Teil 1: Quantitative und qualitative Herangehensweise
Sitzung 1&2: 20/27.09.19 – Einführung in quantitative und quantitative Texte

Prüfung: 08.01.2020 10-12 Uhr!
E-Assessment (am Computer!) Single Choice, Multiple Choice, Kprim
Auf Folien sind Begriffe verlinkt!!
Prüfungsstoff: Diekmann (angegebene Seiten) und Vorlesung

Methodenlehre als disziplinübergreifende Erfahrungswissenschaft.

Quantitative Forschung – strukturierte und standardisierte Verfahren, aus denen quantitative, also numerische (Messwerte) Daten resultieren.

➔ Naturwissenschaften, Historisch älterer und führender Forschungsansatz.
Ursache-Wirkungs-Mechanismus. Welt ist messbar.

Qualitative Forschung – unstrukturierte Datenerhebung, aus denen qualitative, also nicht numerische Daten (Text, Bild, Video) resultieren

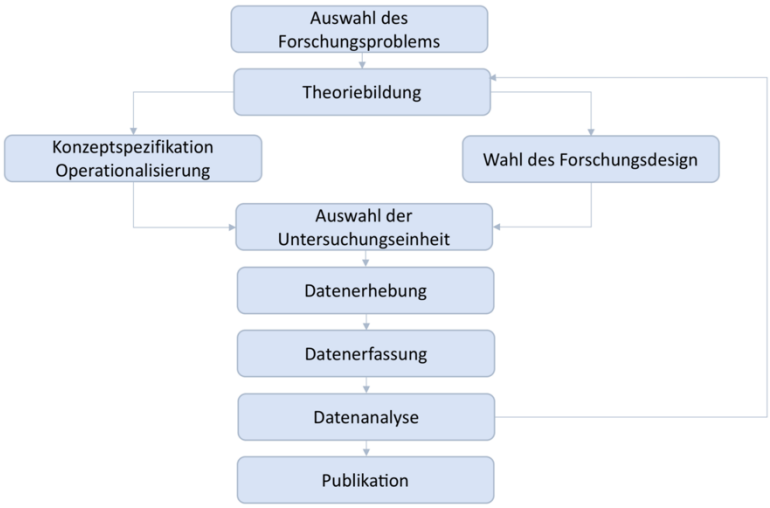
➔ Geisteswissenschaften
Das Verstehen und nicht das Erklären im Zentrum!

Kritik Qualitativ an Quantitativ:
Verfehltes Wissenschaftsverständnis, fälschliche Annahme eine objektive beobachtbare soziale Erfahrungswirklichkeit existiere, bloss Machtverhältnisse bestätigende Befunde werden produziert.
Kritik Quantitativ an Qualitativ:
Zweifel an der Wissenschaftlichkeit und Kritik am Anspruch der Gesellschaftsverbesserung.

➔ In jüngerer Vergangenheit eher Kombination der beiden Forschungsansätze.

Qualitativ(-geisteswissenschaftliches) Paradigma	Quantitativ(-naturwissenschaftliches) Paradigma
Tradition der Geisteswissenschaften	Tradition der Naturwissenschaften
Verstehen des subjektiven Sinns menschlichen Handelns →Generierung von Hypothesen und Theorien am Ende. Induktiv.	Erklären von Sachverhalten aufgrund von Ursache-Wirkungs-Mechanismen, wie in den Naturwissenschaften. →Überprüfen von Theorien und Hypothesen. Deduktiv.
Erfahrungsrealität wird verbalisiert oder grafisch abgebildet.	Erfahrungsrealität wird in Zahlen abgebildet und statistisch ausgewertet.
Nicht-numerisches Datenmaterial → interpretativ-hermeneutische Auswertung Dennoch auch empirisch.	Numerisches Datenmaterial →statistische Auswertung Empirisch.
Wenig Fälle werden untersucht, offene, unstrukturierte und unstandardisierte Fragen	Viele Fälle mit geschlossenen, strukturieren und standardisierten Fragen.
Bsp. Methodologien: ethnografische Feldforschung, Biografieforschung. Realtiv Theoriefrei, denn Datenmaterial wird gesammelt und dadurch geht Theorie hervor.	Bsp. Methodologien: Kausalnachweis im Experiment oder Umfrageforschung auf der Basis von bevölkerungsrepräsentativen Stichproben.
Grundidee: Sozialer Konstruktivismus Soziale Realität als konstruiert, prozesshaft, dynamisch: SUBJEKTIV.	Grundidee: Kritischer Rationalismus. Soziale Realität ist objektiv, denn Welt ist VERMESSBAR.

Quantitative Texte am Beispiel eines Textes zu Populismus.



Gütekriterien der Messung:

- Reliabilität: Zuverlässigkeit
- Validität: Ist bspw. die Frage in einer Umfrage überhaupt brauchbar und zielführend für die Hypothese/Fragestellung zu beginn
- Operationalisierung: Messung populistischer Haltungen Antworten in numerischen Zahlen. (1: *I very much disagree;* 5: *I very much agree*)

➔ Motivation muss immer transparent sein! Und auch alle Forschungsschritte! Wissenschaftlich sein heisst transparent sein!

Teil 2: Grundprobleme empirischer Sozialforschung
Sitzung 3: 04.10.19 – Grundzüge empirischer Sozialforschung

<u>Was sind empirische Sozialwissenschaften?</u>

- Methoden = Systematische(s) Verfahren zur Gewinnung und Überprüfung von Aussagen über die Realität.
  Methodisches Vorgehen gewährleistet, dass die Ergebnisse überprüfbar, kritisierbar und in ihrer Verallgemeinerungsfähigkeit/Gültigkeit abschätzbar sind.

- Methodologie = Lehre von der Methode. (Paradigmen)
  Formulierte Regeln, wie man mit systematischen Verfahren wissenschaftliche Erkenntnisse gewinnen kann. (vgl. qualitativ vs. quantitativ)
  Erhebungs- vs. Auswertungsmethoden:
  Methoden der Datenerhebung (z.B. Befragung oder Beobachtung) stehen Methoden der Datenauswertung (Statistische Verfahren wie Regression) gegenüber.

- Empirie = Auf methodischem Weg gewonnene Erfahrungen (Beobachtungen/Wahrnehmungen der Realität) der Welt.
  → Auf Erfahrung beruhend (≠ Vernunft).
- Empirische Wissenschaft = Erfahrungswissenschaften. Theorien der Empirischen Wissenschaft können deshalb auch an der «Wirklichkeit» überprüft werden (≠ Vernunftissenschaften wie Mathematik)

➔ Ziel der emprisichen Sozialforschung ist die Erkenntnisgewinnung über *soziale* Phänomene/Sachverhalte im Gegensatz zu Naturphänomenen (=Naturwissenschaften)

Grundposition der Erfahrungswissenschaften:
Begriffe werden nicht immer genau gleich verwendet, in dieser Lehrveranstaltung jedoch linear.
Unabhängig von unserer Wahrnehmung existiert es da draussen eine reale Welt, die wir entdecken können. Dies wird (normalerweise) in der Wissenschaft nicht geleugnet.
Hinsichtlich Erfahrbarkeit jedoch zwei Positionen:

- Erkenntnistheoretischer Realismus:
  Eine intersubjektive überprüfbare Erfahrbarkeit (und somit eine objektive Erkenntnis) der Welt ist prinzipiell möglich. Deshalb sind auch Aussagen über die Realität möglich.
- Erkenntnistheoretische Konstruktivismus:
  Die Wirklichkeit ist subjektiv konstruiert («erfunden») und kann daher nicht objektiv «entdeckt» werden. Wahrnehmungen und Beobachtungen der Welt finden immer in einem bestimmten (eigenen) und subjektiven Kontext statt und tragen bereits eine Interpretation in sich. Somit sind Induktionen (Schluss vom Besonderen auf das Allgemeine) und Falsifikation strenggenommen nicht möglich.

Wissenschaft und Alltagswissen
In beiden Fällen wird beobachtet, beschrieben und es wird nach Erklärungen (im Alltag = Alltagstheorien) gesucht.
Unterschiede sind jedoch:

Alltagwissen	Wissenschaft
Bezug auf konkrete Einzelfälle und -situationen (allenfalls: pauschale Verallgemeinerungen)	Generalisierung vieler einzelner Beobachtungen («Induktion»)
Geleitet durch Alltagstheorien/ «Common sense»	Systematisch-methodische Vorgehensweise, Intersubjektivität, Transparenz.
Individuelles (d.h. unvermitteltes) wenig präzises Wissen und gering präzisierte Theorien	Hohe Präzision bei Konzepten und Theorien. Replizierbarkeit, prinzipielle Möglichkeit der Falsifikation. Keine präskriptiven Aussagen.

Merkmale empirischer Wissenschaft
Was zeichnet wissenschaftliche Forschung (nach King et al.) aus:

- Ziel der Forschung ist Inferenz (aufbereitetes Wissen aufgrund logischer Schlussfolgerungen)
- Vorgehensweise der Forschung ist transparent (und somit reproduzierbar)
- Schlussfolgerungen sind prinzipiell *unsicher*
- Das Kennzeichen der Wissenschaft ist die *Methode*, nicht der Inhalt

Ziele wissenschaftlicher Forschung

- Beschreiben
- Erklären (und Vorhersagen)
- Verstehen
- Bewerten (Evaluieren)

Explorative Untersuchungen → Verstehen

- Ziel: *Nachvollziehen* der Bedeutung/Sinns von sozialen Phänomenen: Das Verstehen von Ereignissen/Handlungen durch die Erfassung des Sinnzusammenhangs, in den sie eingebettet sind.

- Vorgehensweise: Bei geringem Vorwissen wird zunächst eine Auslegung des Untersuchungsgegenstandes gemacht. (dimensionale Analyse)
- Methode: qualitative Methoden (z.B. Tiefeninterviews)
- Theoriebildung: Dienen der Bildung von Hypothesen, die anschliessend oft in einer hypothesenüberprüfenden Studie empirisch geprüft werden.

Deskriptive Untersuchung → Beschreiben und Verstehen
- Oft univariate (nur eine Variabel umfassende) Untersuchungen
- Ziel: Schätzung (Statistik) der Verteilung einer VariabeBl (z.B. Geburtenrate, Wähleranteil)
- Ziel generell: *Diagnose* (wie viele Menschen leben unter der Armutsgrenze?) und nicht Ursachenforschung (Wovon ist Armut abhängig?)

Hypothesenprüfende Untersuchung → Erklären
- Vorrangiges Ziel der Sozialwissenschaften
- Herleitung der Hypthesen aus bestehenden Theorien (Deduktion) und ihre anschliessende empirische Überprüfung.

Evaluationsstudien → Bewerten
- Stark anwendungsbezogenen Wirkungsforschung. Ziel ist die Bewertung von politischen und sozialen Massnahmen. (z.B. «Erhöht Online-Voting die Wahlbeteiligung?)
- In der Regel liegt solchen Untersuchungen ein quasi-experimentelles Design zugrunde.

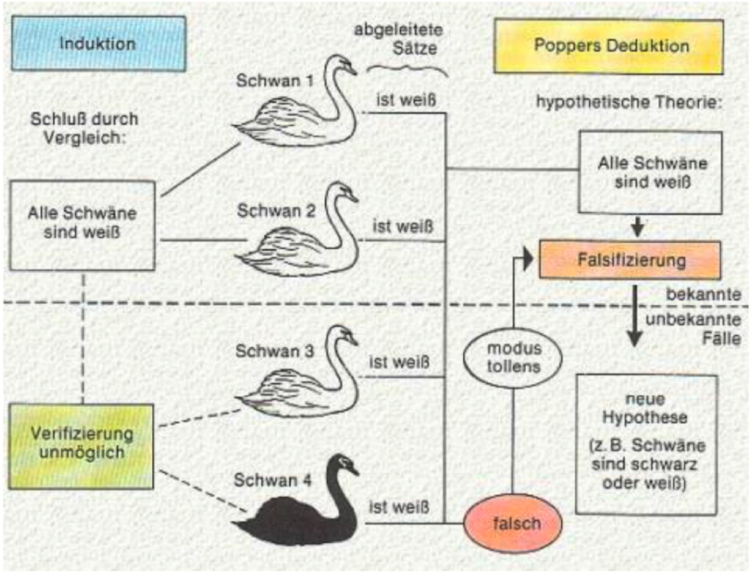
<u>Wissenschaftstheoretische Positionen</u>

		Popper	Frankfurter Schule	
Item	**Positivism**	**Postpositivism**	**Critical Theory et al.**	**Constructivism**
Ontology	Naive realism - real reality and apprehendable	Critical realism - real reality but only imperfectly and probabilistically apprehendable	Historical realism - virtual reality shaped by social, political, cultural, economic, ethnic, and gender values; crystallized over time	Relativism - local and specific constructed realities
Epistemology	Dualist/objectivist; findings true	Modified dualist/objectivist; critical tradition/community; findings probably true	Transactional/subjectivist; value-mediated findings	Transactional/subjectivist/ created findings
Methodology	Experimental/ manipulative; verification of hypotheses; chiefly quantitative methods	Modified experimental/ manipulative; critical multiplism; falsification of hypotheses; may include qualitative methods	Dialogic/dialectical	Hermeneutical/dialectical
Inquiry aim	explanation: prediction and control		critique and transformation; restitution and emancipation	understanding; reconstruction
Nature of knowledge	verified hypotheses established as facts or laws	non-falsified hypotheses that are probable facts or laws	structural/historical insights	individual reconstructions coalescing around consensus

Quelle: Guba & Lincoln (2005). Paradigmatic Controversies, Contradictions, and Emerging Confluences.

Kritischer Rationalismus (Karl Popper) («Postpositivismus»):
- Ausgangslage: Knöpft an naiven Positivismus an, verwirft ihn anschliessend aber.
- Wissenschaft als Falsifikationsprogramm: Allgemeine empirische Sätze und Theorien können demnach nicht verifiziert, sondern falsifiziert werden.
- Alle wissenschaftlichen Aussagen sollen über die Realität informieren und müssen empirisch überprüfbar sein.
    - Konsequenzen:
      Es dürfen nur Konzepte oder Begriffe verwendet werden, die einen Bezug zur Realität bzw. Empirie haben.
    - Die Aussagen müssen so formuliert sein, dass sie prinzipiell widerlegbar sind. Nominaldefinitionen (oder analytisch wahre Sätze wie «Schimmel sind weisse Pferde») sind also nach dem kritischen R. keine empirischen Aussagen.

- Genauso sind Existenzaussagen nicht zugelassen, da sie prinzipiell unwiderlegbar sind. Bspw. «Es gibt schwarze Schwäne»



Verifizierung nicht möglich, Falsifizierung möglich.
(Für Verifizierung müsste jeder einzelne Schwan auf der ganzen Welt angeschaut werden)

Exkurs: Deterministische und probabilistische Erklärungen:
- Deterministische Sätze: sind Aussagen, die immer und überall gültig sind. Also: Wenn x dann immer y. Diese Aussagen haben einen unbeschränkten raumzeitlichen Geltungsanspruch für alle Elemente. Formulieren wir (fast) nie in den Sozialwissenschaften. (Einziges Beispiel wäre: Demokratie führen keine Kriege untereinander. Falsifizierbarkeit abhängig von Definition von Demokratie)
- Probabilistische Aussagen: Gegeben (im Gegensatz zu deterministischen Aussagen) eine Wahrscheinlichkeit an und haben nur einen beschränkten Gültigkeitsbereich.

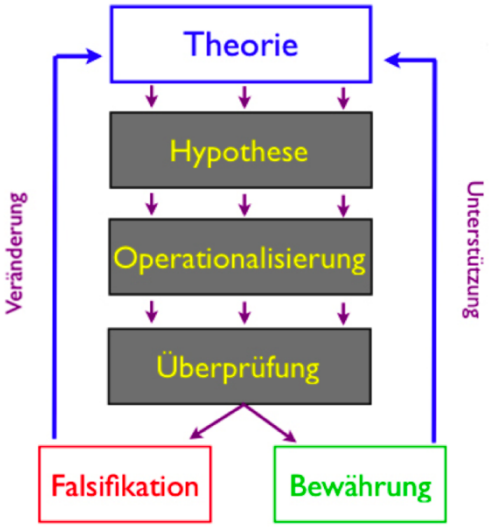
Wissenschaft als Falsifikationsprogramm:
- Konsequenter Fallibilismus: Unser theoretisches Wissen ist grundsätzlich fehlbar, eine sichere Begründung ist nicht möglich.
- Grundlage Falsifikationslogik: Die Möglichkeit der Falsifikation von Allaussagen durch ein einziges Gegenbeispiel.
- Ziel: Falsifikation einer Hypothese.

Reichweiter von sozialwissenschaftlichen Theorien:
- Grundsätzlich: Angestrebt werden All-Aussagen/ nomologische Sätze.
- Da aber Allaussagen in der Sozialwissenschaft (im Vergleich zu Naturwissenschaften) kaum möglich sind, werden «Theorien mittlerer Reichweite» (Robert K. Merton) angestrebt. Sie liegen zwischen «grossen Theorien» und empirischer Bedarfsforschung.

Informationsgehalt von sozialwissenschaftlichen Theorien:
- «Theorien mittlerer Reichweite» haben einen eingeschränkten Objektbereich (z.B. Wahlverhalten von *Jugendlichen*) und/oder raumzeitliche Einschränkungen (z.B. Wahlverhalten in Industriestaaten des *20. Jh.*).
- Potentielle Falsifikatoren = alle Elemente des angegebenen Objektbereichs.
- ➔ Deshalb gilt: Je grösser Anzahl potentieller Falsifikatoren, desto höher Informationsgehalt der Theorien!

- «Ceteris paribus»-Klauseln (Y stimmt in Abhängigkeit zu Z, W, P, U etc.) sind zu vermeiden, denn sie verunmöglichen u.U. die empirische Überprüfbarkeit der Theorie.

Forschungsstrategie nach Karl Popper: «trial and error»

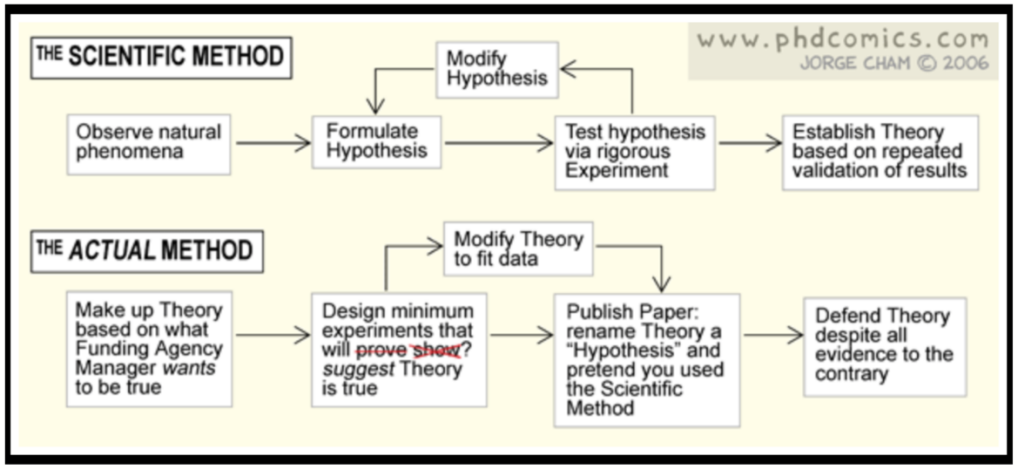


Falsifikation → Verwerfung/Revision Theorie.
Empirische Unterstützung führt zu vorläufiger Bestätigung der Theorie.
→Weitere Überprüfung durch «schärfere Tests» führt zu «bewährten Aussagen».
Ziel: Ausdehnung der Reichweiter der Theorie.

Basissatzproblem:
Ist Falsifikation zweifelsfrei möglich?

- Ausgangslage: Aussagen müssen prinzipiell falsifizierbar sein.
- Problem: Aussagen können nicht direkt mit Beobachtungen verglichen werden, sondern immer nur mit *Aussagen über Beobachtungen*.
- Beobachtungsaussagen ihrerseits können aber falsch sein. Die Gründe sind vielfältig. Bspw.:
  - Keine Erfahrung ist Theorieunabhängig.
  - Messtheorie ist ebenso unsicher wie Kerntheorie. (z.B. «falsche» Wahrnehmung. Vielleicht war der schwarze Schwan eine Gans.)
- Konventionalistische Lösung:
  Einhaltung höchstmöglicher methodischer Standards und transparente Darstellung der Vorgehensweise.

<u>Das Problem der Werturteilsfreiheit in den Sozialwissenschaften:</u>



Werturteilsfreiheit, Max Weber:

- Max Weber: Die Gültigkeit von Werten ist nicht wissenschaftlich beweisbar.
- Webers Forderung: Tatsachenbehauptungen und Werturteile sind in den Wissenschaften deutlich getrennt voneinander darzustellen. Dies, um zu verhindern, dass subjektive Ansichten unter dem Deckmantel der Wissenschaftlichkeit verbreitet wurden.

- Kritik: Es ist unmöglich, empirische Tatsachenbehauptungen und Wertungen voneinander zu trennen
- „Positivismusstreit “: Die kritische Theorie lehnt die Werturteilsfreiheitsthese ab, da die Vorstellung der Totalität der Gesellschaft sowie die Verflochtenheit von Erkenntnis und Gesellschaft die Trennung von Tatsachen und Werten verbietet.
- Frankfurter Schule: jede Beobachtung ist theoriedurchtränkt

➔ Wissenschaft aus einer Perspektive treiben → hier: gesellschaftliche Repression verringern

- Kritik: Eine objektive Gesellschaftskritik ist nicht formulierbar, da es keine Möglichkeit gibt, die Massstäbe der Kritik (z.B. eine weniger repressive Gesellschaft) wissenschaftlich zu begründen.

➔ Wer entscheidet aus welcher Perspektive man die Wissenschaft treibt?

Max Webers Postulate der 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.

➔ Beantwortung der Frage: werturteilsfrei → intersubjektiv

Teil 3: Die theoretische Formulierung des Forschungsproblems
Sitzung 4: 11.10.19 – Konzepte, Definitionen und Analyseeinheiten

Elemente einer Theorie: Konzepte

<u>Konzeptspezifikationen</u>

- In der Regel 2. Schritt im Forschungsablauf. (Nach Formulierung der Forschungsfrage)
- Aussagen enthalten Begriffe bzw. Konzepte. Es muss verständlich sein, was mit den Begriffen bzw. Konzepten gemeint ist.
- Konzeptspezifikation = zentrale Konzepte meiner Aussage definieren.
  → Verknüpfung empirische Wirklichkeit mit sprachlichen Begriffen.

<u>Begriffe</u>

- Begriffe müssen nicht notwendigerweise einen empirischen Bezug vorweisen, in der empirischen Sozialforschung jedoch schon.
- „Begriff ist ein sprachliches Zeichen für eine gedankliche Vorstellung von einer Klasse gleichartiger Sachverhalten oder Gegenständen“ (Kromrey 2009: 138)

- Verbindung zwischen Begriff und verknüpftem Gegenstand:
    - Intensional (Begriffsinhalt) = Zugehörigkeitsmerkmale, Menge aller Kriterien, die einen bestimmten Gegenstand auszeichnen (Zugehörigkeitsmerkmale)
    - Extensional (Begriffsumfang) = Objekte, die mit dem Begriff bezeichnet werden sollen.

Konzept = Gedankliche Vorstellung
Begriff = Sprachliche Bezeichnung für dieses Konzept

<u>Empirischer Bezug der Begriffe</u>
(sofern er einen aufweist, kann dieser sein:)

- **Direkt** (= manifeste Grösse oder Variabel): Grösse, Gewicht, bestimmte Handlungsformen (z.B. Stimmabgabe)
- **Indirekt** (= latente Grösse oder Variabel):Nicht direkt beobachtbar. Wie bspw. Haltungen oder Einstellungen (z.B. Konservatismus/Autoritarismus). Gegenstände oder Objekte, die nur unter Zuhilfenahme von direkten Beobachtungstermini (wie Kleidung) und dazugehörigen Schlussfolgerungen (Messhypothesen) nachweisbar sind. (Bekleidung → politische Haltung)

Beispiel latente Variabel: Erforschung von Autoritarimus.

- Zur Messung von Autoritarismus schlug Stanley Feldman (2003) folgende Batterie von vier Items vor:
    1. "Please tell me which one you think is more important for a child to have: independence or respect for elders?"
    2. "Please tell me which one you think is more important for a child to have: obedience or self-reliance?"
    3. "Please tell me which one you think is more important for a child to have: to be considerate or to be well-behaved?"
    4. "Please tell me which one you think is more important for a child to have: curiosity or good manners?"

    → Idee: Bezug auf Erziehungsstil
- Zur Messung von right wing-authoritarianism schlug Altemeyer (1996) unter anderem folgende Items vor, denen man zustimmen konnte oder die man ablehnen konnte:
    1. "Gays and Lesbians are just as healthy and moral as anybody else."
    2. "There is absolutely nothing wrong with nudist camps."
    3. "A woman's place should be wherever she wants to be. The days when women are submissive to their husbands and social conventions belong strictly to our past."
    4. "There is nothing wrong with premarital sexual intercourse."

    → Idee: Bezug auf die Rolle der Frau in der Gesellschaft/ Sexualmoral

➔ Konzepte sind oft umstritten und haben verschiedene Definitionen!

Weitere Unterscheidungen:

- **Dispositional**: Der Gegenstand/das Phänomen tritt nur unter bestimmten Bedingungen auf.
- **Komplexe Begriffe:** Begriffe die Interaktionen zusammenfassen. Bspw.: Wettbewerb

Deskription und Analyse

- Deskriptive Analyse: (= dimensionale Analyse)
  Bei einer deskriptiven Fragestellung geht es darum, einen empirisch noch kaum erforschten Untersuchungsgegenstand zuerst gedanklich, und dann auch begrifflich vorzustrukturieren.
  →Auslegung des Konzepts, was gehört dazu?
  z.B. Lebensqualität: Wohlstand, Freiheit, Sicherheit, Klima etc.

- Analytische Fragestellung: (= Theorietestende Fragestellung)
  Bereits existierende explizite Begriffe werden sprachlich präzisiert und der Objektbereich eingeengen (= definieren).

Elemente einer Theorie: Definitionen

Problem in der Wissenschaft: Oft werden abstrakte Begriffe verwendet, die nicht ohne Weiteres mit empirischen Sachverhalten in Verbindung gebracht werden können.

➔ Deshalb sollen die beiden Ebenen mit Definitionen verknüpft werden.

Beispiel: Max Webers Definition von «Macht»: „Macht bedeutet jede Chance, innerhalb einer sozialen Beziehung den eigenen Willen auch gegen Widerstreben durchzusetzen, gleichviel worauf diese Chance beruht."

Nominaldefinitionen

- Eine Nominaldefinition ist „die Festlegung der Bedeutung eines Begriffs durch einen oder mehrere bereits bekannte andere Begriffe".
  → tautologische Umformung.
- Der Zusammenhang zwischen Begriff und Gegenstand ist rein konventioneller Art. Das unterscheidet Nominaldefinitionen von Realdefinitionen. Deshalb kann eine Nominaldefinition **niemals falsch** sein, sondern bloss **zweckmässig oder nicht zweckmässig.** → Wie präzise sind sie?

- In der Regel sind Nominaldefinitionen präzise, aber unter Umständen haben sie eine geringe Vergleichbarkeit, weil sich etwas die Definition geändert hat.

- Beispiel: „Als Erwerbslose gelten …
  - Personen im Alter von 15-74, die in der Referenzwoche nicht erwerbstätig waren
  - und die in den vier vorausgegangenen Wochen aktiv eine Arbeit gesucht haben
  - und die für die Aufnahme einer Tätigkeit verfügbar waren."

Realdefinitionen

- Begriffe als Widerspiegelung des Wesens eines Gegenstandes.
- Können also **falsch oder richtig sein**.
- Beispiel: Parteien sollen nicht als a, b und c *gelten* (Nominaldefinition), sondern sie *sind* a, b und c.

➔ Realdefinitionen sind untaugliche für die empirische Sozialforschung, „da sie keine Kriterien angeben können, inwieweit das Wesen einer Sache durch die Definition erfasst wird" (Schnell, Hill, Esser 1999: 50).

Elemente einer Theorie: Untersuchungs- und Beobachtungseinheiten

- Definition: Untersuchungseinheiten sind diejenigen Objekte, über die wir empirische Aussagen machen wollen.
  (Dies ist formuliert wie eine Realdefinition, in empirische Untersuchung müssen ich sagen «seien» anstatt «sind».)
- Diese Objekte weisen in aller Regel unterschiedliche Merkmale auf. Diese stehen im Fokus sozialwissenschaftlicher Analysen. Objekte = Merkmalsträger.
- Beispiele für Einheiten:
  Einzelne Individuen (Umfrage), Gruppen von Individuen (Alterkohorten), Organisationen (Unternehmen), Länder/Kantone, Zeitpunkte (Jahresdurchschnitte)

- Oftmals liegt nur *eine* Untersuchungseinheit, aber meist in *mehreren Fällen* vor.
  Bsp.: Untersuchungseinheit= Schweizer Stimmberechtigter. Jedoch wird nicht nur *ein* Stimmberechtigter befragt, sondern *mehrere.*
- Manchmal gibt es auch mehrere Untersuchungseinheiten.
  Bsp.: Arbeitslosenrate in OECD-Staaten (Untersuchungseinheit 1) in den vergangenen 50 Jahren (Untersuchungseinheit zwei: einzelne Jahre).

- Definition: Beobachtungseinheiten sind/seien diejenigen Objekte, an der Messungen vorgenommen werden.
- Beobachtungseinheit kann = Untersuchungseinheit sein. Muss aber nicht!

Elemente einer Theorie: Variablen

- Definition: Variablen bezeichnen Eigenschaften (oder Merkmale), die mehrere Ausprägungen annehmen können (≠ Konstante).
  Bsp.: Geschlecht (des Individuums), Grösse (einer Organisation), Arbeitslosigkeitsrate (über eine bestimmte Zeit hinweg), Ausprägung des politischen Systems (demokratisch vs. Autoritär)

- Qualitativ: Wenn die Variable verbale Bezeichnungen oder Buchstaben aufweist.
  Bsp.: Augenfarbe, Konfession, Parteizugehörigkeit
- Quantitativ: Wenn die Variable mathematisch interpretierbare Zahlenwerte annimmt.
  Bsp.: Alter in Jahren, Anzahl Kinder, Haushaltseinkommen in CHF.

- Diskret: Merkmale, die eine endliche Zahl von Kategorien innerhalb einer Bandbreite aufweisen. Bsp.: Anzahl Kinder. 1 oder 2 Kinder. 1.5 Kinder nicht möglich.
- Kontinuierlich: Merkmale, die jeden beliebigen Wert innerhalb eines bestimmten Intervalls annehmen können. Bsp.: Körpergrösse

- Spezialfall: binäre/dichotome Variablen
  Eine Variabel mit exakt *zwei* Merkmalsausprägungen.
  In Sozialwissenschaften häufig als Abbildung des Eintretens bzw. Nichteintretens eines bestimmten Ereignisses.
  Oder: Abstimmungsentscheid: Ja/Nein. Partizipation an Wahlen: Teilnahme/Nichtteilnahme. Erfolg: Erfolg/Misserfolg.

- <u>Anforderungen:</u>
  Variablen müssen…
  - Disjunkt sein (= Kategorien/Merkmalsausprägungen dürfen sich nicht überschneiden)
  - Erschöpfend sein (= Variable muss alle möglichen Merkmalsausprägungen abdecken).

  Bsp.: Haushaltseinkommen in CHF.
  (1) 0-3'000 CHF; (2) 3-5'000 CHF; (3) 5-7'000 CHF; (4) 7-10'000 CHF
  Weder disjunkt noch erschöpfend.
  Welcher Kategorie werden Personen mit exakt 3'000 CHF zugewiesen? (z.B. bei (2) sollte stehen 3001-5000)
  Welcher Kategorie werden Personen mit exakt 12'000 CHF zugewiesen? (Es fehlt Kategorie „Über 10'000)

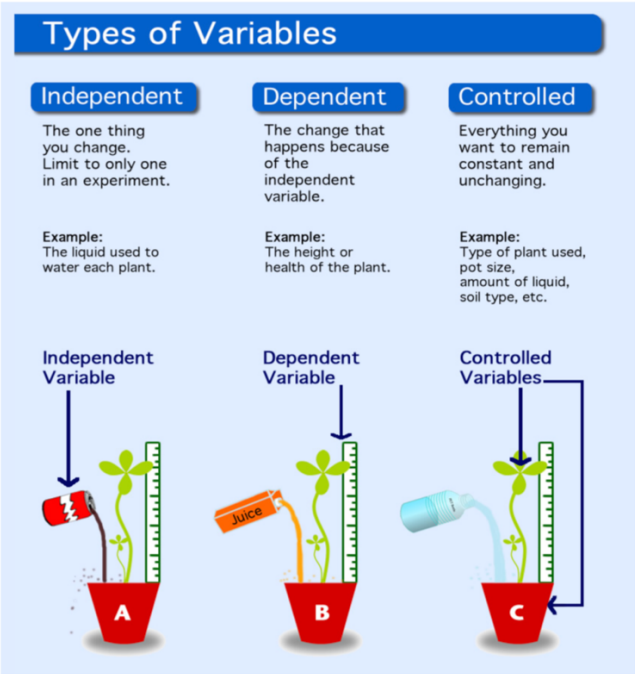
- <u>Individualmerkmale</u>: Variable enthält Informationen über einzelne Individuen
  Bsp.: das Alter der Person xy.
  → Individualdatensatz
- <u>Kollektivmerkmale</u>: Variable enthält Informationen über Kollektive („Aggregate")
  Bsp.: Durchschnittliches Alter in Gemeinde xy.
  → Aggregatdatensatz

Theorien enthalten Konzepte, Definitionen, Variablen und Hypothesen.

Sitzung 5: 18.10.19 – (Theoretische Formulierung des Forschungsproblems:) Hypothesen und Theorien

Elemente einer Theorie: Hypothese

- Definition: Vermutung über einen Zusammenhang zwischen mindestens zwei Sachverhalten/Variablen.
- Elemente einer Hypothese:
  - Unabhängige Variable = Die erklärende Variable oder Ursachenvariable.
  - Abhängige Variable = Die zu erklärende Variable oder Wirkungsvariable.



- Wenn-Dann-Hypothese
  Voraussetzung: beide Variablen haben eine dichotome/binäre Ausprägung.
  - Implikationsbeziehung: Wenn A auftritt, wird B erwartet.
    Wenn A hingegen nicht eintritt, dann kann sowohl B, als auch Nicht-B erfolgen. (Bsp.: «Männer (A) wählen SVP (B)», hier könnte aber auch eine Frau SVP wählen)
  - Äquivalenzbeziehung: Nur dann B, wenn auch A auftritt. A als hinreichende, sowie notwendige Bedingung. (Bsp.: Hier wählen Frauen nie die SVP)

- Je-Desto-Hypothesen
  Voraussetzung: Mehr als zwei Ausprägungen (= polytom) und Rangfolge der Merkmalsausprägungen möglich (= ordinales Skalenniveau)
- Beziehung zwischen den Variablen kann unterschiedlich Form aufweisen, zum Beispiel:
  - Linear (kontinuierlich steigend)
  - Exponentiell
  - Logarithmisch
  - Logistische Funktion (S-förmig)
  - Quadratische Funktion (U-förmig)

- Hypothesenstruktur
  Abhängige und unabhängige Variable können in unterschiedlicher Weise verknüpft sein:
  - Deterministisch (wenn x, dann immer y)
    vs. probabilistisch (wenn x, dann sehr wahrscheinlich y) (→ Wahrscheinlichkeitsaussagen).
  - Reversibel (wenn x, dann y und wenn y, dann x) (→ Wechselseitigkeit: z.B. Je höher das politische Interesse, desto eher nimmt man an Wahlen teil)
    vs. irreversibel (wenn x, dann y, aber wenn y, dann nicht x).
  - Koexistent (wenn x, dann auch y) (keine Aussagen über einen Kausalzusammenhang → Korrelation: zwei Phänomene treten häufig miteinander auf) vs. aufeinanderfolgend (wenn x, dann später y). → chronologisch: nur mit Experiment überprüfbar
  - Hinreichend (wenn x, dann y, ungeachtet alles anderen) vs. bedingt (wenn x, dann y, aber nur, wenn auch z).
  - Notwendig (wenn x, dann y, und nur dann y) vs. substituierbar (wenn x, dann y, aber wenn z, dann auch y).

- Hypothesen: Bezugsebenen (siehe auch Variablen)
  - Individualebenen/-daten: Merkmale von Individuen (z.B. Erwerbstätigkeit einer Person
  - Aggregatebene/-daten: Merkmale von Gruppen/ Mengen von Untersuchungseinheiten.

➔ Aggregatebene: Problem des ökologischen Fehlschlusses
Definition: Schluss von beobachtbaren Beziehungen zwischen Daten über das Kollektive auf Beziehungen zwischen Merkmalen der Elemente der Kollektive.
→ Also von Kollektiv/Aggregatebene auf Individualebene schliessen.
Bsp.: In einem Bundesstaat mit 50% Katholiken wählen 50% eine konservative Partei. Ökologischer Fehlschluss: Katholiken wählen konservative Parteien. Eigentlich müssen diese Zahlen auf der Individualebene aber eben nicht bedeuten, dass Katholiken konservativ Wählen. (Im Extremfall wählt garkeiner konservativ)

- Individualdatenanalyse: Das Ziel der Individualanalyse in der Politikwissenschaft ist, statistische Zusammenhänge zwischen individuellen Merkmalen aufzufinden.
  Vorteile: Analyse von Zusammenhängen von Variablen auf der Untersuchungsebene der einzelnen Individuen (Mikroebene).

- Kontexthypothese: Hypothesen, die beide Ebenen miteinander verknüpfen.
  Bsp.: Je höher die soziale Integration in einer Gruppe, desto geringer ist die Wahrscheinlichkeit, dass sich ein Mitglied dieser Gruppe abweichend verhält.
  → setzt entsprechende (sich auf beide Ebenen beziehende) Daten voraus und bedarf einer entsprechenden (Mehrebenen-) Analysetechnik.
  Varianz auf Kollektivebene nötig.

Abbildung 8: Die Colemansche Badewanne

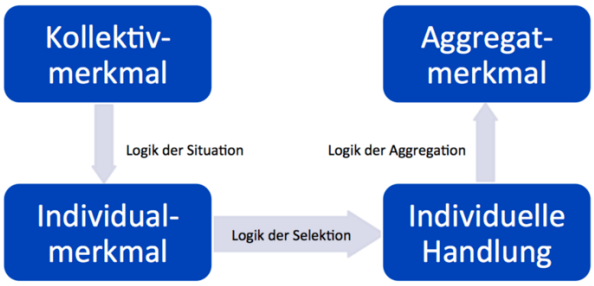
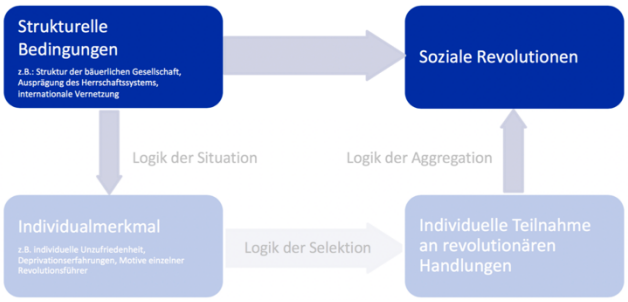


Abbildung 9: Skocpols (1979) strukturtheoretische Revolutionstheorie



Was ist eine Theorie? → Theorienbildung

Drei Definitionen:

- "Ein System logisch widerspruchsfreier Aussagen (Sätze, Hypothesen) über den jeweiligen Untersuchungsgegenstand mit den zugehörigen Definitionen der verwendeten Begriffe." (Kromrey/Roose/Strübing)
- Eine Theorie ist ein "System von Aussagen, das mehrere Hypothesen oder Gesetze umfasst." (Schnell/Hill/Esser)
- "Eine Theorie ist eine Menge von Gesetzen, wenn diese durch logische Ableitbarkeitsbeziehungen miteinander verbunden sind." (Opp)

➔ Ein System, das mindestens zwei Hypothesen beinhaltet + Definitionen der verwendeten Begriffe.

- Elemente/Bestandteile einer Theorie: (Siehe auch PST1)
  Axiome/Grundannahmen
  - Fundamentale (ontologische) Annahmen, die nicht hinterfragt werden. (z.B. wie im jeweiligen Gegenstandsbereich Wissen erzielbar ist)
    Oder Menschenbilder: «Die Individuen versuchen ausschliesslich ihren eigenen Nutzen (in Form von Macht, Einkommen etc.) zu maximieren.»

  Grundbegriffe/Konzepte (und ihre Definitionen)
  - «Eigennutz wird definiert als das Ziel, das zuoberst auf der individuellen Präferenzordnung steht.»

  Hypothesen
  - «Im Zweiparteiensystemen nähern sich die beiden Parteien sukzessiv dem Median-Wähler an.»

- Ziel von Theorien:
    - Wie bei Modellen. Es ist nicht das Ziel die Realität spiegelbildlich abzubilden, sondern einen Ausschnitt der Realität vereinfachend zu abstrahieren, um möglichst allgemeine «Gesetzte» zu erklären und wenn möglich Prognosen zu erstellen.

➔ Reduktion der sozialen Komplexität auf das Wesentliche.
    - Parsimonität (=Sparsamkeitsregel):
      Eine Theorie, die mit wenigen Aussagen auskommt ist einer Theorie mit einer höheren Zahl an Einschränkungen vorzuziehen (auch bei gleicher Erklärungsleistung).
      → Möglichst simpel, wenig Randbedingungen.

Teil 4: Die empirische Umsetzung des Forschungsproblems
Sitzung 6: 24.10.19 – Messung 1: Grundlagen

Operationalisierung
Von der Theorie zur Empirie
Problem: Theoretische Begriffe/Konzepte sind in aller Regel nicht direkt beobachtbar.

➔ Konzepte müssen deshalb mit **beobachtbaren Sachverhalten** verknüpft werden.
Dies erfolgt durch **Korrespondenzregeln**
Korrespondenzregeln bestehen in der Angabe von Messanweisungen, die sich auf **direkt beobachtbare** Sachverhalte beziehen. Diese Sachverhalte wiederum sind Ausprägungen bestimmter Merkmale. Diese (manifeste) Variablen nennen wir:
→ **Indikatoren.**

- Definition: Operationalisierung eines theoretischen Konstrukts besteht aus der Angabe einer Anweisung, wie Objekte, die der theoretische Begriff bezeichnet, beobachtbaren Sachverhalten zugeordnet werden können.
  → «Messbarmachung», legt fest, wie ein theoretisches Konstrukt messbar/beobachtbar gemacht werden soll.

- Indikatoren:
  Mit Indikatoren (manifeste Variablen) können latente Variablen gemessen werden.
  Hilfsmittel, die Informationen an/aufzeigen sollen. (z.B. Blaukrautsaft zur Messung von Säurestärke)
  Rechtfertigung der Zuordnung eines Indikators zu einem theoretischen Begriff:
    - Operationalistische Lösung:
      Operationale Definition des theoretischen Konstrukts («Intelligenz ist, was der Intelligenztest misst.»)
      → Was aus der Messung herauskommt = Konstrukt (keine Validität), Konzeptspezifikation umgehen. → Das Konzept ist stets unklar
    - Typologisch-induktive Lösung:
      Erhebung von Messdaten und Bildung von latenten Gruppen, welche anschliessend als Indikatoren dienen.
      → Schliesst von Empirie auf Theorie.
    - Kausal-analytischer Ansatz: (Häufigster Ansatz!)
      Indikatoren als Manifestationen der latenten Variablen. Aus Konzept x folgt Indikator y. Dazu wird eine Messhypothese (Hilfstheorie) formuliert.
      →Kausale Folge = Indikator. Reflexiv. An bestehende Forschung anknüpfend.

Auswahl von Indikatoren:

- Konzept des Indikatorenuniversums:
  Annahme eines homogenen Indikatorenuniversums, indem eine Vielzahl von Indikatoren existieren, die alle jeweils eine unabhängige Messung desselben Gegenstandes darstellen.
  → Unendliche Anzahl Indiaktoren, die mit einem Sachverhalt verbunden werden können.
- Prinzip der Austauschbarkeit der Indikatoren:
  Es spielt keine Rolle welchen Indikator man aus dem entsprechenden Universum für den jeweiligen Begriff auswählt. Wenn dieses Prinzip nicht erfüllt ist, sind de Ergebnisse der Untersuchung von der Wahl spezielle Indikatoren abhängig.
  → Repräsentative Stichprobe
- Konzept der multiplen Indikatoren:
  Mehrere Indikatoren zur Messung desselben theoretischen Konstrukts. Vorteil: Gleicht Messfehler eher aus und ermöglicht Vergleiche.

Messtheorie

- <u>Definition: Messung</u> ist ist „die Zuordnung von Zahlen zu Objekten nach bestimmten Regeln“ oder „Messung ist eine strukturtreue Abbildung“.
  → Statistische Auswertung.

- Den Messinput bilden die Objekte («empirisches Relativ») und deren Verhältnis zueinander.
  Bsp.: Vergleich zweier Personen bezüglich politisches Interesse.
  Messoutput sind Zahlen und deren Relationen (= «numerisches Relativ)
  → Strukturgetreue Abbildung im numerischen Relativ dem empirischen Relativ getreu.
- Es wird stets von eine Messhypothese ausgegangen, die ihrerseits (genauso wie die «Kerntheorie») richtig oder falsch sein kann.
- Ziel Messung: strukturgetreue Abbildung (= Morphismus) der Objekte durch die zugewiesenen Zahlen.

- <u>Morphismen:</u>
  Zwei Formen:
  - Isomorphismus: ein oder umkehrbar-eindeutige Abbildung

Schüler	Zuordnung	Mathenote
A	⟵ – ⟶	1
B	⟵ – ⟶	2
C	⟵ – ⟶	3
D	⟵ – ⟶	4

  - Homomorphismus: eindeutige Abbildung

A	Schüler	C
↓		↓
1	Deutschnote	2
↑		↑
B	Schüler	D

- <u>Repräsentationsproblem</u>
  Ist ein bestimmtes Objekt überhaupt messbar?
  Bedingungen unter denen sich das empirische durch ein numerisches Relativ repräsentieren lässt.
  Bsp.: Um eine Ordinalskala zu erhalten, muss etwa das Transitivitätsaxiom erfüllt sein.

- <u>Eindeutigkeitsproblem, Transformation:</u>
  Wie können Messwerte verändert werden (Transformation), ohne dass die in ihnen enthaltene Information verloren geht?
  <u>Regel</u>: Nur solche Transformationen sind zulässig, die das empirische Verhältnis zwischen den Werten nicht verändern. (d.h. die Strukturtreue zwischen empirischem und numerischem Relativ bleibt gewahrt.) z.B. bei Ordinalskala durch Multiplikation mit c>0.
  Bsp.: Beim Geschlecht (Nominalskala) dienen die zugeordneten Zahlen lediglich als Namen/Labels und könnten durch beliebige andere (ungleiche!) Werte ersetzt werden, weil ihre numerische Bedeutung nicht benutzt wird.

- <u>Bedeutsamkeitsproblem</u>
  Welche mathematischen Operationen mit Messwerten führen auch zu empirisch sinnvollen Aussagen?
  → praktisches Problem! Nicht Generierung von Messwerten sondern sinnvolle Weiterverarbeitung.
  Welche Rechenoperation (Addition, Division, Logarithmierung…) können durchgeführt werden, so dass daraus empirisch sinnvolle Schlüsse resultieren?
  Bsp.: Ist die Angabe eines arithmetischen (zahlen) Mittelwertes bei ordinalen Werten (Rangordnungsskalen wie Schulnoten, Einstufungsskalen) sinnvoll?
  Beispiel:
  Angenommen, die Variable X steht für das Geschlecht. Weiter gilt:

  $$X = \begin{cases} 0 & \text{if } ♂ \\ 1 & \text{if } ♀ \end{cases}$$

  Mathematisch gesprochen gilt:

  $$0 + 1 = 1$$

  Empirisch gesprochen gilt jedoch:

  $$♂ + ♀ \neq ♀$$

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

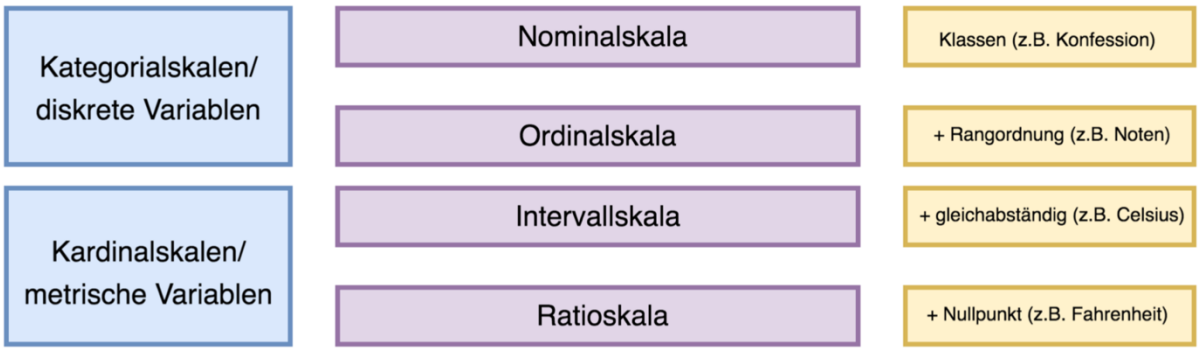
  <u>Regel:</u>
  Das Kriterium für die Bedeutsamkeit einer statistischen Aussage ist, dass sich deren Gültigkeit (Wahrheitswert) unter zulässigen Transformationen der Skalenwerte nicht ändert.

Warum sind diese Grundkenntnisse der Messtheorie wichtig?

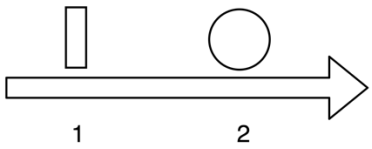
- Um beispielsweise keine sinnlosen Berechnungen durchzuführen (mehr dazu: deskriptive Statistik).
  Beispiel: Es macht keinen Sinn, den arithmetischen Mittelwert einer nominal skalierten Variablen zu errechnen. Denn die Zuweisung von Zahlen zu den

Ausprägungen der nominal skalierten Variablen ist, erstens, willkürlich und, zweitens (daraus folgend), haben die numerischen Abstände zwischen den Ausprägungen keine Aussagekraft.
Dementsprechend beschränken sich mathematisch-statistische Operationen für Nominaldaten auf die Analyse von Häufigkeitsverteilungen.

Skalenniveaus

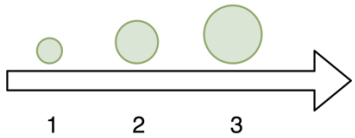


- <u>Nominalskala:</u>



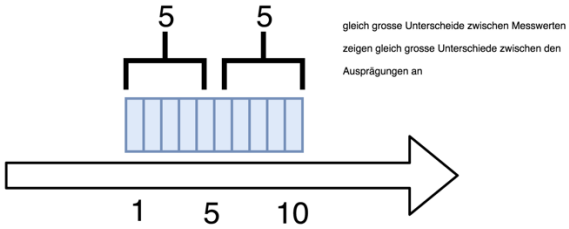
  - Relation: Merkmalsausprägungen folgen keiner natürlichen Reihenfolge, sind demnach beliebig (und gleichberechtigt, «Äquivalenzrelation»)
  - Zuweisung der Werte: Dient nur der Identifikation, für Aussagen der Gleichheit/Ungleichheit.
  - Zulässige Transformation: Ein-eindeutige Transformation.
    d.h. alle T., die gewährleisten, dass die gleichen Merkmalsausprägungen stets die gleichen Messwerte erhalten und dass unterschiedlichen Merkmalsausprägungen auch unterschiedliche Messwerte zugeordnet werden.
  - Statistische Auswertung: Häufigkeitsinformationen, Mittelwert = Modus (häufigster Wert).
  - Beispiel: Geschlecht, Beruf, Religionszugehörigkeit, Matrikelnummer.

- <u>Ordinalskala</u>



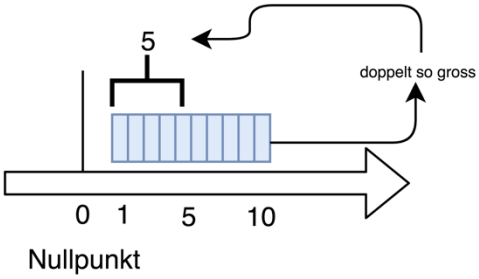
  - Relation: Merkmalsausprägungen folgen einer natürlichen Rangfolge, aber die Abstände sind nicht quantifizierbar. (Transitivitätsaxiom)
  - Zulässige Transformationen: alle rangerhaltenden d.h. monoton steigenden Transformationen. (z. B. Multiplikation)
  - Statistische Auswertung: Häufigkeitsinformationen, Mittelwert = … + Median (der Wert, der die untere von der oberen Hälfte der Messwerte trennt).
  - Beispiel: Noten, Windstärke, Einstufungsskalen («sehr einverstanden», «eher einverstanden» etc.)

- <u>Intervallskala</u>



  - Relation: Merkmalsausprägungen folgen einer natürlichen Rangfolge, die Abstände sind quantifizierbar. (D.h. Differenzen dazwischen sind gleich) Somit sind Aussagen möglich wie: Heute ist es 10 grad wärmer. Nicht aber Verhältnisaussagen: heute ist es doppelt so warm wie gestern.
  - Relation II: Kein natürlicher Nullpunkt!
  - Zulässige Transformationen: Alle linearen Transformationen.
  - Statistische Auswertung: … + arithemtischer Mittelwert.
  - Beispiel: Temperatur in Celsius (willkürlich gewählter Nullpunkt) oder Intelligenzquotient.

- <u>Ratioskala (Verhältnisskala)</u>



  - Relation: Merkmalsausprägungen folgen einer natürlichen Rangfolge, Abstände sind quantifizierbar und die Skala hat einen natürlichen Nullpunkt! →Verhältnisaussagen sind möglich!
  - Zulässige Transformationen: Alle Ähnlichkeitstransformationen, (z.B. Kilometer in Meter)
  - Statistische Auswertung: alle. Mittelwert = … + geometrisches Mittel.
  - Beispiel: Temperatur in Kelvin. Lebensalter.

Messniveaus: 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? Sind Ratingskalen intervall- oder ordinalskaliert?

Testtheorie
Worum geht es?
In welchem Zusammenhang steht das zu testende latente Merkmal mit dem in einem Test tatsächlich beobachtbaren Testverhalten (d.h. den Testantworten) ?
Treffen von Annahmen, wie aus den empirisch ermittelten Testergebnissen Aussagen über die latenten Merkmale der getesteten Personen gemacht werden können.
Frage der Anforderungen, denen ein Test genügen muss, um aufgrund eines Testergebnisses auf die tatsächliche Ausprägung des getesteten Merkmales schliessen zu können.

➔ Wird das theoretische Produkt, dass man prüfen möchte, überhaupt geprüft?
➔ (Interesse am wahren Wert)

Zwei Messmodelle:

- Klassische Testtheorie (KTT, *classical test theory, true score theory,* TST)
  - In der Forschung und Praxis bis heute dominierend.
  - Testergebnis entspricht direkt dem wahren Ausprägungsgrad des untersuchten Merkmals. Aber jede Messung oder jedes Testergebnis weist zusätzlich einen Messfehler auf.
- Probabilistische Testtheorie (PTT, *probabilistic test theory, latent trait theory,* Item-Response-Theorie (IRT))
  - Die Wahrscheinlichkeit einer Antwort auf ein einzelnes Item ist von Merkmalen des Items und latenten Merkmalen der Person abhängig.
  - Bsp.: Die Schulleistungstests der international vergleichenden PISA-Studien basieren auf der IRT.

Annahmen der klassischen Testtheorie:

- Voraussetzung: Beide Testtheorien lassen sich nur anwenden, wenn *mehrere* Items verwendet werden, die *alle dasselbe Konstrukt* messen und zu einem Test- bzw. Skalenwert zusammengefasst werden.
  Nicht anwendbar, wenn als Messinstrumente *Einzel*items bzw. *Einzel*indikatoren verwendet werden.
- Grundgedanken: Fünf Axiome (daraus sind anschliessend die drei Testgütekriterien Objektivität, Reliabilität und Valdidität ableitbar.)
  - Verknüpfungsaxiom
    - Jede Messung weist Messfehler auf. Deshalb setzt sich die Messung für eine Person m eines Tests i ($=x_{mi}$) setzt sich additiv zusammen aus „wahrem" Wert (τ_{mi}) und einem unsystematischen Messfehler (ϵ_{mi}):

$$x_{mi} = \tau_{mi} + \epsilon_{mi}$$

  - Existenzaxiom
    - Messfehler streuen um den wahren Wert, wobei der Erwartungswert des Messfehlers null beträgt:

$$\mathrm{E}(\epsilon_{mi}) = 0$$

    - Zum Beispiel: willkürliches, zufällig erfolgendes falsches Ablesen der Zeit.[1]
    - Weil $\mathbb{E}(\epsilon_{mi}) = 0$, gilt daher:

$$\mathrm{E}(x_{mi}) = \tau_{mi}$$

    Gegenbeispiel: Systematische Fehler = Frageeffekte bei Befragungen (z.B. soziale Erwünschtheit, beeinträchtigt die Validität)

- Unabhängigkeitsaxiom
  - Die Höhe des Messfehlers ist unabhängig vom Ausprägungsgrad des getesteten Merkmals, d.h. wahrer Wert und Fehlerwert sind unkorreliert.[2]

$$\mathrm{Corr}(\tau_{mi}, \epsilon_{mi}) = 0$$

  Gegenbeispiel: Je höher das Einkommen, desto eher wird bei der Angabe des Einkommens untertrieben.

- Unabhängigkeit zwischen Messfehler der Items
  Die Höhe des Messfehlers zwischen zwei Messwertreihen (oder zwei Items) i und j ist nicht korreliert.[3]

$$\mathrm{Corr}(\epsilon_{mi}, \epsilon_{mj}) = 0$$

  Gegenbeispiel: Messfehler bei denen Motive für den Entscheid und den Inhalt der Vorlage korrelieren wegen allgemeiner Erinnerungsschwäche.

- Unabhängigkeit der Messfehler zwischen Personen
  Die Messfehler zweier beliebiger Personen m und v sind jeweils unkorreliert, wenn sie dasselbe Item i wiederholt bearbeiten.

$$\mathrm{Corr}(\epsilon_{mi}, \epsilon_{vi}) = 0$$

Anhang

Empirisches Relativ

- Ein empirisches Relativ ist eine beliebige Menge an Objekten, (z.B. Personen, Ereignisse). In jedem empirischen Relativ gibt es eine oder mehrere Relationen, die die Beziehung zwischen den Objekten charakterisieren (z.B. gleich oder ungleich, grösser oder kleiner als; etc.).

Eindeutigkeitsproblem

Was sind zulässige Transformationen?

- Beispiel: Es liegt eine Ordinalskala für das Merkmal `Grad des politischen Interesses` vor (Merkmalsausprägungen i = tiefes Interesse, j= eher tiefes Interesse, k= eher hohes Interesse, l= sehr hohes Interesse) mit der folgenden Abbildungsfunktion: $\phi(i) = 1$, $\phi(j) = 2$, $\phi(k) = 3$ und $\phi(l) = 4$.
- Diese Abbildungsfunktion ist zunächst strukturerhaltend, denn je höher das Interesse, desto höher der zugewiesene Wert.
- Wenn wir nun zu diesen Werten 1-4 eine konstante Zahl c *addieren*, so bleibt der Informationsgehalt erhalten: $4 + c$ ist nach wie vor höher als $3 + c$.
- Wenn wir die Werte 1-4 mit einer konstante Zahl c *multiplizieren* (wobei $c > 0$), so bleibt der Informationsgehalt erhalten: $4 \cdot c$ ist nach wie vor höher als $3 \cdot c$. Zwar ist der Abstand zwischen $4 \cdot c$ und $3 \cdot c$ durch die Multiplikation grösser geworden. Aber bei Ordinalskalen sind die Abstände zwischen den Werten irrelvant.
- Generell: Alle Transformationen, welche die Grösser-Kleiner-Relationen zwischen den ursprünglichen Rangzahlen nicht verändert, sind zulässig. Transformationen mit dieser Eigenschaft bezeichnet man allgemein als monotone Transformationen.

Sitzung 7: 01.11.19 – Messung II: Skalen und Indizes

Indizes (Sg. Index)

- Definition: Eine Variable, die sich aus mehreren Einzelindikationen zusammensetzt.
- Indexbildung empfiehlt sich bei multidimensionalen Konstrukten. Teilindikatoren = verschiedene Dimensionen des Konstrukts.
  Bsp.: Demokratiebarometer, Human Development Index (HDI = $\sqrt[3]{LEI \cdot BI \cdot EI}$ → drei Faktoren)

Zwei grundlegende Fragen:

- Festlegung der Dimension
  Welche Dimensionen sollen in den Index einfliessen?
  Dimensionen bilden eine Merkmalsraum, in dem jedes Objekt der Forschung eine spezifische Position belegt. Herleitung erfolgt theoretisch.
  Bei diskreten Variablen (mit wenig Ausprägungen): Typologietabelle
  Bei kontinuierlichen Variablen: Vektorraum.

- Kombination der Dimension
  Wie sollen die Dimensionen kombiniert werden?
  Generell: Ein Index fasst viele Kombinationen des Merkmalsraumes zu einigen wenigen Kombinationen zusammen.
  Indexbildung = Reduktion des Merkmalsraum
  Dafür spricht:
  - Gewisse Kombinationen kommen selten bis nie vor.
  - Verschiedene Kombinationen könne gleichgesetzt werden.
  - Verschiedene Kombinationen könne theoretisch zusammengefasst werden.

  Drei Grundlegende Varianten:
  - ungewichtet-additiv Indizes
    Additive Indizes werden durch die Addition der Werte der Einzelindikatoren berechnet: → Summe der Einzelindikatoren.
    $Index_k = Ind_1 + Ind_2 + ... + Ind_n$.
    Alle Einzelindikatoren müssen denselben Wertebereich haben (**gleich skaliert/gewichtet**), 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.

  - gewichtet-additiv Indizes
    Wenn ein bestimmter Indikator stärker in den Indexwert eingehen soll als die anderen Indikatoren, muss ein gewichteter additiver Index gebildet werden.
    Erlaubt die unterschiedliche Bedeutung einzelner Merkmalsdimensionen zu berücksichtigen (z.B.: Eine Gesamtnote aus unterschiedlich wichtigen Teilprüfungsleistungen).
    Partiell substituierende Wirkung:
    $Index_k = a_1 \cdot Ind_1 + a_2 \cdot Ind_2 + ... + a_n \cdot Ind_n$.
    Die Gewichte können entweder theoretisch oder statistisch (z.B. durch Faktorenanalyse) bestimmt werden. Auch möglich: eine Kombination von beidem (inhaltliche Überlegungen empirisch begründen, vgl. Verbraucherpreisindex).

- Multiplikative Indizes
  Multiplikation der Indikatorenwerte: $Index_k = Ind_1 \cdot Ind_2 \cdot ... \cdot Ind_n$.
  Verstärkende Wirkung bei Vorliegen mehrerer positiver Indikatorwerte.
  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).
  Beispiel: HDI (nicht gleich skaliert, geometrischer ($\sqrt[3]{}$) Mittelwert)
  → Keine Substituierung (Ersetzten) mehr, sondern die Faktoren können sich sogar gegenseitig verstärken.

Skalen (Sg. Skala)

- Definition: Messung mit mehreren Items eines Eindimensionalen theoretischen Konstrukts.
  → **Definition Item:** Fragen, Statements oder Aussagen, mit denen die Befragten einverstanden oder nicht einverstanden sein können.

- Hauptvorteile:
  - Skalen werden in der Regel dann gebildet, wenn ein einzelner Indikator die latente Variable nicht mit ausreichender Genauigkeit misst.
    Dadurch sollte die Validität gesteigert werden.
    → Messfehler können eher ausgeglichen werden.
- Hauptnachteil:
  - Grosser Zeitaufwand für Befragte und Forschende. (praktisch)

- Auswahl der Items
  Es gibt keine formalen Regeln zur Formulierung und Entdeckung von Items.
  In der Praxis aber oft Wiederverwendung bereits existierender Items.
  In der Regel empirisch Überprüfung der Items, hierzu einige (ausgewählte) Kriterien zur Formulierung von Items:
  - 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 (nicht) zustimmt.
    Bsp.: «Die Armeeausgaben sind zu hoch Der Staat sollte stattdessen mehr für Soziales ausgeben.»
  - Keine *valence issues:* z.B. «Die Arbeitslosigkeit soll bekämpft werden.»
    Diesem Statement werden vermutlich die meisten Befragten zustimmen, es besitzt demnach keine Trennschärfe.
  - Positive und negative Items

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

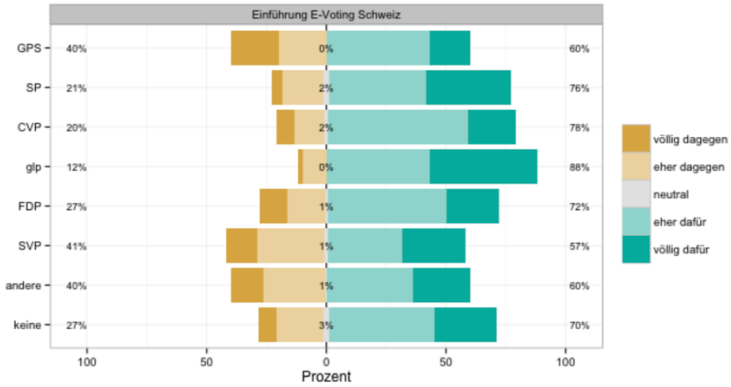
<u>Skalierungsverfahren</u>

Gebräuchliche (es gibt noch weitere, weniger gebräuchliche):

- Likert-Skala
- Guttmann-Skala
- Thurstone-Skala (nicht behandelt)
- Rasch-Skala (nicht behandelt)

Generell: Die verschiedenen Skalierungsverfahren unterscheiden sich hinsichtlich der Annahme über das Antwortverhalten, sowie die Verrechnung der Antworten zu den entsprechenden Skalenwerten.

- Likert-Skala
  - Mit Abstand am häufigsten verwendetes Skalierungsverfahren
  - Voraussetzung: monotone Itemcharakteristik
    (mit Anstieg latente Variabel auch Anstieg Wahrscheinlichkeit Zustimmung)
  - Grundprinzip: Summiertes Rating (oder arithmetischer Mittelwert)
  - Item-Analyse bzw. -Selektion: Nicht-Berücksichtigung ungeeigneter Items mittels bestimmter Verfahren.
  - In Forschungspraxis wird der Skalenbildungsprozess oft abgekürzt



- Guttman-Skala
  - Voraussetzung: Items, die in Bezug auf die interessierende Variabel unterschiedlich radikal formuliert sind.
    Zudem: dichotome Itemcharakteristika (erst mit Überschreitung eines bestimmten Schwellenwertes der latenten Variabel erfolgt Zustimmung)
  - Grundprinzip: Skalenwert entspricht der Nummer des Items, das zuletzt akzeptiert wurde.
  - Problem: Konstruktion von Items, die empirisch nachweisbar eine schrittweise stärkere Merkmalsausprägung widerspiegeln. Folge: Perfekte Guttmann-Skalen gibt es in der Realität kaum.
  - Item-Analyse: Berechnung des Reproduzierbarkeits-Koeffizenten.

TABLE 4-4

Guttman Scale of Attitudes toward Abortion

Schrittweise höherer Schwierigkeitsgrad der Frage. Aufeinander aufbauend. Wenn nr 3 zugestimmt wird, wurde wahrscheinlich auch 1 und 2 zugestummen.

Respon-dent	Life of Woman	Incest or Rape	Un-healthy Fetus	Father	Afford	Any-time	No. of Agree Answers	Revised Scale Score
1	A	A	A	A	A	A	6	5
2	A	A	A	D	A	D	4	4
3	A	A	A	D	D	D	3	3
4	A	A	D	A	D	D	3	2
5	A	D	D	A	D	D	2	1
6	D	A	D	D	D	D	1	0

Note: Hypothetical data. A = Agree, D = Disagree.

Gemeinsamkeiten/Unterschiede zwischen Indizes und Skalen

- Gemeinsamkeit: Beide sind Zusammenfassungen von Einzelindikatoren zur Messung einer latenten Variabel. (Zusammengesetzte Messungen)

- Indizes werden mit mehrdimensionalen Konzepten gebildet, bei welchen aber eine zugrundeliegende, latente Variabel vermutet wird.

- Skalen werden dann gebildet, wenn ein einzelner Indikator die latente Variabel nicht mit ausreichender Genauigkeit misst. Skalen messen zudem nur *eine* Dimension.

- Konstruktion eines Index erfolgt anhand von theoretischen Überlegungen.
  Skalen können anhand von empirischen Daten überprüft werden.

Sitzung 8: 08.11.19 – Messung 3: Gütekriterien der Messung

Gütekriterien der Messung

<u>Objektivität</u>: Ist Messung unabhängig von der Person, die sie durchführt?

- Ergebnisse sollen unabhängig von der Person sein, die sie durchführt.
- 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 objektiver Messung 1 oder zumindest nahe 1 betragen muss (z.B. Korrektur der Prüfungsergebnisse)

- Zu unterscheiden sind Durchführungsobjektivität und Auswertungsobjektivität:
  - DO: Verzerrungen, die durch die messende Person ausgelöst werden: z.B. Befragungseffekte. Deshalb: standardisierte Befragungen
  - AO: Verzerrungen, die bei der Auswertung entstehen (z.B. qualitative Bewertung eines Aufsatzes)

<u>Reliabilität</u>: Ist unsere Messung zuverlässig/reproduzierbar?

- Reliabilität ist ein Mass für die Reproduzierbarkeit von Messergebnissen
  Gibt die Präzision eines Messinstrumentes an. Dabei ist die Reliabilität umso höher, je geringer die Messfehler schwanken.

Reliabilität gibt demnach die Präzision eines Messinstruments an. Dabei ist die Reliabilität umso höher, je geringer die Messfehler (ϵ) schwanken.

Aufbauend auf den Annahmen der KTT (siehe vorhergehende Lektion) wissen wir:

$$\sigma_x^2 = \sigma_\tau^2 + \sigma_\epsilon^2$$

Demnach lässt sich Reliabilität als Anteil der wahren Varianz an der beobachteten Varianz des Testwertes σ_x^2 definieren:[1]

$$\rho_{xx'} = \frac{\sigma_\tau^2}{\sigma_x^2} = 1 - \frac{\sigma_\epsilon^2}{\sigma_x^2}$$

Wir können *systematische* Messfehler ausschliessen, sofern die Annahmen der KTT zutreffen

- Problem: Für die Berechnung eines Reliabilitätskoeffizienten brauchen wir eine Schätzung für die unbekannte wahre Varianz.
  Vier Methoden der Schätzung der Reliabilität:
  - Paralleltest
    Messung mit zwei vergleichbaren Messinstrumenten (am selben Messobjekt)
    Bsp.: Zeitmessung mit analoger und digitaler Uhr.
    Problem: Wirklich parallele Tests zu finden.
    $$\rho_{xx'} = \frac{\mathrm{COV}(x_A, x_B)}{\sigma_x^2}$$ (Korrelation Messinstrument A und B)
  - Test-Retest-Methode
    Wiederholte Messung (zu mind. *zwei* Testzeitpunkten) mit *demselben* Instrument und Objekt.
    Voraussetzung: Zeitliche Stabilität des Messobjekts und Paneldesign.
    Problem: Meist Unterschätzung aber auch Überschätzung (Erinnerungseffekt) möglich.
    $$\rho_{xx'} = \frac{\mathrm{COV}(x_{t1}, x_{t2})}{\sigma_x^2}$$
  - Testhalbierungs-Methode
    Aufteilung der Items in zwei Hälften (x1, x2), wobei jede Hälfte einen Paralleltest zur anderen Hälfte darstellt.
    $$\rho_{xx'} = \frac{\mathrm{COV}(x_{x1}, x'_{x2})}{\sigma_x^2}$$
    Unterschätzung der Reliabilität wegen halber Testlänge. Korrektur durch Spearman-Brown-Prohecy-Formula. Wobei I = Anzahl Items, k = Faktor, um den der Test verlängert wurde und Rel = pxx' des älteren Tests:
    $$\mathrm{Rel}(k \cdot I) = \frac{k \cdot \mathrm{Rel}}{1 + (k-1) \cdot \mathrm{Rel}}$$
  - Intern Konsistenz-Methode
    Erweiterung der Testhalbierungsmethode: Mittelwert aller möglichen Testvarianten/ split-half-Koeffizienten.

Am gebräuchlichsten ist Cronbachs Alpha (α)[4] (wobei N die Anzahl der Items und $\bar{r}$ die durchschnittliche Korrelation zwischen den Items bezeichnet):

$$\alpha_{st} = \frac{N \cdot \bar{r}}{1 + (N-1) \cdot \bar{r}},$$

<u>Validität:</u> Messen wir das, was wir messen wollen?

- Validität gibt an, ob das Instrument das theoretische Konstrukt misst, das es messe soll
- Wichtigstes Testgütekriterium!
- Reliabilität und Objektivität sind notwendige, aber nicht hinreichende Voraussetzungen für Validität. (Messung kann nicht valid sein, wenn sie nicht auch reliabel ist. Jedoch kann sie reliabel sein, aber nicht valid.)

- Achtung: Validität hier als Gütekriterium der Messung. *Nicht* dasselbe wie *interne und externe Validität.* Letzte sind Gütekriterien des Untersuchungsdesigns bzw. von wissenschaftlichen Aussagen über Studienergebnisse.

Formen der Validität:
Wird eine operationale Definition des theoretischen Konstrukts genommen («Intelligenz ist genau das, was der Intelligenztest misst») ist Validität zwar stets gegeben, aber nach woe vor unklar wofür das Konstrukt steht.

Vier Formen:
- Augenscheinvalidität: Selbstevidente Messung (die Fragestellung kann nur valide untersucht werden)

- Inhaltsvalidität (*content validity)*:
  Ist gegeben, wenn die Items das Zielkonstrukt in all ihren inhaltlichen Bedeutungsdimensionen vollständig und sinngemäss reflektieren.
  Erfolgt meist theoretisch-argumentativ und nicht empirisch.
  Bsp.: Der Intelligenztest im Sinne von Problemlösungsfähigkeit misst nicht bloss die Rechtfertigung, sondern auch die Fähigkeit des räumlichen Denkens, Artikulationsfähigkeit, Logik etc.

- Kriteriumsvalidität (*criterion validity*):
  Bezieht sich auf ein bestimmtes Aussenkriterium (in der Regel ein vom eigenen Messinstrument unabhängiges, aber bewährtes Messinstrument (*Benchmark test*)) und misst die Korrelation mit diesem Aussenkriterium.
  Hierbei muss ein reliables und valides Aussenkriterium vorliegen. Als Validitätsmass werden Korrelationseffizienten verwendet.
  Verschiedene Formen der Kriteriumsvalidität (abhängig vom Zeitpunkt der Erfassung des Aussenkriteriums):
  - Übereinstimmungsvalidität
    Der Kriteriumswert wird zum gleichen Zeitpunkt wie der zu validierende Testwert erhoben.
    Bsp.: Umweltbewusstsein und Mitgliedschaft bei einer Umweltorganisation

Oder Links-Rechts-Einstufung und Wahlverhalten

- Vorhersagevalidität
  Kriteriumsdaten werden zeitlich später erhoben.
  → Prognosekraft des Messinstrumentes wird gemessen.
  Bsp.: Akademischer Erfolg messen mit Prüfungsnote des Methodenkurses, verglichen mit Abschlussnote. Hohe Übereinstimmung = Hohe prognostische Validität.
  Problem: Selten ist eine solche Kriteriumsvariable bekannt und wenn ja, wozu dann eine neue Messung?

- Konstruktvalidität
  Gibt an, wie genau ein Messinstrument ein anderes, verwandtes Konstrukt misst (oder daraus ableitbare Hypothesen bestätigt) und nicht auch weiter, damit verwandte bzw. gegensätzliche Konstrukte.
  Unterschied Kriteriumsvalidität:
  Es werden aus dem Zielkonstrukt weitere verwandte Konstrukte oder empirisch erhärtete Hypothesen ab und überprüft, ob die eigene Messvariante damit korreliert oder nicht.
  Verschiedene Formen der Konstruktvalidität:
  - Konvergente Validität
    Ein Test, der belegen soll, dass Konstrukte, die mit dem eigenen Konstrukt nahe verwandt sind, auch empirisch stark mit diesem korrelieren.
    Bsp.: politisches Interesse, Wissen, Involvierung.

  - Diskriminante Validität (Gegensatz!)
    Ein Test, der belegen soll, dass Konstrukte, die mit dem eigenen Konstrukt nichts oder nur wenig zu tun haben, damit auch nicht korrelieren.
    Bsp.: Liberal/konservativ vs. Links/rechts
    (Überprüfung mit Multitrait-Multimethod-Matrizen)

<u>Zusammenfassung</u>

- 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.

- Konstrukt = (physische oder mentale) Stärke
- Indikator der SI ist die Platzierung beim Ironman.
- Reliabilität: Würde bei wiederholter Messung dasselbe Resultat erfolgen? Was, wenn Ryf bei wiederholter Messung Zweite anstatt Erste wird (zeitliche Stabilität, etc.)?
- Validität: Misst der Ironman-Wettbewerb physische Kraft oder etwas anderes? Was ist mit Stärke genau gemeint (Definition des Konzepts)?

Beispiel:

Varianz: Streuung einer Variable.
Reliabilitätstest können nur dann durchgeführt werden, wenn wir ein eindimensionales Konstrukt messen. Bei Mehrdimensionalen Konstrukten nicht wirklich möglich.
Für Test müssen sich Objekt und Messung sehr ähneln!!
Formeln müssen wahrscheinlich nicht gekonnt werden, sonst stellt er noch Formelsammlung online. Aber Dinge wie bspw. Cronbachs Alpha müssen wir können!

Teil 5: Untersuchungsformen
Sitzung 9&10: 15/22.11.19 – Kausalanalyse und experimentelle und Nicht-experimentelle Designs
Was ist ein Untersuchungsdesign?

- Das Forschungs- oder Untersuchungsdesign legt fest, wann, wo, wie oft und auf welche Art die Messung erfolgen soll.
- Ziel (Messvorschrift): ein möglichst genauer Test der Hypothese ermöglichen Hierbei ist vor allem wichtig, dass alternative Erklärungen ausgeschlossen werden können.
- Das Design ist abhängig von verschiedenen Kriterien, z.B. Erkenntnisinteresse der Untersuchung.

Tabelle 1: Neun Klassifikationskriterien für Untersuchungsdesigns
Heute vor allem die letzten 5 Punkte

Kennzeichen des Untersuchungsdesigns	Varianten von Untersuchungsdesigns
Wissenschaftstheoretischer Ansatz der Studie Wähle ich einen Quantitativen oder Qualitativen Ansatz	▪ Quantitative Studie ▪ Qualitative Studie ▪ Mixed-Methods-Studie
Erkenntnisziel der Studie	▪ Grundlagenwissenschaftliche Studie ▪ Anwendungswissenschaftliche Studie
Gegenstand der Studie	▪ Empirische Studie (Original bzw. Replikation) ▪ Methodenstudie ▪ Theoriestudie
Datengrundlage bei empirischen Studien Werden selber Daten erhoben oder verwenden sie Daten, die bereits erhoben worden sind	▪ Primäranalyse ▪ Sekundäranalyse Anaylse von Daten, die von jemand anderem erhoben worden sind ▪ Metaanalyse
Erkenntnisinteresse bei empirischen Studien	▪ Explorative (gegenstandsbeschreibende/theoriebildende) Studie ▪ Deskriptive (populationsbeschreibende) Studie ▪ Explanative (hypothesenprüfende) Studie „Königsdisziplin“
Art der Untersuchungsgruppen (Heute im Detail)	▪ Experimentelle Studie bzw. randomisierte kontrollierte Studie (RCT) ▪ Quasi-experimentelle Studie bzw. nicht-randomisierte kontrollierte Studie ▪ Nicht-experimentelle Studie z.B. auf Befragungen basierende Studien
Untersuchungsort bei empirischen Studien	▪ Laborstudie ▪ Feldstudie
Anzahl der Untersuchungszeitpunkte bei empirischen Studien	▪ (Quasi-)Experimentelle Studien mit und ohne Messwiederholungen ▪ Nicht-experimentelle Studien mit und ohne Messwiederholungen ▪ Querschnittstudie ▪ Trendstudie ▪ Längsschnittstudie
Anzahl der Untersuchungsobjekte bei empirischen Studien	▪ Gruppenstudie ▪ Einzelfallstudie

Erkenntnisinteressen sozialwissenschaftlicher Forschung:

- Deskription: Beschreiben
- **Explanativ: Erklären (und Vorhersagen) → Nachweis von Kausalitäten**
- Verstehen
- Bewerten (Evaluieren)

<u>Wie sind kausale Wirkungsmechanismen überprüfbar?</u>

- Assoziation: Zwischen x und y besteht ein statistischer Zusammenhang
- Zeitliche Abfolge: Ursache (x) geht der Wirkung (y) zeitlich voraus.
- Kontrolle/Isolation: Ausschaltung von Störfaktoren oder bewusste Manipulation des Ursachenfaktors (kontrafaktische Aussagen). Bsp.: Dass ich selbst die Gewichtung setze und nicht der Befragte. An bestimmte Gruppe Placebo, nicht Selbstmedikamentation.

Kontrolle der Untersuchungssituation

- = Ausschluss (bzw. Konstanthaltung) von Störfaktoren und Manipulation der Ursachenvariabel. (Wenn Probanden zwischen den Untersuchungseinheiten bspw. nachhause gehen sind das Faktoren, die ich nicht kontrollieren kann)
- Störfaktoren = Faktoren, die neben dem eigentlichen (von Forscherin postulierten) Wirkungszusammenhang Einfluss auf die abhängige Variabel haben könn(t)en («Drittvariabel»).

- Hierbei gibt es zwei Gruppen von Störfaktoren:
  - **Interne Validität** = vollständige Kontrolle der Untersuchungssituation. Intern Valide ist ein Ergebnis, welches nur aufgrund meiner Untersuchung geschehen ist, ohne andere Faktoren!
    Bsp.: Gesundungszustand nur aufgrund meiner Behandlung.
  - **Externe Validität** = Möglichkeit der Generalisierung der Resultate über die speziell experimentelle Situation hinaus auf andere (lebensweltlichere) Kontexte. → Sind die Ergebnisse des (künstlichen) Experiments auf die reale Situation übertragbar?

➔ Graduelles Konzept! Es gibt natürlich auch Zwischenstufen.
➔ Achtung: Unterschied von Validität, die die Messung betrifft (S8) und hier interner o. externer Validität, die das Ergebnis betreffen!

- Störfaktoren interne Validität:
  - Zwischenzeitliches Geschehen: Ereignisse zwischen erster und zweiter Messung, die neben dem eigentlichen Stimulus beeinflussen könn(t)en.
    → Betrifft vor allem Langzeitstudien.
  - Reifungsprozesse: (Bei Individuen) Intrapersonale Prozesse, die nicht durch den Stimulus in Gang gesetzt werden, sondern «natürlich» erfolgen.
    → z.B. der natürliche Gesundungsprozess
    Mess- oder Testeffekte: «Lerneffekte» bei den Probanden durch wiederholte Messung. Bsp.: Lerneffekte bei kognitiven Tests wie dem IQ-Test.
    → Messwert ≠ "true score": Verzerrung des wahren Verhaltens, an dem man interessiert ist
    → wenn die Probanden wissen, dass sie untersucht werden: Verhalten verändert sich deswegen
  - Hilfsmitteleffekte: Effekte des Messinstrumentariums, Verzerrung durch unterschiedliche Verwendung von Messinstrumenten.
  - Auswahl- und Ausfallverzerrungen: Verzerrung der Ergebnisse dadurch, dass sich Experimental- und Kontrollgruppen nicht nur hinsichtlich des Stimulus, sondern auch im Hinblick auf andere Merkmale unterscheiden.
    Eine asymmetrische Gruppenzusammensetzung kann Folge von Auswahlprozessen wie z.B. Selbstselektion bei Onlineumfragen sein oder von Ausfällen wie Verweigerung bei Telefonumfragen.
- Störfaktoren externe Validität
  - Messreaktivität: Effektkumulierung durch wiederholte Messung. Bspw. Wirkungen die nur aufgrund der Testsituation entstehen.
  - Situationsreaktivität: Alltagssituationen können im Labor nicht vollumfänglich simuliert werden. Hinzu kommt, dass Versuchspersonen den Zweck der Untersuchung zu ergründen versuchen und sich den vermuteten Erwartungen des Versuchsleiters entsprechend verhalten könn(t)en.
    → Laborergebnisse sind nie komplett übertragbar auf Alltagssituationen.

Kontrolltechniken

Die beiden Gruppen müssen sich gleichen!

Zuweisung der Untersuchungseinheiten zu Experimental- oder Kontrollgruppe: Randomisierung oder Matching

Vorher-Messung

Setzen des Stimulus/ Treatments (z.B. Medikament für Experimentalgruppe, Placebo für Kontrollgruppe)

Nachher-Messung

- Elimination: Ausschaltung aller potentiellen Störfaktoren bei der Durchführung des Experiments. Im Prinzip nur bei Laborexperimenten denkbar.
  Welches sind die potentiellen Störfaktoren?
  Bsp.: Wahlverhalten beeinflusst durch: Parteizugehörigkeit → ja; Augenfarbe → nein (dieses Merkmal hat keinen Einfluss auf das politische Verhalten)

- Konstanthaltung: Weil Elimination oft nicht möglich ist, wird versucht alle möglichen Störgrössen konstant zu halten, indem die Experimentsituation für alle möglichst gleich gestaltet wird.

- Randomisierung: Zufallzuweisung der Untersuchungseinheiten zur Experimental- und Kontrollgruppe. Damit sollen systematische Unterschiede in der Gruppenzusammensetzung verhindert werden.

- Matching: Die Zuweisung von «gleichen» Personen zu je einer Gruppe (d.h. zu Experimental bzw. Kontrollgruppe). Dadurch soll gewährleistet werden, dass sich die jeweiligen Gruppen nicht hinsichtlich weiterer möglicher Störgrössen unterscheiden. Jeh mehr potentielle Störfaktoren wichtig sind, desto schwieriger wird es, sich entsprechende Gruppen zu finden.
  → die Gruppen sollten Spiegelbilder voneinander sein ≠ jede einzelne Person von einer anderen

Tabelle 3: Zuweisung zu Experimental- und Kontrollgruppe durch parallelisierte Gruppen

Störvariable	Ausprägung	Gesamt	Exper.gruppe	Kontrollgruppe
Geschlecht	♂	110	55	55
	♀	90	45	45
Alter	18-35	40	20	20
	36-60	110	55	55
	ü60	50	25	25
Bildung	Berufsschule	30	15	15
	Matur	90	45	45
	Uni	80	40	40

Untersuchungsdesigns
Die Kontrolle von Störfaktoren erfolgt durch Varianzkontrolle.
Varianzkontrolle wiederum wird gewährleistet, dadurch dass die Untersuchungseinheiten einer der beiden Gruppen (Versuchsgruppe oder Kontrollgruppe) zugewiesen wurde.
Die Zuweisung kann entweder **vor** (**Ex-ante-Anordnungen**) der Behandlung bzw. Datenerhebung erfolgen oder **danach** (**Ex-post-facto-Anordnung**).

Ex-ante-Anordnungen: experimentelle Designs
- Nur Experimente erlauben eine systematische Kontrolle von Störfaktoren.
- Die Zuweisung der Versuchspersonen erfolgt durch den Forscher/die Forscherin (und nicht durch Selbstselektion oder -rekrutierung).
- Die Versuchsbedingungen werden durch eine der genannten Kontrolltechniken (siehe oben) 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 werden nicht gleichzeitig, sondern aufeinanderfolgend gemessen.
- Vorteile:
    - Generell: Hohe interne Validität. (Kontrolle Untersuchungssituation)
    - Überprüfung kausaler Zusammenhänge möglich.
    - Randomisierung neutralisiert Drittvariablen.
- Nachteile:
    - Eher geringe externe Validität. (Generalisierungsmöglichkeit)
    - Reaktivität
    - Hoher Aufwand
    - Randomisierung aufgrund der Natur sozialwissenschaftlicher Variablen einerseits und ethisch-moralischen Bedenken andererseits schwer möglich.

Quasi-Experimentelle Designs = Difference-in-Difference-Design
(Zwischending ex-ante und ex-post-facto)
- Quasi-Experimente sind Experimente
    - ohne Randomisierung
    - ohne die Stimulus-Kontrolle durch den/die Wissenschaftler*in.
- Voraussetzung: Daten derselben Untersuchungseinheit müssen über mindestens zwei Zeitpunkte vorliegen und zwar vor und nach einer Einflussnahme.
- Solche Situationen ergeben sich oft durch politisch-administrative Veränderungen, verfolgen also keinen wissenschaftlichen Erkenntnisgewinn per se, sondern ergeben sich ohne das bewusste Zutun des Forschers.
- Bsp.: Einführung von neuen Unterrichtsformen, Einführung von brieflicher Stimmabgabe oder Online-Voting, soziale Reformen irgendwelcher Art. (gerade in CH mit föderaler Struktur häufig)

Ex-post-facto-Anordnungen
Was sind ex-post-facto/Nicht-experimentelle Untersuchungen/Anordnungen?
- Keine (aktive) Manipulation des Treatments durch Forschende möglich
- Keine Randomisierung (Zuteilung Vergleichsgruppen *nach* Datenerhebung)
- Effekte werden im Nachhinein konstatiert, untersucht. Daten sind bereits gegeben.

Synonyme: Beobachtungsstudie, korrelative (wechselseitige) Studien, Natürliche Experimente (Uraschenvariablen liegen «natürlicherweise» vor).
→ Diese Synonyme sind auch Charakteristika der Untersuchung.

- 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.
- Auch die Zuweisung der Untersuchungseinheiten 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-Nacher-Messung), weshalb die kausale Reihenfolge unklar bleibt. (Ursache und Wirkung nur rein theoretisch nicht aus Aufbau des Designs folgend)

- Hauptproblem: interne Validität

➔ Problem der Scheinkorrelation/ Konfundierungsproblem
  - Weder die Richtung der Kausalität noch ob überhaupt eine Kausalität vorliegt, kann strenggenommen überprüft werden. → Scheinkorrelation.
  - Konfundierung: Zusammenhang zwischen einer unabhängigen und einer abhängigen Variabel verzerrt. (bspw. Aggressive Videospiele und Gewalttätigkeit bei Jugendlichen -> wir wissen nicht wirklich, ob ein Zusammenhang besteht. Untersuchung im Alltag, deswegen wissen wir nicht ob die Gewalttätigkeit eigentlich andere Gründe hat)
  - Kontrolle von Störfaktoren erfolgt über die statistische Kontrolle von Drittvariablen. Hierfür muss aber bekannt sein, welche Drittvariabel potentiell Einfluss auf die abhängige Variabel ausüben und zudem müssen diese Variablen im gegebenen Datensatz enthalten sein.

(Es gibt also Möglichkeiten die Störfaktoren auch bei einer nicht-experimentellen Untersuchungen zu berücksichtigen.)

- Vorteil: Geringer Aufwand.

Warum überhaupt nicht-experimentell?

- Die interessierte unabhängige Variable ist eine «personen- oder umweltgebundene» Variable, die gar nicht manipulierbar ist.
- Forschungsökonomische Gründe: Es wäre im Prinzip möglich, so ein Treatment experimentell zu setzten, aber viel zu teuer.
- Forschungsethik: Experimentelle Variation/ Manipulation forschungsethisch nicht vertretbar. (z.B.: Einfluss Naturkatastrophen, Konsum gewalttätiger Videospiele von Jugendlichen, oder generell, wenn illegales Verhalten untersucht werden soll) (Wandelt sich auch über die Zeit)

Kausalanalysen mit nicht-experimentellen Designs:

- Prospektive Kohortenanalyse: Gruppenvergleiche zwischen einer Gruppe, die aktuell einer Ursache ausgesetzt sind und einer die ihr nicht ausgesetzt ist, über Zeit hinweg. Bsp.: Vergleich Rauche und Nichtraucher über die Zeit hinweg.
- Retroperspektive Kohortenanalyse: Wirkung ist eingetroffen, nun wird retrospektiv nach den Ursachen geforscht.
- Fall-Kontroll-Studie: Randomisierung und anschliessender, retroperspektiver Vergleich der möglichen Ursachen. Unterschied zur retroperspektiven Kohortenanalyse: Zusätzliche Kontrollgruppe!

Spielarten von Ex-post-facto-Anordnungen:

- Querschnittdesigns (Einmalige Erhebung)
  Nur ein Messzeitpunkt, nur eine Stichprobe.
  Ein Querschnittdesign ist eine einmalige Erhebung bestimmter Variablenwerte bei einer bestimmten Zahl von Untersuchungseinheiten mit einer Stichprobe.
  Bsp.: Bevölkerungsbefragung («Querschnitt der Bevölkerung»)

- Trenddesigns (Längschnittdesign)
  mehrere Messzeitpunkte (mit der gleichen Variablen)
  unterschiedliche Stichproben (bspw. verschiede Individuen)
  Ein Trenddesign ist eine Erhebung der Werte der gleichen Variablen zu mehreren Zeitpunkten mit jeweils unterschiedlichen Stichproben.
  → erlaubt Aussage auf der Kollektivebene, nicht auf Individualebene, weil unterschiedliche Stichproben.

- Paneldesign
  mehrere Messzeitpunkte
  gleiche Stichprobe (dieselben Individuen)
  Mit einem Paneldesign werden Werte der gleichen Variablen zu mehreren Zeitpunkten, aber auf der Grundlage einer identischen Stichprobe (dieselben Individuen) erhoben.
  → erlaubt Aussagen über Kollektive und auch über Individuen
  (Es können auch in einem Messzeitpunkt gleich zwei gefragt werden, z.B. wird 2015 gefragt, wie 2011 gewählt wurde)

- Zeitreihenstudien
  mehrere Messzeitpunkte
  Wiederholte Erhebung von Werten der gleichen Variablen beim gleichen Merkmalsträger (wobei Anzahl Merkmalsträger =1)
  Bsp.: Die Arbeitslosenquote in der Schweiz.
  Merkmalsträger = Schweiz. Arbeitslosenquote wird zu verschiedenen Zeitpunkten erhoben.

Daten und Design

- Designtyp und Datentyp stimmen überein, wenn der Zeitpunkt der Erhebung identisch ist mit dem Messzeitpunkt.
  → Umfragen: nachträgliche Befragung einer Wahl (rückerinnertes Verhalten): kein Paneldesign, sondern Paneldaten
- 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

Beste Daten (mit höchstem Informationsgehalt) haben Panel*daten*!
Sie erlauben sowohl Querschnitt- wie auch Längsschnittanalysen.
Trend*daten* erlauben Querschnittanalysen, aber keine Panelanalysen.
Querschnitt*daten* lassen sich nur auf Querschnittanalysen anwenden.

Nachteil Paneldaten:
Sie werden selten erhoben, weil:
- sehr grosser Erhebungsaufwand,
- Panelmortalität (Probanden machen nicht mehr mit),
- Repräsentativitätsprobleme (Wandel der Bevölkerungsstruktur über Jahre),
- Edukationseffekte (Bei mehrfacher Teilnahme an einer Befragung lernen die Probanden und informieren sich bspw.)

Eine einzelne Studie sollte nie als beleg für eine Theorie genommen werden. Erst mehrere Datensätze und Studien können eine Theorie wirklich beweisen.

Teil 6: Datenerhebung
Sitzung 11: 29.11.19 – Beobachtungen und Inhaltsanalysen

Erhebungsverfahren im Generellen:
Unterscheidung anhand folgender Punkte:
- Reaktivität: Reaktionen bei Untersuchungsobjekt
  - Persönliche Befragung: Reaktion der Respondenten möglich und wahrscheinlich
  - Schriftliche Befragung: Reaktion der Respndenten kaum möglich.
- Standardisierung: Grad der Strukturierung einer Erhebung.
  Standardisierter Fragebogen vs. offene Beobachtung.

Beobachtungen
- Definition: «Die systematische und kontrollierte Beobachtung menschlicher Handlungen, sprachlicher Äusserungen und nonverbaler Reaktionen oder anderer sozialer Merkmale.» (Diekmann)
- Jedes empirische Erhebungsverfahren ist Beobachtung. Aber in einem engeren Sinne wird unter Beobachtung die Observation menschlichen Handelns verstanden.
-

Vorteile:
- Das Verhalten der Untersuchungsobjekte wird **unmittelbar** beobachtet.
  (nicht wie bei Befragungen über Erinnerungsfragen erfragt)

Nachteile:
- Möglichkeit der selektiven Wahrnehmung: Es werden nur Handlungen wahrgenommen, die die Hypothese bestätigen (*Halo*-Effect).
- *Going native* vs. *othering.*
- Gefahr der Fehlinterpretation (z.B. ein Augenzwinkern kann vielerlei Bedeutungen haben). Kultur als «selbstgesponnenes Bedeutungsgewebe».

Beobachtungstechnik:

Die Rolle des Beobachters; teilnehmend vs. Nicht-teilnehmend
Die Beobachtungstechnik ist von der Fragestellung abhängig. Unterschiede gibt es im Bezug auf die **Rolle des Beobachters**.
- Teilnehmend:
  Beobachter nimmt innerhalb des Beobachtungsfeldes selbst eine Rolle ein – entweder passiv (z.B. der Ethnograf als «Besucher» eines autochthonen Stammes) oder aktiv (der Lehrer, der das Verhalten seiner Schüler beobachtet).

➔ Probleme:
  - Identifikation mit den Untersuchungseinheiten (*going native*). Wissenschaftliche Distanz zum Beobachtungsfeld geht verloren.
  - Beeinflussung der Beobachtung durch Teilnahme (Doppelfunktion, Intrarollenkonflikte)
  - Kein Zugang zu eigentlichen Schlüsselpersonen

- Nicht-Teilnehmend:
  - Keine (bzw. nur geringe) Einflussmöglichkeiten
  - Protokollierungsmöglichkeiten eher gegeben, weil keine Doppelfunktion

<u>Die (wahrgenommene) Präsenz des Beobachters</u>
Was wissen die Untersuchungsobjekte von der Beobachtung?
- Offen: Die «Probanden» wissen über die Präsenz des Beobachters/sind sich ihrer bewusst, («Big Brother»). Problem der externen Validität.
- Verdeckt-teilnehmend: Beobachter ist aktiv, aber als solcher unerkannt. (wie «Undercover-Agent)
- Verdeckt aber nicht teilnehmend: Beobachter ist inaktiv und unbemerkt («Schlüssellochmethode»).
  - Vorteil: Verhalten wird nicht durch den Beobachtungsvorgang beeinflusst (nicht reaktiv)
  - Nachteil: ethisch-moralisch höchst fragwürdig.

<u>Beobachtungstechnik: Feld- vs. Laborforschung</u>
- Anwendbarkeit ist von Forschungsfrage und dem Untersuchungsgegenstand abhängig. (Arbeitslosigkeit kann bspw. nicht als «Treatment» gesetzt werden)
- Vor- und Nachteile Laborexperimente:
  - Der Forscher setzt die Stimuli und kontrolliert die Randbedingungen.
  - Kontrolle von Störfaktoren durch Randomisierung.
  - Hohe interne Validität, aber niedrige externe Validität.
    → «Natürliche» Langzeitfolgen nicht messbar.
- Vor- und Nachteile Feldbeobachtungen:
  - Langfristige Untersuchung von Effekten in «natürlicher» Umgebung.
  - Nicht durch Forscher manipulierbare Variablen beobachtbar.
  - Komplexe, in Laborexperimenten nicht simulierbare soziale Sachverhalte können erfasst/beobachtet werden.
  - Hohe externe, aber tiefe interne Validität.

- Das Feldexperiment: Eine Kombination von Experiment und Feldbeobachtung!
  - In «natürlicher» Umgebung (d.h. nicht im Labor)
  - In der Regel wissen die Probanden nicht, dass sie Teil eines Experimentes sind.
  - Triangulation möglich: Kombination von Labor- und Feldexperiment, um Resultate zu validieren.

<u>Beobachtungstechniken: Strukturierungsgrad</u>
- Strukturierte Beobachtung:
  - Präzise und strikte Beobachtungsschemata (Leitfaden).
  - Bedarf eines ausgearbeiteten Leitfadens
  - Ziel: Höhere Durchführungsobjektivität und Reliabilität. Gefahr der verzerrten Wahrnehmung soll eingedämmt werden.

- Unstrukturierte Beobachtung:
  - Nur grober Rahmen und einige Leitlinien werden vorgegeben. Weniger Beobachtungskriterien als bei strukturierter Beobachtung.
  - Ermöglicht gewisse Flexibilität und Offenheit des Beobachters für den Beobachtungsgegenstand. Verringert aber so die Reliabilität.
- Kombination möglich: Leitfadenbeobachtung

Inhaltsanalyse

- Definition: «Systematische Erhebung und Auswertung von Texten, Bildern und Filmen.» Systematische Identifikation von Text-Elementen und ihre Zuordnung zu Kateogiren.

➔ Gemäss Quantitativer Herangehensweise.
➔ ≠ Hermeneutik, Textinterpretation, «Verstehen des Sinns».

- Beschränkt sich nicht notwendigerweise nur auf Inhalte, sondern kann auch formale Texteigenschaften (Länge, stilistische Elemente etc.) umfassen.
  - Bsp.: Aktionsquotient = Verhältnis zwischen Verben und Adjektiven.

- Vorteile:
  - Keine Beschränkung auf Gegenwart, Vergangenheitsbezug ist möglich.
  - In der Vergangenheit produziertes Material kann untersucht werden und unterliegt zumindest keinen Erinnerungsirrtümern (wie bei Befragung).
  - Soziale Veränderungen (wie bspw. Wertewandel) sind erforschbar.
  - Nichtreaktivität: Das gesamte Datenmaterial ist nicht reaktiv.
- Nachteile:
  - Hoher Interpretationsaufwand und hohes Kontextwissen nötig. (Bedeutung von Wörtern hat sich möglicherweise gewandelt)

Formen und Zwecke der Inhaltsanalyse

"Who says what, to whom, why, to what extent and with what effect?" (H. Lasswell)

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 Kommunikatonstrends
	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 Kategoriesystem.
- Es muss vor der Analyse erstellt sein. (dimensionale/deskriptive Analyse)
- Operationalisierung = Suche nach Textindikatoren für die interessierenden Sachverhalte.
- Formulierung von Oberbegriffen/Kategorien, weiter Untergliederung in Teildimensionen. Kategorien sind Äquivalent zu den Fragen in eine Befragung.
- Codierschema als Abfrageschema für Texte.

Beispiel:

a53e/z	Welches sind die Hauptgründe, dass Sie die Volksinitiative "Gegen Masseneinwanderung" abgelehnt haben?
10	**Allgemeines**
11	allgemeine Antworten (z.B. schon mal gehört)
12	allgemeine negative Äusserungen (z.B. unfair, unsolidarisch, zu extrem, der falsche Weg)
13	Bauchgefühl
14	ein Zeichen setzen/Bewegung in die Sache bringen
15	sollte es so belassen, wie es ist/Status Quo gut, nicht klar wie es rauskommt
16	löst die vorhandenen Probleme nicht
19	**Anderes zu Allgemeines**
20	**Gegen Kontingentierung**
21	Grundsätzlich gegen jede Art von Kontingentierung
22	Generell gegen Beschränkungen/Abschottung
23	Zuwanderung kann nicht über Quoten/Zahlen reguliert werden
24	Unterscheidung nach Herkunftsregionen ist unfair
25	Die Schweiz braucht eine weitere Öffnung/Initiative führt zu weiterer Abschottung der Schweiz
29	**Anderes zu gegen Kontingentierung**
30	**Wirtschaftliche Gründe**
31	Initiative schlecht für die Wirtschaft, schadet der Wettbewerbsfähigkeit
32	Wirtschaft profitiert von Zuwanderung, profitiert von Zugang zu Fachkräften
33	Wohlstandsgewinn durch Zuwanderung
34	Tourismuskantone sind auf Ausländer angewiesen
35	Der Wirtschaft geht es gut/Situation ist nicht so dramatisch
36	Ausländer machen Arbeiten, die Schweizer nicht machen (wollen)
37	brauchen Ausländer/ausländische Arbeitskräfte/sind auf sie angewiesen
38	Gewisse Bereiche sind auf ausländische Arbeitskräfte angewiesen (z.B. Pflege/Spital, Gastronomie)
39	**Anderes zu wirtschaftliche Gründe**
40	**Initiative ist ungeeignet**
41	Initiative bringt nichts/ungeeignet um Ausländerprobleme zu lösen, Scheinlösung/Zuwanderung kann nicht via Gesetzen verhindert werden
42	Es gibt keine Masseneinwanderung, Initiative unnötig
43	Initiative ist zu restriktiv/zu radikal/zu starr
44	Initiative ist rechtswidrig, verstösst gegen die bilateralen Verträge, ist nicht umsetzbar/bin für die Bilateralen Verträge
45	Differenzierung zwischen Fachkräften und Asylanten etc. fehlt
46	Personenfreizügigkeit wichtig/bin für die Personenfreizügigkeit
49	**Anderes zu Initiative ist ungeeignet**

Zuverlässigkeitskriterien:

- **Stabilität (Intracoder-Reliabilität)**: Grad der Übereinstimmung zwischen verschiedenen Kodiervorgängen derselben Kodierer.
- **Wiederholbarkeit (Intercoder-Reliabilität)**: Grad der Übereinstimmung zwischen zwei verschiedenen Kodierern am selben Material.
- Forscher-Kodierer-Reliabilität: Übereinstimmung zwischen Forscher und Kodierer.

- Holsti-Index: Reliabilitätsmass für Intercoder-Reliabilität
  - Es gibt keine fixen cut-offs values. Werte nahe 1 sind anzustreben.
  - Wobei wobei „M is the number of coding decisions on which the two judgees are in agreement, and N1 and N2 refer to the number of coding decisions made by judges 1 and 2, respectively“.

  → M = Anzahl Übereinstimmung der Codierenden. Dividiert durch Codierentscheidungen N1 und N2

$$CR = \frac{2\,C_Ü}{C_1 + C_2}$$

CR = Codierer-Reliabilität
$C_Ü$ = Anzahl der übereinstimmenden Codierungen
C_1 = Anzahl der Codierungen von Codierer 1
C_2 = Anzahl der Codierungen von Codierer 2

Zwei Codierer verschlüsseln die Variable „Bezugnahme auf Wirtschaft?" mit den Ausprägungen 0 „Kommt nicht vor", 1 „Kommt als Nebenaspekt vor", 2 „Kommt als Hauptaspekt vor" in 18 Artikeln und kommen zu folgenden Ergebnissen…

	Bezugnahme auf Wirtschaft?																	
C1	1	1	1	1	2	0	0	0	1	0	2	0	0	1	1	1	1	0
C2	1	1	2	1	2	0	0	1	1	0	2	0	0	1	1	0	1	0

Zahl der Übereinstimmungen: 15

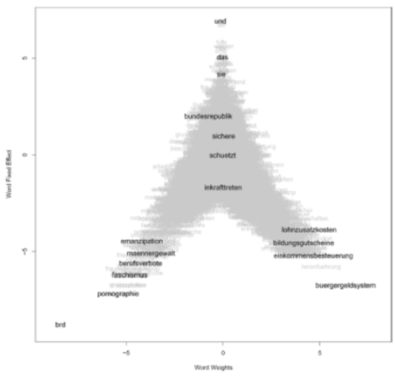
$$CR_{\text{Bezugnahme Wirtschaft}} = \frac{2 * 15}{(18 + 18)} = \frac{30}{36} = 0.83$$

$$CR = \frac{2M_0}{N_1 + N_2}$$

<u>Techniken:</u>

- Frequenzanalyse
  - Häufigkeit von ausgewählten Textelementen (Name von PolitikerInnen, Parteinamen…)
    Annahme: Häufigkeit korreliert mit Bedeutung
  - Können auch rein formale Texteigenschaften sein
  - Heute: Zumeist computergestützte Inhaltsanalysen.

FIGURE 2 Word Weights vs. Word Fixed Effects, Left-Right Dimension, Germany 1990–2005 (Translations given in text)



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

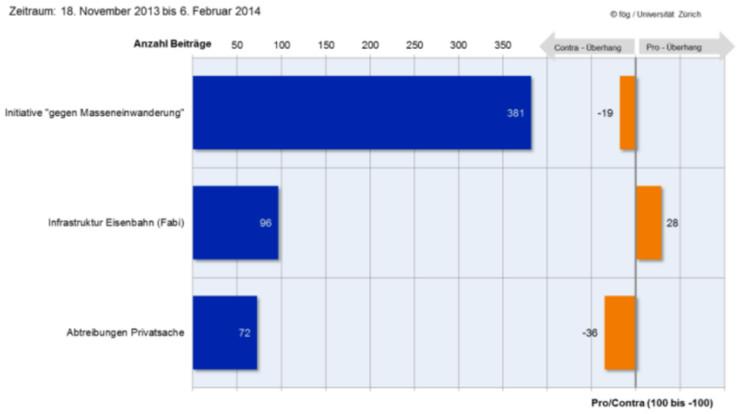
Selbstbeschreibung von Inserenten in Kontaktanzeigen (% der Inserate)

	jung	*attraktiv*	*groß*	*schlank*	*...*	*wohlhabend*
Frauen	34	12	17	64		8
Männer	27	24	46	11		33

Gemeinsames Auftreten von Paaren von Kategorien (% der Inserate)

	jung	*attraktiv*	*groß*	*schlank*	*...*
attraktiv	24				
groß	12	10			
schlank	33	8	19		
...	...	...	...	...	

- Valenzanalyse
  - Bewertungstendenz von Textinhalten (bspw. negative oder positive Bewertung, pro/contra)
  - Setzt zumindest ordinalskaliertes Niveau voraus.
  - In der Regel Schätzurteile, aber heutzutage auch mit entsprechend computerunterstützten Analysen möglich.
  - Intensitätsanalysen: Nicht nur die Richtung wird hier bewertet, sondern auch die Intensität (wie stark positiv bzw. negativ?)



- 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/Instagram, Tweets auf Twitter, Wikipedia-Traffic auf bestimmten Seiten etc.
  - Oftmals fehlen jedoch Theorien darüber, wer diese Daten erzeugt hat, zu welchem Zweck und wer sie nutzt.

(a) Supports layer, $|V| = 2,740$, $|E| = 16,657$ (b) Likes layer, $|V| = 1,147$, $|E| = 6,386$ (c) Comments layer, $|V| = 1,091$, $|E| = 6,800$

Figure 2: Visualization of network layers of `supports`, `likes` and `comments` excluding unaligned politicians. Colours of the nodes are labelled according to the parties self-reported by politicians. Party colors are reported in Table 1. The networks are drawn using the Fruchterman-Reingold layout algorithm (Fruchterman and Reingold, 1991).

(Diekmann Text anschauen zu Inhaltsanalysen)

Sitzung 12: 06.12.19 –Befragung und Interview
(Hinweis nebenbei: Wenig strukturiertes Interview vs. Standardisierte Befragung)

Formen der Befragung:

- Das persönliche Interview
  Einst die am weitesten verbreitete Befragungsform
  Vorteile:
    - Meist bessere Verständlichkeit und höhere Kontrolle
    - Niedrige Abbruchquote; Einfluss auf die Teilnahmequote ist umstritten.
    - Visuelle Hilfsmittel möglich
    - Befragungssituation weitestgehend kontrollierbar
    - Zusätzliche Beobachtungen möglich (Reaktionen des Befragten)
  Nachteile:
    - Reaktivität, da der Interviewer persönlich anwesend ist. (Befragter reagiert evtl. auf gewisse Charakteristika des Befragers)
    - Aufwand und Kosten sind vergleichsweise hoch.

- Das telefonische Interview
  Vorteile:
    - Kurzfristig einsetzbbar
    - Vergleichsweise tiefe Abbruchquote.
    - Befragungssituation bis zu einem gewissen Grad kontrollierbar
    - Befrager kann unterstützend eingreifen bei Verständnisfragen
  Nachteile:
    - Abnehmender Abdeckungsgrad bezüglich Strichprobenziehung; keine repräsentative Strichprobe mehr möglich (siehe Sampling)
    - Kosten und Aufwand steigen zunehmend
    - Visuelle Hilfsmittel nicht möglich
    - Reaktivität (auch wenn kein Mensch direkt gegenübersitzt)

- Schriftliche Befragung
  Früher schriftlich-postalische Befragung, heute immer häufiger: Online-Befragung
  Vorteile:
    - Grösse Anonymität, geringer Intervieweffekt
    - Kommt dem Zeitmanagement des Befragten entgegen und er kann Fragen besser durchdenken.
    - Untersuchungsergebnisse zu Rücklaufquoten fallen unterschiedlich aus.
    - Geringe Kosten und hohe Erreichbarkeit (sofern Adressliste vollständig)
  Zusätzliche Vorteile Online-Befragung:
    - Schnell und kostengüstig
    - Durch Software können auch Antwortverhalten und Geschwindigkeit gemessen werden.
    - Spielerisches Element
    - Experimentelle Studien sind einfacher realisierbar.
  Nachteile:
    - Kognitive Test (Wissensfragen) nur schwer möglich (ausser: Online-Software, die das bei Onlinebefragungen überwacht)
    - Identifikation der Zielperson nicht möglich
    - Keine Hilfe möglich bei Verständnisfragen
    - Keine Kontrolle der Interviewsituation möglich, dadurch kann der Befragungsablauf nicht garantiert werden.

    - Hohe Abbruchquote
- Zusätzlicher Nachteil Online-Befragung:
    - Meist Selbstrekrutierung (jedoch nicht zwingend).
      →Allerdings Problem der Strichprobenziehung und nicht der Form der Befragung.

Voraussetzungen einer Befragung:
- Kooperationsbereitschaft
    - Bereitschaft ist von vielen Faktoren abhängig, wie bspw. Interesse am Thema, finanzielle Anreize.
    - Die Kennzahl, die die Kooperationsbereitschaft misst = **Ausschöpfungsquote** (Teilnehmende aller Angefragten)
- Existenz einer «wahren Haltung» und die Bereitschaft, diese mitzuteilen
    - Verschiedene Wissenschaftler bezweifeln die Existenz solch einer «wahren Haltung» (*true attitudes*) und gehen eher von eine Kontinuum bzw. situativ abhängigen *considerations* aus.
    - Höflichkeitsnormen oder Furcht vor juristischen oder sozialen Sanktionen könne ebenfalls Hintergründe sein.
- Gemeinsame Sprache

Theorien des Antwortverhaltens:
- Probabilistische Ansätze
    - Menschliches Verhalten ist generell stochastisch, d.h. Menschen entscheiden in 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
- Die Theorie der rationalen Entscheidung
    - Menschen streben nach sozialer Anerkennung und fürchtet sich vor sozialen Sanktionen.
- Kognitionspsychologische Antworttheorie:
    - Memory-based/File-drawer-model: 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.
    - On-line-processing model
      Die Valenz einer Information wird sofort in ein evaluatives Zwischentotal integriert. Im Anschluss wird bloss dieses *summary tally* im Langzeitgedächnis abgespeichert, nicht aber die einzelnen, spezifischen Informationselemente. (Bei Vertrauten Themen hat die Befragte Person bereits eine Antwort parat, ohne genau die Informationen «wieso» abzurufen.)

Erhebungsgegenstände:
- Einstellungen (*attitudes, opinions*)
    - Einstellung als(*likes* und *dislikes*) ist ein wertendes Urteil über ein Einstellungsobjekt.
      Es sollte jedoch auch *weiss nicht* geben, Befragte haben nicht zwingend eine Einstellung zu einem Thema. (Bspw. «Mögen sie Pizzas mit Sardellen?»)
    - Präskriptive Aussage (Soll, normative Werturteile, soll)

- Überzeugungen (*beliefs*):
    - Umfasst alle Informationen und Kognitionen, welche mit dem entsprechenden Einstellungsgegenstand verknüpft sind. (Überzeugungen, «Weltwissen») →geglaubtes Wissen
    - Streng genommen noch nicht wertend, Ansichten darüber, wie die Welt geschaffen ist.
    - Deskriptive Aussagen, (Ist, Beschreibung)
- Verhalten (*behavior*):
    - Achtung: Bei Befragungen wird ausnahmslos berichtetes Verhalten erhoben (denn es wird ja nicht selbst beobachtet, wie bei einer Beobachtung).
- Sozialstatistik:
    - Soziale Merkmale (objektive Eigenschaften der befragten Person)

Fehlerquellen/Antwortverzerrungen:
Es gibt unterschiedliche Kategorien von Fehlerquellen bzw. Quellen von Antwortverzerrungen:

- (Persönlichkeits)merkmale des Befragten
- Merkmale des Befragenden bzw. Reaktionen des Interviewten darauf(Interviewefekt)
- Fragebogen bzw. Frageformulierung, d.h. Reaktionen auf formale Aspekte von Fragen (Frageeffekte)
- Interviewsituation

- <u>Merkmale des Befragten</u>
    - Soziale Erwünschtheit
      Die Angabe einer Antwort, die nicht der «wahren» Haltung entspricht, sondern dem, was als sozial erwünscht wahrgenommen wird.
      Entweder die persönlich vom Befragten wahrgenommene soziale Erwünschtheit oder auch als situationsspezifische Reaktion
      Gegenmassnahmen: Antworten (zusätzlich) anonymisieren, Enttabuisierung durch spezifische Frageformulierung. (Randomized Response Technique) (müssen wir für Prüfung nicht kennen).
    - Zustimmungstendenz (unabhängig vom Inhalt der Fragen)
      → immer mit Ja beantworten. Erkennbar an spezifischen Antwortmustern.
      Gegenmassnahmen: Umpolung der Frageformulierung oder der Antwortkategorie (vgl. Paralleltest)
    - Non-Attitude:
      Befragte beantwortet Frage obwohl sie keine inhaltliche Meinung dazu hat.
      Gegenmassnahme: «Weiss nicht»-Kategorie. (Kann unter Umständen aber auch «falsche» Meinungslosigkeit fördern.)

- <u>Merkmale des Befragers und Interviewsituation</u>
    - Frageformulierung (auch bei online-befragungen)
      Formale wie auch inhaltliche Aspekte der Frageformulierung (bspw. Suggestivformulierungen, psychologische Effekte)
    - Positionseffekte
      Antworten zu Fragen sind möglicherweise von gewissen Primingeffekten abhängig, d.h. davon, ob zuvor bestimmte kognitive Assoziationen aktiviert wurden oder nicht.
    - Vorgaben von Antwortkategorie.
      Meistens wählen die Befragten die Mitte, wenn sie wählen können. (zb. 1-1)
      «Weiss nicht», «keine Meinung» können auch verzerren

- Interviewermerkmale:
  Stark abhängig vom Inhalt der Frage. Wenn bspw. illegales Verhalten abgefragt wird, ist mit erheblichem Interviewereffekten zu rechnen.
- Interviewsituation
  Anwesenheit Dritter.
- Sponsorship-Effekt:
  Befragte richten ihre Antworten danach, wer der Auftraggeber der Studie ist.

Fragebogenkonzeption:

Kriterien Fragenformulierung:

- Verständlich, einfach
- Keine doppelten Verneinungen, keine Suggestivfragen, einfache Frageformulierung.
- Eindimensionalität (nur eine Dimension abfragen, selbst wenn ich mehrdimensionale Konstrukte untersuchen will)
- Möglichst konkreter Bezug, nicht hochabstrakte Wertevorstellungen, oder sehr allgemeine Fragen. (Wertevorstellungen vs. Einstellungen zu konkreten Sachfragen.)
- Antwortkategorien, erschöpfend (müssen das ganze Spektrum abdecken), disjunkt (Kategorien überschneiden sich nicht) und ausgewogen.
- Neutrale Formulierungen und (wenn möglich) keine hypothetischen Fragen (weil dann auch nur hypothetische Antwort).
- Einbau von Filterfragestruktur: Fragen die filtern, z.B. Voranstellend, «haben sie an der Abstimmung teilgenommen, ja-> Fragen. Nein-> überspringt Block.. Vorteil der Onlinebefragung.
- Vergleichbarkeit mit anderen Untersuchungen. (z.B. der Vergangenheit) verlangt identische Fragestellungen (zum Beispiel politische Wissen). Nur mit gutem Grund abändern!
- Vereinbarkeit mit geplantem Auswertungsverfahren (Auswahl eines ausreichenden Skalenniveaus).

Fragebogenaufbau:

- Einleitungsfragen („Eisbrecherfragen“), Interesse wecken und Vertrauen aufbauen.
- Positionierung der Fragen im Fragebogen: Kernfragen möglichst früh stellen (wegen Abbrechern). 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?“)
  → Vorteil der Onlinebefragung, bei schriftlicher Befragung etwas schwieriger
- 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

Bei geschlossenen Fragen sind Antwortkategorien vorgegeben. (z.B. Multiple Choice)

- Vorteile:
  - Reliabilität höher. Hohe Vergleichbarkeit und Durchführungsobjektivität.

    - Geringer Zeitaufwand; für Befragte sowie Auswertungsteam, da keine inhaltsanalytische Codierung der Antworten erforderlich.
    - Geringer Interpretationsspielraum für Codierer.
- Nachteile:
    - Antworten werden Vorstrukturiert und so vorgespurt: Kann im schlimmsten Fall Antwort verfälschen, weil «richtige» bzw. passende Antwort nicht vorgesehen ist.
      Oder das Gegenteil: Angabe einer Kategorie, die dem Befragten selbst nicht spontan in den Sinn gekommen wäre (Ex-Post-Rationalisierung)

Offene Fragen sind keine Antwortkategorien vorgegeben. Befragte artikuliert seine Antworten selbständig.
- Vorteile:
    - Der Befragte antwortet innerhalb seines eigenen Referenzsystems.
- Nachteile:
    - (Siehe Vorteile geschlossener Fragen)
      →Hoher Zeitaufwand, das Kodierung sehr lange dauert, grosser Interpretationsspielraum, Reliabilität tiefer.

Gute Alternative: halboffene Fragen:
- Fragen bei denen gewisse Antwortkategorien vorgegeben sind, aber auch eigenen Antworten formuliert werden können (bspw. mit freiem Feld).
  Mehrheit wird mit vorformulierten Fragen abgefangen, Rest kann selber ergänzen.

Sitzung 13: 13.12.19 – Befragungen und Stichprobentheorie

Stichprobentheorie
Voll- und Teilerhebungen
- Ziel: Aussagen über eine Grundgesamtheit (Bsp.: alle Schweizer Stimmberechtigten)
- Problem: In aller Regel liegen nur Daten einer Teilerhebung vor (z.B. kann man nur einen Teil aller Schweizer Stimmberechtigten 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

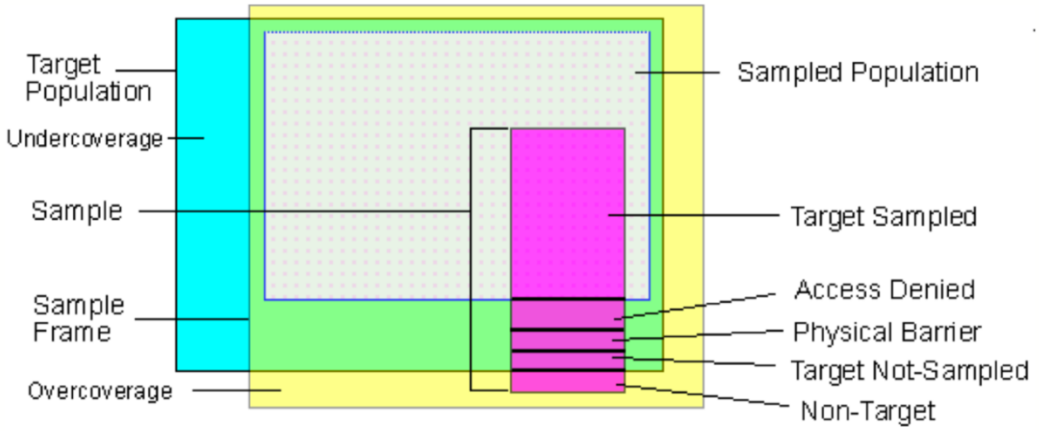
- <u>Die Grundgesamtheit</u> (*target population*)
  Die Menge derjenigen Individuen, über die eine Aussage gemacht werden soll, wird die Grundgesamtheit genannt.
  → Von der jeweiligen Fragestellung abhängig
  Bsp.: Wer hat wie gewählt bei den Nationalratswahlen 2015? Gilt die Nicht-Teilnahme ebenfalls 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.

- <u>Die Auswahlgesamtheit</u> (*frame population*)
  Die Menge der Elemente, die eine prinzipielle Chance haben, in die Stichprobe zu gelangen.
  Bsp.: Bildet das offizielle Telefonverzeichnis der Swisscom die Auswahlgesamtheit,

so haben nur die über die darin registrierten Telefonnummern (bzw. die darunter Erreichbaren) die prinzipielle Chance, in die Stichprobe zu gelangen.

- Die Inferenzpopulation (*inference population*)
  wird aus der Auswahlgesamtheit abgeleitet. Sie ist die Population, über die auf der Grundlage der vorliegenden Stichprobe tatsächlich Aussagen gemacht werden können. Im Idealfall Auswahlgesamtheit und Inferenzpopulation = Grundgesamtheit. (identisch oder nahezu identisch). Je weiter die beiden Grössen auseinanderfallen (siehe unte *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 keinen Teil der Grundgesamtheit sind.
    Bsp.: Grundgesamtheit = Schweizer Stimmberechtigte. Befragt werden indes alle, die im Swisscom-Verzeichnis enthalten sind (z.B. auch AusländerInnen)
    In der Praxis selten ein Problem, «überzählige» Stichproben können einfach wieder entfernt werden.
  - Undercoverage = wenn Elemente der Grundgesamtheit nicht in der Auswahlgesamtheit enthalten sind.
    Bsp.: Grundgesamtheit = Schweizer Stimmberechtigte. Befragt werden jedoch nur jene, die im Swisscom-Verzeichnis registriert sind.
    Ein weitaus schwierigeres Problem als Overcoverage. Denn die Verzerrungen lassen sich kaum, bzw. nur unter Zuhilfe bestimmter Annahmen korrigieren.



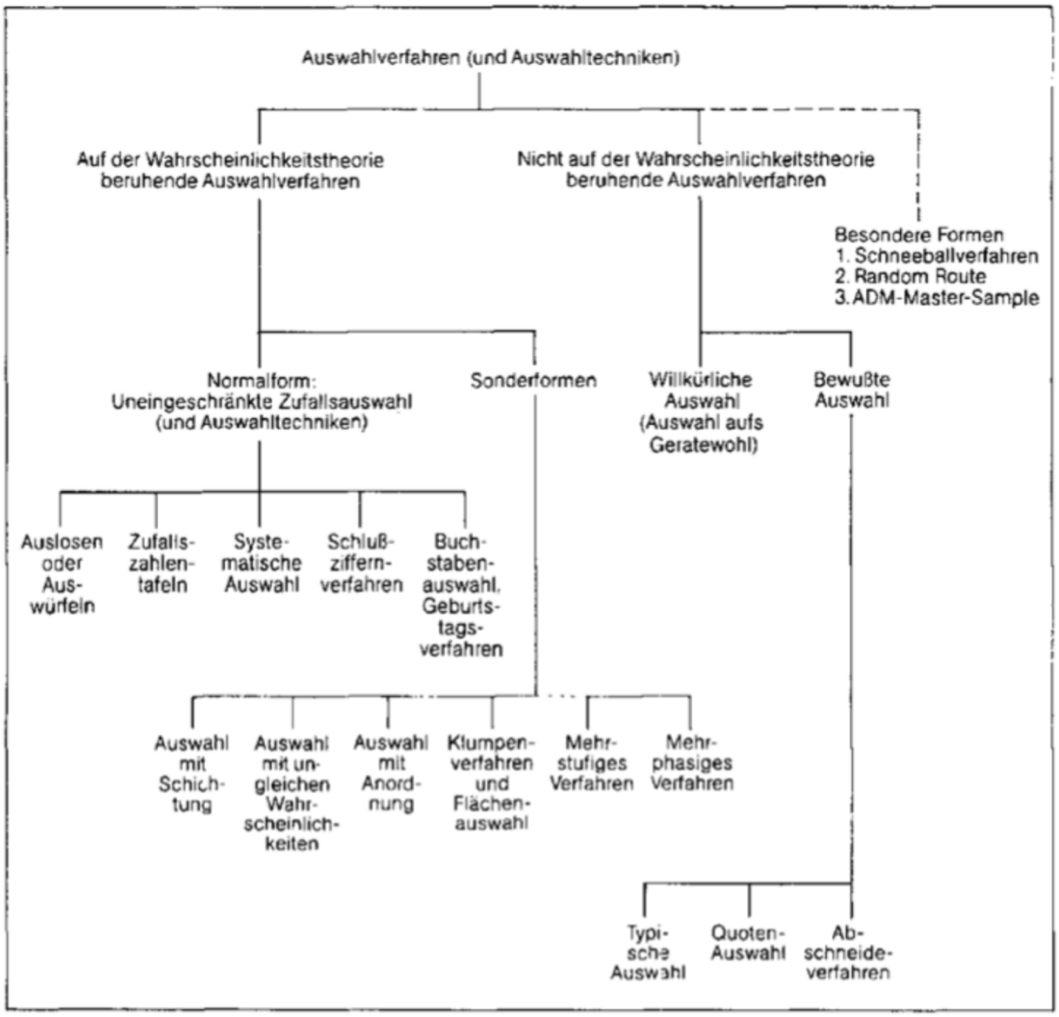
- Auswahl-, Erhebungs- und Untersuchungseinheiten
  Auswahleinheiten = Auswahlplan
  Erhebungseinheiten = Die Einheiten, bei denen die Informationen erhoben werden. Oftmals nicht identisch, dies stellt die Umfrageforschung vor gewisse Probleme.

  Untersuchungseinheiten = Die Einheit, auf welche sich die Untersuchung bezieht, bzw. an welcher die relevante Messung als solches vorgenommen wird.

- <u>Repräsentativität</u>
  Ein (geschützter) Fachbegriff «Repräsentativität» existiert nicht. Es gibt unterschiedliche Bedeutungen dieses Begriffs in der nicht-wissenschaftlichen, der wissenschaftlichen und der statistischen Literatur.
  1. **„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. Repräsentativität als **Strukturkonzept**: Stichprobe und Grundgesamtheit sind gleich zusammengesetzt. Das Sample ist ein verkleinertes, aber ansonsten strukturgleiches Abbild der Grundgesamtheit. Hohe Suggestivwirkung, aber nicht immer sinnvoll.
  3. Repräsentativität 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.
     → Zufallsauswahl = repräsentative Stichprobe

<u>Typologien von Auswahlverfahren</u>

Auswahlverfahren und Auswahltechniken



Quelle: *Hüttner*, 1989, S. 87.

Nicht-zufällige Auswahlverfahren
→ Prozess der Auswahl ist nicht kontrollierbar und kann mit den statistischen Modellen nicht beschrieben werden.
Zufällige Auswahlverfahren:
→ Nur Zufallsstichproben sind für den Induktionsschluss geeignet, weil nur in solchen Auswahlen jedes Element der Grundgesamtheit mit der gleichen (bzw. berechenbaren) Wahrscheinlichkeit in die Stichprobe aufgenommen wird (analog zum Zufallsexperiment).
(Bemerkung: Normalerweise erfolgt der Ziehungsprozess bei Bevölkerungsbefragungen ohne das erforderliche Zurücklegen. Dadurch sind die gezogenen Elemente jedoch nicht unabhängig voneinander und haben auch nicht dieselben Ziehungswahrscheinlichkeiten (siehe Voraussetzungen für Binomialverteilungen). Aber: Bei grossen Stichproben und grossen Grundgesamtheiten fällt dies kaum ins Gewicht)

- <u>Einfache Zufallsauswahl</u>
  zeichnet sich dadurch aus, dass jede Stichprobe (vom Umfang n) mit gleicher Wahrscheinlichkeit gezogen wird.
  → daraus folgt, dass alle Elemente der Grundgesamtheit die gleichen Chancen haben, ins Sample zu kommen. Chance >0
  Probleme bei der Realisierung:
    - Konzeptionell zwar am einfachsten zu verstehen, aber meist liegt keine (vollständige) Liste der Elemente der Grundgesamtheit vor oder die Durchführung einer solchen Datenerhebung ist zu teuer.

- <u>Komplexe Zufallsauswahl</u>
  die einzelnen Elemente haben nicht mehr dieselben, aber eine *vorab berechnete*, Chance, in die Stichprobe zu gelangen. Chance >0
  Aspekte der Realisierung:
    - Komplexe Zufallsauswahlen bieten sich an, wenn interessierende Untergruppen der Grundgesamtheit nur sehr geringe Anteile aufweisen.
    - Eine Schichtung (Spielart der komplexen Zufallsauswahl) bietet sich an, wenn die Schichten in sich möglichst *homogen* sind, sich aber in Hinsicht auf die interessierenden Variable möglichst *stark unterscheiden* (bspw. Sprachregion und Abstimmungsverhalten in der Schweiz)
  - Geschichtete Zufallsauswahl
    Bei einer geschichteten Zufallsstichprobe werden die Elemente der Grundgesamtheit zunächst einer (und nur einer) Schicht zugewiesen, woraus anschliessend jeweils unabhängige Zufallsauswahlen gezogen werden.
    Vorteil:
    Geschichtete Zufallsstichproben haben in der Regel eine geringeren Standardfehler, dadurch erhöht sich die Schätzgenauigkeit.

Abbildung 4: Schematische Darstellung einer komplexen Zufallsauswahl

Warum bietet sich eine geschichtete Auswahl an?

- Proportionale Schichtung. Wenn sichergestellt werden muss, dass die Häufigkeitsanteile der Subpopulationen ihren Anteilen in der Population exakt entsprechen.
- Disproportionale Schichtung. Wenn man an spezifischen Subpopulationen interessiert ist, deren Gesamtpopulation zu klein ist und mit einer einfachen Zufallsstichprobe durch eine zu geringe Fallzahl vertreten wäre.

Voraussetzungen für eine Schichtung:

- Es müssen (vorab) Grundinformationen über die Auswahleinheiten vorhanden sein, die es ermöglichen, sie in die Schichten zu unterteilen. Deshalb sind die gängigsten Schichtungsfaktoren räumliche Variablen.

- <u>Klumpen- und Flächenstichproben</u>
  Klumpen (engl. *cluster*) sind «natürlich» auftretende Gruppen wie Schulen oder Unternehmen.
  Flächen sind in der Regel geographische Einheiten wie Gemeinden.
  - Klumpenstichproben sind häufig, weil eine Liste aller Elemente der Grundgesamtheit und selten existiert, eine Liste von Klumpen aber oft bzw. öfter 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.

  - Problem: Die Mitglieder eines Klumpens (z.B. Schulklassen) sind sich in vielen Merkmalen oftmals ähnlicher als die Mitglieder verschiedener Klumpen.
  - Klumpenstichproben liefern deshalb ungenaue Schätzungen der Populationsparameter (höhere Standardfehler) als Schätzungen einer einfachen Zufallsstichprobe gleichen Umfangs. Die daraus resultierende Verzerrung nennt man des «Design-» oder «Klumpeneffekt».
  - Grundsätzlich: Klumpen sollten idealerweise maximal heterogen sein und sich nur geringfügig von anderen Klumpen unterscheiden.

- <u>Willkürliche Auswahlen</u>
  Einer willkürlichen Stichprobe fehlt ein Auswahlplan. Die Befragten werden aufgrund ihrer einfachen Verfügbarkeit oder einer anderen nicht-systematischen Methode ausgewählt. Bsp.: Strassenbefragungen («Was denkt der Mann auf der Strasse») oder optionale Online Umfragen, bei denen diejenigen, die sich (zufällig) auf der

entsprechenden Internetseite aufhalten, selbst entscheiden, ob sie teilnehmen wollen oder nicht.

- <u>Nicht-Zufallsgesteuerte Auswahlen: Bewusste Auswahl/Quotenauswahl</u>
    - Bei einer Quotenauswahl wird die Population 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 Zufallsstichproben wird keine Zufallsstichprobe innerhalb der jeweiligen Schichtung gezogen, sondern der Interviewer entscheidet selbst, wie er die vorgegebene Quote 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:
- Zufallsfehler der Stichprobe (*sampling error*):
  Der Zufallsfehler ist allerdings berechenbar
- Non sampling errors:
    - Systematische Fehler aufgrund des Auswahlverfahrens bzw. -rahmens (*coverage error*)
    - Non Response: Verweigerung (*item* und *unit nonresponse*)
    - Messfehler (*measurment error*): zum Beispiel das Overreporting bei der Angabe der Wahlteilnahme.
    - Gewichtung und Datenaufarbeitung (*processing errors*)

- Problem I: Eine Liste der Grundgesamtheit existiert nicht (Ausnahme: Bfs-Registerdaten, SRPH)
    - Deshalb weichen fast alle Befragungsinstitute auf Telefonlisten aus. Diese sind jedoch *nie* eine vollständige Liste der Grundgesamtheit.
      Deshalb ist kaum eine Schweizer Befragung repräsentativ (wenn Repräsentativitätsbegriff an Existenz von vollständiger Liste der Grundgesamtheit geknüpft).
    - Praxis: Telefonlisten. Nur Festnetzanschlüsse und «Geburtstagsmethode» («Darf ich bitte mit derjenigen Person in Ihrem Haushalt sprechen, die zuletzt Geburtstag hatte?»)
    - Undercoverage und overcoverage-Probleme
- Problem II: Abdeckung des öffentlichen Telefonverzeichnisses beträgt rund 70%, Tendenz fallend. Handys sind immer mehr verbreitet, aber Handynummern sind viel seltener registriert als Festnetzanschlüsse.
- Problem III: Interviewverweigerung ist kein Problem der Stichprobenziehung, denn diese Fälle sind als Untersuchungseinheiten ja vorgesehen. (?)
  Ausschöpfungsquoten betragen mittlerweile 20-25%. Allerdings: höhere Ausschöpfungsquote ≠ (notwendigerweise) genauere Ergebnisse.

- Systematische Verteilung von Interviewverweigerern
    - Bekennereffekt: Verlierer einer Abstimmung verweigern Interviews häufiger als Gewinner.
    - Generell: Teilnehmende sind überrepräsentiert in der Stichprobe (aus unterschiedlichen Gründen)

Sitzung 14: 20.12.19 – Abschlusssitzung

Item: In der Regel in der Befragung, eine manifeste Variable. Antwortverhalten
Antwort auf das Item ist dann ein Indikator

Indikator als Oberbegriff, Item als spezifischer Indikator für Aussagen oder Statements bei Befragungen.

Messfehler und Testtheorie anschauen!!!!!!! Wichtig für Prüfung.

Kriteriumsvalidität: Benchmark der sich bewährt hat, ich vergleiche meinen Indikator mit diesem Indikator (dieser ist dann das Kriterium)

Konstruktvalidität: Zwei Konstrukte die stark miteinander verknüpft sind sollte mit ähnlichen Mitteln erforscht werden. Indikator für Populismus sollte ähnlich sein wie Indikator zur Elitenforschung.
Klumpenstichprobe?