Tracking und Datenspenden, (6) Datenbanken und Datenarchive.

**Aufgabe 19.4:** Was ist die Zielsetzung einer Sentimentanalyse?

**Aufgabe 19.4:** Ziel der Sentimentanalyse ist es, natürlichsprachliche Texte computerbasiert dahingehend zu untersuchen, ob bzw. in welcher Intensität sie positive und negative Stimmungen ausdrücken. Die Analyse kann dabei auf der Ebene von Wörtern, Sätzen oder längeren Texten erfolgen. Fortgeschrittene Methoden der Sentimentanalyse erlauben mit entsprechenden Diktionären auch die Differenzierung unterschiedlicher negativer (z. B. Angst, Wut) und positiver (z. B. Freude, Vertrauen) Emotionen.

**Aufgabe 19.5:** Welche Bedeutung haben APIs im Kontext computationaler Forschung?

**Aufgabe 19.5:** Mit APIs (Application Programming Interfaces) sind im Kontext computationaler Forschung zumeist die Schnittstellen von Social-Media-Plattformen oder auch von digitalen Geräten gemeint. Bieten Unternehmen offene APIs an, können digitale Daten für die Forschung relativ komfortabel in großem Stil erhoben werden. Dies ist jedoch oft nicht der Fall. So fehlen bei vielen führenden Social-Media-Plattformen (z. B. Instagram, TikTok) offene APIs, was den Datenzugang erschwert.

**Aufgabe 19.6:** Bei der automatischen Klassifikation von Twitter-Tweets als beleidigend oder nicht-beleidigend hat der Klassifikationsalgorithmus nach dem Training mit  $N = 1\,000$  authentischen Tweets einen Recall-Score von .18 und einen Precision-Score von .86 erreicht. Was bedeutet dieses Ergebnis? Und wie ist es ursächlich zu erklären?

**Aufgabe 19.6:** Diese beiden Gütemaße für überwachtes Maschinenlernen besagen, dass der Klassifikator nur relativ selten einen Tweet als beleidigend einstuft (18 %), also viele beleidigende Tweets übersieht und somit eine geringe Trefferquote hat. Wenn er jedoch einen Tweet als beleidigend einstuft, ist dies meist korrekt (86 %). Das heißt, der Klassifikator hat eine hohe Genauigkeit. Ein Grund für die geringe Trefferquote kann darin liegen, dass authentische Twitter-Datensätze unbalanciert sind, also deutlich weniger beleidigende als nicht-beleidigende Tweets enthalten. Beim Training mit solchen unbalancierten Datensätzen kann der Algorithmus das Erkennen von Beleidigungen nicht so gut erlernen.

**Aufgabe 19.7:** Welche forschungsethischen Probleme sind bei computationaler Forschung besonders vordringlich?

**Aufgabe 19.7:** Ethische Fragen rund um die Erhebung, Auswertung und Nutzung von Social-Media-Daten ohne informierte Einwilligung der Social-Media-Nutzenden und teilweise auch ohne ausdrücklich Zustimmung der Social-Media-Unternehmen.

**Aufgabe 19.8:** Werden durch computationale, induktiv arbeitende Verfahren der Textanalysen (z.B. Topic Modeling) das Arbeiten mit Theorien sowie menschliche Kodierende überflüssig (Begründung)?

**Aufgabe 19.8:** Nein, sie werden nicht überflüssig. Denn computationale Analysen benötigen zur Einordnung und Interpretation der Befunde immer auch Domänenwissen (Theorie- und Forschungsstand) sowie auch eine Überprüfung der vom Algorithmus vorgenommenen Kategorisierungen des Materials durch Vergleich mit manuellen Textcodierungen.

**Aufgabe 19.9:** Was ist das Tidyverse?

**Aufgabe 19.9:** Das Tidyverse ist eine Sammlung von Paketen für die Software R, die alle einem ähnlichen Grundprinzip folgen und der automatischen Textanalyse dienen.

**Aufgabe 19.10:** Welche Bedeutung haben Lemmatisierung und Stemming im Zusammenhang mit automatischer Textanalyse?

**Aufgabe 19.10:** Beides sind Verfahren der Vorverarbeitung im Kontext automatischer Textanalysen. Bei der Lemmatisierung werden Wörter auf ihre Grundform gebracht, die als Lemma auch im Wörterbuch steht. Beim Stemming werden Wörter dagegen auf ihren Wortstamm komprimiert, der meist nicht im Wörterbuch steht.