

Elementarwissen medizinische Psychologie und medizinische Soziologie Methodische Grundlagen

Inhalt

Artikel

Elementarwissen medizinische Psychologie und medizinische Soziologie: Methodische Grundlagen	1
Elementarwissen medizinische Psychologie und medizinische Soziologie: Literatur	22

Referenzen

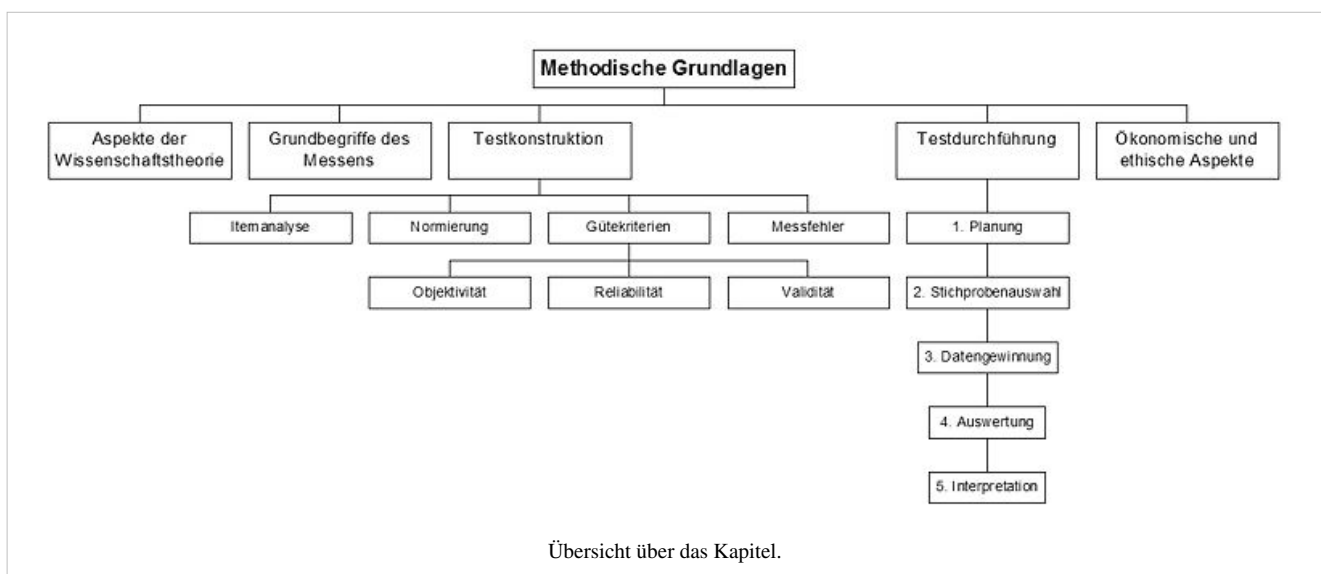
Quellen und Bearbeiter des Artikels	23
Quellen, Lizenzen und Autoren des Bildes	24

Artikel Lizenzen

Lizenz	25
--------	----

Elementarwissen medizinische Psychologie und medizinische Soziologie: Methodische Grundlagen

Vorbemerkung: Dieses Kapitel behandelt zum Teil recht spezielle Themen, die erst im klinischen Abschnitt relevant werden – diese Themen sind an der verkleinerten Schriftart zu erkennen. Den Studierenden im vorklinischen Studienabschnitt sei daher geraten, sich in erster Linie mit den groß gedruckten Textteilen zu beschäftigen, um ein elementares Verständnis von statistischen und epidemiologischen Zusammenhängen und Denkweisen zu bekommen, und sich dabei nicht von den recht detaillierten Ausführungen zur Statistik behelligen zu lassen. Die kleingedruckten Abschnitte sind hingegen vor allem für Studierende im klinischen Studienabschnitt interessant – als Einführung oder als Ergänzung zu Vorlesungen und Praktika des Fachs *Biometrie und Epidemiologie (Q1)*.



Aus einer Theorie gewinnt man Forschungshypothesen und überprüft sie, wobei das Falsifikationsprinzip gilt

In der Psychologie geht es wie in jeder anderen echten Wissenschaft darum, Beobachtungen in den Gesamtzusammenhang einer Theorie zu bringen, aus dieser Theorie Hypothesen, d. h. überprüfbare Aussagen abzuleiten, die Hypothesen zu überprüfen und somit festzustellen, ob die Theorie noch Gültigkeit besitzt. Einige **Grundbegriffe** sollte man hierbei kennen:

- Eine *Theorie* ist eine Menge von Sätzen zur zusammenfassenden erklärenden und vorhersagenden Beschreibung von Sachverhalten.
- Überprüft wird aber nicht direkt die Theorie, sondern aus ihr abgeleitete *Hypothesen*:
 - Definition: Hypothesen sind aus Theorie abgeleitete falsifizierbare Sätze, die in Studien geprüft werden.
 - Hypothesen unterscheiden sich in ihrem Geltungsbereich:
 - Deterministische Hypothese (= nomologische Hypothese): ein Zusammenhang zwischen zwei Merkmalen gilt in allen Fällen, also zeitlich und räumlich unbegrenzt.
 - Probabilistische Hypothese: ein Zusammenhang zwischen zwei Merkmalen gilt nur mit einer gewissen Wahrscheinlichkeit (unperfekte Korrelation) und ist nur für eine Gesamtheit feststellbar, für das einzelne Individuum ist dagegen keine sichere Aussage möglich, allenfalls eine Wahrscheinlichkeitsaussage.

- Hypothesen können auf verschiedene Weisen gewonnen werden:
 - Deduktives Verfahren: aus der Theorie werden Hypothesen abgeleitet.
 - Induktives Verfahren: man geht von Beobachtungen über Hypothesen zur Theorie; POPPER versteht unter "induktivem Vorgehen" lediglich eine Methode der Theoriefindung (Heuristik).
- In der Psychologie hat man es meist mit *Konstrukten* zu tun. Ein Konstrukt ist ein theoretisches Konzept für etwas, das nicht direkt beobachtet werden kann, sondern aus (beobachtbaren, messbaren) Indikatoren erschlossen werden muss. Beispielsweise wird das Konstrukt "Intelligenz" aus Leistungen im Intelligenztest (Indikatoren) erschlossen. Die Intension des Konstrukts, d. h. sein für uns fassbarer Inhalt, hängt daher von der Art der Indikatoren ab. Konstrukte können durch Operationalisierung erfassbar gemacht werden.

Merke: Bei vielen Gegenständen der Psychologie handelt es sich um Konstrukte, d. h. um nicht direkt beobachtbare, sondern nur aus Beobachtungen erschließbare Konzepte (z. B. Intelligenz, Motivation, Emotion etc.).
- Man unterscheidet grundsätzlich zwei Arten von *Fehlern*:
 - Zufällige Fehler (zufällige Streuungen um den wahren Wert) vermindern lediglich die Präzision (verminderte Reliabilität) und mitteln sich bei vielen Einzelmessungen heraus.
 - Systematische Fehler verfälschen das Ergebnis dagegen in eine bestimmte Richtung (verminderte Validität); sie kommen zustande durch Erfassungsfehler (z. B. fehlerhafte Messinstrumente), Confounder oder Selektions- und Informationsbias.
- Seit POPPER gilt in der Wissenschaft nicht mehr das Verifikationsprinzip, sondern das *Falsifikationsprinzip*. Grob formuliert besagt es, dass eine Theorie niemals bestätigt, sondern nur vorläufig gültig ist, weil in der Zukunft Fälle gefunden werden können, welche die Theorie widerlegen. Eine Theorie lässt sich jedoch mittels Exhaustion "retten", d. h. man modifiziert den Geltungsbereich der Theorie, indem man mit der Theorie unvereinbare Fälle ausgrenzt.

Merke: Aus Theorien abgeleitete Hypothesen sind nur so lange gültig, wie keine Fälle gefunden werden, die sie und mithin die zugrundeliegende Theorie widerlegen, d. h. falsifizieren. Forschung führt daher immer nur zu vorläufigen Ergebnissen.
- Die *standardisierte Datenerhebung* besteht aus drei Phasen: Durchführung, Auswertung, Interpretation.
- Daten werden erhoben, indem die Ausprägungen von Merkmalen (Variablen) erfasst werden, welche den Beobachtungseinheiten (kleinste Einheiten, an denen Veränderungen feststellbar sind) und somit den Untersuchungseinheiten (Merkmalsträger, z. B. Probanden) zueigen sind. Merkmale lassen sich dabei einteilen nach ihrer Funktion bei der statistischen Analyse (Zielgrößen, Einflussgrößen), nach ihrem Skalenniveau (nominal-, ordinal-, intervall- oder ratioskaliert) und nach der Struktur ihrer Merkmalsausprägungen (diskret oder stetig). Am Ende der Datenerhebung besitzt man eine vollständige, disjunkte (d. h. keine Überlappungen enthaltende) Auflistung der Merkmalsausprägungen, die als Liste oder Tabelle (Spalten: Merkmale, Zeilen: Beobachtungseinheiten; Legende, Kennzeichnung fehlender Daten etc.) darstellbar ist.

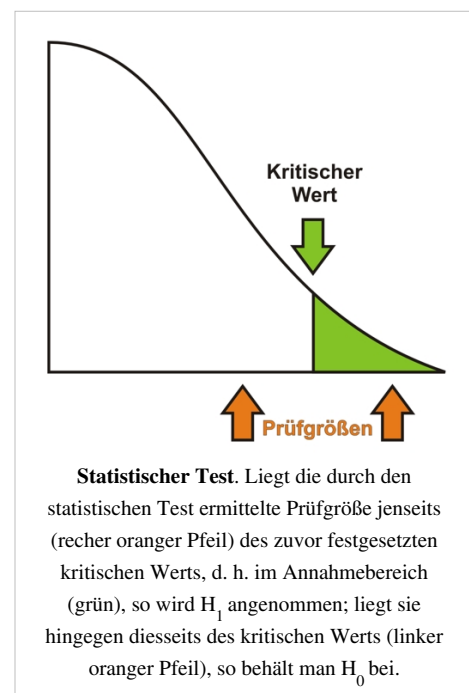
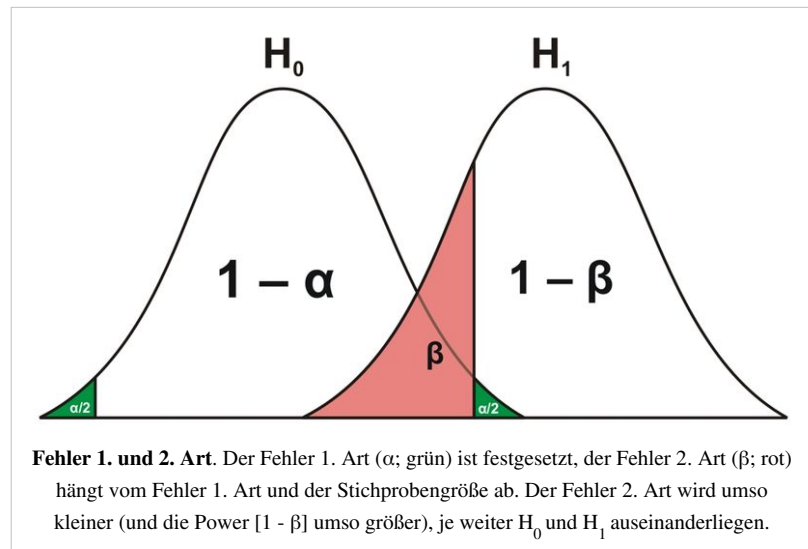
Hypothesen werden mittels eines **statistischen Messverfahrens** überprüft.

- *Hypothesen*:
 - Methodisch
 - Die Nullhypothese ($= H_0$) behauptet einen als bestehend angenommenen Sachverhalt (z. B. "Das Gehirn von Skilangläufern wiegt durchschnittlich 1500 g"). Ihr gegenüber steht die zu ihr komplementäre Forschungshypothese (= Alternativhypothese, H_1), die einen neuen Sachverhalt behauptet und den in der Nullhypothese artikulierten Sachverhalt negiert (z. B. "Das Gehirn von Skilangläufern wiegt *nicht* durchschnittlich 1500 g, sondern mehr oder weniger als 1500 g").
 - Damit Hypothesen statistisch geprüft werden können, müssen inhaltliche Hypothesen in statistische Hypothesen umgewandelt werden. Hierzu gibt man Kriterien an, bei deren Erfülltsein man H_1 als zutreffend anerkennt.
 - Inhaltlich
 - Unterschiedshypothese: wird durch Häufigkeits- und Mittelwertvergleich erfasst

- Zusammenhangshypothese: wird durch eine Korrelationsrechnung erfasst

Merke: Die Nullhypothese (H_0) behauptet einen konkreten, als bestehend angenommenen Sachverhalt, die Forschungshypothese (H_1) einen zu ihr komplementären Sachverhalt, der mittels Forschung erst noch herausgefunden und bestätigt werden muss.














- Der *statistische Test* erlaubt festzustellen, ob Unterschiede zwischen Kennwerten einer Stichprobe einerseits und bestimmten Vergleichswerten (Kennwerte anderer Stichproben oder definierte Sollwerte) andererseits rein zufällig zustande gekommen sind (etwa weil die Stichprobe schlecht ausgewählt worden ist) oder aber *überzufällig*, d. h. signifikant sind. (Beispiel: durchschnittliches Gehirngewicht von Skilangläufern: empirischer, anhand einer Stichprobe ermittelter Mittelwert 1400 g vs. Sollwert laut H_0 1500 g → ist der Unterschied zwischen Mittelwert und Sollwert signifikant, besteht also wirklich ein Unterschied, so dass H_1 berechtigterweise angenommen und H_0 berechtigterweise verworfen wird?). Er gibt also an, mit welcher Wahrscheinlichkeit aufgrund der durch die Stichprobe erlangten Erkenntnisse H_1 angenommen wird, obwohl in der Gesamtpopulation in Wirklichkeit H_0 gilt, d. h. mit welcher Wahrscheinlichkeit anhand der Stichprobe ein Unterschied zwischen Kennwert und Vergleichswert festgestellt wird, obwohl in der Gesamtpopulation ein solcher Unterschied überhaupt nicht existiert.



- Zwei Arten von Fehlern sind dabei möglich:
 - Fehler 1. Art (α -Fehler): falsch-positives Ergebnis, d. h. H_1 wird angenommen, obwohl H_0 gilt. Dieser Fehler ist schwerwiegender als der Fehler 2. Art und wird daher durch Festsetzung eines niedrigen Maximalwertes (meist 5 %) kontrolliert (d. h. auf eine Falschannahme in 20 Hypothesenprüfungen beschränkt); H_1 wird also nur dann akzeptiert, wenn H_0 sehr unwahrscheinlich ist. Da H_0 eindeutig formuliert ist, ist dieser Fehler kontrollierbar.

- Fehler 2. Art (β -Fehler): falsch-negatives Ergebnis, d. h. H_1 wird verworfen, obwohl es in der Population gilt. Da H_1 als Negation von H_0 nicht eindeutig formuliert ist, lässt sich dieser Fehler nur ungenau abschätzen. Der β -Fehler ist umso größer, je kleiner der α -Fehler und je kleiner der Stichprobenumfang n .
- Der statistische Test ermöglicht die algorithmische Berechnung einer Prüfgröße, die mit der festgesetzten Irrtumswahrscheinlichkeit α (in Kauf genommene Wahrscheinlichkeit für den α -Fehler; meist 5 %, manchmal auch 1 % oder 0,1 %) verglichen wird. Die Irrtumswahrscheinlichkeit ist identisch mit dem Signifikanzniveau. Ist die Prüfgröße kleiner als das Signifikanzniveau (kritischer Wert), liegt sie also im überzufälligen Annahmebereich, so wird H_1 akzeptiert, man sagt: "Die Alternativhypothese wird auf dem Signifikanzniveau α angenommen"; im anderen Fall wird H_1 abgelehnt, man sagt: "Die Nullhypothese kann auf dem Signifikanzniveau α nicht verworfen werden." Da durch Festlegen des Signifikanzniveaus eine maximale Wahrscheinlichkeit eingeräumt wird, mit der in einer bestimmten Studie *fälschlicherweise* H_1 behauptet wird, sind Studien potentiell fehlerhaft. Diese potentielle Fehlerhaftigkeit ist jedoch erforderlich, um überhaupt einen Zusammenhang postulieren zu können. Denn würde man die Irrtumswahrscheinlichkeit und somit das Signifikanzniveau auf 0 festsetzen, so wäre der Annahmebereich für die Prüfgröße ebenfalls gleich 0 – man könnte mithin keine überzufälligen Ergebnisse feststellen und H_1 somit gar nicht behaupten. Dabei hat die Stichprobengröße n einen Effekt auf die Entscheidung für oder gegen H_1 : bei kleinem n wird H_1 eher verworfen, bei großem n eher angenommen – durch eine ausreichend große Stichprobe lassen sich auch sehr kleine Unterschiede als statistisch signifikant einstufen, wenn diese Unterschiede mitunter auch praktisch wenig relevant sind.
- Power (= Güte, Teststärke, Trennschärfe eines Tests): Vermögen eines Tests zu richtig-positiven Entscheidungen, d. h. ein in der Population gültiges H_1 als solches zu erkennen; Power = $1 - \beta$
- p-Wert: Eine rationalere Methode der Signifikanz-Beurteilung stellt die Berechnung des p-Werts dar (mittels des exakten Tests nach Fisher). Hierbei wird anhand der Daten ein p-Wert ermittelt, der gewissermaßen die durch die Daten bestimmbare kleinstmögliche Irrtumswahrscheinlichkeit angibt. Ist der p-Wert kleiner als das Signifikanzniveau α , so ist der Unterschied zwischen Kennwert und Sollwert bzw. zwischen verschiedenen Kennwerten signifikant.

Merke: Der statistische Test ist ein spezifischer Test, mit dessen Hilfe sich Aussagen über die Signifikanz, d. h. Überzufälligkeit eines mit Hilfe der Stichprobe festgestellten Unterschieds machen lassen. Ein Unterschied ist im statistischen Sinne signifikant, wenn die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen (Annahme von H_1 , obwohl in der Grundgesamtheit H_0 gilt), sehr gering ist (meist kleiner als 5 %).

Weblinks:  Theorie,  Hypothese,  Konstrukt,  Falsifikation,  Karl Raimund Popper,  Zufälliger Fehler,  Systematischer Fehler,  Datenerhebung,  Forschungshypothese,  Nullhypothese,  Statistischer Test,  Irrtumswahrscheinlichkeit,  p-Wert

Selbsttest:

1. Welche der folgenden Hypothesen probabilistisch, welche deterministisch? Hypothese A: "Alle Schwäne sind weiß.", Hypothese B: "Die meisten Elefanten sind grau."
2. Was ist ein Konstrukt? Nennen Sie einige Beispiele für Konstrukte!
3. Jens hat sein Regal mit einem falsch geeichten Metermaß gemessen, nun passt es nicht in die neue Wohnung. Welche Art von Messfehler wurde begangen?
4. Welche Phasen umfasst die standardisierte Datenerhebung?
5. Was sagt H_0 aus, was H_1 ?
6. Was ist der Fehler 1. Art?
7. Wozu dient ein statistischer Test?

Die abhängige Variable ist eine Funktion der unabhängigen Variable und möglicherweise anderer Variablen

Um ein Konstrukt (z. B. Intelligenz) zu erfassen, benötigt man ein bestimmtes Messverfahren, mit dem man bestimmten Variablen bestimmte Werte zuordnen kann.

- *Operationalisierung* ist die Angabe eines Messverfahrens, mit dem man ein Konstrukt erfassen kann.
- *Messung*
 - Definition: eine Messung ist eine Zuordnung von Zahlenwerten zu einzelnen Merkmalsausprägungen (z. B. Zahlenverständnis → 80)
 - Die Messung im Laborexperiment hat im Vergleich zur Feldbeobachtungen drei entscheidende Eigenschaften: Willkürlichkeit, Variierbarkeit, Wiederholbarkeit.
- *Variablen* sind Entitäten mit unterschiedlicher Ausprägung (vgl. Konstante), wobei die Ausprägung diskret oder stetig sein kann. Den Forscher interessieren dabei folgende Variablen: Ursache, Wirkung und dazwischenliegende Variablen.
 - Ursache und Wirkung bilden die "Endpunkt-Variablen" und tragen spezielle Bezeichnungen.
 - Die unabhängige Variable ist die Ursache und besteht aus den vom Forscher variierten Einflussfaktoren; sie sollte konstant gehalten werden.
 - Die abhängige Variable ist die Wirkung.
 - "Zwischenvariablen" liegen zwischen unabhängiger und abhängiger Variable und sollten eliminiert oder kontrolliert werden.
 - Mediator-Variable: Glied in der Ursachenkette, die auf die abhängige Variable einwirkt, d. h. sie vermittelt die Wirkung und fungiert gewissermaßen als Zwischenstation der Ursachen-Wirkungs-Kette.
 - Moderator-Variable: beeinflusst den Zusammenhang zwischen unabhängiger und abhängiger Variable und generiert einen Interaktionseffekt, d. h. die abhängige Variable und die Moderator-Variable interagieren miteinander, so dass die Moderator-Variable den Effekt der unabhängigen Variable auf die abhängige Variable verändert.
 - Störvariable (Confounder): nicht berücksichtigte intervenierende Variable, die zufällig mit der unabhängigen Variable auftritt und kausal auf die abhängige Variable einwirkt, während die unabhängige Variable nur als Indikator fungiert.

Merke: Die abhängige (y-)Variable ist abhängig von der unabhängigen (x-)Variable, d. h. die unabhängige Variable verursacht die abhängige Variable.

Die Zuordnung der Werte zu den Variablen hängt von der Art und Weise der **Skalierung** ab. Die Variablen haben ihrerseits eine Struktur, die mittels einer Skala abgebildet wird.

- *Skalierung* ist die Anwendung eines Regelkalküls, um der Ausprägung eines Merkmals einen bestimmten Wert oder ein bestimmtes Symbol zuzuordnen.
- *Absolute und relative Skalierung:*
 - Bei der absoluten Skalierung wird ein Gegenstand ohne Bezug auf einen anderen Gegenstand beurteilt. Der Cut-off-Wert ist dabei derjenige Wert, ab dem ein einzelnes Testergebnis als außergewöhnlich gilt.
 - Bei einer einstufigen Skala können die Items nur einen von zwei Werten ("wahr" oder "falsch") annehmen. Die Summenskala ist dabei eine Skala bei der die Items zu einem Gesamtwert aufsummiert werden. Die Indexbildung ist hier ein Spezialfall: hierbei vergibt man bei Vorliegen eines Sachverhalts einen Punkt, die Punkte werden dann zusammengezählt.
 - Bei einer mehrstufigen Skala können die Items einen von mehr als zwei Werten (meist fünf) annehmen. Die Likert-Skala ist ein mehrstufiges verbal verankertes Antwortmodell (z. B. "nicht, wenig, mittel, ziemlich, stark")
 - Bei der relative Beurteilungsskala erfolgt die Bewertung eines Gegenstandes durch den Vergleich mit einem anderen Gegenstand (z. B. "x ist beliebter als y").














- *Skalen* sind Bezugssysteme zur Messung von Merkmalsausprägungen; sie können qualitativ oder quantitativ sein. Beispiele für Skalen sind folgende:

- Paarvergleich: jedes Item wird jeweils mit jedem anderen Item verglichen (A vs. B, A vs. C, A vs. D, B vs. C etc.)
- Rangordnung (Rangreihenvergleich): Objekte werden hinsichtlich eines Kriteriums geordnet (z. B. X ist größer als Y, Y ist größer als Z)
- Rating-Skala: Abstufungen von qualitativen Merkmalen (z. B. "gut – mittel – schlecht")
- Polaritätsprofil (= semantisches Differenzial): profilartige Erfassung der Bewertung von Objekten über Einordnung des Objekts in verschiedene Gegensatzpaare (z. B. "das Wohnheim ist laut vs. leise, billig vs. teuer etc.").
- Visuelle Analogskala: stetige Skala, wobei die Endpunkte die Extremausprägungen der einen oder anderen Seite darstellen (z. B. Schmerzmessung: linker Endpunkt = kein Schmerz, rechter Endpunkt = maximaler Schmerz)
- Guttman-Skala: Items und Personen werden gleichzeitig skaliert (z. B. 1,76 m großer Proband soll auf folgende Fragen antworten: "Sind Sie größer als 1,60 m?", "Sind Sie größer als 1,70 m?", "Sind Sie größer als 1,80 m?" etc.); plötzlicher "Bewertungssprung" (Beispiel: Proband antwortet bis Item X mit "Nein", ab Item X mit "Ja"); Ordinalskalenniveau

Merke: Der Begriff "Skala" hat je nach Kontext andere Bedeutung: Fragebogen, Messniveau oder Messinstrument.

- *Index* ist eine Skala, die aus mehreren Indikatoren zusammengesetzt ist (z. B. Körpermasse-Index [BMI]).
- Man unterscheidet verschiedene *Skalenniveaus*, die mit entsprechenden Maßen der mittleren Tendenz assoziiert sind.
- Skalen haben unterschiedliches Niveau. Je höher dabei das Niveau ist, mit desto aufwendigeren mathematischen Methoden lässt sich die Skala bearbeiten. Höhere Niveaus implizieren zudem niedrigere Niveaus und lassen sich in niedrigere Niveaus transformieren, was aber mit einem Informationsverlust einhergeht. Man unterscheidet nonmetrische und metrische Skalen.
 - Nonmetrische Skalen für qualitative (kategoriale) Merkmale
 - Nominalskala: bloße Zuordnung zu Kategorien, keine Beziehung zwischen den Werten (z. B. "Fisch – Fleisch – Gemüse – andere Nahrungsmittel"); zugehöriges Maß der mittleren Tendenz: Modalwert (= Modus, d. h. höchstbesetzte Kategorie)
 - Ordinalskala (Rangskala): Rangordnung (z. B. "klein – mittelgroß – groß"); die in der Ordinalskala gleichen Stufenabstände können vom Befragten als ungleich wahrgenommen werden; zugehörige Maße der mittleren Tendenz: Median (= derjenige Wert, der die Stichprobe genau halbiert), außerdem Quartile, Prozentrangwert, Spannweite
 - Metrische Skalen für quantitative Merkmale
 - Intervallskala (Abstandsskala): gleiche Abstände zwischen den Stufen, so dass man Differenzen bilden darf; der Nullpunkt ist aber variabel, weshalb keine Quotienten gebildet werden dürfen; zugehörige Maße der mittleren Tendenz: arithmetischer Mittelwert, außerdem Standardabweichung, Schiefe
 - Ratioskala (Verhältnisskala): gleiche Verhältnisse, d. h. gleiche Abstände und derselbe Nullpunkt, weshalb man Quotienten und Vielfache der Merkmalsausprägung bilden darf; zugehöriges Maß der mittleren Tendenz: geometrisches Mittel

Merke: Nach aufsteigendem Niveau unterscheidet man folgende Skalen: Nominalskala (besteht aus voneinander unabhängigen Kategorien), Ordinalskala (Rangordnung mit ungleichen Abständen), Intervallskala (Rangordnung mit gleichen Abständen), Ratioskala (Rangordnung mit gleichen Abständen und definiertem [absolutem] Nullpunkt). Mit zunehmendem Niveau sind mehr Rechenoperationen erlaubt.

Weblinks:  Operationalisierung,  Messung,  Unabhängige Variable,  Abhängige Variable,  Confounder,  Skalierung,  Skala,  Index,  Skalenniveau,  Nominalskala,  Ordinalskala,  Intervallskala,  Ratioskala

Selbsttest:

1. Was bedeutet "Operationalisierung"?

2. Ein Forscher möchte mittels verschiedener Antibiotika herausfinden, welches Antibiotikum eine gegebene Bakterienpopulation am besten abtötet. Was ist die abhängige Variable, was die unabhängige Variable?
3. Ordnen sie folgende Skalenniveaus nach absteigendem (!) Niveau: Intervallskala, Rationalkala, Nominalskala.
4. Bestimmen Sie die Niveaus der folgenden Skalen: 1. Celsius-Temperaturskala, 2. Schulzensuren, 3. Kelvin-Temperaturskala, 4. Statistik über die Besetzungszahlen von Praktikumsgruppen.

Ein guter Test besteht aus hochwertigen Items, genügt den Gütekriterien der Objektivität, Reliabilität und Validität und beugt erwartbaren Messfehlern vor

Eine qualitativ hochwertiger Test erfüllt verschiedene Kriterien:

- Die Items sind angemessen
- Der Test ist geeicht
- Der Test genügt den Kriterien Objektivität, Reliabilität und Validität
- Messfehler werden so gut wie möglich kontrolliert
- Der Test ist ökonomisch

Einen Test konstruiert man, indem man geeignete Items auswählt und zu Skalen zusammenfasst. Die Items sollten dabei bestimmten Qualitätsstandards genügen, was man mittels der **Itemanalyse** ermittelt.

- Die Schwierigkeit sollte angemessen sein (mittlerer Schwierigkeitsgrad).
- Die Trennschärfe sollte ausreichend sein, d. h. die Items sollten die Individuen mit starker Merkmalsausprägung von den Individuen mit schwacher Merkmalsausprägung gut trennen können; das Gesamtergebnis sollte positiv mit den Ergebnissen der Einzelitems korrelieren. Beispielsweise weisen manche Items (Einzelfragen) im Physikum eine schlechte Trennschärfe auf, d. h. sie werden von denjenigen, die eine gute Gesamtleistung erbringen genauso gut oder schlecht beantwortet wie von denjenigen, die eine schlechte Gesamtleistung erbringen. Solche Items sind somit nicht geeignet, zwischen "Guten" und "Schlechten" zu differenzieren und sind testtheoretisch daher nahezu wertlos.

Merke: Gute Testitems sind trennscharf (d. h. sie ermöglichen eine zuverlässige Trennung zwischen Individuen mit starker Merkmalsausprägung und Individuen mit schwacher Merkmalsausprägung) und besitzen einen angemessenen Schwierigkeitsgrad.

Unter "**Testnormierung**" versteht man den Vergleich der Untersuchungsgruppe mit einer Eichstichprobe (= Normstichprobe; Mittelwert und Standardabweichung müssen bekannt sein). Man berechnet also Kennwerte, die aussagen, wie sehr das Ergebnis der Untersuchungsgruppe von der statistischen Norm abweicht.

- Bei Normalverteilung der Eichstichprobe: Normierung anhand von Mittelwert und Standardabweichung
 - T-Wert (40-60)
 - Stanine (4-6)
- Bei nicht normalverteilten Eichstichproben: Normierung anhand von Prozenträngen (wieviel Prozent der Eichstichprobe liegen höher oder niedriger).

Ein guter Test erfüllt drei **Gütekriterien**, wobei das jeweils zuerst genannte Gütekriterium notwendige Voraussetzungen für das nächstgenannte ist.

- **Objektivität:** Unabhängigkeit des Testergebnisses von Situation und Testleiter
 - Objektivität ist erreichbar durch Standardisierung, d. h. durch standardisierte Instruktionen und streng formalisierte Auswertungsprozeduren.
 - Sie lässt sich prüfen, indem man Testergebnisse unterschiedlicher Prüfungsleiter vergleicht, d. h. indem man das Ausmaß der Interrater-Reliabilität bestimmt.
 - Entsprechend der drei Phasen eines Tests unterscheidet man drei Arten von Objektivität:
 - Durchführungsobjektivität

- Auswertungsobjektivität; beispielsweise besitzt das schriftliche Physikum eine fast perfekte Auswertungsobjektivität (Multiple-Choice-Fragen, Auswertung mittels Automaten etc.)
- Interpretationsobjektivität
- **Reliabilität:** Genauigkeit, formale Zuverlässigkeit einer Messung; bei einem sehr reliablen Test ist also der Standardmessfehler (= zufälliger Fehler) sehr klein.
- Die Reliabilität lässt sich prüfen, indem man die Reproduzierbarkeit erfasst, d. h. die Genauigkeit, mit der Testergebnisse bei verschiedenen Messungen übereinstimmen. Man erhält dadurch einen Korrelationskoeffizienten, und zwar den den Reliabilitätskoeffizienten.
- Es gibt dabei verschiedene Methoden zur Reliabilitätsabschätzung:
 - Test-Retest-Reliabilität (Testwiederholungsreliabilität): ein gleichbleibendes Merkmal wird 2x hintereinander gemessen (aber: es können Übungs- und Lerneffekte auftreten; es liegt zudem die Annahme zugrunde, dass das Merkmal über die Zeit konstant bleibt)
 - Paralleltest-Reliabilität (= Äquivalenz-Reliabilität): zwei gleiche Versionen eines Tests, d. h. zwei Tests, die die gleichen Merkmale mit Hilfe von unterschiedlichen Fragen erfassen, werden miteinander verglichen, d. h. miteinander korreliert (aber: es lässt sich schwer feststellen, ob die Tests tatsächlich gleich sind)
 - Testhalbierungs-Reliabilität (= Split-Half-Reliabilität): die Skala wird zweigeteilt und beide Hälften werden miteinander korreliert.
 - Interne Konsistenz: die einzelnen Items werden miteinander korreliert (dem liegt die Annahme zugrunde, dass jedes Item exakt das gleiche Merkmal misst; die interne Konsistenz ist zudem stark abhängig von der Zahl der Items, sie steigt bei Hinzunahme von Items); Cronbachs α dient als Maß der Homogenität eines Tests.
- Als Maß für die Reliabilität dient Cohens κ , das sich aus dem Quotienten der Differenz zwischen beobachteter und erwarteter Übereinstimmungswahrscheinlichkeit ($p_b - p_e$) einerseits und der zur erwarteten Übereinstimmungswahrscheinlichkeit komplementären Wahrscheinlichkeit ($1 - p_e$) andererseits errechnet; es kann Werte zwischen 1 (perfekte Übereinstimmung) und 0 (rein zufällige Übereinstimmung) annehmen.
- **Validität:** Ausmaß, mit dem der Test das misst, was er messen soll (beispielsweise misst ein hochvalider Intelligenztest die Intelligenz sehr gut); die Validität kann nicht quantifiziert werden, sondern wird mittels verschiedener Strategien ungefähr ermittelt.
- Validität von Messinstrumenten
 - Bei der internen Validierung geht man nach eigenem Ermessen vor.
 - Augenscheinvalidität (als Kriterium dient die Evidenz)
 - Expertenrating (Experten beurteilen die Validität des Tests)
 - Bei der externen Validierung bezieht man sich auf ein Außenkriterium.
 - Kriteriumsbezogene Validität: der Test wird mit einem Referenzstandard in Beziehung gesetzt, d. h. mit einem Merkmal, von dem man glaubt, dass es mit dem zu messenden Merkmal stark zusammenhängt ("Goldstandard").
 - Konkurrente Validität (zu messendes Kriterium liegt in der Gegenwart)
 - Prädikative/prognostische Validität (zu messendes Kriterium liegt in der Zukunft, s. u.)
 - Diskriminative Validität: die Validität wird dadurch beurteilt, indem man erfasst, wie stark sich unterschiedliche Gruppen im Test unterscheiden.
 - Prognostische Validität: aus den Testergebnissen leitet man Vorhersagen ab und prüft dann, ob die Vorhersagen eintreffen.
 - Konstruktvalidität: der fragliche Test wird auf einen anderen Test bezogen
 - Konvergente Validität: Vergleich mit Indikatoren desselben Konstrukts (z. B. wird ein Intelligenztest mit einem anderen Intelligenztest verglichen; Gefahr: Zirkelschlüsse, Kreisvalidierung).
 - Divergente/diskriminante Validität: Vergleich mit Indikatoren eines anderen Konstrukts, um die Unabhängigkeit eines Tests zu prüfen (z. B. wie unabhängig ein Intelligenztest von einem Prüfungsangst-Test ist).
 - Unter "Änderungssensitivität" versteht man die Fähigkeit des Tests, Veränderungen (z. B. im Krankheitsverlauf) zu erfassen.
- Validität von Aussagen über Kausalzusammenhänge:










- Interne Validität: die Änderung der abhängigen Variable ist eindeutig auf die Änderung der unabhängigen Variable zurückzuführen; Ergebnisse und ihre Konsequenzen sind für die Stichprobe korrekt.
- Externe Validität: die Ergebnisse, die in der untersuchten Population gefunden worden sind, lassen sich auf größere (Teil-)Populationen übertragen.

Merke: Gute Tests erfüllen die Kriterien der Objektivität (Kontextunabhängigkeit), Reliabilität (Genauigkeit, Zuverlässigkeit) und Validität (Gültigkeit).

Testergebnisse weisen Streuungen auf und können durch bestimmte psychische Dispositionen der Versuchspersonen verfälscht werden. Derartige **Messfehler** muss man einkalkulieren:

- Der *Standardmessfehler* ergibt sich aus der Streuung des zufälligen Fehlers, der auf inter- und intraindividuelle Variabilität beruht (vgl. systematischer Messfehler).
 - Es lässt sich ein Vertrauensbereich (Konfidenzintervall) berechnen, in dem der wahre Wert sehr wahrscheinlich liegt; Unterschiede bei Testergebnissen sind erst dann "richtige" Unterschiede, wenn sie das Konfidenzintervall übersteigen. Je kleiner das Konfidenzintervall, desto geringer ist der Standardmessfehler.
 - Der Standardmessfehler wird reduziert, indem man entweder mehr Items hinzunimmt (→ Vergrößerung von n; Gesetz der großen Zahlen) oder Stichprobenteilnehmer selektiert oder die Stichprobe stratifiziert (→ Verkleinerung der Standardabweichung).
- *Antworttendenzen* können dazu führen, dass der Proband etwas anderes sagt als er eigentlich meint:
 - Soziale Erwünschtheit: der Proband sagt das, was erwartet wird (Abhilfe: Abschätzung mittels einer in den Test integrierten Lügenskala).
 - Ja-sage-Tendenz (Abhilfe: 50 % der Fragen negativ [d. h. als "Nicht"-Aussagen] formulieren)
 - Tendenz zur Mitte (Abhilfe: gerade Anzahl von Fragen)

Tests sollten auch **ökonomisch** gestaltet sein. Sie sollten zum einen eine möglichst große Reliabilität und Validität bei möglichst wenigen Items gewährleisten; jedoch erhöht sich die Reliabilität, je mehr Items man verwendet. Zum anderen sollten Tests auch leicht auszuwerten sein.

Weblinks:  Itemanalyse,  Trennschärfe,  Normierung,  Objektivität,  Reliabilität,  Validität,  Standardmessfehler,  Operationalisierung,  Soziale Erwünschtheit

Selbsttest:

1. Definieren Sie kurz die Begriffe "Objektivität", "Reliabilität" und "Validität" im Hinblick auf Messinstrumente/Tests.
2. Ist ein Schulaufsatz im Fach Deutsch objektiv, reliabel und valide? Und wie sieht es mit dem schriftlichen Physikum hinsichtlich dieser Kriterien aus?
3. Was bedeutet "externe Validität", was "interne Validität"?
4. Was ist ein Konfidenzintervall?
5. Welche psychischen Dispositionen des Probanden können eine Messung verfälschen?

Die Messung von Merkmalen hat einen standardisierten Ablauf: Planung der Studie, Stichprobenbildung, Datengewinnung, Auswertung

Planung

Am Anfang einer Messung steht die **Planung**, d. h. die Entscheidung, welches Studiendesign man verwendet. Diese Designs sind dabei nicht gleichwertig, sondern lassen sich hinsichtlich ihrer internen Validität in einer Hierarchie bringen, wobei die Randomisierte kontrollierte Studie an der Spitze dieser Hierarchie steht.

Arten von Studiendesigns

Experimentelle Studiendesigns:

- Experiment
 - Definition: ein Experiment ist gekennzeichnet durch planmäßige Beeinflussung einer Situation und Überprüfung der Wirkungen.
 - Voraussetzungen für ein Experiment:
 - Wiederholbarkeit
 - Kausalanalyse
 - Kontrolle der Bedingungen
- Die *Randomisierte kontrollierte Studie* (RCT [randomized controlled trial], Interventionsstudie, klinisch kontrollierte Studie mit experimentellem Design) gilt als "Prototyp" einer experimentellen Studie und ist am besten geeignet, um die Wirksamkeit einer Maßnahme zu prüfen. Sie wird im Folgenden daher näher erläutert.
 - Die Studie nennt sich "kontrolliert", weil der Experimentalgruppe (EG; Intervention) eine strukturgleiche Kontrollgruppe (KG; ohne Intervention) gegenübersteht und somit ungeplante Einflussfaktoren *kontrolliert* werden (z. B. Spontanremission im natürlichen Krankheitsverlauf, Plazebo-Effekt, Effekte von anderen Behandlungen, statistische Regression zur Mitte).
 - Die Studie nennt sich "randomisiert", weil die Verteilung der Probanden auf EG und KG *zufällig* erfolgt; ein Sonderfall ist die Blockrandomisierung (Zufallsverteilung innerhalb eines definierten Blocks).
 - Für die RCT gelten diverse Postulate:
 - Jede RCT muss von einer Ethikkommission genehmigt werden.
 - EG und KG müssen äquivalent sein: die Gruppen müssen strukturell gleich zusammengesetzt sein, um personenbezogene Einflussfaktoren (Alter, Geschlecht, Schicht, Motivation, unbekannte Faktoren) auf das Ergebnis gleichmäßig zu verteilen. Dies lässt sich erreichen mittels Randomisierung (muss unvorhersehbar sein; externe/zentrale Randomisierung, um Fälschungsversuche des Forschers zu vermeiden) und Parallelisierung (je einem Mitglied der KG steht ein Mitglied der EG mit gleichen Merkmalen gegenüber).
 - Es kommt auf Vollständigkeit und aufs "Durchhalten" von Studienprinzipien an.
 - Die Katamnese (Follow-up-Messung) muss vollständig sein, um eine selektive Verzerrung der Stichprobe durch Nichtberücksichtigen der Ausfälle (drop-out) zu verhindern.
 - Intention-to-treat: um die Strukturgleichheit zwischen EG und KG zu wahren, müssen alle Patienten – auch Studienabbrecher – in derjenigen Gruppe analysiert werden, der sie zugeteilt worden sind. Die Prinzipien As-Treated und Per-Protocol schließen Studienabbrecher aus, was zwar Unterschiede zwischen EG und KG deutlicher hervortreten lässt, aber zu Lasten der Strukturgleichheit geht.
 - Die Studie muss mindestens doppelt verblindet sein (d. h. Proband und Versuchsleiter wissen beide nicht, ob der Proband zur KG oder zur EG gehört), um Erwartungseffekte der Versuchspersonen (Plazebo-Effekt, Hawthorne-Effekt) und Versuchsleiter-Effekte (Rosenthal-Effekt) zu vermeiden. Für Notfälle gibt es einen Notfallumschlag, der über die Gruppenzugehörigkeit des Patienten informiert.
 - Bei mehrmaliger Messung gleicher Versuchspersonen sollte man Verzerrungen aufgrund der Testreihenfolge vermeiden, und zwar mittels Ausbalancieren (Variieren der Reihenfolge).

Nichtexperimentelle Designs: naturalistische Studien, Beobachtungs-Studien

Quasiexperimentelle Designs: Rückgriff auf eine vorhandene Vergleichsgruppe (d. h. es findet keine Randomisierung statt)

Merke: Kausale Zusammenhänge (z. B. Wirkungsnachweise) lassen sich am besten mit Hilfe der randomisierten kontrollierten Studie prüfen. Sie ist die "Königin" unter den Studien.

Hierarchie der Studiendesigns

Die Studiendesigns lassen sich hinsichtlich ihrer internen Validität in aufsteigender Folge anordnen:

- Ex-post-facto-Studie: nachträgliche Erklärung vorliegender Daten
- Querschnittsstudie: Zu einem Zeitpunkt T werden alle interessierenden Daten erhoben; dadurch lassen sich die Häufigkeit eines Merkmals und Korrelationen zwischen Merkmalen ermitteln
- Einzelfallstudie: Qualitative/quantitative Analyse von Einzelfällen
- Evaluationsstudie
- Ein-Gruppen-Prä-Post-Design zur Beschreibung von Veränderungen, die im Verlauf einer Intervention auftreten (keine Kausalität!)
- Fall-Kontroll-Studie (deskriptiv, analytisch): Vergleich jedes "Falls" aus der Experimentalgruppe mit einem passenden (gematchten) "Nichtfall" aus der zur Experimentalgruppe möglichst strukturgleichen Kontrollgruppe

- Prinzip: "Fälle" (haben Merkmal) und "Nichtfälle" (haben Merkmal nicht) werden retrospektiv nach Risikofaktoren befragt – man geht also vom Diagnosestatus zum Expositionsstatus; als Kennziffer dient die

$$\text{Odds Ratio: } Odds\ Ratio = \frac{\frac{Exponierte_F}{Nichtexponierte_F}}{\frac{Exponierte_N}{Nichtexponierte_N}}$$

Die Odds Ratio ist das Verhältnis zweier Chancen, d. h. das Verhältnis der

"Chance", bei Exposition zum Fall zu werden (z. B. eine bestimmte Krankheit zu bekommen), zur "Chance" bei fehlender Exposition zum Fall zu werden. Ist die Odds ratio größer oder kleiner als 1, so besteht ein Chancen-Ungleichgewicht, beträgt sie hingegen 1 (oder beinhaltet das Konfidenzintervall die 1), so besteht kein Zusammenhang zwischen dem Fall-Merkmal und dem Expositionsstatus.

- Fehlerquellen: Wahl der Kontrollgruppe, selektive Erinnerung (recall bias, d. h. künstliche Assoziation zwischen Risikofaktor und Erkrankung)

	Exponierte	Nichtexponierte
Fälle (F)	Exponierte_F	Nichtexponierte_F
Nichtfälle (N)	Exponierte_N	Nichtexponierte_N

- Kohortenstudie (prospektiv, longitudinal): Eine Untersuchungsgruppe (Kohorte, d. h. Gruppe von Personen mit gemeinsamem Merkmal) mit Mitgliedern, deren Expositionsstatus bekannt ist, wird prospektiv (daher kein recall bias!) über einen längeren Zeitraum beobachtet, um den Zeitpunkt herauszufinden, an dem ein bestimmtes Ereignis (Krankheit, Tod) eintritt. Es wird also die Inzidenz eines Ereignisses ermittelt, durch Bezug auf den Expositionsstatus lassen sich dann verschiedene Risikokennziffern berechnen.

- Methode:

1. Erfassen von Risikofaktoren der Kohorten-Mitglieder,
2. Abwarten, bis Erkrankung eintritt.
3. Auswertung:

- Kennziffern: Relatives Risiko (Risiko, z. B. an einer Krankheit zu erkranken, wenn man den Risikofaktor trägt:

$$\text{Relatives Risiko} = \frac{\text{Risiko der exponierten Gruppe}}{\text{Risiko der nicht exponierten Gruppe}}$$

Absolute Risikoreduktion (ARR;

durch Ausschaltung des Risikofaktors mögliche Reduktion der Erkrankungswahrscheinlichkeit; entspricht dem attributablen Risiko), Number Needed to Treat (NNT: = 1/ARR)

- Überlebenszeitanalyse: man erstellt eine treppenartig verlaufende Überlebenskurve, aus der sich verschiedene Überlebensraten (z. B. 5-Jahres-Überlebensrate, mediane Überlebensrate) bei gegebener Krankheit ermitteln lassen; mit Hilfe der Kaplan-Meier-Methode gelingt es dabei, zensierte Daten (Studienabbrecher, Drop-Outs) zu berücksichtigen.
- Spezielle Kohortenstudien:
 - Historische (retrospektive, konkurrierende) Kohortenstudie: hierbei liegt der Startpunkt der Studie in der Vergangenheit.
 - Eingebettete (= nested) Fall-Kontrollstudie: die Studie beginnt als Kohortenstudie, im Verlauf der Studie – wenn Untersuchungsmaterial gesammelt und der Krankheitsstatus der Teilnehmer bekannt ist (Aufteilung der Kohorte in Fälle und Nichtfälle) – lässt sich die Kohortenstudie zusätzlich als Fall-Kontroll-Studie auswerten.
- Mittels Kohortenstudien werden allerdings nur Korrelationen gemessen (Kriterien der biologischen Plausibilität [biologische Plausibilität, Chronologie, Dosis-Wirkung-Beziehung etc.] können nur auf kausale Zusammenhänge *hinweisen*; Confounder!); Risikofaktoren wirken als "Marker" (= prognostischer Indikator), d. h. sie lassen eine Vorhersage über das Krankheitsrisiko zu, wenn sie auch nicht kausal wirken. Die Kohortenstudie bildet die wichtigste Alternative zur randomisierten kontrollierten Studie, die bei Menschen ethisch meist sehr problematisch ist.
- Randomisierte kontrollierte Studie.

Merke: Bei der Fall-Kontroll-Studie geht man retrospektiv vor (Diagnosestatus → Expositionsstatus), bei der Kohortenstudie prospektiv (Expositionsstatus → Diagnosestatus).

Besondere Studienarten

- Einzelfallstudie: Baseline-Erhebung → Intervention mit Messwert-Registrierung → Absetzen der Intervention und prüfen, ob Messwerte auf Niveau der Baseline zurückfallen → nochmalige Intervention mit Messwert-Registrierung; der Patient dient somit gewissermaßen als sein eigener Kontroll-Patient.
- Multizentrische Studie (= gemeinsames Forschungsprojekt mehrerer Standorte).

Stichprobenauswahl

Um Merkmale einer Population zu untersuchen, wäre es sehr unökonomisch, sämtliche Mitglieder der Population zu befragen. Vielmehr bildet man eine **Stichprobe**, die gewissermaßen eine Miniaturausgabe der Gesamtpopulation darstellt.

- Zunächst gilt es, die *Grundgesamtheit* festzulegen ("wen will ich überhaupt untersuchen?"), und zwar mittels Ein- und Ausschlusskriterien.
- Anschließend ist eine *Stichprobe* zu bilden, die die Grundgesamtheit/Population möglichst gut repräsentiert.
 - Je größer die Stichprobe, desto mehr ähnelt sie der Population; das Konfidenzintervall wird kleiner, Stichprobenfehler (etwa aufgrund schlechter Stichprobenauswahl) minimieren sich.
 - Die Auswahl der Stichproben-Mitglieder kann mittels verschiedener Verfahren geschehen, und zwar per...
 - ...Zufall, d. h. jedes Populationsmitglied hat die gleiche Chance, in die Stichprobe zu kommen, was die Repräsentativität (= Ähnlichkeit mit der Grundgesamtheit) der Studie gewährleistet; eine Sonderform ist die geschichtete (stratifizierte) Zufallsauswahl: dabei wird die Grundgesamtheit in Untergruppen aufgeteilt, innerhalb dieser Untergruppen erfolgt dann eine Zufallsauswahl.
 - ...Konsekution, d. h. eintreffende Patienten werden nacheinander in die Stichprobe aufgenommen.
 - ...Quote, d. h. der Interviewer wählt anhand der Ein- und Ausschlusskriterien die Probanden selbst aus; hier können jedoch leicht Verzerrungseffekte auftreten.
- *Weitere Begriffe und Aspekte*
 - Quotastichprobe: die Eigenschaften der Grundgesamtheit sind bekannt, durch gezielte Auswahl von Personen anhand von zu erfüllenden Quoten (*keine* Randomisierung/Parallelisierung) bildet man eine Miniaturausgabe dieser Grundgesamtheit.
 - Klumpenauswahl: Zusammenfassung von Gruppen zu Klumpen (z. B. Stadtteile).
 - Extremgruppe: Gruppe aus Probanden mit extremer Merkmalsausprägung (± 2 Standardabweichungen).

Merke: Die Stichprobe ist eine Miniaturausgabe der zu untersuchenden Grundgesamtheit. Sie kann gewonnen werden per Zufall (ermöglicht die beste Repräsentativität der Stichprobe), Konsekution oder Quote. Bei der Quotastichprobe hingegen wird die Stichprobe durch gezielte Probandenauswahl gebildet.

Datengewinnung

Aus der Stichprobe **gewinnt man nun Daten**, und zwar mittels Befragung oder mittels Beobachtung.

- *Befragung* (= Durchführen von Interviews):
 - Man unterscheidet Fremdbeurteilung (externe Beobachter beschreiben Merkmale des Individuums X) und Selbstbeurteilung (Individuum X beschreibt eigene Merkmale):
 - Fremd- und Selbstbeurteilung weichen oft stark voneinander ab, wobei die Selbstbeurteilung meist günstiger ausfällt; eine Ausnahme bilden hier Beurteilungen stark subjektiv zugänglicher Merkmale wie Schmerzintensitäten
 - Die Übereinstimmung hinsichtlich eines Merkmals ist umso größer, je besser das Merkmal von außen wahrnehmbar ist.
 - Das Interview kann verschiedene Fragetypen enthalten:
 - Offene Frage (Sonderfall: Sondierungsfrage)
 - Geschlossene Frage
 - Katalogfrage; Sonderfall: Alternativfrage ("ja oder nein")
 - Skalierungsfrage
 - Suggestivfrage
 - Oftmals sind bei den Antworten des Interviewten psychische Antworttendenzen im Spiel, die man beachten sollte, und zwar unter anderem:
 - Tendenz zur Mitte.
 - Soziale Erwünschtheit.
 - Ja-sage-Tendenz.
 - Es gibt verschiedene Arten von Interviews:
 - Quantitative Verfahren (unterschiedlich stark standardisiert)
 - Unstandardisiert (nur das Thema ist vorgegeben): Offenes Interview (offene Frage, Antwort allein vom Antwortenden abhängig)
 - Standardisiert:
 - Teilstrukturiertes (halbstandardisiertes) Interview: offene Eingangsfrage, dann Abarbeitung eines Themenkatalogs (Interviewleitfaden); dies gewährleistet eine größere Flexibilität bei gleichen Themen, ist allerdings mit einer aufwendigen Auswertung verbunden
 - Strukturiertes Interview: Fragen und oft auch Antwortmöglichkeiten sind vorgegeben; dies garantiert eine leichte Auswertung, ist jedoch mit einer starken Einengung auf spezifische Themen verbunden.
 - Qualitative (= hermeneutische) Verfahren:
 - Biografisches Interview (→ individuelle Lebensgeschichte)
 - Ethnografisches Interview (→ kulturspezifische Besonderheiten)
 - Narratives Interview (etwas über einen bestimmten Gegenstand erzählen)
 - Tiefeninterview (→ Psychoanalyse)
 - Der Interviewer bestimmt den Interviewstil dadurch, dass er dem Interviewten wenig (direktiv) oder viel (nondirektiv) Gestaltungsspielraum lässt:
 - Direktiv: der Interviewer, der klar strukturierte Konzepte und Ziele verfolgt, bestimmt Interviewablauf; auf klare Fragen (v. a. geschlossene Fragen) folgen klare Antworten.
 - Nondirektiv: der Interviewte erhält stärkeren Einfluss; dieser Stil ist gekennzeichnet durch offene Fragen (als Anregung zum Nachdenken), Empathie und Rückmeldung des Interviewers.
 - Die erhobenen Daten lassen sich nach unterschiedlichen Kriterien klassifizieren:

- Primärdaten (nur für Studie X erhoben), Sekundärdaten (bei einer anderen Studie Y erhoben, aber für Studie Z verwendet; haben eine geringe Aussagekraft, weil evtl. Confounder in der früheren Studie nicht erfasst worden sind).
- Individualdaten (Daten über ein einzelnes Individuum), Aggregatdaten (Zusammenfassung von Daten mehrerer Individuen; ökologische Studien: aggregierte Daten aus großen Gebieten); ein Aggregat ist eine Gruppe von Individuen, die ein gemeinsames Merkmal besitzen, aber nicht miteinander in Kontakt treten können.
- Selbstbeurteilung, Fremdbeurteilung
- *Beobachtung:*
 - Systematische Beobachtungen haben festgelegte Rahmenbedingungen, ein festgelegtes Beobachtungssystem (Zeichensystem oder Kategoriensystem) und festgelegte Beobachtungskriterien (d. h. es ist festgelegt, was beobachtet wird und wie es eingeordnet wird); eine hohe Interrater-Übereinstimmung lässt sich durch Training der Beobachter erreichen.
 - Teilnehmende Beobachtung: der Beobachter begibt sich in die zu beobachtende Situation hinein; diese Art der Beobachtung ist weniger standardisiert

Feldstudie: Studie, bei der der Proband in seinem Umfeld beobachtet wird ohne es zu wissen

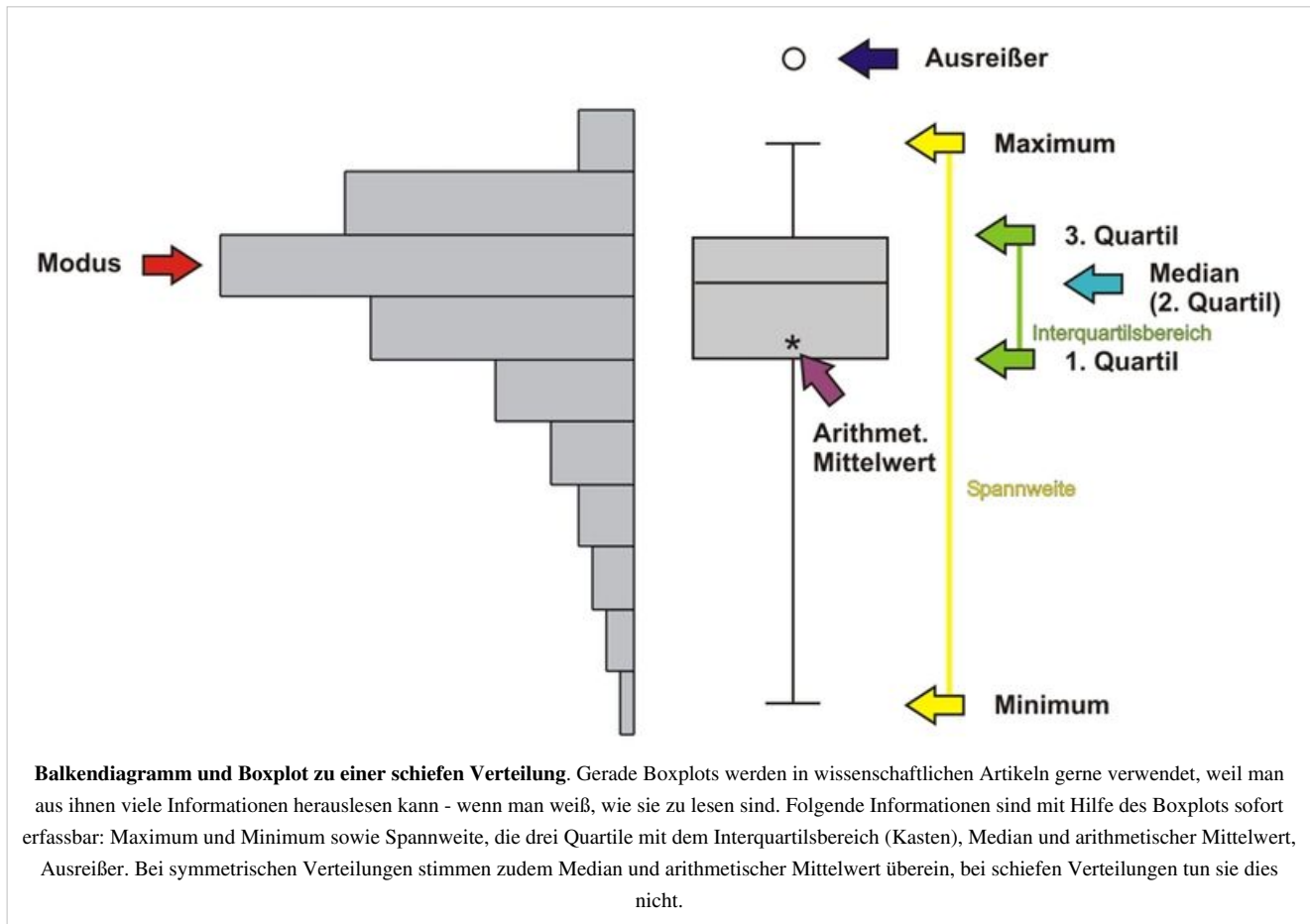
Merke: Sekundärdaten sind Daten, die ursprünglich zu einem anderen Zweck als der Studie erhoben worden sind.

Auswertung

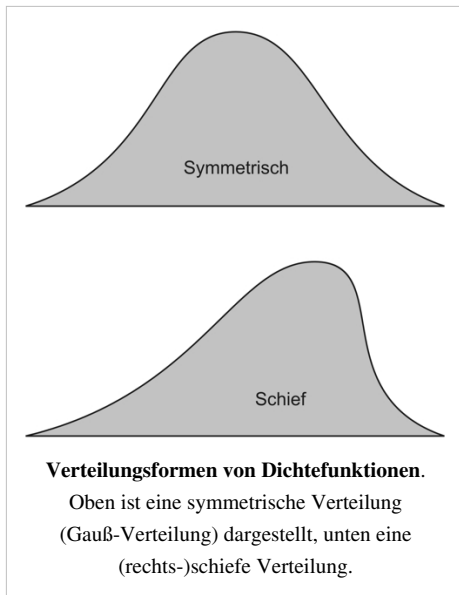
Um aus den gewonnenen Daten Erkenntnisse zu gewinnen, müssen sie **ausgewertet** werden. Quantitative Daten kann man direkt auswerten, qualitative Daten müssen erst in quantitative Daten transformiert werden.

Quantitative Auswertung

Im Rahmen der **deskriptiven Statistik** werden die individuellen Daten zusammengefasst und in charakteristischen Kennwerten repräsentiert; man stellt also dar, wie die Verhältnisse innerhalb der Stichprobe beschaffen sind.



Mittelwerte der Standardnormen	
Skala	Mittelwert
z	0
Z	100
C (Stanine)	5
Prozentrang	0,5



- **Univariate Analyse:** jede einzelne Variable wird für sich analysiert.
 - Formen der Verteilung von Merkmalsausprägungen:
 - Schief
 - Symmetrisch (= normalverteilt): Glockenförmig Dichtekurve mit asymptotischer Annäherung an die x-Achse, zwischen den beiden Wendepunkten liegen 2/3 der Gesamtfläche; jede normalverteilte Zufallsvariable ist mittels der z-Transformation in eine *Standardnormalverteilung* (Mittelwert = 0, Varianz = 1) umwandelbar – unterschiedliche Stichproben lassen sich somit vergleichen. Normalverteilt sind Mittelwerte (auch bei schief verteilter Grundgesamtheit), Zufallsvariablen (bei vielen unabhängigen Einflüssen) und Summen von vielen Zufallsvariablen (zentraler Grenzwertsatz).
 - Merkmale können diskret (abzählbar viele Merkmalsausprägungen) oder stetig (potentiell unendlich viele Merkmalsausprägungen) sein:
 - Diskrete Merkmale: es lassen sich absolute und relative Häufigkeiten ermitteln und mittels Kreis-, Rechteck-, Balken- oder Punktediagrammen darstellen.
 - Stetige Merkmale: die Ausprägungen lassen sich zu einer endlichen Zahl von Klassen zusammenfassen und graphisch darstellen, meist mittels eines Histogramms (gibt Aufschluss über Lage und Streuung der Werte sowie über die Form der Verteilung).
 - Häufigkeiten:
 - Eindimensionale Häufigkeiten:
 - Absolute Häufigkeit: Summe der Vorkommnisse einer bestimmten Merkmalsausprägung.
 - Relative Häufigkeit: Quotient aus absoluter Häufigkeit und Stichprobenumfang
 - Kumulative Häufigkeit (= Summenhäufigkeit): durch Aufaddieren der absoluten oder relativen Häufigkeiten ergibt sich eine monoton ansteigende Treppenfunktion mit Werten zwischen 0 und 1 (empirische Verteilungsfunktion).
 - Zweidimensionale Häufigkeiten: werden mittels einer Kontingenztafel dargestellt, bei der die Ausprägungen zweier (diskreter) Merkmale miteinander kombiniert werden; die bekannteste Kontingenztafel ist die Vierfeldertafel, aus der sich u. a. die Odds ratio errechnen lässt.
 - Maße: je nach Skala, abhängig von der Verteilung der Werte
 - Lagemaße (Lokalisationsmaße):
 - Wenn keine Gauß-Verteilung vorliegt:
 - Median (= Zentralwert): Punkt, an dem Zahl der Darüberliegenden gleich der Zahl der Darunterliegenden ist und somit die Stichprobe in zwei gleich große Hälften teilt
 - Modus (= Modalwert, Dichtemittel): Ausprägung mit der größten Häufigkeit (unimodale Verteilung; z. B. Ausprägung mit längstem Balken im Balkendiagramm) oder den größten Häufigkeiten (multimodale Verteilung); bei nominal skalierten Daten ist der Modus das einzige erlaubte Lagemaß.

- Minimum und Maximum: niedrigste bzw. höchste Ausprägung, geben groben Aufschluss über die Streubreite der Daten.
 - Wenn eine Gauß-Verteilung vorliegt, ist zusätzlich der arithmetische Mittelwert (= Durchschnitt) bestimmbar, wobei jedoch Ausreißer berücksichtigt werden müssen (beeinflussen den Mittelwert stark)
 - Streuungsmaße (= Dispersionsmaße):
 - Wenn keine Gauß-Verteilung vorliegt:
 - Dezilabstand (Interdezilbereich: Abstand vom ersten Dezil bis zum letzten [= neunten] Dezil, d. h. zwischen 10 % und 90 %) und Quartilabstand (Interquartilbereich: Abstand vom ersten Quartil bis zum letzten [= dritten] Quartil, d. h. Bereich zwischen 25 % und 75 %, hierin liegen die mittleren 50 %)
 - Spannweite (= Variationsbreite): Abstand zwischen Minimum und Maximum
 - Wenn eine Gauß-Verteilung vorliegt, sind zudem folgende Streuungsmaße sinnvoll:
 - Standardabweichung (= SD, Streuung): 68 % aller Werte liegen im Intervall "Mittelwert \pm 1 SD", 95% im Intervall "Mittelwert \pm 2 SD"; die Streuung macht also Aussagen über die Homogenität einer Stichprobe (SD = 0 wäre perfekt homogen); entspricht der Wurzel der Varianz
 - Varianz: Abweichung der Stichprobenwerte vom arithmetischen Mittelwert; entspricht SD^2 ;

$$Var = \frac{\text{Summe der Abstandsquadrate}}{n - 1}$$
 - Variationskoeffizient (nur für rational skalierte Merkmale)
 - Formmaße
 - Schiefe: gibt Aufschluss über Symmetrie oder Asymmetrie der Verteilung.
 - Wölbung (= Kurtosis, Exzess): lässt Massenanhäufungen um den Mittelwert oder um die Extremwerte erkennen.
 - Aussagen über Abweichungen von der Vergleichspopulation sind möglich mittels z-Transformation, T-Werten, Staninen, Prozenträngen (Rangperzentile; auch bei schiefer Verteilung möglich!)
 - Bivariate Analyse: man analysiert, wie stark zwei Variablen miteinander zusammenhängen, man ermittelt also das Ausmaß der Korrelation zweier Variablen; man beachte, dass sich dabei nur Korrelationen und nicht Kausalzusammenhänge ermitteln lassen, denn Kausalzusammenhänge kann man nur mittels einer Kausalanalyse anhand des Untersuchungsdesigns feststellen.
 - Ein erster Eindruck hinsichtlich der Korrelation lässt sich anhand der Punktwolke gewinnen.
 - Anhand der Kovarianz s_{xy} lässt sich berechnen, ob die beiden Merkmale in die gleiche Richtung ($s_{xy} > 0$), in verschiedene Richtungen ($s_{xy} < 0$) miteinander variieren, oder ob zwischen beiden kein linearer Zusammenhang besteht ($s_{xy} = 0$).
 - Sind beide Variablen quantitativ, sind die Beobachtungseinheiten voneinander unabhängig, sind die Merkmale mindestens intervallskaliert und besteht ein in etwa linearer Zusammenhang zwischen den Merkmalen, so lässt sich der PEARSONSche Korrelationskoeffizient (Produkt-Moment-Korrelationskoeffizient) r bestimmen:

$$r = \frac{\text{Kovarianz}}{\text{Standardabweichung}_{\text{Variable 1}} * \text{Standardabweichung}_{\text{Variable 2}}}$$
 Er liegt zwischen +1 (perfekter positiver Zusammenhang) und -1 (perfekter negativer Zusammenhang)
 - $r = 0$ bedeutet, dass zwischen den beiden Variablen kein Zusammenhang besteht, die Messwerte sind zufällig verteilt, es herrscht maximale Streuung.
 - $r = 0,1$ steht für einen schwachen positiven Zusammenhang, $r = 0,3$ für einen mittelstarken positiven Zusammenhang, $r = 0,5$ für einen starken positiven Zusammenhang.
- Zudem lässt sich der – auf weniger Voraussetzungen beruhende – SPEARMANSche Korrelationskoeffizient (Rangkorrelation) errechnen, wobei man die Werte der jeweiligen Variable sortiert, ihnen Ränge zuordnet und dann die Korrelation zwischen den Rängen ermittelt.
- Mit Hilfe der Regressionsanalyse lässt sich eine Geradengleichung für eine Gerade bestimmen, welche die Art des Zusammenhangs zweier Merkmale am besten wiedergibt; dabei schließt man von der unabhängigen (x-)Variable auf die abhängige (y-)Variable, man sagt also die Ausprägungen der abhängigen Variable durch Werte der unabhängigen Variable voraus (unabhängige Variable als *Prädiktor* auf die abhängige Variable als *Kriterium*). Sind die Ausprägungen der x-Variable bekannt, so spricht man von Regression 1. Art, ist jedoch auch die x-Variable eine Zufallsvariable, so spricht man von Regression 2. Art. Bei mehreren x-Variablen, die auf die y-Variable Einfluss

nehmen, verwendet man die multiple Regressionsanalyse.

- Durch die Regressionsanalyse lässt sich das Bestimmtheitsmaß (Determinationskoeffizient) berechnen. Es sagt aus, wie viel Prozent der Varianz einer Variable durch die andere Variable determiniert wird und beträgt das Quadrat des Korrelationskoeffizienten, d. h. Bestimmtheitsmaß = r^2 . Das Bestimmtheitsmaß ergibt sich aus dem Quotienten der durch die Regressionsanalyse ermittelten erklärten Varianz und der Gesamtvarianz (= erklärte Varianz + Varianz der Residuen [Residuen = Abweichungen der tatsächlichen Werte der Punktwolke von den durch die Regressionsgerade vorhergesagten Werten]), d. h.

$$r^2 = \frac{\text{erklärte Varianz}}{\text{erklärte Varianz} + \text{Varianz der Residuen}}$$

- **Multivariate Analyse:** man analysiert, wie mehr als zwei Variablen zusammenhängen (multivariat: mehrere abhängige Variablen; multivariabel: mehrere unabhängige Variablen).
 - Multiple Regressionsanalyse
 - Lineare Strukturgleichungsmodelle, Pfadanalysen, Kausalanalysen: Bildung von Modellen des Zusammenhangs zwischen unabhängigen und abhängigen Variablen, um zu prüfen, ob der Zusammenhang zwischen den Variablen mit einem bestimmten Kausalmodell vereinbar ist.
 - Diskriminanzanalyse
 - Faktorenanalyse
 - Multidimensionale Skalierung
 - Clusteranalyse

Merke: Die deskriptive Statistik macht lediglich Stichproben-interne Aussagen. Dabei kann jede Variable für sich analysiert werden (univariate Analyse: Eigenschaften der Verteilung von Merkmalsausprägungen) oder der Zusammenhang zweier (bivariate Analyse: Korrelationskoeffizienten, Regressionsanalyse) oder mehrerer Variablen (multivariate Analyse).

Bei der **induktiven Statistik** (schließende, analytische, beurteilende Statistik) setzt man die Stichprobe mit der Grundgesamtheit in Beziehung, d. h. man extrapoliert die Erkenntnisse aus der Stichprobe auf die ihr zugrundeliegende Population.

- **Schätzverfahren:** anhand der Stichprobe lassen sich Werte der Grundgesamtheit schätzen.
 - **Punktschätzung:** ein Parameter der Grundgesamtheit (z. B. Erwartungswert) wird durch einen Stichprobenwert (z. B. arithmetischer Mittelwert) geschätzt.
 - Hierbei gelten folgende Postulate:
 - Erwartungstreue: der Durchschnitt aller möglichen Schätzwerte stimmt mit dem gesuchten Parameter der Grundgesamtheit überein.
 - Konsistenz: die Varianz der Schätzfunktion (Schätzer; Funktion zur Berechnung des unbekanntem Parameters) geht bei großen Stichproben gegen 0.
 - Effizienz: auch bei kleinen Stichproben ist die Varianz des Schätzers klein.
 - Exhaustivität: zur Schätzung werden alle Informationen der Stichprobe berücksichtigt.
 - Mittels bestimmter Schätzfunktionen lassen sich aus der Stichprobe gewisse Parameter der Grundgesamtheit näherungsweise errechnen: Erwartungswert (entspricht dem arithmetischen Mittelwert der Stichprobe), Median, Varianz, Kovarianz, relative Häufigkeit etc
 - Die Schätzung des Erwartungswerts erfolgt dabei umso genauer, je größer die Stichprobe ist (Gesetz der großen Zahlen)
 - Auch die erwähnten statistischen Kenngrößen sind zufällig verteilt, und zwar in Form von Prüfverteilungen mit charakteristischen Dichtefunktionen (z. B. t-Verteilung, χ^2 -Verteilung, F-Verteilung)
 - **Intervallschätzung:** Mit Hilfe der Stichprobenkenngrößen (und der Kenntnis ihrer Verteilung) lässt sich ein Konfidenzintervall (Vertrauensbereich) für die Grundgesamtheit errechnen, in welchem ein bestimmter Parameter mit einer hohen Wahrscheinlichkeit enthalten ist, und zwar mit der Wahrscheinlichkeit $1 - \alpha$ (d. h. mit der Irrtumswahrscheinlichkeit α liegt der gesuchte Parameter nicht in dem Konfidenzintervall); je schmaler dieses Intervall ist, desto genauer ist die Schätzung. Die Breite des Konfidenzintervalls hängt dabei ab von der

Irrtumswahrscheinlichkeit α , der Standardabweichung SD sowie dem Stichprobenumfang n.

- Konfidenzintervalle für einen Erwartungswert: mittels aufwändiger Formeln lassen sich die Grenzen dieses Konfidenzintervalls angeben, wobei endliche Grundgesamtheiten eine Endlichkeitskorrektur erforderlich machen.
- Konfidenzintervalle für eine Wahrscheinlichkeit (relative Häufigkeit)
- Konfidenzintervalle für Korrelationsmaße: für die Korrelationskoeffizienten nach PEARSON und SPEARMAN sowie für die Steigung der Regressionsgeraden.
- Hypothesenprüfung:
 - Signifikanzprüfung mittels statistischer Tests: "Sind die ermittelten Unterschiede überhaupt *signifikant*?"
 - Tests für Zusammenhangshypothesen: Prüfung, ob Korrelationskoeffizienten ungleich 0 sind.
 - Tests für Unterschiedshypothesen: Vergleich zweier (oder von mehr als zwei) Gruppen (z. B. Experimental- und Kontrollgruppe) hinsichtlich einer (stetigen) Variable, ob sich ihre Mittelwerte signifikant voneinander unterscheiden, oder Vergleich des Mittelwerts einer Gruppe mit einem vorgegebenen Sollwert. Je nach Stichprobenzahl und -art sowie abhängig von der Art des Merkmals und Verteilung seiner Ausprägungen verwendet man unterschiedliche Tests, welche die algorithmische Berechnung einer Prüfgröße erlauben, die dann mit einem bestimmten kritischen Wert verglichen wird, so dass sich die Signifikanz des Unterschieds feststellen lässt.
 - Effektstärke-Prüfung: "Wie *groß* sind die Unterschiede?"
 - Maß der Effektstärke: Cohens d ($d = 0,2 \rightarrow$ kleiner Effekt, $d = 0,5 \rightarrow$ mittlerer Effekt, $d = 0,8 \rightarrow$ starker Effekt)
 - Beurteilung von Signifikanz und Effektgröße eines Mittelwertsunterschieds durch Angabe eines 95%-Konfidenzintervalls

Stichprobenzahl und -art		Merkmalseigenschaften		
		Quantitativ (intervallskaliert)		Qualitativ oder dichotom
		Normalverteilung	Unbekannte Verteilung	
1		t-Test für 1 Stichprobe		Binomialtest
2	verbunden (Wertepaare)	t-Test für 2 verbundene Stichproben	Wilcoxon-Test	McNemar-Test
	unverbunden	t-Test für 2 unverbundene Stichproben	U-Test (Mann, Whitney, Wilcoxon)	χ^2 -Test, Exakter Test nach Fisher
> 2	verbunden	Varianzanalyse mit Messwiederholungen	Friedman-Rangvarianzanalyse	
	unverbunden	1-faktorielle Varianzanalyse	Kruskal-Wallis-Test	

Erläuterung der Tests:

- t-Test: Vergleich von Mittelwert mit Sollwert bzw. von zwei Erwartungswerten
- Wilcoxon-Test: Vergleich von Median der Grundgesamtheit mit Sollwert bzw. Vergleich von zwei Medianen
- U-Test: Alternative zum t-Test für zwei unverbundene Stichproben
- Binomialtest: Vergleich der relativen Häufigkeit einer Merkmalsausprägung mit einer vorgegebenen Wahrscheinlichkeit
- McNemar-Test: Vergleich zweier verbundener Stichproben hinsichtlich eines Alternativmerkmals
- χ^2 : Vergleich von beobachteten Häufigkeiten mit erwarteten Häufigkeiten
- Exakter Test nach Fisher: direkte Berechnung des p-Werts, für Kontingenztafeln beliebiger Größe anwendbar
- Logrank-Test: Vergleich der Überlebensfunktionen zweier unverbundener Stichproben; hierbei wird die Differenzfläche zwischen den beiden Überlebensfunktionen gemessen
- 1-faktorielle Varianzanalyse: Erweiterung des t-Tests
- Varianzanalyse mit Messwiederholungen: wie t-Test, allerdings können mehr als zwei Zeitpunkte verglichen werden
- Friedman-Rangvarianzanalyse: Verallgemeinerung des Wilcoxon-Tests
- Kruskal-Wallis-Test: Erweiterung des U-Tests

Merke: Bei der induktiven Statistik bezieht man die Ergebnisse aus der Stichprobe auf die Grundgesamtheit. Dabei werden Verfahren zur Schätzung (Punktschätzung, Intervallschätzung) sowie Methoden zur Signifikanzprüfung von

Unterschieden (statistische Tests) relevant.

Auswertung von nicht quantitativen Daten

Qualitative Daten können in quantitative Daten transformiert werden, d. h. man ordnet bestimmte Merkmalsausprägungen einer bestimmten Kategorie zu und bestimmt ihr Ausmaß.

- Inhaltsanalyse: z. B. Dokumentenanalyse
 - Kategorien finden und prüfen: induktives Auffinden von Kategorien (Einteilungsschemata), dann deduktives Überprüfen des Kategorien-Katalogs an anderen Interviewtexten
 - Kategorien ausarbeiten: Definition der Kategorie, Musterbeispiel, Kodierregeln (zur Anwendbarkeit bei Zweifelsfällen)
 - Analyse-Methoden:
 - Frequenzanalyse ("wie häufig kommen bestimmte formale/inhaltliche Elemente vor?")
 - Valenzanalyse ("wie bewertet der Interviewte die Inhalte?")
- Tiefeninterviews (offenes qualitatives Interview, das auch Unbewusstes zutage fördern soll):
 - Gruppendiskussion (z. B. Fokusgruppe; dient der Exploration eines Themas),
 - Soziometrie (dient dazu, die informelle Struktur einer Gruppe herausfinden): jedes Gruppenmitglied bewertet alle anderen hinsichtlich einer bestimmten Fragestellung, wodurch sich ein Soziogramm bilden lässt.

Weblinks: 🌐 Randomisierte kontrollierte Studie, 🌐 Fall-Kontroll-Studie, 🌐 Odds ratio, 🌐 Kohortenstudie, 🌐 Stichprobe, 🌐 Quotastichprobe, 🌐 Fragetypen, 🌐 Primärdaten, 🌐 Sekundärdaten, 🌐 Aggregatdaten, 🌐 Deskriptive Statistik, 🌐 Korrelationskoeffizient, 🌐 Arithmetischer Mittelwert, 🌐 Median, 🌐 Modus, 🌐 Standardabweichung, 🌐 Varianz, 🌐 Regressionsanalyse, 🌐 Bestimmtheitsmaß, 🌐 Induktive Statistik, 🌐 Punktschätzung, 🌐 Intervallschätzung, 🌐 Konfidenzintervall, 🌐 Statistischer Test, 🌐 Effektstärke

Selbsttest:

1. Was bedeuten die Worte "randomisiert" und "kontrolliert" bei der randomisierten kontrollierten Studie? Erläutern Sie das Grundprinzip dieses Studientyps!
2. Erklären Sie die Methodik der Kohortenstudie am Beispiel Raucher – Nichtraucher – Diagnose Lungenkrebs!
3. Was sind die Vorteile und was die Nachteile einer großen Stichprobe (großes n)?
4. Was sind Sekundärdaten, was Aggregatdaten?
5. Im Rahmen einer (fiktiven) bivariaten Analyse wurde der Zusammenhang zwischen dem Merkmal "Körpergröße" und dem Merkmal "Schuhgröße" analysiert. Man fand einen PEARSONSchen Korrelationskoeffizienten von $r = +0,7$. Was bedeutet dies?
6. Eine andere (fiktive) Analyse ergab $r = -0,5$ für die Merkmale "Intelligenz" und "Lesen von Comics". Was sagt dies aus?

Ein guter Test erfüllt bestimmte Qualitätsstandards und berücksichtigt ökonomische und ethische Aspekte

Die Güte eines Tests lässt sich anhand bestimmter Kriterien feststellen. So genügt ein guter Test bestimmten **Qualitätsstandards**:

- *Replizierbarkeit*, d. h. Ergebnisse sollen durch Wiederholbarkeit überprüft werden können
- *Übertragbarkeit (Generalisierbarkeit)* auf andere Situationen, Populationen, Interventionen; dies ist abhängig von Datenmenge, Datenqualität und Fragestellung; in der Praxis geht man pragmatisch vor: zunächst erfolgt der Nachweis der Efficacy (= Wirksamkeit unter Idealbedingungen, Treatment-Integrität), dann der Nachweis der Effectiveness (= Wirksamkeit unter Praxisbedingungen); je höher Effectiveness, desto höher ist die externe Validität.

Gute Tests berücksichtigen auch **ökonomische** und **ethische Implikationen**:

- Gesundheitsökonomie: Mittels Kosten-Wirksamkeits-Analysen wird die Effizienz bestimmt; im Gegensatz zur Effektivität, die nur die Wirksamkeit betrachtet, setzt die Effizienz Wirksamkeit ins Verhältnis zu den Kosten (direkte, indirekte [z. B. Produktionsausfall wegen Krankheit], intangible Kosten ["psychologische Kosten"]); man unterscheidet dabei technische Effizienz (maximale Wirksamkeit bei gegebenen Kosten) und Kosteneffizienz (minimale Kosten bei gegebener Wirksamkeit).
- Kosten-Nutzen-Analyse: Kosten und Nutzen werden monetär bewertet und dann miteinander verglichen.
- Kosten-Effektivitäts-Analyse: die Kosten werden monetär, der Nutzen nicht monetär bewertet (Werturteilsfrage: wofür soll man das Geld ausgeben?).
- Kosten-Nutzwert-Analyse: wie Kosten-Effektivitäts-Analyse, aber der Nutzen einer Maßnahmen wird durch Rekurs auf einen gemeinsamen Maßstab (z. B. QUALY) mit dem Nutzen einer anderen Maßnahme vergleichbar gemacht.
- Ethische Aspekte:
 - Die utilitaristische Ethik beurteilt Handlungen anhand ihres Nutzens; man unterscheidet u. a. Nutzensummen-Utilitarismus und Einzelnutzen-Utilitarismus
 - Die deontologische Ethik (griechisch *deón*: Pflicht) beurteilt Handlungen unter Rückgang auf allgemeingültige Werte (z. B. Pflicht zum Guten); klassisches Beispiel hierfür ist KANTs kategorischer Imperativ, den er aus der Tatsache herleitet, dass Menschen endliche Vernunftwesen sind: Handle stets so, dass die Maxime, die deiner Handlung zugrunde liegt, ein allgemeines Gesetz sein könnte.
 - Die Verantwortungsethik beurteilt Handlungen anhand ihrer Ergebnisse, während die Gesinnungsethik Handlungen anhand der zugrundeliegenden Motive bewertet.
 - Zukunftsethik: auf die Zukunft ausgerichtet (vgl. "nach mir die Sintflut")



Immanuel Kant. * 1724, † 1804, "Kritik der reinen Vernunft" (1781), "Kritik der praktischen Vernunft" (1788), "Zum ewigen Frieden" (1795).

Weblinks: Effektivität, Effizienz, Efectiveness, Utilitarismus, Deontologische Ethik, Immanuel Kant, Verantwortungsethik

Zusammenfassung

Die Psychologie möchte das Wissen über (hauptsächlich) menschliches Erleben und Verhalten mehren. Dabei geht sie wie jede andere echte Wissenschaft vor: aus *Theorien* werden *prinzipiell falsifizierbare Hypothesen* abgeleitet und an der Realität getestet. Da es sich bei den Untersuchungsgegenständen der Psychologie um *Konstrukte* handelt, also um nicht direkt messbare Entitäten, sind besondere Verfahrensweisen nötig (*Operationalisierung*). Zudem macht die Psychologie meist *probabilistische* Aussagen, die nur auf Gesamtheiten von Individuen anwendbar sind. Die Hypothesen müssen daher einem *statistischen Test* unterzogen werden und sollten eine bestimmte *Irrtumswahrscheinlichkeit* nicht überschreiten.

Den zu messenden *Variablen* muss gemäß eines Kalküls je ein Wert zugeordnet werden (*Skalierung*). Die Variablen ihrerseits besitzen gewisse Eigenschaften, die durch *Skalen* darstellbar sind. Diese Skalen unterscheiden sich hinsichtlich ihres *Niveaus*, wobei umso komplexere Rechenoperationen erlaubt sind, je höher das Skalenniveau ist.

Gute Tests zeichnen sich durch verschiedene Faktoren aus: *mittelschwere Items* mit ausreichender *Trennschärfe*, *Eichung* anhand einer Normstichprobe, ein hohes Maß an *Objektivität* (Unabhängigkeit von den Testbedingungen), *Reliabilität* (Zuverlässigkeit, Wiederholbarkeit), *Validität* (Beziehung zum zu messenden Merkmal), Kontrolle von vorhersehbaren *Messfehlern* wie etwa Antworttendenzen des Probanden sowie leichte Durchführbarkeit.

In der Praxis steht am Beginn einer Studie die Auswahl des *Studiendesigns*. Dabei unterscheiden sich die Studiendesigns hinsichtlich ihrer internen Validität –die *randomisierte kontrollierte Studie* gilt als das beste aller Studiendesigns, wenngleich ihre Durchführung recht aufwändig ist (*Kontrollgruppe, Randomisierung* etc.). Mittels

Befragung oder *Beobachtung* gewinnt man Daten über die Probanden. Diese Daten werden anschließend *ausgewertet*, wobei man jede einzelne Variable für sich (*univariate* Analyse) oder den Zusammenhang zwischen mehreren Variablen untersuchen kann (*bivariate* oder *multivariate* Analyse). Eine Studie kann dann als nützlich gelten, wenn sie von anderen Forschergruppen *repliziert* werden kann und in der Praxis anwendbar ist. Eine gute Studienplanung berücksichtigt zudem *ökonomische* und *ethische* Aspekte.

Wichtige Prüfungsthemen (alphabetisch): Fall-Kontroll-Studie, Fehler 1. und 2. Art, Fragetypen, Hypothese, Individual- und Aggregatdaten, Kohortenstudie, Konstrukt, Korrelationskoeffizient, Messfehler, Operationalisierung, Primär- und Sekundärdaten, Randomisierte kontrollierte Studie, Skalenniveaus, Skalierung, Standardabweichung, Statistischer Test, Stichprobenbildung, Testgütekriterien (Objektivität, Reliabilität, Validität)

Elementarwissen medizinische Psychologie und medizinische Soziologie: Literatur

Verwendete Literatur

- Benesch H, von Saalfeld H (2002), *dtv-Atlas Psychologie*, München
- Bierbrauer G (2005), *Sozialpsychologie*, Stuttgart
- Brenner C (1978), *Grundzüge der Psychoanalyse*, Frankfurt am Main
- Brunnhuber S, Frauenknecht S, Lieb K (2008), *Intensivkurs Psychiatrie und Psychotherapie*, München
- Faller H, Lang H (2006), *Medizinische Psychologie und Soziologie*, Heidelberg
- Fink G (2003), *Who's who in der antiken Mythologie*, München
- Golenhofen K (2006), *Basiswissen Physiologie*, München
- Kant I (2003), *Kritik der reinen Vernunft*, Stuttgart
- Kasten E, Sabel BA (2005), *GKI Medizinische Psychologie Medizinische Soziologie*, Stuttgart
- Legewie H, Ehlers W (2000), *Handbuch moderne Psychologie*, Augsburg
- Mutschler E et. al. (2008), *Arzneimittelwirkungen – Lehrbuch der Pharmakologie und Toxikologie*, Stuttgart
- Myers D (2005), *Psychologie*, Heidelberg
- Institut für Medizininformatik, Biometrie und Epidemiologie (2008), *Querschnittsbereich I – Teil Biometrie/Epidemiologie WS 2008/09*, Erlangen
- Paschke D (2004), *Grundlagen der Volkswirtschaftslehre anschaulich dargestellt*, Heidenau
- Schmidt R, Lang F, Thews G (Hg.) (2005), *Physiologie des Menschen mit Pathophysiologie*, Heidelberg
- Schön M (2007), *Medizinische Psychologie und Soziologie* In: Prieue J, Tümmers D (Hg.), *Das Erste – Kompendium Vorklinik*, Heidelberg
- Schulz von Thun F (2008), *Miteinander Reden 2*, Hamburg
- Teichert D (2006), *Einführung in die Philosophie des Geistes*, Darmstadt
- Weiß C (2008), *Basiswissen medizinische Statistik*, Heidelberg
- Wittgenstein L (2003), *Tractatus logico-philosophicus – Logisch-philosophische Abhandlung*, Frankfurt am Main
- Zimmer G (2006), *Prüfungsvorbereitung Rechtsmedizin*, Stuttgart
-

Quellen und Bearbeiter des Artikels

Elementarwissen medizinische Psychologie und medizinische Soziologie: Methodische Grundlagen *Quelle:* <http://de.wikibooks.org/w/index.php?oldid=488362> *Bearbeiter:* OnkelDagobert, Tschäfer

Elementarwissen medizinische Psychologie und medizinische Soziologie: Literatur *Quelle:* <http://de.wikibooks.org/w/index.php?oldid=488369> *Bearbeiter:* OnkelDagobert, 1 anonyme Bearbeitungen

Quellen, Lizenzen und Autoren des Bildes

File:Uebersicht methodische grundlagen.jpg *Quelle:* http://de.wikibooks.org/w/index.php?title=Datei:Uebersicht_methodische_grundlagen.jpg *Lizenz:* Creative Commons Attribution-Sharealike 3.0 *Bearbeiter:* User:OnkelDagobert

File:H0 h1 fehler.jpg *Quelle:* http://de.wikibooks.org/w/index.php?title=Datei:H0_h1_fehler.jpg *Lizenz:* Creative Commons Attribution-Sharealike 3.0 *Bearbeiter:* User:OnkelDagobert

File:Pruefgroesse kritischer wert.jpg *Quelle:* http://de.wikibooks.org/w/index.php?title=Datei:Pruefgroesse_kritischer_wert.jpg *Lizenz:* Creative Commons Attribution-Sharealike 3.0 *Bearbeiter:* User:OnkelDagobert

Bild:Wikipedia-logo.png *Quelle:* <http://de.wikibooks.org/w/index.php?title=Datei:Wikipedia-logo.png> *Lizenz:* logo *Bearbeiter:* Abigor, Bastique, Guillom, Mike.lifeguard, Richie, Rocket000, Schaengel89

File:Statistik boxplot balken.jpg *Quelle:* http://de.wikibooks.org/w/index.php?title=Datei:Statistik_boxplot_balken.jpg *Lizenz:* Creative Commons Attribution-Sharealike 3.0 *Bearbeiter:* User:OnkelDagobert

File:Verteilungsformen.jpg *Quelle:* <http://de.wikibooks.org/w/index.php?title=Datei:Verteilungsformen.jpg> *Lizenz:* Creative Commons Attribution-Sharealike 3.0 *Bearbeiter:* User:OnkelDagobert

Datei:DBP - 250 Jahre Immanuel Kant - 90 Pfennig - 1974.jpg *Quelle:* http://de.wikibooks.org/w/index.php?title=Datei:DBP_-_250_Jahre_Immanuel_Kant_-_90_Pfennig_-_1974.jpg *Lizenz:* unbekannt *Bearbeiter:* User:NobbiP

Lizenz

Creative Commons Attribution-Share Alike 3.0 Unported
<http://creativecommons.org/licenses/by-sa/3.0/>
