

Methoden

Vorlesung 1: Grundzüge empirischer Sozialforschung

- **Begriffe:**

1. **Methoden** = Systematische(s) Verfahren zur Gewinnung wissenschaftlicher Erkenntnisse. Sie gewährleistet, dass die Ergebnisse überprüfbar sind.
2. **Empirie** = Auf methodischem Weg gewonnene Erfahrung der Welt. Empirische Wissenschaften sind demnach Erfahrungswissenschaften, deren Theorien an der «Wirklichkeit» überprüft werden.

Die Unterscheidung zwischen Alltagswissen und Wissenschaft ist sehr wichtig.

- **Grundpositionen der Erfahrungswissenschaften**

1. **Erkenntnistheoretischer Realismus**: Eine intersubjektiv überprüfbare Erfahrbarkeit (und damit eine objektive Erkenntnis) der Welt ist prinzipiell möglich. Deshalb sind auch Aussagen über die Realität möglich.
2. **Erkenntnistheoretischer Konstruktivismus**: Auffassung, wonach die Wirklichkeit subjektiv konstruiert («erfunden») und nicht etwa objektiv «entdeckt» wird. Die Welt «draussen» ist immer nur über Beobachtung zugänglich, Beobachtungen tragen jedoch immer schon eine Interpretation in sich. Damit sind Induktion (Schluss vom Besonderen auf das Allgemeine) und Falsifikation (siehe nächste Lektion) strenggenommen nicht möglich.

Vorlesung 2: Grundbegriffe und wissenschaftstheoretische Kontroversen

- **Konzeptspezifikation**

Der Schritt, in welchem man die zentralen Konzepte seiner Aussage definiert, wird Konzeptspezifikation genannt. Die Konzeptspezifikation selbst dient dazu, die empirische Wirklichkeit mit sprachlichen Begriffen zu verknüpfen.

Verbindung zwischen Begriff und damit verknüpften Gegenstand:

1. **Intensional (Begriffsinhalt)**: die Menge aller Kriterien, die einen bestimmten Gegenstand auszeichnen (Zugehörigkeitsmerkmale). **Beispiel Staat: Staatsform, Staatsinstitutionen etc...**
2. **Extensional (Begriffsumfang)**: meint hingegen die Menge aller Objekte, die mit dem Begriff bezeichnet werden sollen. **Beispiel Staat: USA, Schweiz etc...**

Der empirische Bezug kann...

1. **direkt sein** (= manifeste Grösse oder Variable): Grösse, Gewicht, aber auch bestimmte Handlungsformen, z.B. die Stimmabgabe.
2. **indirekt sein** (= latente Grösse oder Variable). Diese Gegenstände oder Objekte sind nur unter Zuhilfenahme von direkten Beobachtungstermini und dazugehörigen Schlussfolgerungen (Messhypothesen) nachweisbar. Beispiele: Haltungen oder Einstellungssyndrome wie «Autoritarismus» oder «Konservatismus».
3. **dispositional**: Der Gegenstand/das Phänomen tritt nur unter bestimmten Bedingungen auf. Beispiel: Mut tritt nur unter bestimmten Bedingungen auf. **Nicht dispositional** wäre z.B. die Augenfarbe oder das Geschlecht
4. **komplexe Begriffe**: Begriffe, die Interaktionen zusammenfassen, z.B. «Wettbewerb».

• Deskription und Analyse

Deskriptive Analyse: Bei einer deskriptiven Fragestellung geht es darum, den empirisch noch kaum erforschten Untersuchungsgegenstand zuerst gedanklich, sodann auch begrifflich vorzustrukturieren. Dies wird als dimensionale Analyse bezeichnet /d.h. die Benennung einer bestimmten Dimension/ Aspekts der Realität). Zerteilen eines Begriffs in Unterbegriffe.

Analytische Fragestellung: Bei einer theorientestenden Fragestellung geht es darum, die bereits existierenden, expliziten Begriffe sprachlich zu präzisieren und den Objektbereich einzuengen (= definieren).

• Definitionsarten

1. **Nominaldefinition**: «die Festlegung der Bedeutung eines Begriffs durch einen oder mehrere bereits bekannte andere Begriffe». Mit anderen Worten: eine tautologische Umformung. (*Ein Schimmel ist ein weisses Pferd*). Eine **Nominaldefinition kann niemals falsch sein**, sondern bloss zweckmässig oder nicht zweckmässig.

Beispiel: Die (intensionale) Definition politischer Parteien bei Kromrey 2009: 151: «Eine politische Partei soll sein eine Organisation mit ...

- 1 eingeschriebenen Mitgliedern und
- 2 demokratischer Binnenstruktur,
- 3 die an Wahlkämpfen teilnimmt und
- 4 sich um Regierungsbeteiligung bewirbt.»

Werden gewisse dieser Definitionsmerkmale gestrichen, wächst die Extensionalität des Begriffes.

2. **Realdefinition**: unterscheiden sich von Nominaldefinitionen dadurch, dass sie Begriffe als Widerspiegelung des Wesens eines Gegenstands erachten. Realdefinitionen können falsch oder richtig sein. Realdefinition sind **untauglich für die empirischen Sozialforschung**, «da sie keine Kriterien angeben können, inwieweit das Wesen einer Sache durch die Definition erfasst wird»

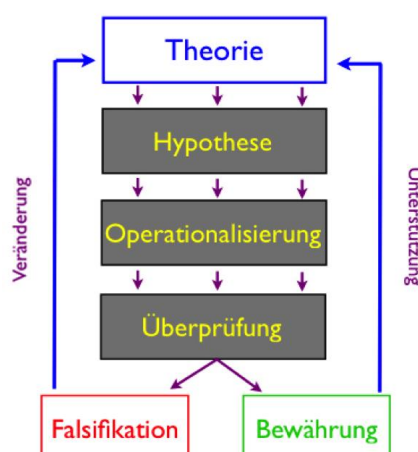
- **Kritischer Rationalismus**

1. Es sollen Gesetzmässigkeiten gefunden werden. Die formulierten Kausalzusammenhänge sind sodann an der Empirie zu überprüfen. Angestrebt werden All-Aussagen/ nomologische Sätze.
2. Es dürfen nur Konzepte oder Begriffe verwendet werden, die auch einen Bezug zur Realität bzw. Empirie aufweisen. (Keine Begriffe wie Engel oder Gott)
3. Die Aussagen müssen so formuliert sein, dass sie prinzipiell widerlegbar sind. Nominaldefinitionen (oder: analytisch wahre Sätze) sind demnach keine empirischen Aussagen i.S. des Kritischen Rationalismus.
4. Es sind keine Existenzaussagen zugelassen, da sie prinzipiell unwiderlegbar sind. Beispiel: «Es gibt schwarze Schwäne».

- **Deterministische und probabilistische Erklärungen**

1. **Deterministische Sätze** sind Aussagen, die immer und überall gültig sind. Also: Wenn x, dann **IMMER** y. Diese Aussagen haben einen unbeschränkten raumzeitlichen Geltungsanspruch und sie gelten für alle Elemente.
2. **Probabilistische Aussagen** geben im Gegensatz zu deterministischen Aussagen eine Wahrscheinlichkeit an und haben nur einen beschränkten Gültigkeitsbereich.

- In den Sozialwissenschaften werden Theorien mittlerer Reichweite angestrebt, da es im Unterschied zu den Naturwissenschaften kaum möglich ist, Allaussagen aufzustellen. Diese Theorien haben einen **eingeschränkten Objektbereich** (z.B. eine Theorie, die nur das Wahlverhalten von Jugendlichen erklärt) und/oder eine raumzeitliche Einschränkung (z.B. eine Theorie, die das Wahlverhalten in Industriestaaten des 20. Jahrhunderts erklärt). Durch diese **Randbedingungen** wird die Theorie eingegrenzt wodurch die Wahrscheinlichkeit grösser wird dass die Theorie stimmt. **Ziel ist es, so nahe wie möglich an Allaussagen zu kommen, indem man Schritt für Schritt die Randbedingungen aufhebt und überprüft ob die Theorie nach wie vor stimmt.** (Forschungsstrategie Karl Popper). **Potentielle Falsifikatoren** sind alle Elemente des angegebenen Objektbereichs. Deshalb gilt: Je grösser die Anzahl der potentiellen Falsifikatoren, desto höher der Informationsgehalt der Theorie.



Falsifikation führt zur Verwerfung der Theorie (u.U. zu ihrer Revision).

Empirische Unterstützung einer Theorie führt zu ihrer vorläufigen Bestätigung.

Ziel: Weitere Überprüfung durch «schärfere» Tests: «bewährte Aussagen».

Ziel: Ausdehnung der Reichweite der Theorie.

Bild: Forschungsstrategie Karl Popper: «trial and error»

- **Basissatzproblem**

Ausgangslage: Aussagen müssen prinzipiell falsifizierbar sein.

Problem: Aussagen können nicht direkt mit Beobachtungen verglichen werden, sondern immer nur mit Aussagen über eine Beobachtung.

Beobachtungsaussagen ihrerseits können aber falsch sein. Die Gründe dafür sind vielfältig. Ein ganz elementares Problem ist, dass keine Erfahrung theorieunabhängig ist.

Zudem: die Messung war allenfalls fehlerhaft (z.B. «falsche» Wahrnehmung eines schwarzen Schwans – vielleicht war es eine Gans). Messtheorie ist ebenso unsicher wie Kerntheorie.

Konventionalistische Lösung: Einhaltung höchstmöglicher methodischer Standards und transparente Darstellung der Vorgehensweise.

- **Werturteilsfreiheit**

Bei der Auswahl der Fragestellung sind Wertungen unvermeidbar. Denn es geht dabei, aus der unendlichen Zahl von möglichen Fragestellungen **eine bestimmte Fragestellung auszusuchen**. Dieser **Entdeckungszusammenhang wird von subjektiven Wertevorstellungen motiviert**.

Auch der **Verwertungszusammenhang** ist zwangsläufig wertend. Mit Verwertungszusammenhang ist gemeint, dass die Ergebnisse einer empirischen Studie in aller Regel verwertet werden, d.h., in Handlungsanweisungen zur Erreichung eines bestimmten Ziels umgesetzt werden. Aus den wissenschaftlichen Erkenntnissen alleine folgt jedoch noch kein Hinweis, welche Zielvorstellungen zu verfolgen sind.

Werte selbst können Gegenstand wissenschaftlicher Untersuchungen sein.

Im **Begründungszusammenhang** ist auf **Werturteile** zu verzichten. Der Begründungszusammenhang umfasst alle Forschungsschritte, die zwischen Fragestellung und Verwertung der Erkenntnisse liegen. D.h.: **Die Beschreibung, Messung, die Auswahl und auch das Erklären von Tatsachen soll wertfrei erfolgen**.

Vorlesung 3: Die theoretische Formulierung des Forschungsproblems

- **Was sind Variablen?**

Variable: «ist ein Merkmal oder eine Eigenschaft von Personen, Gruppen, Organisationen oder anderer Merkmalsträger.» Das **Merkmal** hat verschiedene Funktionen (Variable: Augenfarbe, Variablenausprägung: Blau) Eine Variable muss mindestens zwei Merkmalsausprägungen (= Konstante) haben.

Dichotome Variable: Eine Variable mit exakt zwei Merkmalsausprägungen. Sie stellt einen **Spezialfall** dar, ist in den Sozialwissenschaften als Abbildung des Eintretens bzw. Nichteintretens bestimmter Ereignisse aber häufig. (**Beispiel** dichotome Variable: Geschlecht: männlich/weiblich)

• Skalenniveaus von Variablen

Kontinuierlich: Merkmale, die jeden beliebigen Wert eines bestimmten Intervalls annehmen können (z.B. Körpergröße, Alter, etc..) (Alter könnte auch in Monaten, Tagen, Sekunden etc. gemessen werden.)

Diskret: Merkmale, die eine endliche Zahl von Kategorien innerhalb einer Bandbreite aufweisen (Zahl der Kinder, aber auch Bildungsniveau, etc.).

Variablen müssen **disjunkt** (= die einzelnen Kategorien (bzw. Merkmalsausprägungen) dürfen sich nicht überschneiden) und **erschöpfend** sein (=die Variable muss alle möglichen Kategorien abdecken). Variablen werden oft in Klassen/ Kategorien eingeteilt. Diese Kategorien dürfen sich **nicht** überschneiden. **Beispiel:**

Klasse 1: 18-29 jährige

Klasse 2: 30-39 jährige

Diese zwei Klassen wären jetzt zwar disjunkt, aber nicht erschöpfend, da alle unter 18 und über 39-jährigen nicht abgedeckt sind.

Weitere Unterscheidungen:

Individualmerkmale: Variable enthält Informationen über einzelne Individuen (z.B. das Alter der Person xy).

Kollektivmerkmale: Variable enthält Informationen über Kollektive («Aggregate»), beispielsweise: durchschnittliches Alter in der Gemeinde xy.

• Was sind Hypothesen?

Eine **Hypothese** ist eine Aussage über Merkmalszusammenhänge. In Hypothesen wird demnach ein Zusammenhang zwischen mindestens zwei Variablen postuliert.

Dabei bildet die vermutete Ursachenvariable die **unabhängige Variable**, während die vermutete Wirkungsvariable als die **abhängige Variable** bezeichnet wird.

1. Wenn-Dann-Hypothese

Wenn-Dann-Hypothesen setzen voraus, dass **beide Variablen eine dichotome Ausprägung** haben. (**Beachte:** Variablen mit mehr als zwei Ausprägungen können unter Umständen in eine dichotome Variable transformiert werden; z.B: Wahl einer Partei (=viele mögliche Merkmalsausprägungen) in Wahl einer bestimmten Partei (=zwei Ausprägungen: Wahl oder Nicht-Wahl der betreffenden Partei)

- **Implikationsbeziehung:** Wenn A auftritt, dann wird B erwartet. (**Beispiel:** Wenn man ein Mann ist wählt man SVP.) Wenn A hingegen nicht eintritt, dann kann sowohl B, aber auch «Nicht-B» erfolgen.
- **Äquivalenzbeziehung:** Nur wenn A (männliches Geschlecht) auftritt dann B (Wahl der SVP). Wenn A nicht eintritt, dann nie B. In einer Äquivalenzbeziehung wählen im Beispiel Frauen die SVP nie. In diesem Fall ist das männliche Geschlecht eine hinreichende und notwendige Bedingung dafür, SVP zu wählen. Die Äquivalenzbeziehung tritt praktisch nie ein.

2. Je-Desto-Hypothese

Voraussetzung: Mehr als zwei Ausprägungen (=polytom) und Rangfolge der Merkmalsausprägungen möglich (=ordinales Skalenniveau). Beziehung zwischen den Variablen kann unterschiedliche Formen aufweisen (zum Beispiel):

- linear (kontinuierlich steigend)
- Exponentiell (konvex)
- Logarithmisch (konkav)
- Logistische Funktion (S-förmig)
- Quadratische Funktion (U-förmig)

(Beispiele für die Formen der Kurven befinden sich in den Folien der Lektion 3, Seite 14-16.)

• Hypothesenstruktur

Abhängige und unabhängige Variable können in unterschiedlicher Weise verknüpft sein:

1. **Deterministisch** (wenn x, dann immer y) vs. **probabilistisch** (wenn x, dann sehr wahrscheinlich y).
2. **Reversibel** (wenn x, dann y und wenn y, dann x) vs. **irreversibel** (wenn x, dann y, aber wenn y, dann nicht x).
3. **Koexistent** (wenn x, dann auch y) vs. **aufeinanderfolgend** (wenn x, dann später y).
4. **Hinreichend** (wenn x, dann y, ungeachtet alles anderen) vs. **bedingt** (wenn x, dann y, aber nur, wenn auch z).
5. **Notwendig** (wenn x, dann y, und nur dann y) vs. **substituierbar** (wenn x, dann y, aber wenn z, dann auch y).

• Hypothesen: Bezugsebenen

- **Individualdaten**: Merkmale von Individuen, z.B. Arbeitslosigkeit einer Person.
- **Aggregatdaten**: Merkmale von Gruppen/Mengen von Untersuchungseinheiten, können aus Merkmalen der einzelnen Untersuchungseinheiten abgeleitet sein, z.B. Arbeitslosenquote in einer Gebietseinheit.
- **Aggregatdatenanalyse**: (ökologische Analyse); Ziel: statist. Zusammenhang zw. Regionaler Struktur des polit. Verhaltens und der regionalen Bevölkerungsstruktur aufzufinden und damit polit. Verhalten zu erklären.
- **Ökologischer Fehlschluss**: «Schluss von beobachteten Beziehungen zwischen Daten über Kollektive auf Beziehungen zwischen Merkmalen der Elemente der Kollektive»; Bsp: Arbeitslose haben Hitler gewählt. (Elemente der Kollektive = Individuen)
- **Vorteile der Individualdatenanalyse**: erlauben die Analyse von Zusammenhängen von Variablen auf gleicher Untersuchungsebene (Mikroebene), Datenerhebung lässt sich «designen».

- **Definition Theorie**

1. Eine Theorie ist eine «Aussagenmenge, die aus einer Reihe von Gesetzen, deskriptiven Aussagen und logischen Ableitungen besteht».
2. Eine Theorie ist ein «System von Aussagen, das mehrere Hypothesen oder Gesetze umfasst»
3. Eine Theorie ist eine Menge von Gesetzen, wenn diese durch logische Ableitbarkeitsbeziehungen miteinander verbunden sind

- **Struktur Theorie**

1. Ein System logisch widerspruchsfreier Aussagen. Elemente einer Theorie:
 - **Definitionen:** «Eigennutz wird definiert als das Ziel, das zuoberst auf der individuellen Präferenzordnung steht.»
 - **Axiome:** „Die Individuen versuchen ausschliesslich ihren eigenen Nutzen (in Form von z.B. Einkommen, Prestige und Macht) zu maximieren.» «Politik bzw. Wahlen sind ein Markt, auf dem Unternehmer (Parteien) Käufern (Bürger) Waren (politische Programme) anbieten.»
 - **Hypothesen:** «In Zweiparteisystemen nähern sich die beiden Parteien sukzessive dem Median-Wähler an.»

- **Ziel von Theorien**

- Das **Ziel** von Modellen und Theorien besteht nicht darin, die Realität spiegelbildlich abzubilden, sondern **Gesetzmässigkeiten zu formulieren**. Deshalb sind Theorien zwingend **eine Reduktion der sozialen Komplexität** – eine Reduktion auf das Wesentliche.
- **Parsimonität** (=Sparsamkeitsregel). Eine Theorie, die mit wenigen Aussagen auskommt, ist – bei gleicher Erklärungsleistung – einer Theorie mit einer höheren Zahl an Einschränkungen vorzuziehen

Vorlesung 4: Methodenübung

Vorlesung 5: Forschungsdesigns und Untersuchungsformen

- **Zweck Untersuchungsdesign**

- Das Forschungsdesign legt fest, wann, wo, wie oft und auf welche Art die Messung erfolgen soll.
- FD = Messvorschrift, wie und wie oft die Indikatoren an Objekten erfasst werden sollen. Diese Messvorschrift hat das Ziel, einen möglichst genauen Test der Hypothese zu ermöglichen. Dabei ist vor allem darauf zu achten, dass alternative Erklärungen ausgeschlossen werden können.
- Ein gutes Untersuchungsdesign erlaubt den Ausschluss anderer, alternativer Erklärungsansätze.
- Abhängig vom generellen Ziel der Fragestellung.

• Ziele sozialwissenschaftlicher Forschung

1. Explorative Untersuchungen

- Bei geringem Vorwissen wird zunächst eine **Auslegeordnung** des Untersuchungsgegenstandes gemacht (siehe auch dimensionale Analyse in Lektion 2).
- Dabei kommen vorzugsweise **qualitative Methoden** zur Anwendung (z.B. Tiefeninterviews).
- Solche Untersuchungen dienen zur **Generierung von Hypothesen**, die anschliessend oft in einer hypothesenüberprüfenden Studie empirisch geprüft werden

2. Deskriptive Untersuchungen

- In der Regel **univariate** (=nur eine Variable umfassende) **Untersuchungen**.
- Ziel ist die Schätzung der Verteilung einer Variablen (Geburtenrate, Wähleranteile, etc.).
- Ziel generell: **Diagnose** («wie viele Menschen leben unter der Armutsgrenze?») und nicht Ursachenforschung («Wovon ist Armut abhängig?»).

3. Hypothesenprüfende Untersuchungen

- **Vorrangiges Ziel der Sozialwissenschaften**.
- Herleitung von Hypothesen aus bestehenden Theorien (Deduktion) und ihre anschliessende empirische Überprüfung.

4. Evaluationsstudien

- Stark anwendungsbezogene Wirkungsforschung. Ziel ist die Bewertung von politischen oder sozialen Massnahmen (z.B.: «Erhöht Online-Voting die Wahlbeteiligung?»).
- In der Regel liegt solchen Untersuchungen ein quasi-experimentelles Design (siehe Untersuchungsdesigns) zugrunde.

• Kontrolle der Untersuchungsdesigns

- Kontrolle der Untersuchungssituation heisst im Wesentlichen Ausschluss von Störfaktoren. Störfaktoren sind all diejenigen Faktoren, die neben dem eigentlichen (vom Forscher postulierten) Wirkungszusammenhang Einfluss auf die unabhängige Variable haben könn(t)en («Drittvariablen»).
- Dabei können zwei Gruppen von Störfaktoren unterschieden werden: solche, welche die interne Validität betreffen und solche, welche die externe Validität betreffen.
 1. **Interne Validität:** vollständige **Kontrolle der Untersuchungssituation**. Intern valide ist demnach ein Ergebnis, wenn aufgrund der Untersuchungsanlage die Variation der AV einzig und allein auf die Manipulation der UV zurückgeführt werden kann.
 2. **Externe Validität:** Möglichkeit der **Generalisierung der Resultate** über die spezielle experimentelle Situation hinaus auf andere (lebensweltlichere) Kontexte. Anders formuliert: Sind die Ergebnisse eines (künstlichen) Experiments auch auf reale Situationen übertragbar?

- **Störfaktoren (interne Validität I)**

- Beispiel Medikamententests:

1. **Zwischenzeitliches Geschehen:** Ereignisse zwischen erster und zweiter Messung, die neben dem eigentlichen Stimulus die AV beeinflussen könn(t)en. (Was geschieht in der Woche, in der das Medikament genommen wurde.)
2. **Reifungsprozesse:** Intrapersonale Prozesse, die nicht durch den Stimulus in Gang gesetzt werden, sondern «natürlich» erfolgen (z.B. der natürliche Gesundungsprozess, der unabhängig von der Einnahme von Medikamenten ist).
3. **Messeffekte:** «Lerneffekte» bei den Probanden durch wiederholte Messungen (z.B. Lerneffekte bei kognitiven Tests wie dem IQ-Test).
4. **Hilfsmittelleffekte:** Effekte des Messinstrumentariums, Verzerrungen durch die unterschiedliche Verwendung von Messinstrumenten (siehe hierzu Sitzung zur Operationalisierung)
5. **Auswahl- und Ausfallverzerrungen:** Verzerrung der Ergebnisse dadurch, dass sich Experimental- und Kontrollgruppe nicht nur hinsichtlich des Stimulus, sondern auch im Hinblick auf andere Merkmale unterscheiden. Eine asymmetrische Gruppenzusammensetzung kann dabei die Folge von Auswahlprozessen (z.B. Selbstselektion), aber von Ausfällen (z.B. Verweigerung) sein.

- **Störfaktoren (externe Validität)**

1. **Messreaktivität:** Effektkumulierung durch wiederholte Messung. Beispielsweise Wirkungen, die sich einzig aufgrund der Testsituation (z.B. zweimalige Messung) ergeben.
2. **Situationsreaktivität:** Alltagssituationen können im Labor nicht vollumfänglich simuliert werden. Laborexperimente weichen stets von Alltagssituationen ab. Hinzu kommt, dass **Versuchspersonen** den Zweck der Untersuchung zu ergründen versuchen und sich **den vermuteten Erwartungen des Versuchsleiters entsprechend verhalten könn(t)en**.
 - *Laborexperimente haben den Vorteil, dass sie praktisch alle Störfaktoren ausschliessen können, was jedoch den Nachteil hat, dass in der Realität praktisch nie solch eine künstliche Situation vorhanden ist.*

- **Störfaktoren Varianzkontrolle**

- **Kontrolle der Störfaktoren = (v.a.) Varianzkontrolle.**
 - These: Mit 200 kmh fahren ist viel sicherer als mit 100 kmh, da es viel mehr Unfälle mit 100 kmh gibt.
 - Problem: Keine Berücksichtigung der Varianz, dass viel mehr Leute mit 100 kmh fahren, daher auch viel mehr Unfälle.
- Varianzkontrolle bedeutet zudem, dass Varianz auf der unabhängigen wie auch auf der abhängigen Seite vorhanden sein muss (gute Beispiele dazu: Diekmann 2014:331 ff.).

- Generelles zur Varianzkontrolle:
 - **Bildung von Vergleichsgruppen** (in der Regel: Experimental- und Kontrollgruppe) und Zuweisung der Untersuchungseinheiten zu den beiden (oder allenfalls: mehreren) Vergleichsgruppen. Dadurch wird gewährleistet, dass auf unabhängiger Seite Varianz vorhanden ist.
 - **Vorher-Nachher-Messung**: Wirkungsforschung ist nur möglich, wenn das Wirkungsmerkmal vor und nach dem Setzen des Treatments gemessen wird.

● Störfaktoren: Kontrolltechniken

1. **Elimination**: Ausschaltung aller potentiellen Störfaktoren bei der Durchführung des Experiments. Im Prinzip nur bei Laborexperimenten denkbar.
2. **Konstanthaltung**: Weil Elimination oft nicht möglich ist, wird versucht, alle möglichen Störgrößen konstant zu halten, indem die Experimentsituation für alle möglichst gleich gestaltet wird. (Experiment immer zur gleichen Tageszeit durchführen, damit beispielsweise der Strassenlärm von aussen immer gleich gross ist.)
3. **Matching**: Die Zuweisung von «gleichen» Personen zu je einer Gruppe (d.h. zur Experimental- bzw. Kontrollgruppe). Dadurch soll gewährleistet werden, dass sich die jeweiligen Gruppen nicht hinsichtlich weiterer möglicher Störgrößen unterscheiden.
4. **Randomisierung**: Zufallszuweisung der Untersuchungseinheiten zur Experimental- und Kontrollgruppe. Damit sollen systematische Unterschiede in der Gruppenzusammensetzung verhindert werden

● Untersuchungsdesign

- Die Kontrolle von Störfaktoren erfolgt durch Varianzkontrolle.
- Varianzkontrolle wiederum wird dadurch gewährleistet, dass die Untersuchungseinheiten einer der beiden Gruppen – Versuchsgruppe oder Kontrollgruppe – zugewiesen werden.
- Diese Zuweisung kann entweder vor der Behandlung bzw. Datenerhebung erfolgen oder danach.
- Designs, bei denen die Zuweisung vor der Datenerhebung erfolgt, nennt man **Ex-ante-Anordnungen**. Solche, bei denen die Zuweisung nach der Datenerhebung erfolgt, nennt man **Ex-post-facto-Anordnungen**.

- **Ex-ante-Anordnungen: experimentelle Designs**

- Nur Experimente erlauben eine systematische Kontrolle von Störfaktoren.
- Die Zuweisung der Versuchspersonen erfolgt durch den Forscher bzw. die Forscherin (und nicht durch Selbstselektion oder –rekrutierung).
- Die **Versuchsbedingungen** werden durch eine der genannten **Kontrolltechniken** (Elimination, Konstanthaltung, etc.) **kontrolliert**.
- Der Stimulus wird vom Forschenden gesetzt und liegt bei den Befragten nicht bereits vor.
- Eine **Vorher-Nachher-Messung** wird durchgeführt, d.h. **Ursachenvariable (Stimulus)** und **Wirkungsvariable (Wirkung)** werden nicht gleichzeitig, sondern aufeinanderfolgend gemessen.
- **Vorteile:**
 - Generell: hohe interne Validität.
 - Überprüfung kausaler Zusammenhänge möglich.
 - Randomisierung neutralisiert Drittvariablen.
- **Nachteile:**
 - Eher geringe externe Validität.
 - Reaktivität.
 - Hoher Aufwand.
 - Randomisierung aufgrund der Natur sozialwissenschaftlicher Variablen einerseits und ethisch-moralischen Bedenken andererseits schwer möglich.

- **Quasi-experimentelle Designs**

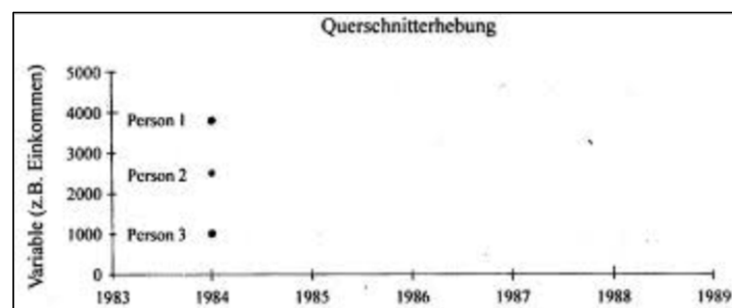
- Quasi-Experimente sind Experimente . . .
- . . . **ohne Randomisierung**
- . . . und **ohne die Stimulus-Kontrolle** durch den Wissenschaftler.
- Quasi-experimentelle Situationen ergeben sich oft durch politisch-administrative Veränderungen, verfolgen also keinen wissenschaftlichen Erkenntnisgewinn per se, sondern **ergeben sich ohne das bewusste Zutun des Forschers**.
- Beispiele: Einführung von neuen Unterrichtsformen, die Einführung von brieflicher Stimmabgabe oder von Online-Voting, aber auch die Erhöhung von Bussen oder soziale Reformen irgendwelcher Art. Gerade in der Schweiz mit dem starken Föderalismus kommen solche Situationen oft vor.

- **Ex-post-facto-Anordnungen**

- Der Stimulus wird nicht vom Forscher gesetzt, sondern liegt ohne Zutun des Forschenden vor (oder auch nicht). Durch die **Befragung einer möglichst hohen Zahl** wird Varianz angestrebt.
- Zuweisung der UE zu den Gruppen wird nicht vom Forschenden vorgenommen. Dadurch ist die **Kontrolle von Drittvariablen viel schwieriger** zu gewährleisten als bei Experimenten.
- Alle Daten werden gleichzeitig erhoben (keine Vorher-Nachher-Messung), weshalb die kausale Reihenfolge unklar bleibt.
- Der **grosse Vorteil** solcher Untersuchungsdesigns ist: Sie sind mit geringem Aufwand realisierbar.

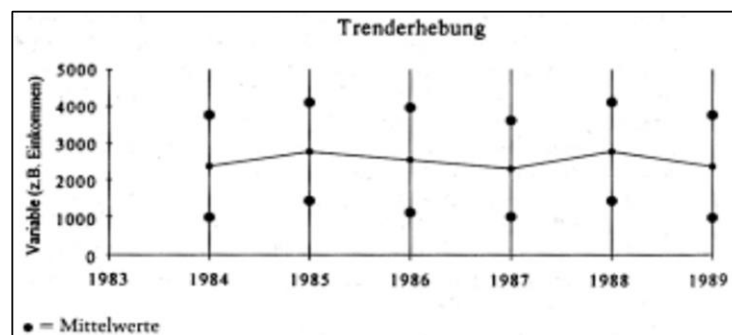
- 1. **Querschnittsdesigns**

- Ein Querschnittsdesign ist eine einmalige Erhebung bestimmter Variablenwerte bei einer bestimmten Zahl von Untersuchungseinheiten mit einer Stichprobe.
- Ein **Beispiel** für Querschnittsdaten sind die Daten einer einmaligen Bevölkerungsbefragung («Querschnitt der Bevölkerung»).



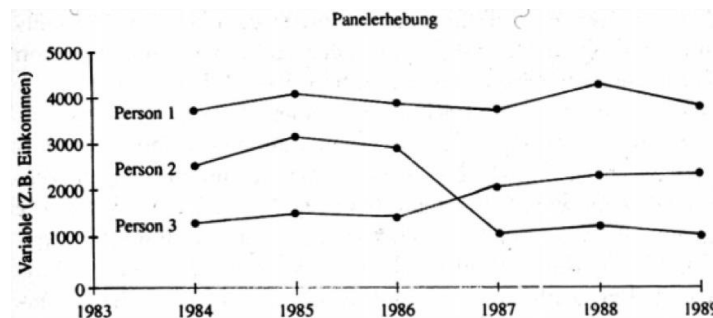
- 2. **Trenddesigns**

- Das Trenddesign ist eine *Erhebung der Werte der gleichen Variablen zu mehreren Zeitpunkten mit jeweils unterschiedlichen Stichproben*.
- Trenddaten erlauben **nur Aussagen auf der Kollektivebene**, nicht auf der Individualebene (weil nicht dieselben Individuen wiederholt befragt werden). (Beispiel Selects, 1995-2015)



3. Paneldesign

- Mit einem Paneldesign werden **Werte der gleichen Variablen** zu mehreren Zeitpunkten, aber **auf der Grundlage einer identischen Stichprobe** (dieselben Individuen werden wiederholt befragt) erhoben.
- Das erlaubt Aussagen über Kollektive, aber auch über Individuen.



4. Zeitreihenstudien

- Eine Besonderheit sind **Zeitreihenstudien**. Unter Zeitreihenstudien versteht man die **wiederholte Erhebung von Werten der gleichen Variablen beim gleichen Merkmalsträger**, wobei die **Anzahl Merkmalsträger Eins** beträgt.
- Ein Beispiel dafür ist etwa die Arbeitslosenquote in der Schweiz. Einziger Merkmalsträger ist die Schweiz. Die Arbeitslosenquote wird zu verschiedenen Zeitpunkten erhoben.

• Daten und Designs

- Designtyp und Datentyp stimmen überein, wenn der **Zeitpunkt der Erhebung identisch** ist mit dem **Messzeitpunkt**.
- Wenn dies nicht der Fall ist, dann müssen auch Designtyp und Datentyp nicht zwingend übereinstimmen.
- Beispielsweise ist es möglich, in einer Querschnittbefragung Retrospektivfragen zu stellen und damit Längsschnittdaten zu generieren. Allerdings dürfte die Qualität von Retrospektivfragen tiefer sein als diejenige «echter» Trend- bzw. Paneldaten.

• Datenhierarchie

- Paneldaten haben den höchsten Informationsgehalt. Sie erlauben zudem sowohl Querschnitt- wie auch Längsschnittanalysen. Trenddaten erlauben Querschnittanalysen, aber keine Paneldaten. Mit Querschnittsdaten lassen sich nur Querschnittanalysen erbringen.
- **Nachteile von Paneldaten:**
 - Hoher Erhebungsaufwand
 - «Panelmortalität»
 - Repräsentativitätsprobleme
 - Edukationseffekte

Vorlesung 6: Mess- und Testtheorie

• Operationalisierung: Von der Theorie zur Empirie

- Die **Operationalisierung** eines theoretischen Konstrukts besteht aus der Angabe einer Anweisung, wie Objekten, die der theoretische Begriff bezeichnet, **beobachtbare Sachverhalte zugeordnet** werden können.
- Das wesentliche **Problem** besteht darin, dass theoretische Begriffe in aller Regel nicht direkt beobachtbar sind. **Konzepte müssen deshalb mit beobachtbaren Sachverhalten verknüpft werden**. Diese Verknüpfung erfolgt durch Korrespondenzregeln.
- **Korrespondenzregeln** bestehen in der Angabe von Messanweisungen, wobei sich diese auf direkt beobachtbare Sachverhalte beziehen. Die Sachverhalte wiederum werden als Ausprägungen bestimmter Merkmale betrachtet. Man nennt diese (manifesten) Variablen auch „**Indikatoren**“.
(**Beispiel politische Macht**: Indikator: Anteil Sitze im Parlament = politische Macht)

• Indikatoren

- Wie kann die Zuordnung eines Indikators zu einem theoretischen Begriff gerechtfertigt werden? **Drei Ansätze**:
- 1. **Die operationalistische Lösung**: Operationale Definition des theoretischen Konstrukts („Intelligenz ist, was der Intelligenztest misst“).
- 2. **Die typologisch-induktive Lösung**: Erhebung von Messdaten und Bildung von latenten Gruppen („latent classes“), welche anschliessend als Indikatoren dienen. (Daten aller möglichen Aspekte sammeln)
- 3. **Der kausal-analytische Ansatz**: Die Indikatoren sind Manifestationen der latenten Variablen. Aus dem theoretischen Konzept x folgt der Indikator y . Dazu wird eine Messhypothese („Hilfstheorie“) formuliert. (geläufigster Ansatz)

• Auswahl von Indikatoren

- **Konzept des Indikatorenuniversums**:
 - Annahme eines homogenen Indikatorenuniversums: Es existiert eine Vielzahl von Indikatoren, die alle jeweils eine unabhängige Messung desselben Gegenstandes darstellen. (**Gefahr**, dass mehr über den Indikator für sich ausgesagt wird und weniger über die eigentliche Forschungsfrage.)
- **Prinzip der Austauschbarkeit der Indikatoren**:
 - Es spielt keine Rolle, welchen Indikator man aus dementsprechenden Universum für den jeweiligen Begriff auswählt. Wenn dieses Prinzip nicht erfüllt ist, sind die Ergebnisse der Untersuchung von der Wahl spezieller Indikatoren abhängig.

- **Konzept der multiplen Indikatoren:**

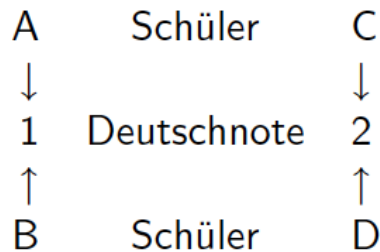
- Mehrere Indikatoren zur Messung desselben theoretischen Konstrukts.
- **Vorteil:** Gleicht Messfehler eher aus und ermöglicht Vergleiche.

- **Messtheorie**

- Messung ist „die **Zuordnung von Zahlen zu Objekten nachbestimmten Regeln**“ (Stevens 1951) oder „Messung ist eine „strukturtreue Abbildung“ (Schnell, Hill und Esser). Messung ist Datenerhebung, denn bei der Datenerhebung messen wir die Merkmalsausprägungen von interessierenden Variablen.
- Messungen = Grundlage für die mögliche Falsifizierung der Theorie
- **Qualitativen Merkmalen** muss man eine Zahl zuweisen um damit rechnen zu können (**Beispiel** männlich/ weiblich oder 1: überhaupt nicht interessiert, 5: sehr interessiert)
Quantitativen Merkmalen (Beispiel: Alter) muss man keine Zahl zu weisen, weil es bereits eine Zahl ist
- Der Messinput bilden die Objekte und deren Verhältnis zueinander (Relationen). Dies wird auch als „empirisches Relativ“ bezeichnet.
- Das **Messoutput sind Zahlen und deren Relationen** (=„Numerisches Relativ“).
- Bei der Messung gehen wir stets von einer Messhypothese aus, die ihrerseits (genauso wie die eigentliche „Kerntheorie“) falsch oder richtig sein kann.
- Ziel der Messung ist eine **strukturgetreue Abbildung** (= Morphismus) der Objekte durch die zugewiesenen Zahlen.
- Zwei Formen von Morphismen:
 1. **Isomorphismus:** ein- oder umkehrbar- eindeutige Abbildung (Umkehrbar, eindeutig: weil man umkehrend von der Mathenote auf den Schüler schliessen kann)
 - In beide Richtungen messbar

| Schüler | Zuordnung | Mathenote |
|---------|-----------|-----------|
| A | ← — → | 1 |
| B | ← — → | 2 |
| C | ← — → | 3 |
| D | ← — → | 4 |

2. **Homomorphismus:** eindeutige Abbildung
 (Wenn Zwei Schüler die gleiche Note haben ist das nicht mehr möglich. Man kann zwar von Schüler A aus schliessen, dass er die Note 1 hat, aber nicht von der Note 1 auf Schüler A, da auch Schüler B die Note 1 hat.)
 ➤ Nur in eine Richtung messbar



• Drei Kardinalprobleme

1. Repräsentationsproblem:

- Grundsätzlich: Ist ein bestimmtes Objekt überhaupt messbar?
 Dabei muss – damit Messung überhaupt möglich ist – zumindest ein Homomorphismus vorliegen. **Damit ist eine Repräsentation des empirischen Relatives durch ein numerisches Relativ gewährleistet.**
- Das Repräsentationsproblem fragt zudem nach den *Bedingungen, unter denen sich das empirische durch ein numerisches Relativ repräsentieren lässt.*
- Um beispielsweise eine **Ordinalskala** zu erhalten, muss etwa das **Transitivitätsaxiom** erfüllt sein (weitere Relationen wie Äquivalenz und Ordnungsrelation, symmetrische und reflexive Relation, etc.).
- **Beispiel:** Beweis, dass Abbildung einer empirischen Relation «zufriedener als» zwischen zwei Personen als mathematische Relation «größer als» zwischen Zahlen überhaupt möglich ist und unter welchen Bedingungen.

2. Eindeutigkeitsproblem:

- Wie können Messwerte verändert werden, ohne dass die in ihnen enthaltene Information verloren geht?
- Welche Transformationen sind zulässig? Damit eng verknüpft ist die **Bestimmung des Skalenniveaus einer Variablen.**
- Ergibt sich eine zweite, strukturverträgliche Skala durch eine bestimmte Transformation – zum Beispiel durch Multiplikation – aus der ersten Skala, so ist diese Transformation zulässig.
- Multipliziert man beispielweise die numerischen Relative einer Ordinalskala, so bleibt die Rangordnung bestehen. **Multiplikation ist also als Transformation für Rangfolgeskalen zulässig**

3. Bedeutsamkeitsproblem

- **Problem:** Welche mathematischen Operationen mit Messwerten führen auch zu empirisch sinnvollen Aussagen?
- Das **Bedeutsamkeitsproblem** ist weniger ein theoretisches als vielmehr ein **praktisches Problem**: es geht nicht um die Generierung von Messwerten, sondern um ihre sinnvolle Weiterverarbeitung.
- Z.B.: welche Rechenoperationen (Addition, Division, Logarithmierung, Bildung von Mittelwerten) mit Messwerten eines bestimmten Skalenniveaus können durchgeführt werden, so dass **daraus empirisch sinnvolle Schlüsse resultieren**.
- Konkretes Beispiel: Ist die Angabe des arithmetischen Mittelwertes bei ordinalen Werten (Rangordnungsskalen, z.B.: Schulnoten, Einstufungsskalen) sinnvoll?

- Angenommen, die Variable X steht für das Geschlecht. Weiter gilt:

$$X = \begin{cases} 0 & \text{if } \sigma \\ 1 & \text{if } \varphi \end{cases}$$

- Mathematisch gesprochen gilt:

$$0 + 1 = 1$$

- Empirisch gesprochen gilt jedoch:

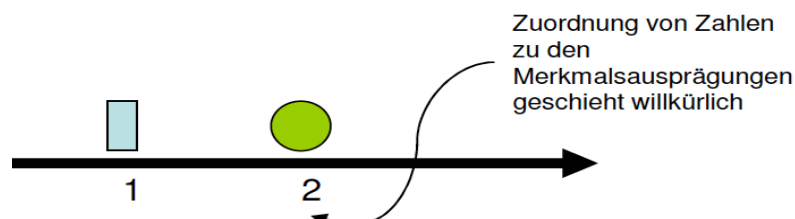
$$\sigma + \varphi \neq \varphi$$

- Die Addition beider Werte macht in diesem Beispiel keinen Sinn, weil die Werte 0 und 1 bloss dazu dienen, die beiden Geschlechter voneinander zu unterscheiden. Sie enthalten darüber hinaus jedoch keine weiteren Informationen zum *Verhältnis zwischen den Geschlechtern*.

• Skalenniveaus

1. Nominalskala

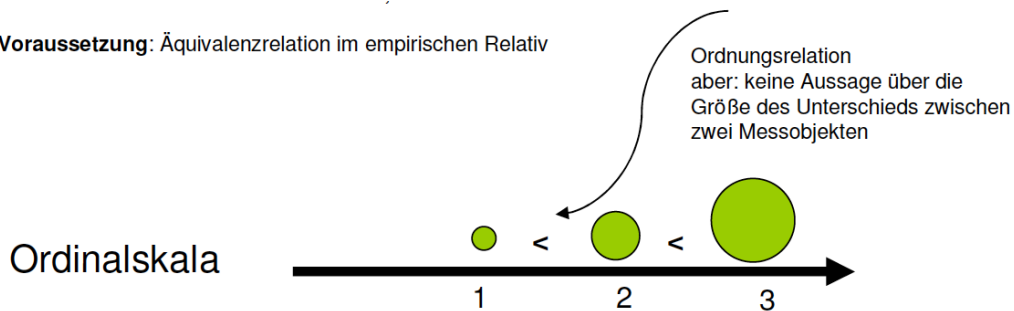
- Die Merkmalsausprägungen folgen **keiner natürlichen Reihenfolge**, sind demnach beliebig (existieren gleichberechtigt nebeneinander; „Äquivalenzrelation“).
- **Zuweisung** der Werte zu den Objekten dient lediglich der **Identifikation**; damit sind nur Aussagen zu Gleichheit/Ungleichheit möglich.
- Zulässige Transformationen: so gut wie alle, so lange sie die Unterschiede beibehalten.
- **Statistische Auswertungen**: Häufigkeitsinformationen, Mittelwert = Modus (häufigster Wert).
- **Beispiel**: Geschlecht, Beruf, Religionszugehörigkeit, aber auch Nummerierung von Fussballspielern, Matrikelnummer



2. Ordinalskala

- Die Merkmalsausprägungen folgen einer **natürlichen Rangfolge**, aber *die Abstände sind nicht quantifizierbar* (Voraussetzung: Transitivitätsaxiom).
- Zulässige Transformationen: alle rangerhaltenden Transformationen, d.h. monoton steigende Transformationen (z.B. die Multiplikation).
- Statistische Auswertungen: Häufigkeitsinformationen, Mittelwert = ... + Median (der Wert, der die untere von der oberen Hälfte der Messwerte trennt).
- **Beispiel**: Noten, Windstärke, Einstufungsskalen („sehr einverstanden“, „eher einverstanden“, „gar nicht einverstanden“, etc.)

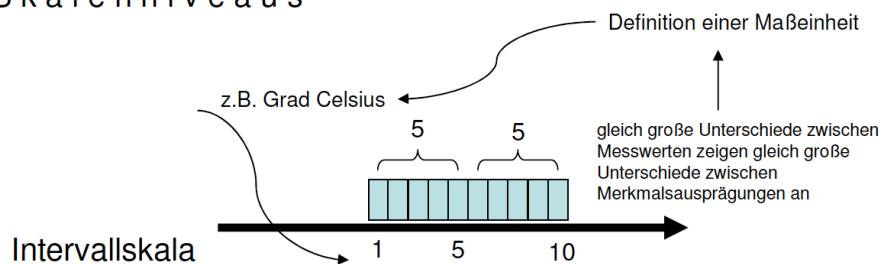
Voraussetzung: Äquivalenzrelation im empirischen Relativ



3. Intervallskala

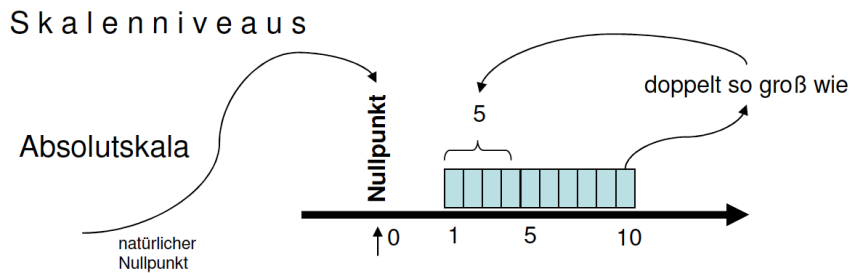
- Die **Merkmalsausprägungen** folgen einer natürlichen Rangfolge, die Abstände sind quantifizierbar. Mit anderen Worten: **Die Differenzen** zwischen den Merkmalsausprägungen sind **gleich**. Damit sind Aussagen möglich wie: Heute ist es 10 Grad wärmer (nicht aber Verhältnisaussagen: heute ist es doppelt so warm wie gestern). Intervallskalen haben keinen natürlichen Nullpunkt.
- Zulässige Transformationen: alle linearen Operationen.
- Statistische Auswertungsverfahren: ... + arithmetischer Mittelwert.
- **Beispiel**: Temperatur in Celsius oder Intelligenzquotient.

Skalenniveaus



4. Ratioskala

- Die **Merkmalsausprägungen** folgen einer natürlichen Rangfolge, die Abstände sind quantifizierbar und die Skala hat einen natürlichen Nullpunkt. Verhältnisaussagen sind möglich.
- Zulässige Transformationen: alle Ähnlichkeitstransformationen ($y=a*x$; also z.B. Kilometer in Meter).
- Statistische Auswertungen: alle, Mittelwert = ... + geometrisches Mittel.
- **Beispiel**: Temperatur in Kelvin.



5. Absolutskala

- Die **eindeutigste Skala** ist die absolute Skala. Sie **erlaubt keine Transformationen** mehr (mit Ausnahme der Identitätstransformation, d.h. Multiplikation mit der Zahl Eins).
- Auch die Masseinheit ist vorgegeben.
- **Beispiele**: Alle Skalen, die auf Abzählen beruhen. Z.B. absolute oder relative Häufigkeiten (etwa Geburts- und Todesraten, Konsumquote usw.), Wahrscheinlichkeiten.

• Zusammenfassung

- Je höher das Messniveau, desto weniger Transformationen der Messwerte sind zulässig.
Aber: Je höher das Messniveau, umso mehr mathematische Verfahren können auf die Messwerte angewendet werden.
- **Generell:** mit steigendem Messniveau steigt auch der Informationsgehalt.
- Lösung des Repräsentations- und Eindeutigkeitsproblem in den Sozialwissenschaften selten.
- Für die meisten sozialwissenschaftlichen Messverfahren liegen keine entsprechenden Theoreme vor. Das liegt daran, dass die meisten Messungen auf vermuteten Zusammenhängen zwischen latenter Variable und interessierendem Merkmal beruht.
- Intervallskalen in den Sozialwissenschaften? **Nein**, hauptsächlich haben wir **Ordinalskalen**

• Klassische Testtheorie

- Grundgedanken: Fünf Axiome (daraus sind anschliessend die drei Testgütekriterien **Objektivität**, **Reliabilität** und **Validität** ableitbar):
 1. Jede Messung weist auch **Messfehler** auf. Die Messung setzt sich zusammen aus „wahrem“ Wert und einem Störterm.
Beispiel: Menschliches Verhalten ist grundsätzlich stochastisch.
 2. *Messfehler streuen um den wahren Wert, wobei der **Erwartungswert des Messfehlers** null beträgt.* Das heisst: Bei wiederholten Testanwendungen heben sich die Messfehler auf, der Mittelwert der Fehler ist gleich Null.
Zum Beispiel: willkürliches, zufällig erfolgendes falsches Ablesen der Zeit.
Gegenbeispiel: Systematische Fehler = Frageeffekte bei Befragungen (z.B. soziale Erwünschtheit, beeinträchtigt die Validität → **Beispiel**: Frage nach Umweltbewusstsein, jeder möchte umweltbewusst sein).
 3. *Die Höhe des Messfehlers ist unabhängig vom Ausprägungsgrad des getesteten Merkmals,* d.h. wahrer Wert und Fehlerwert sind unkorreliert.
Gegenbeispiel: Je höher das Einkommen, desto eher wird bei der Angabe des Einkommens untertrieben..
 4. *Die Höhe des Messfehlers zwischen zwei Messwertreihen ist nicht korreliert.*
Gegenbeispiel: Messfehler bei den Motiven für Entscheid und dem Inhalt der Vorlage korrelieren wegen allg. Erinnerungsschwäche.
 5. *Der Messfehler einer Testanwendung ist unabhängig vom wahren Wert einer zweiten Messung.*
Beispiel: Falschangaben bei der Wahlteilnahme sollten nicht mit der wahrgenommenen Erwünschtheit der Wahlteilnahme korrelieren.

• Gütekriterien der Messung

1. **Objektivität**

- Ergebnisse sollen **unabhängig von der Person sein, welche die Messung durchführt** („Sind die Messergebnisse unabhängig von der messenden Person?“).
- Vollständige Objektivität liegt dann vor, wenn – unabhängig von der messenden Person – mit **dem gleichen Testinstrument die gleichen Resultate** erzielt werden.
- Ein *Objektivitätsmass* ist der *Korrelationskoeffizient*, der bei objektiven Messungen 1 oder zumindest nahe 1 betragen muss (z.B. Korrektur der Prüfungsergebnisse).
- Zu unterscheiden sind **Durchführungsobjektivität** (DO) und **Auswertungsobjektivität** (AO):
 - DO: Verzerrungen, die durch die messende Person ausgelöst werden: z.B. **Befragereffekte**.
Deshalb: standardisierte Befragungen (geringer Spielraum).
 - AO: Verzerrungen, die bei der Auswertung entstehen (z.B. qualitative Bewertung eines Aufsatzes)

2. Reliabilität

- Reliabilität ist ein Mass für die Reproduzierbarkeit von Messergebnissen („Misst das Instrument verlässlich, was es messen soll?“)
- **Drei Messmethoden:**
- **Paralleltest:** Messung mit zwei vergleichbaren Messinstrumenten (z.B. Messung der Zeit mit analoger und digitaler Uhr). **Problem:** wirklich parallele Tests zu finden.
 - Parallele, tau-äquivalente und kongenerische Tests:
 - Eine latente Variable wird **durch je drei Items** gemessen.
 - **Paralleler Test:** Die latente Variable wirkt auf alle drei Items gleich stark und die Störgrößen ebenfalls.
 - **Tau-äquivalenter Test:** Die latente Variable wirkt gleich stark auf die Items, aber die Störgrößen wirken unterschiedlich.
 - **Kongenerischer Test:** Latente Variable und Störgrößen wirken unterschiedlich.
- **Test-Retest-Methode:** wiederholte Messung mit demselben Instrument. Setzt jedoch Stabilität des Messobjekts voraus (z.B. die Stabilität von Einstellungen) und ein Paneldesign voraus (das Messobjekt muss dasselbe sein). **Meist:** Unterschätzung, aber auch Überschätzung möglich.
- **Testhalbierung:** Setzt sich ein Messinstrument aus mehreren Indikatoren (z.B. Items bei einer Befragung) zusammen („Index“), können die Items in zwei Hälften aufgeteilt werden, wobei die Korrelation der Messwerte der zwei Messungen über ihre Reliabilität informiert.

3. Validität

- Validität gibt an, ob das Instrument das theoretische Konstrukt misst, das es messen soll („Misst das Instrument, was es messen soll?“). **Validität ist unabhängig von der Reliabilität oder der Objektivität.**
- Nimmt man eine operationale Definition des theoretischen Konstruktes vor („Intelligenz ist genau das, was der Intelligenztest misst“), ist Validität zwar stets gegeben, aber nach wie vor unklar, wofür das Konstrukt steht.
- **Drei Formen der Validität:**
- Augenscheinvalidität („Face validity“): selbstevidente Messungen.
- **Inhaltsvalidität**
- **Kriteriumsvalidität**
- **Konstruktvalidität**

- **Inhaltsvalidität („content validity“)**

- „Stichprobenmodell“: Angenommen wird einerseits eine **unendliche oder zumindest hohe Anzahl an möglichen Messinstrumenten** (Items, Indikatoren, etc.), aus denen eine repräsentative Stichprobe gezogen wird.
- Angenommen wird ausserdem, dass das theoretische Konstrukt mehrere, unterschiedliche Dimensionen aufweist.
- *Inhaltsvalidität ist sodann gegeben, wenn eine genügend hohe Zahl an repräsentativen Aspekten bei der Operationalisierung eines Konzeptes berücksichtigt wurde.*
- **Beispiel:** Der **Intelligenztest** im Sinne von Problemlösungsfähigkeit misst nicht bloss die Rechenfertigkeit, sondern auch die Fähigkeit des räumlichen Denkens, Artikulationsfähigkeiten, Logik, etc..

- **Kriteriumsvalidität**

- Die Kriteriumsvalidität bezieht sich auf ein bestimmtes Aussenkriterium (in der Regel eine vom eigenen Messinstrument unabhängiges, aber bewährtes Messinstrument („benchmark test“)) und misst die Korrelation mit diesem Aussenkriterium.
 - **Übereinstimmungsvalidität** (gleicher Zeitpunkt, concurrent validity): siehe Beispiel oben. Aber auch: Umweltbewusstsein und Mitgliedschaft bei einer Umweltorganisation (Diekmann 2014: 259) oder Links-Rechts-Einstufung und Wahlverhalten.
 - **Vorhersagevalidität** (unterschiedlicher Zeitpunkt, prognostische oder prädiktive Validität): In diesem Fall werden die Kriteriumsdaten zeitlich später erhoben und man misst die „Prognosekraft“ des Messinstruments. Beispielsweise liessen sich die gemessenen Prüfungswerte von Absolventen des Methodenkurses mit ihrem späteren akademischen Erfolg (z.B. Abschlussnote) vergleichen. Hohe Übereinstimmung = hohe prognostische Validität. **(Diese Art kommt kaum jemals vor!)**
- Problem: Selten ist einmal eine solche Kriteriumsvariable bekannt und wenn ja, wozu denn eine neue Messung (Schnell, Hill & Esser 1999:150)?

- **Konstruktvalidität**

- Konstruktvalidität gibt an, wie genau ein Messinstrument ein spezifisches Konstrukt misst und nicht auch weitere, damit verwandte bzw. gegensätzliche Konstrukte.
 - **Convergent validity** ist ein Test, der belegen soll, dass Konstrukte, die mit dem eigenen Konstrukt nahe verwandt sind, auch empirisch stark mit diesem korrelieren (**Beispiel:** politisches Interesse, politisches Wissen, politische Involvierung).
 - **Discriminant validity** ist ein Test, der belegen soll, dass Konstrukte, die mit dem eigenen Konstrukt nichts oder nur wenig zu tun haben, damit auch nicht korrelieren (liberal/ konservativ vs. „links/rechts“).
- Überprüfung mit Multitrait-Multimethod-Matrizen.

- **Zusammenfassung**

- Reliabilität und Objektivität sind notwendige, aber nicht hinreichende Bedingungen für Validität.
- Ein Messinstrument ist umso reliabler, je weniger zufällige Fehler die Messung beeinflussen.
- Ein Messinstrument ist um so valider, je weniger systematische Fehler die Messung beeinflussen.

Vorlesung 7: Skalen und Indizes

- **Indizes und ihre Konstruktion**

- Index = Variable, die sich aus mehreren Einzelindikatoren zusammensetzt.
- Eine **Indexbildung** empfiehlt sich bei **multidimensionalen Konstrukten**, denen aber eine latente Variable zugrunde liegt. Die einzelnen Teilindikatoren decken dabei die verschiedenen Dimensionen des Konstrukts ab.
- Beispiel: Der Demokratiebarometer (siehe nachfolgende Folie).
- Bei der Indexkonstruktion sind zwei Fragen zu beantworten:
 - *Welche Dimensionen sollen in den Index einfließen?*
 - *Wie sollen die Dimensionen kombiniert werden?*

- **Indizes: Festlegung der Dimensionen**

- **Dimensionen bilden einen Merkmalsraum**, in dem jedes Objekt der Forschung eine spezifische Position belegt.
- Bei diskreten Variablen mit wenigen Ausprägungen: Typologietabelle. (**diskrete Variable**: abzählbar endliche Variable)
- Bei kontinuierlichen Variablen: Vektorraum.
- Herleitung der Dimensionen erfolgt theoretisch.

- **Indizes: Kombination der Dimensionen**

- Generell gilt: ein Index fasst viele Kombinationen des Merkmalsraumes zu einigen, wenigen (neuen) Kombinationen zusammen. Deshalb **Indexbildung = Reduktion des Merkmalsraum**.

- **Drei Gründe** sprechen dafür:
 1. **Gewisse Kombinationen kommen nie vor** oder sind sehr selten. (Beispiel: Hohes Bildungsniveau und tiefes Einkommen.)
 2. Verschiedene Kombinationen **können durch Gewichtung zusammengefasst** werden.
 3. Verschiedene Kombinationen können theoretisch zusammengefasst werden.
- Wie sollen die Dimensionen kombiniert werden?
 1. **ungewichtet-additiv:**
 - einfache Addition der Indikatorenwerte (Index = Ind1+ Ind2+ ... + Indn).
 - Alle **Einzelindikatoren müssen denselben Wertebereich haben**, ansonsten fließen sie ungleichgewichtig in die Analyse.
 - Voll substituierende Wirkung: D.h., ein **niedriger Punktwert auf einem Indikator** kann durch einen **hohen Punktwert auf einem anderen Indikator** vollumfänglich ausgeglichen werden.

Beispiel: Man hat die **drei Dimensionen Beruflicher Status (X1), Nettoeinkommen (X2) und Schulbildung (X3)** die alle mit demselben Wertebereich charakterisiert werden. Durch ungewichtete Addition kann der Merkmalsraum auf sechs verschiedene Schichten reduziert werden (Siehe unten).

| Variable X ₁ : Beruflicher Status. | | | \tilde{x}_1 |
|---|--|--|---------------|
| <i>Arbeiter:</i> | ungelernte Arbeiter | | 1 |
| | angelernte Arbeiter | | 4 |
| | gewöhnliche Facharbeiter | | 9 |
| | höchstqualifizierte Facharbeiter | | 13 |
| <i>Angestellte:</i> | ausführende Angestellte | | 10 |
| | qualifizierte Angestellte | | 16 |
| | leitende Angestellte | | 27 |
| <i>Beamte:</i> | untere Beamte | | 10 |
| | mittlere Beamte | | 16 |
| | leitende Beamte | | 23 |
| <i>Selbständige:</i> | kleine Selbständige | | 15 |
| | mittlere Selbständige | | 20 |
| | führende Selbständige | | 30 |
| | freie Berufe, intellektuelle Berufe | | 25 |
| | kleine Landwirte (5ha / 20 Morgen) | | 15 |
| | mittlere Landwirte (20 ha / 80 Morgen) | | 20 |
| | große Landwirte | | 23 |

Bei Arbeitslosen wird von der jeweiligen Punktzahl für den Beruf 1 Punkt abgezogen.

| Nettoeinkommen: | | \tilde{x}_2 |
|-----------------|--------------|---------------|
| | unter 150 DM | 1 |
| 150 — | 299 DM | 3 |
| 300 — | 399 DM | 5 |
| 400 — | 499 DM | 6 |
| 500 — | 599 DM | 8 |
| 600 — | 699 DM | 9 |
| 700 — | 799 DM | 10 |
| 800 — | 899 DM | 13 |
| 900 — | 999 DM | 13 |
| 1000 — | 1499 DM | 18 |
| 1500 — | 1999 DM | 19 |
| 2000 und mehr | | 20 |

Der Interviewer hatte durch Nachfragen das Nettoeinkommen des Haupterzhäuers festzustellen.

| Variable X ₃ : Schulbildung. | | | \tilde{x}_3 |
|---|---|--|---------------|
| <i>Schulbildung:</i> | Volksschule, unvollständig | | 0 |
| | Volksschule ohne Lehre | | 2 |
| | Volksschule mit Lehre | | 4 |
| | Handelschule oder Mittelschule ohne Abschluß | | 5 |
| | Höhere Schule bis Obertertia | | 7 |
| | Mittlere Reife | | 9 |
| | Höhere Schule länger als Unterschule, ohne Abitur | | 11 |
| | Höhere Fachschule mit Abschluß | | 12 |
| | Abitur | | 14 |
| | Hochschule ohne Abschluß | | 16 |
| | Hochschule mit Abschluß | | 20 |

| Bildung von Schichten | | | | | |
|-----------------------|---|----|---|----|------------------------|
| 0 | ≤ | X* | ≤ | 14 | untere Unterschicht |
| 15 | < | X* | < | 22 | obere Unterschicht |
| 23 | ≤ | X* | ≤ | 29 | untere Mittelschicht |
| 30 | < | X* | < | 39 | mittlere Mittelschicht |
| 40 | ≤ | X* | ≤ | 49 | obere Mittelschicht |
| 50 | < | X* | < | 63 | Oberschicht |

2. gewichtet-additiv:

- partiell substituierende Wirkung: $(\text{Index} = a_1 \cdot \text{Ind}_1 + a_2 \cdot \text{Ind}_2 + \dots + a_n \cdot \text{Ind}_n)$.
- Erlaubt die **unterschiedliche Bedeutung einzelner Merkmalsdimensionen** zu berücksichtigen (z.B.: Gesamtnote aus unterschiedlich wichtigen Teilprüfungsleistungen).
- Die Gewichte können entweder theoretisch oder statistisch (z.B. Faktorenanalyse) bestimmt werden.
Auch möglich: eine Kombination von beidem (inhaltliche Überlegungen empirisch begründen, vgl. Verbraucherpreisindex).

3. multiplikativ:

- Multiplikation der Indikatorenwerte $(\text{Index} = \text{Ind}_1 \cdot \text{Ind}_2 \cdot \dots \cdot \text{Ind}_n)$.
- Verstärkende Wirkung bei Vorliegen mehrerer positiver Indikatorenwerte.
- **Absicht:** Das **Fehlen eines einzelnen Indikators** soll sich überproportional stark auswirken (z. B.: Wahlerfolg = Sympathie * Programm, d.h., wenn Sympathie = 0, dann auch kein Wahlerfolg, ungeachtet des Programms).

• Skalen

- **Reihe von Items** (Aussagen, z.B. Fragen eines Fragebogens), die **entlang einer Dimension** misst.
- Skalen werden in der Regel dann gebildet, wenn ein einzelner Indikator die latente Variable nicht mit ausreichender Genauigkeit misst.
- **Items = Fragen, Statements oder Aussagen**, mit denen die Befragten einverstanden oder nicht einverstanden sein können.
- Skalen werden in der Regel für die Messung von Einstellungen verwendet.
- Es gibt keine formalen Regeln zur Formulierung und Entdeckung von Items. In der Praxis werden bereits existierende Items wieder (oder neu) verwendet.
- Einige (ausgewählte) Kriterien zur Formulierung von Items nach Schnell/Hill/Esser:
 - **Gegenwartsbezug:** Rückerinnerungsfragen messen oftmals die Erinnerungsfähigkeit und nicht das eigentliche Konstrukt.
 - **Eindeutige Interpretation, eindimensional:** Enthält ein Statement mehrere Dimensionen, ist unklar, welchem Aspekt der Befragte zustimmt bzw. nichtzustimmt. Z.B.: „Die Armeeausgaben sind zu hoch. Der Staat sollte stattdessen mehr für Soziales ausgeben.“
 - **Keine „valence issues“:** Z.B. „Arbeitslosigkeit soll bekämpft werden.“ Diesem Statement werden vermutlich die meisten Befragten zustimmen, es besitzt demnach keine Trennschärfe.

- Die Zustimmung zu einem Item hängt dabei von **zwei Faktoren** ab:
 1. Die **Ausprägung** der zugrundeliegenden latenten Variablen. **Beispiel**: Je „rechter“ die Position auf der latenten Variablen „ideologische Selbsteinstufung“, desto eher wird dem Item „Verschärfung des Ausländerrechts“ zugestimmt.
 2. Die „**Itemschwierigkeit**“: Je „schwieriger“ das Item, desto geringer die Zustimmung (bei gleichbleibender latenter Haltung). **Beispiel**: Eine Person, die sich gemässigt rechts einstuft, wird einer moderaten Verschärfung des Ausländerrechts (Item 1) zustimmen, nicht aber einer deutlichen Verschärfung (Item 2).

- **Skalierungsverfahren:**

- **Generell**: Skalierungsverfahren unterscheiden sich hinsichtlich der Annahmen über das Antwortverhalten sowie die Verrechnung der Antworten zu den entsprechenden Skalenwerten.

1. **Likert-Skala**

- Mit Abstand am häufigsten verwendetes Skalierungsverfahren.
- Unterscheidung zwischen **trifft voll zu** bis **trifft gar nicht zu**
- **Voraussetzung**: monotone Itemcharakteristik (mit Ansteigen der latenten Variablen steigt auch die Wahrscheinlichkeit der Zustimmung) **Beispiel: Skala 1-10**
- **Grundprinzip**: Summiertes Rating
- **Item-Analyse**: Nicht-Berücksichtigung ungeeigneter Items mittels bestimmter Verfahren (Trennschärfe-Index oder Trennschärfe-Koeffizient).

2. **Guttman-Skala**

- Selten verwendetes Skalierungsverfahren.
- **Voraussetzung**: Items, die in Bezug auf die interessierende Variable unterschiedliche radikal formuliert sind. Zudem: **dichotome Itemcharakteristik** (erst mit Überschreiten eines bestimmten Schwellenwertes der latenten Variablen erfolgt Zustimmung).
- **Grundprinzip**: Skalenwert entspricht der Nummer desjenigen Items, das zuletzt akzeptiert wurde.
- **Problem**: Perfekte Guttman-Skalen gibt es in der Realität kaum.
- **Item-Analyse**: Berechnung des Reproduzierbarkeits-Koeffizienten.

Vorlesung 9: Beobachtungen und Inhaltsanalysen

- **Worin unterscheiden sich Erhebungsverfahren ganz prinzipiell?**
- **Reaktivität:** Löst das Erhebungsverfahren Reaktionen bei den Untersuchungsobjekten aus?
 - **Persönliche Befragung:** Reaktionen der Respondenten möglich und wahrscheinlich. (Möglichkeit, dass der Interviewer die Befragungsperson beeinflusst.)
 - **Schriftliche Befragung:** Reaktionen der Respondenten kaum möglich. (Keine soziale Interaktion. Beeinflussung unwahrscheinlich)
- **Standardisierung:** Grad der Strukturierung einer Erhebung.
- **Beobachtung: Definition und Beispiele**
- Was ist Beobachtung? → Mutter aller Datenerhebungstechniken
- Jede empirische Methode ist „Beobachtung“.
- Im engeren Sinn: „**Die systematische und kontrollierte Beobachtung** menschlicher Handlungen, sprachlicher Äusserungen und nonverbaler Reaktionen oder anderer sozialer Merkmale“ (Diekmann 2014:548).
- **Vorteile:**
- Das Verhalten der Untersuchungsobjekte **wird unmittelbar beobachtet** (und nicht wie bei Befragungen über Erinnerungsfragen erfragt).
→ Kann direkt protokolliert werden
- **Nachteile:**
- **Möglichkeit der selektiven Wahrnehmung** (es werden nur Handlungen wahrgenommen, die die Hypothese bestätigen).
- Gefahr der Fehlinterpretation (vgl. Clifford Gertz „Dichte Beschreibungen“ und das Beispiel des Augenzwinkerns, das vielerlei Bedeutungen haben kann). Kultur als „selbstgesponnenes Bedeutungsgewebe“.

• Beobachtungstechniken: die Rolle des Beobachters

- Die Beobachtungstechnik ist von der Fragestellung abhängig. Unterschiede gibt es zunächst in Bezug auf die Rolle des Beobachters:

1. **Teilnehmend:**

- Beobachter **nimmt innerhalb des Beobachtungsfeldes selbst eine Rolle ein** – entweder **passiv** (der Ethnograf als „Besucher“ eines autochthonen Stammes) oder **aktiv** (Lehrer, der Verhalten der Schüler beobachtet).

➤ **Probleme:**

- **Identifikation mit den Untersuchungseinheiten** („going native“). Dadurch geht die Distanz des Aussenstehenden zum Beobachtungsfeld verloren.
- **Beeinflussung der Beobachtung durch Teilnahme** („Doppelfunktion“, „Intrarollenkonflikt“).
- **Kein Zugang zu „Schlüsselpersonen“**.

2. **Nicht-teilnehmend:**

- Nicht teilnehmende Beobachtungstechnik ist zu bevorzugen, ist allerdings nicht immer möglich.
- Protokollierungsmöglichkeit eher gegeben, weil keine Doppelfunktion.
- Keine (bzw. nur geringe) Einflussmöglichkeiten.

• Beobachtungstechniken: Die (wahrgenommene) Präsenz des Beobachters

- Was wissen die Untersuchungsobjekte von der Beobachtung?

1. **Offen:** Die „Probanden“ sind sich der Präsenz des Beobachters bewusst („Big Brother“). **Problem der externen Validität.**

2. **Verdeckt-teilnehmend:** Beobachter ist aktiv, aber als solcher unerkannt („**Undercover-Agent**“). → moralisch fragwürdig

3. **Verdeckt-nicht-teilnehmend:** Beobachter ist inaktiv und unbemerkt („**Schlüssellochmethode**“).

➤ Der **Vorteil** verdeckter Beobachtung ist, dass das Verhalten durch den Beobachtungsvorgang nicht beeinflusst wird (**nicht-reaktiv**).

➤ **Allerdings** ist diese Methode aus **ethisch-moralischer Perspektive höchst fragwürdig**, erst recht, wenn irreführende Stimuli gesetzt werden.

- **Beobachtungstechniken: Feld- vs. Laborforschung**

- Anwendbarkeit ist von der Forschungsfrage und dem Untersuchungsgegenstand abhängig (Arbeitslosigkeit kann z.B. nicht als „Treatment“ gesetzt werden).

1. **Vor- und Nachteile von Laborexperimenten:**

- Der Forscher setzt die Stimuli und **kontrolliert die Randbedingungen**.
- **Kontrolle von Störfaktoren** durch Randomisierung.
- **Hohe interne Validität**, aber... **niedrige externe Validität**.
- „Natürliche“ Langzeitfolgen sind nicht messbar.

2. **Vor- und Nachteile von Feldbeobachtung:**

- Langfristige Untersuchung von Effekten in „natürlicher“ Umgebung.
- Nicht durch Forscher manipulierbare Variablen beobachtbar.
- Komplexe, in Laborexperimenten nicht simulierbare soziale Sachverhalte können erfasst werden.
- Hohe externe Validität, aber **niedrige interne Validität**.
- **Unsicherheit ob das Ergebnis auf die Studie zurückzuführen ist.**

3. **Das Feldexperiment:** eine Kombination von Experiment und Feldbeobachtung

- Das Feldexperiment findet in „natürlicher“ Umgebung (d.h. nicht im Labor) statt.
- In der Regel wissen die „Probanden“ nicht, dass sie teil eines Experiments sind.
- **Triangulation** möglich: Kombination von Labor- und Feldexperiment, um Resultate zu validieren

- **Beobachtungstechniken: Strukturierungsgrad**

1. **Strukturierte Beobachtung:**

- Präzise und strikte Beobachtungsschemata (Leitfaden).
- Bedarf eines ausgearbeiteten Leitfadens.
- **Ziel: Höhere Durchführungsobjektivität und Reliabilität.** Gefahr der verzerrten Wahrnehmung soll eingedämmt werden.

2. Unstrukturierte Beobachtung:

- **Vorteil:** Ermöglicht eine gewisse **Flexibilität und Offenheit** des Beobachters für den Beobachtungsgegenstand.
- **Nachteil:** Es werden nur ein grober Rahmen und einige Leitlinien vorgegeben. Zudem ist die **Zahl der vorgegebenen Beobachtungskategorien geringer** als bei der strukturierten Beobachtung.
- Kombination möglich: **Leitfadenbeobachtung**

• Inhaltsanalyse: Definition und Beispiele

- «Systematische Erhebung und Auswertung von Texten, Bildern und Filmen.»
- Nicht zu verwechseln mit Hermeneutik (=Textinterpretation, „Verstehen des Sinns“), während **Inhaltsanalyse systematische Identifikation von Text-Elementen** und ihre Zuordnung zu Kategorien meint.
- Beschränkt sich nicht notwendigerweise nur auf Inhalte, sondern kann auch formale Texteigenschaften (Länge der Sätze, stilistische Elemente wie Verwendung von Konjunktiv, etc.) umfassen.
- Zum Beispiel **Aktionsquotient** = Verhältnis zwischen Verben und Adjektiven.
- **Type-Token-Ratio:** Reichhaltigkeit des Vokabulars.

• Vor- und Nachteile

1. Vorteile:

- **Keine Beschränkung auf Gegenwart.** Vergangenheitsbezug möglich: In der Vergangenheit produziertes Material kann untersucht werden und **unterliegt zumindest keinen Erinnerungsirrtümern** wie bei der Befragung.
- Soziale Veränderungen (zum Beispiel Wertewandel, Verschiebung von Präferenzen) sind erforschbar.
- **Nichtreaktivität:** Das gesamte Datenmaterial ist nichtreaktiv

2. Nachteile:

- **Hoher Interpretationsaufwand und hohes Kontextwissen nötig** (Bedeutung von Wörtern hat sich möglicherweise gewandelt, vgl. etwa „links“ und „rechts“ in der politischen Kommunikation).

- **Formen und Zwecke der Inhaltsanalyse**

| Zweck | Analyseeinheit | Frage (nach Lasswell) | Beispiele |
|--|---------------------|-----------------------|--|
| Diagnostisch: Inferenz bezüglich Sender und sozialen Umfelds | Sender | Who? | «Autorship analysis» |
| | Kodierung | Why? | Qualitative Inhaltsanalyse, Parteiprogramme im Wandel der Zeit |
| Formal-deskriptiv: Inferenz bezüglich formaler Texteigenschaften | Kommunikationskanal | How? | Analyse stilistischer oder persuasiver Elemente |
| | Botschaft | What? | Analyse von Kommunikationstrends |
| | Empfänger | To whom? | Analyse zielgruppengerechter Werbebotschaften |
| Prognostisch: Inferenz bezüglich Wirkung auf den Empfänger | Dekodierung | With what effect? | Wirkungsforschung (zum Beispiel in der pol. Kommunikationsforschung) |

- **Konzeptspezifikation und Operationalisierung**

- Zentrales Werkzeug der Inhaltsanalyse ist ein **Kategoriensystem**. Es muss vor der Analyse erstellt sein (siehe auch dimensionale Analyse).
- **Operationalisierung** = Suche nach Text-Indikatoren für die interessierenden Sachverhalte.
- Formulierung von Oberbegriffen (Kategorien), **weitere Untergliederung in Teildimensionen**. Kategorien sind das Äquivalent zu den Fragen in einer Befragung.
- **Codierschema als „Abfrageschema“ für Texte.**

- **Zuverlässigkeitskriterien:**

- **Stabilität** (Intracoder-Reliabilität): Grad der Übereinstimmung zwischen verschiedenen Kodierergängen derselben Kodierer.
- **Wiederholbarkeit** (Intercoder-Reliabilität): Grad der Übereinstimmung zwischen zwei verschiedenen Kodierern am selben Material.
- **Text hat hohe Stabilität, da er sich nicht verändert.** Lebende Objekte (Befragungsperson) kann bereits innerhalb von 3 Tagen die Haltung ändern. Die Messung kann zwar reliabel sein, aber das "Forschungsobjekt" ist unberechenbar.
- **Forscher-Kodierer-Reliabilität:** Übereinstimmung zwischen Forscher und Kodierer.
- **Holsti-Index:** Siehe Vorlesung 9, Seite 20!

- **Techniken: Frequenzanalyse**

- **Rein quantitative Analyse**
- Häufigkeit von ausgewählten Textelementen (Namen von PolitikerInnen, Parteinamen).
Annahme: Häufigkeit korreliert mit Bedeutung.
- Können auch rein formale Texteigenschaften sein (Type-Token-Ratio).
- **Heute:** Zumeist computerunterstützte Inhaltsanalysen: zum Beispiel Wordscore und Wordfish.

- **Techniken: Kontingenzanalyse**

- **Kontingenzanalyse:** Gemeinsames Auftreten bestimmter Textmerkmale.
- **Beispiel:** Kontaktanzeigen im Wandel der Zeit bzw. wie beschrieben sich die Inserenten selbst?

Selbstbeschreibung von Inserenten in Kontaktanzeigen (% der Inserate)

| | <i>jung</i> | <i>attraktiv</i> | <i>groß</i> | <i>schlank</i> | ... | <i>wohlhabend</i> |
|--------|-------------|------------------|-------------|----------------|-----|-------------------|
| Frauen | 34 | 12 | 17 | 64 | | 8 |
| Männer | 27 | 24 | 46 | 11 | | 33 |

Gemeinsames Auftreten von Paaren von Kategorien (% der Inserate)

| | <i>jung</i> | <i>attraktiv</i> | <i>groß</i> | <i>schlank</i> | ... |
|------------------|-------------|------------------|-------------|----------------|-----|
| <i>attraktiv</i> | 24 | | | | |
| <i>groß</i> | 12 | 10 | | | |
| <i>schlank</i> | 33 | 8 | 19 | | |
| ... | .. | ... | ... | ... | |

- **Techniken: Valenzanalyse**

- **Valenzanalyse:** Bewertungstendenz von Textinhalten (z.B. negative oder positive Bewertung, pro oder contra) → *Ist der Beitrag von Journalist X eher positiv oder negativ?*
- **Setzt zumindest ordinalskaliertes Niveau voraus.**
- In der Regel Schätzurteile, aber heutzutage auch mit entsprechenden computerunterstützten Analysen möglich.
- **Intensitätsanalysen:** Nicht nur die Richtung wird hier bewertet, sondern auch die Intensität (wie stark positiv bzw. wie stark negativ?)

- **Techniken: Verhaltensspuren (im Internet)**

- Nutzung von **prozessgenerierten Daten** (d.h. Daten, die nicht zu wissenschaftlichen Zwecken erhoben wurden).
- Solche Daten gab es schon immer (z.B. Stimmregisterdaten), aber **mit der Nutzung des Internets werden diese Datenmengen und ihre Vielfalt immer grösser („Big Data“)**.
- Dazu gehören: Google-Suchbegriffe, Leserkommentare in Newsforen, Likes in Facebook, Tweets auf Twitter, Wikipedia-Traffic auf bestimmten Seiten, etc.
- Oftmals fehlen jedoch noch Theorien darüber, wer diese Daten erzeugt hat, zu welchem Zweck und wer sie nutzt.
- **Beispiel:** Ausmass einer Grippe konnte ermessen werden, durch Google, indem man geschaut hat, wie oft ein Medikament als Suchbegriff eingegeben wurde.

Vorlesung 10: Befragungen und Interviews

- **Formen der Befragung: das persönliche Interview**

- Einst die am weitesten verbreitete Befragungsform.
- Das persönliche Interview („face-to-face“):
- **Vor- und Nachteile:**
- In der Regel **bessere Verständlichkeit und höhere Kontrolle**.
- **Niedrige Abbruchquote**; Einfluss auf die Teilnahmequote ist hingegen umstritten
- **Reaktivität**, da der Interviewer persönlich anwesend ist.
- Kann das Interview beeinflussen, weil man dann Antworten gibt die gesellschaftlich eher erwünscht sind.
- **Aufwand und Kosten** sind vergleichsweise hoch.

- **Formen der Befragung: das telefonische Interview**

- **Vor- und Nachteile:**
- siehe „face-to-face-Interview“.
- **Kosten und Aufwand steigen** zunehmend.
- **Abnehmender Abdeckungsgrad.**
 - Deckt höchstens 70 Prozent der Bevölkerung ab. Nicht mehr so repräsentativ.
- Zu den Problemen, die mit der Stichprobenziehung aus Telefonregistern verbunden sind, siehe nächste Lektion (Sampling).

- **Formen der Befragung: die schriftliche Befragung**

- Formen der schriftlichen Befragung (selbstadministriert):
 - Die Schriftlich-postalische Befragung.
 - Immer häufiger (von der Struktur her jedoch dasselbe): Online-Umfragen.
- **Vor- und Nachteile:**
- **Grössere Anonymität, geringere Interviewereffekte.**
- Kommt dem Zeitmanagement des Befragten eher entgegen, der Befragte kann Fragen besser durchdenken. **Aber keine Hilfe möglich** bei Verständnisfragen.
- Die Untersuchungsergebnisse zu Rücklaufquoten fallen unterschiedlich aus.
- Kognitive Tests („Wissensfragen“) nur schwerlich möglich (obwohl: Online-Software, die das bei Onlinebefragungen überwacht). **Identifikation der Zielperson nicht möglich.**

- **Formen der Befragung: die Online-Befragung**

- **Vor- und Nachteile:**
- Meist selbstrekrutiert (jedoch nicht zwingend), was ein **grosser Nachteil** ist, da Selbstrekrutierung zu grossen Verzerrungen führen kann. Ist allerdings ein Problem der Stichprobenziehung und nicht der Form der Befragung.
- **Schnell und kostengünstig.**
- Das Antwortverhalten und die Aufzeichnung der Geschwindigkeit der Antwortreaktionen („latency“) u.a. sind mit Software messbar.
- Spielerisches Element.
- Experimentelle Studien (Conjoint-Analysen) sind viel einfacher realisierbar.

- **Voraussetzungen für alle Formen der Befragung**

- **Kooperationsbereitschaft**

- Die Bereitschaft, an Interviews teilzunehmen ist von vielerlei Faktoren abhängig: V.a. Interesse am Thema, aber auch finanzielle Anreize (z.B. 10 Franken-Gutschein) spielen (selbstredend) eine Rolle.

- Die Kennzahl, welche die Kooperationsbereitschaft misst, lautet: „Ausschöpfungsquote“.

- Die Existenz einer „**wahren**“ **Haltung** und die Bereitschaft, diese mitzuteilen:

- Verschiedene Wissenschaftler bezweifeln die Existenz von „true attitudes“ (Converse 1964, Zaller 1992) und gehen eher von einem Kontinuum bzw. von situativ abhängigen „considerations“ aus. Die Haltung zu gewissen Themen kann sich ziemlich schnell ändern.

- **Höflichkeitsnormen oder Furcht vor juristischen oder sozialen Sanktionen können ebenfalls Hinderungsgründe sein.**

- **Gemeinsame Sprache**

- **Theorien des Antwortverhaltens**

1. **Probabilistische Ansätze** (Achen 1975):

- Menschliches Verhalten ist generell stochastisch (**zufällig**) (**Menschen entscheiden in ein und derselben Entscheidungssituation nicht identisch**).

- Deshalb sind Attitudes konzeptionell nicht als ein singulärer Punkt auf einer Dimension vorzustellen, sondern als eine begrenzte Bandbreite von Punkten. Das Antwortverhalten von Befragten wiederum variiert innerhalb dieser Bandbreite – **abhängig von der Frageformulierung**.

2. **Die Theorie der rationalen Entscheidung** (Esser 1975):

- Mensch **strebt nach sozialer Anerkennung** und «fürchtet» sich vor sozialen Sanktionen. → Mensch ist ein rationaler Nutzenmaximierer

3. Kognitionspsychologische Antworttheorien:

- **«Memory-based»** oder **«file-drawer-model»** (Feldman und Zaller 1992):
Wird nach einer Evaluation gefragt, bemüht sich das Individuum alle (relevanten), mit dem Einstellungsgegenstand verknüpften Erwägungen in Erinnerung zu rufen und auf der Basis dieser «recollections» eine Einschätzung abzugeben.
 - Man muss sich Gedanken dazumachen, wie die eigene Haltung zu einem Thema ist
- **«On-line-processing model»** (Lodge, Steenbergen und Brau 1995):
Die Valenz einer Informationen wird sofort in ein **evaluatives Zwischentotal** («summary tally») integriert. Im Anschluss wird bloss dieses «summary tally» im Langzeitgedächtnis abgespeichert, nicht aber die einzelnen, spezifischen Informationsbits.
 - Man weiss direkt was man mag oder nicht mag

• Erhebungsgegenstände

1. Einstellungen («attitudes», «opinions»)

- Einstellungen sind «likes and dislikes» (Eagly and Chaiken 1993: The Psychology of Attitudes)
- **Demnach:** «Attitudes» sind ein **wertendes Urteil über ein Einstellungsobjekt**.
- **Allerdings:** Der Befragte muss nicht zwingend eine substantielle Meinung zu allen Einstellungsgegenständen aufweisen («Weiss nicht»). → Beispiel: «*Mögen Sie Pizzas mit Sardellen?*»

2. Überzeugungen («beliefs»)

- «Beliefs» umfassen alle Informationen und Kognitionen, welche mit dem entsprechenden Einstellungsgegenstand verknüpft sind (Überzeugungen, «Weltwissen»).
- «Beliefs» sind geglaubtes Wissen («Die Erde ist eine Scheibe»).
- **Unterschied zu «Attitudes»:** «Beliefs» sind *deskriptive Statements*, Einstellungen («attitudes») hingegen sind *präskriptive Aussagen*.
 - deskriptiv: Der Haschischkonsum **ist** legalisiert worden
 - präskriptiv: Der Haschischkonsum **soll** legalisiert werden.

3. Verhalten («behavior»)

- **Achtung:** Bei Befragungen wird ausnahmslos berichtetes Verhalten erhoben (nicht jedoch bei der Beobachtung!).

4. Sozialstatistik: soziale Merkmale (objektive Eigenschaften der befragten Person).

- **Fehlerquellen, Formen der Antwortverzerrungen**

- Unterschiedliche Kategorien von Fehlerquellen bzw. Quellen von Antwortverzerrungen:
- Persönlichkeitsmerkmale des Befragten.
- Merkmale des Befragenden bzw. die Reaktionen der Interviewten darauf (Interviewereffekte).
- Fragebogen bzw. Frageformulierungen, d.h. Reaktionen auf formale Aspekte von Fragen (Frageeffekte).
- Die Interviewsituation.

- **Antwortverzerrungen: Merkmale des Befragten**

- **Soziale Erwünschtheit:**
 - Die Angabe einer Antwort, die nicht der „wahren“ Haltung entspricht, sondern dem, was als sozial erwünscht wahrgenommen wird. Soziale Erwünschtheit ist **der wichtigste Punkt bei Befragungen**. (Vor allem bei sensiblen Themen/ Fragestellungen)
 - Die Angabe sozial erwünschter Antworten ist einerseits von **Persönlichkeitsmerkmalen** des einzelnen Befragten **abhängig**, kann andererseits aber auch Resultat einer situationspezifischen Reaktion sein.
 - **Gegenmassnahmen:** Antworten (zusätzlich) anonymisieren, Enttabuisierung durch spezifische Frageformulierung, → Randomized Response Technique
- **Zustimmungstendenz** (unabhängig vom Inhalt der Fragen):
 - Erkennbar an spezifischen Antwortmustern (Response-Set).
 - **Gegenmassnahmen:** Umpolung der Frageformulierung oder der Antwortkategorien (vgl. Paralleltest).
- **Non-Attitude:**
 - Angabe einer inhaltlichen Antwort, obwohl keine substanzielle Meinung zum Einstellungsgegenstand vorhanden ist.
 - **Gegenmassnahmen:** Angabe einer „Weiss nicht“-Kategorie (kann unter Umständen wiederum «falsche» Meinungslosigkeit fördern → **Gefahr der Bequemlichkeit:** Leute haben zwar eine Meinung aber finden es einfacher mit "weiss nicht" zu antworten)

- **Antwortverzerrungen: Frageeffekte**

- **Frageformulierung:**

- Formale wie auch inhaltliche Aspekte der Frageformulierung haben möglicherweise **Einfluss auf das Antwortverhalten**. Z.B. Suggestivformulierungen, psychologische Effekte (**siehe Beispiel in Vorlesung 10, Seite 17**).
- **Rating oder Ranking.**

- **Positionseffekte:**

- Antworten zu Fragen sind möglicherweise von gewissen Primingeffekten abhängig, d.h. davon, ob zuvor bestimmte kognitive Assoziationen aktiviert wurden oder nicht.
- Gegenmassnahme: Randomisierung.

- **Vorgabe von Antwortkategorien:**

- Mitte-Kategorie («Mitte-Tendenz»).
- «Weiss nicht», «keine Meinung».

- **Antwortverzerrungen: Interviewer bzw. Interviewsituation**

- **Interviewermerkmale:**

- Stark abhängig vom Inhalt der Frage. Wenn z.B. illegales oder sozial geächtetes Verhalten abgefragt wird, ist mit erheblichen Interviewereffekten zu rechnen.

- **Interviewsituation:**

- Z.B. Anwesenheit Dritter.

- **Sponsorship-Effekt:**

- Befragte richten ihre Antworten danach aus, wer der Auftraggeber der Studie ist.

• Kriterien für Frageformulierung

- **Verständlich, einfach.**
- **Keine doppelten Verneinungen**, keine Suggestivfragen und einfache Frageformulierung (An einer Volksbefragung 1997 in Graz (Österreich) wurde die Frage gestellt: «*Treten Sie dafür ein, dass die von der Stadt Graz geplante Verlängerung der Linie 6, die in dieser Form nicht zur Lösung der bestehenden Verkehrsprobleme beiträgt, nicht zur Ausführung gelangt?*»).
- **Eindimensionalität** (nur eine Dimension abfragen). **Nicht:** *Sind sie für die Reduktion der Militärausgaben und die Erhöhung der Sozialausgaben*
- Möglichst **konkreter Bezug** (Wertevorstellungen vs. Einstellungen zu konkreten Sachfragen).
- Antwortkategorien erschöpfend, disjunkt und ausgewogen. **Nicht überschneiden:** Kategorie 1: 300-500/ Kategorie 2: 501-1000. *Bei Überschneidungen kann es passieren, dass sich eine Person gar nicht einteilen lässt, bzw. in zwei Kategorien fällt!*
- **Neutrale Formulierung** und (wenn möglich) keine hypothetischen Fragen.
- Einbau von **Filterfragestrukturen**. → Vorsehen, dass wenn bei einer gewissen Frage etwas gesagt worden ist, alle Fragen die ein anderes Verhalten ausschliessen aus dem Fragebogen herausgenommen werden.
- **Vergleichbarkeit** mit anderen Untersuchungen (z.B. der Vergangenheit) **verlangt identische Fragestellungen** (zum Beispiel politisches Wissen).
- Vereinbarkeit mit geplantem Auswertungsverfahren (Auswahl eines ausreichenden Skalenniveaus).

• Fragebogaufbau

- Einleitungsfragen („Eisbrecherfragen“), **Interesse wecken** und Vertrauen aufbauen.
- Positionierung der Fragen im Fragebogen: **Kernfragen** möglichst **früh stellen** (wegen Abrechnen). Sozialstatistische Fragen vorzugsweise am Ende der Befragung (wenn möglich).
- Filterfragen („In welchem Kanton sind Sie stimmberechtigt?“). **Vermeidung überflüssiger Fragen** („*Waren Sie 2011 im selben Kanton stimmberechtigt wie heute?*“ besser als: „*In welchem Kanton waren Sie 2011 stimmberechtigt?*“)
- Strukturierte Themenblöcke mit Überleitungen.
- Vermeidung von Saliency-Effekte (vorherige Fragen beeinflussen Antworten auf spätere Fragen). **Fragen sollten so platziert werden, dass sie nicht von den vorherigen abhängig sind.**
- Vermeidung von Ex-Post-Rationalisierung und Konsistenzeffekte: Befragte neigen zur Ex-Post-Rationalisierung und „Harmonisierung“ von Antworten. **Deshalb:** Fragen, die Konsistenz testen, weit voneinander setzen.

• Geschlossene Fragen vs. offene Fragen

1. Bei **geschlossenen Fragen** sind die Antwortkategorien vorgegeben.

- **Vorteile:**

- **Hohe Vergleichbarkeit** und **hohe Durchführungsobjektivität**.
- **Geringerer Zeitaufwand** für Befragte und für das Auswertungsteam, da keine inhaltsanalytische Codierung der Antworten erforderlich ist.
- **Geringer Interpretationsspielraum** für Codierer.

- **Nachteile:**

- **Antworten** werden durch die Vorstrukturierung „**vorgespart**“: Im schlimmsten Fall erhält man eine falsche Antwort, weil die richtige bzw. passende Antwort gar nicht vorgesehen ist. Aber auch das Gegenteil möglich: Angabe einer Kategorie, die dem Befragten spontan nie in den Sinn gekommen wäre (Ex-Post-Rationalisierung).

2. Bei **offenen Fragen** sind die Antwortkategorien nicht vorgegeben. Der Befragte artikuliert seine Antworten „selbständig“.

- **Vorteile:**

- Der Befragte antwortet innerhalb seines eigenen Referenzsystems.

- **Nachteile:**

- Siehe Vorteile geschlossener Fragen.

3. **Hybridform**: halboffene Fragen.

- Fragen, bei denen gewisse Antwortkategorien vorgegeben sind, bei denen der Befragte jedoch auch eine eigene Antwort formulieren kann.

Vorlesung 11: Auswahlverfahren

• Einleitung: Voll- und Teilerhebungen

- **Ziele von Umfragen**

- Das Ziel einer Umfrage ist es, **Aussagen über das Verhalten oder die Haltungen einer Grundgesamtheit** (z.B. «alle Schweizer Stimmberechtigten») machen zu können. Allerdings lässt sich in aller Regel **nur ein Teil dieser Grundgesamtheit befragen**. Wie muss nun diese Teilmenge (oder Stichprobe) zusammengesetzt bzw. beschaffen sein, so dass sich die Stichprobenergebnisse verallgemeinern lassen? Wie zieht man eine solche, „**repräsentative**“ **Stichprobe**?

• Grundbegriffe der Umfrageforschung

• Die Grundgesamtheit („target population“)

- Die Menge derjenigen Individuen, über die eine Aussage gemacht werden soll, wird die Grundgesamtheit genannt.
- Die Grundgesamtheit ist demnach von der jeweiligen **Fragestellung abhängig**.
- **Beispiel:** Wer hat wie gewählt bei den Nationalratswahlen 2015? Gilt die Nicht-Teilnahme ebenfalls als Wahloption, so bilden alle Schweizer Wahlberechtigten die Grundgesamtheit. Interessiert man sich indessen nur für das substanzielle Wahlverhalten, so bilden strenggenommen nur die an den Wahlen 2015 Teilnehmenden die Grundgesamtheit.

• Die Auswahlgesamtheit („frame population“)

- Die **Menge der Elemente, die eine prinzipielle Chance haben, in die Stichprobe zu gelangen**. Bildet beispielsweise das offizielle Telefonverzeichnis der Swisscom die Auswahlgesamtheit, so haben nur die über die darin registrierten Telefonnummern Erreichbaren die prinzipielle Chance, in die Stichprobe zu gelangen.
- Die **Inferenzpopulation** wird aus der Auswahlgesamtheit abgeleitet. Sie ist die Population, über die auf der Grundlage der vorliegenden Stichprobe tatsächlich Aussagen gemacht werden können.
- Auswahlgesamtheit (bzw. Inferenzpopulation) und Grundgesamtheit **sollten im Idealfall identisch** oder zumindest nahezu identisch sein. Je weiter diese beiden Grössen auseinanderfallen (siehe overcoverage und undercoverage), desto grösser ist die Wahrscheinlichkeit, dass die Umfrageresultate verzerrt sind.

• Undercoverage und Overcoverage

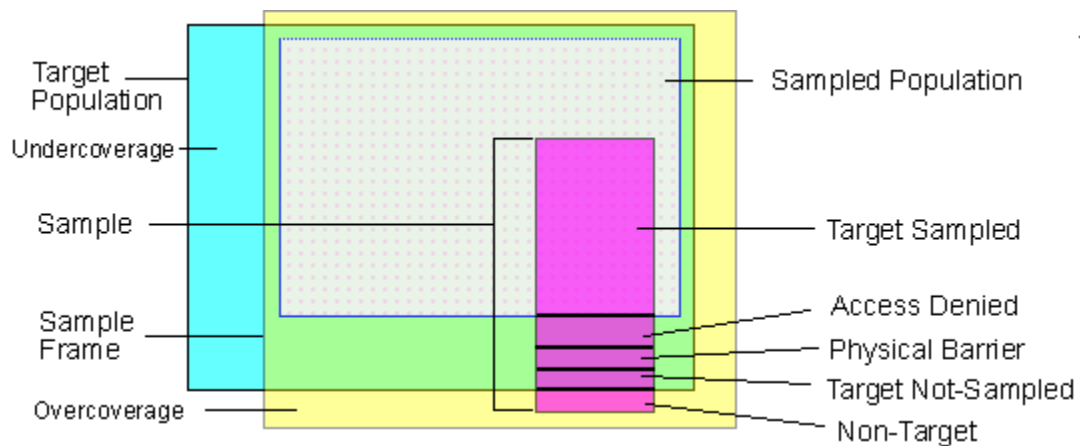
• „Overcoverage“

- Die Auswahlgesamtheit enthält Elemente (d.h. Befragte), die kein Teil der Grundgesamtheit sind.
- **Beispiel:** Die Schweizer Stimmberechtigten bilden die Grundgesamtheit. Man befragt indes alle, die im Swisscom-Verzeichnis enthalten sind (z.B. auch AusländerInnen oder unter 18-jährige). → Kann die Umfrage verzerren
- **Overcoverage** ist in der Praxis **selten ein Problem**. Die «überzähligen» Fälle können nachträglich aus der Stichprobe entfernt werden.

• „Undercoverage“

- Wenn Elemente der Grundgesamtheit nicht in der Auswahlgesamtheit enthalten sind.
- **Generell gilt:** **undercoverage ist ein weitaus schwerwiegenderes Problem** als „overcoverage“. Denn die Verzerrungen, die von „undercoverage“ ausgehen, lassen sich kaum bzw. nur unter Zuhilfenahme bestimmter Annahmen korrigieren.

Abbildung: Grundgesamtheit und Auswahlgesamtheit (sample frame)



• Auswahl-, Erhebungs- und Untersuchungseinheit

- **Auswahleinheiten** beziehen sich auf den Auswahlplan, während **Erhebungseinheiten** diejenigen Einheiten sind, bei denen die Informationen erhoben werden. Diese Einheiten sind oftmals nicht identisch und dies stellt die Umfrageforscher vor gewisse Probleme.
 - **Beispiel:** Auswahleinheit (Beispiel Telefonnummer) kann verschiedenen Erhebungseinheiten zugewiesen werden.
- Die **Untersuchungseinheit** ist diejenige Einheit, auf welche sich die Untersuchung bezieht bzw. an welcher die relevante Messung als solches vorgenommen wird.

• Repräsentivität

- Ein („geschützter“) Fachbegriff „Repräsentivität“ existiert nicht. Es gibt **unterschiedliche Bedeutungen** dieses Begriffs in der nichtwissenschaftlichen, der wissenschaftlichen und der statistischen Literatur
1. Repräsentivität gilt landläufig als „**Gütesiegel**“ für demoskopische Resultate („Here the investigator gives the data a pat on the back by using a seemingly scientific term to rise its stature.“). **Wenig sinnvoll.**
 2. Repräsentivität als Strukturkonzept: Stichprobe und Grundgesamtheit sind gleich zusammengesetzt. Das Sample ist ein verkleinertes, aber ansonsten strukturgleiches Abbild der Grundgesamtheit. Hohe Suggestivwirkung, aber ebenfalls wenig sinnvoll.
 - **allgemein bekannte Vorstellung der Repräsentivität**
 3. Repräsentivität in der statistischen Literatur: bezieht sich nicht auf die Struktur der Stichprobe, sondern auf den **Prozess der Auswahl**. Nur Zufallsauswahlen ermöglichen die inferenzstatistischen Verfahren zur Berechnung der Verlässlichkeit, mit welcher man vom Sample auf die Grundgesamtheit schliessen kann.

- **Auswahlverfahren: Zufallsgesteuerte Auswahlen**
 - **Einfache Zufallsauswahl**
 - Eine **Zufallsauswahl** zeichnet sich dadurch aus, dass **alle Elemente** der Grundgesamtheit **dieselbe Chance** haben, ins Sample zu gelangen (und diese Chance höher als Null ist).
 - **Komplexe Zufallsauswahlen**
 - In **komplexen Zufallsauswahlen** haben die einzelnen Elemente nicht mehr dieselbe, aber **eine berechenbare Chance**, in die Stichprobe zu gelangen. Komplexe Zufallsauswahlen werden dann bevorzugt, wenn beispielsweise **interessante Untergruppen der Grundgesamtheit nur sehr geringe Anteil aufweisen** und demnach – bei einer einfachen Zufallsauswahl – nicht gesondert ausgewertet werden können, da die Fallzahlen schlicht zu gering sind.
 - **Geschichtete Zufallsauswahlen**
 - Bei einer geschichteten Zufallsstichprobe werden die **Elemente der Grundgesamtheit** zunächst **einer (und nur einer) Schicht zugewiesen** woraus anschliessend eine Zufallsauswahl gezogen wird. Eine geschichtete Stichprobe bietet sich insbesondere in den folgenden beiden Situationen an:
 1. Wenn sichergestellt werden muss, dass die Häufigkeitsanteile der Subpopulationen ihren Anteilen in der Population exakt entsprechen (**proportionale Schichtung**).
 2. Wenn man an spezifischen Subpopulationen interessiert ist, deren Anteil in der Gesamtpopulation aber klein ist. Eine einfache wie auch eine proportional geschichtete Zufallsstichprobe würden wohl eine **zu geringe Fallzahl für die interessierende Subpopulation** realisieren. Hier bietet sich die **disproportional geschichtete Stichprobe** an.
 - **Klumpenauswahl** („cluster sample“)
 - Klumpen sind "natürlich" auftretende Gruppen wie beispielsweise Schulen.
 - **Klumpenstichproben** sind in der Praxis ziemlich häufig, weil eine Liste aller Elemente der GG nur selten, eine Liste von Klumpen aber oft vorhanden ist.
 - **Vorgehen**: Aus der Liste der Klumpen wird eine **einfache Zufallsauswahl von Klumpen** gezogen. Danach wird eine Vollerhebung der Elemente der so ausgewählten, einzelnen Klumpen durchgeführt.
 - Im Unterschied zum geschichteten Zufallsverfahren wird in der untersten Ebene jeder befragt.

- **Auswahlverfahren: Nicht-Zufallsgesteuerte Auswahlen**

- **Willkürliche Auswahlen**

- **Unterschied** von **willkürlich** und **zufällig** ist die **Rekrutierung**: bei zufälligem Auswahlverfahren wird systematisch rekrutiert, bei willkürlichem wird selbst entschieden ob man teilnimmt.
- Einer willkürlichen Stichprobe fehlt ein eigentlicher Auswahlplan. Die Befragten werden aufgrund ihrer **einfachen Verfügbarkeit** oder einer **anderen nicht-systematischen Methode** ausgewählt.
- **Beispiele: Strassenbefragungen**, welche vom Fernsehen ab und an durchgeführt werden, um zu erfahren, was der Mann auf der Strasse über ein Problem denkt.

Aber auch **opt-in Onlineumfragen**, bei denen diejenigen, die sich auf der entsprechenden Internetseite aufhalten, selbst entscheiden, ob sie teilnehmen wollen oder nicht.

- **Bewusste Auswahlen/ Quotenauswahlen**

- Bei einer Quotenauswahl wird die Population, wie beim Schichtungsverfahren zunächst nach bestimmten Schichtungsmerkmalen in verschiedene Subpopulationen unterteilt. Darauf aufbauend **erhalten die Interviewer entsprechende "Quoten", die erfüllt werden müssen**. Diese Quoten können u.U. auch Merkmalskombinationen sein (z.B. „25% müssen 18-29-jährige Frauen sein“)
- Im **Gegensatz zu geschichteten Zufallsstichprobe** wird **keine Zufallsstichprobe** innerhalb der jeweiligen Schichten gezogen, sondern der **Interviewer entscheidet selbst, wie er die vorgegebenen Quoten erfüllt**. Aus diesem Grund sind Quotenstichproben keine zufallsgesteuerten Auswahlen.

- **Vor- und Nachteile der einzelnen Auswahlverfahren**

- **Zufallsauswahl:**

- Anwendung **inferenzstatistischer Verfahren** nur bei Zufallsverfahren möglich.
- **Nachteil**: Praktikabilität (in aller Regel fehlt eine Liste der Grundgesamtheit).

- **Quotenauswahl:**

- Setzt Vorwissen über die Grundgesamtheit voraus. Quotenmerkmale sind mit den anderen interessierenden Merkmalen (oftmals) korreliert.
- Quotierung kann (muss aber nicht) zu **höherer Strukturgleichheit** führen.
- Wird Quotenplan sorgfältig durchgeführt, wird der Schätzfehler geringer (weil äquivalent zu einer geschichteten Zufallsstichprobe)
- Verletzung von Quotenanweisungen **kaum kontrollierbar**

- **Geschichtete Zufallsauswahl:**
 - Verringerung des Fehlerintervalls der Schätzung.
 - Setzt Vorwissen über die Grundgesamtheit voraus und hohe Heterogenität eines Merkmals in der Grundgesamtheit.
- **Klumpenauswahl:**
 - Bei fehlender Liste der Grundgesamtheit, aber vollständiger Liste der „Klumpen“ empfehlenswert.
 - Schätzfehler in aller Regel grösser als bei einfachen Zufallsstichproben (abhängig von Homogenität innerhalb der Klumpen und Heterogenität zwischen den Klumpen).
- **Probleme und Herausforderungen von Bevölkerungsumfragen.**
 - **Fehlerquellen 1: «sampling errors»**
 - **Zufallsfehler der Stichprobe (sampling error):** ist allerdings berechenbar (siehe nachfolgende Lektion)
 - **Fehlerquellen 2 «non sampling errors»**
 - Systematischer Fehler aufgrund des Auswahlverfahrens bzw. -rahmens (**coverage error**)
 - **Non Response:** Verweigerung (**item und unit nonresponse**)
 - Gesamtverweigerung einer Befragung (unit non response) oder Verweigerung gewisser Fragen einer Befragung (item)
 - **Messfehler (measurement error):** zum Beispiel das „Overreporting“ bei der Angabe der Wahlteilnahme. → **Falschaussagen**
 - Gewichtung und Datenaufbereitung (**processing errors**)
 - **Problem 1:**
 - Eine **Liste der Grundgesamtheit existiert nicht** (Ausnahme: SRPH, BfS-Registerdaten). Deshalb weichen so gut wie alle Befragungsinstitute auf Telefonlisten aus. Diese sind nie (selbst unter Beimischung von RDD) eine vollständige Liste der Grundgesamtheit. Verknüpft man den Repräsentativitätsbegriff an die Existenz einer vollständigen Liste der Grundgesamtheit, kann kaum eine Schweizer Befragung als repräsentativ bezeichnet werden.
 - Praxis: Telefonlisten (bilden dann die Auswahlgesamtheit, vgl. vorangegangene Folie)
 - Nur Festanschlüsse.
 - „Geburtstagmethode“

- Undercoverage und overcoverage-Probleme (vgl. Schnell et al. 2005)
- **Problem 2:**
- **Abdeckung des öffentlichen Telefonverzeichnis:** rund 70 Prozent, Tendenz fallend. Aber Handys sind immer weiter verbreitet. Handynummern sind viel seltener registriert als Festnetzanschlüsse. → Es ist unmöglich eine genügend grosse Grundgesamtheit zu erreichen.
- **Problem 3:**
- **Interviewverweigerung.** Ist kein Problem der Stichprobenziehung, denn diese Fälle sind als Untersuchungseinheiten ja vorgesehen! **Ausschöpfungsquoten** betragen mittlerweile 20-25 Prozent.
- Allerdings zeigen Untersuchungen, dass höhere Ausschöpfungsquoten nicht notwendigerweise genauere Ergebnisse liefern.