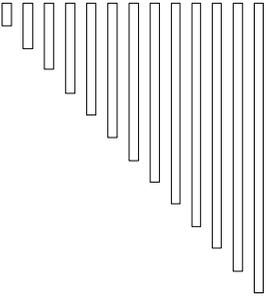


Quantitativ forschen: Fragebögen entwickeln und auswerten

Prof. Dr. Sieghard Beller
Fakultät für Psychologie
Universität Bergen, Norwegen





Fragebögen

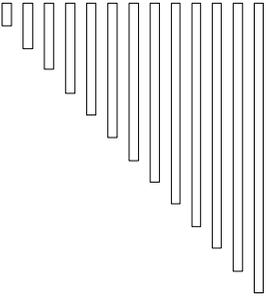
Beispiele

- Meinungsumfrage
- Marktforschung
- Lehrevaluation
- Online-Tests
- etc.

Ziele: Information einholen;
Interventionen begründen

Zentrales Element: Fragen

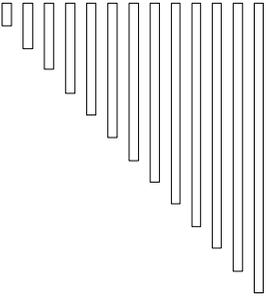




Übersicht

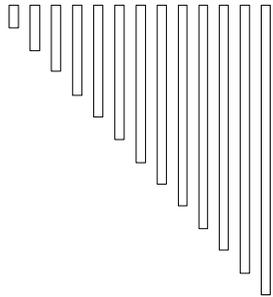
- Von der Planung bis zum Rücklauf
- Kategorien und Ratings als Antwortformat
- Güte von Fragebogendaten

Ziel: Aufzeigen von Gestaltungsspielräumen, (*researchers degrees of freedom*) und der methodischen Konsequenzen der getroffenen Entscheidungen



Übersicht

- Von der Planung bis zum Rücklauf
 - Planung und Umsetzung
 - Schriftliche vs. online-Fragebögen
 - Was der Rücklauf aussagt
 - Mögliche „Antwortfehler“
- Kategorien und Ratings als Antwortformat
- Güte von Fragebogendaten



Planung und Umsetzung

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- Makroplanung

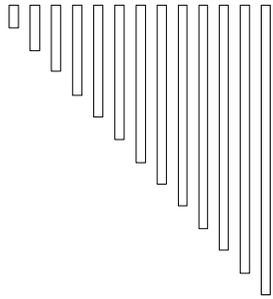
- Planung der Gesamtstruktur

- Mikroplanung

- Ausdifferenzierung der Inhaltsbereiche
- Formulierung der Fragen
- Allgemeine Instruktion

- Umsetzung

- Papier-und-Bleistift (postalisch) vs. online (Internet)
- Layout → möglichst übersichtlich



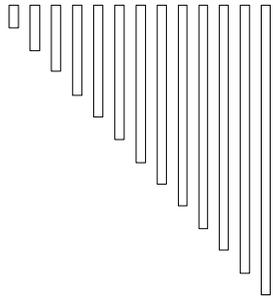
Entwicklung von Fragen

- Planung und Rücklauf

- Kategorien und Ratings
- Güte der Daten

Zu entscheiden

- Inhalte der Fragen
- Formulierung der Fragen
- Grad der Standardisierung der Antworten



Was sind „gute“ Fragen?

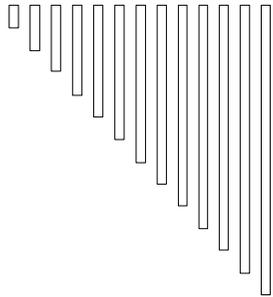
- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

Einige Richtlinien: Eine gute Frage ...

- ist sprachlich einfach formuliert
- ist eindeutig formuliert
- ist nicht zu allgemein formuliert
- ist nur auf einen Sachverhalt gerichtet
- bringt nicht in Verlegenheit
- ist nicht suggestiv (→ legt keine bestimmte Antwort nahe)
- hat eine angemessene Antwortvorgabe
- ist im Prinzip beantwortbar



Fragen mit Negationen

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- sind schwerer verständlich (Skopus der Negation?)
- die Antworten sind oft nicht eindeutig zu interpretieren

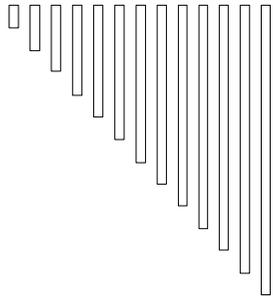
Beispiel

Frage: Gehen Sie nicht häufig unvorbereitet in eine Stunde?

Antwort: Nein.

Was meint diese Antwort?

- Nein, ich gehe *selten* unvorbereitet ...
- Falsch, ich gehe *häufig* unvorbereitet ...



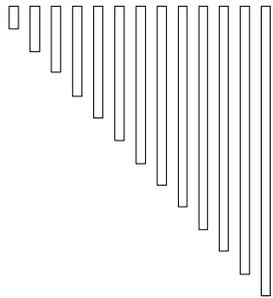
Gute Fragen?

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- Sie meinen doch auch, dass Herr X untragbar geworden ist, oder? → suggestiv
- Was würden Sie wählen, wenn am Sonntag Wahl wäre: (a) CDU (b) SPD? → unvollständige Antwortvorgabe
- Welcher Kandidat ist Ihnen am meisten sympathisch: Steinbrück oder Merkel? → halbwegs okay
- Glauben Sie nicht, dass die SPD gewinnt? → Negation



Standardisierung der Antwort

- Planung und Rücklauf

- Kategorien und Ratings
- Güte der Daten

Vorgabe der Antwortmöglichkeiten

- fest: Optionen vorgegeben
 - → Ratingskalen → Multiple-Choice
 - Wann sinnvoll? Wenn die Optionen vollständig bekannt sind
 - Vorteil: Leicht auswertbar
- offen: Keine Optionen vorgegeben
 - Wann sinnvoll? Wenn die Optionen kaum bekannt oder die erwarteten Antworten umfangreich sind
 - Auswertung: Inhaltsanalyse → Kategorisierung

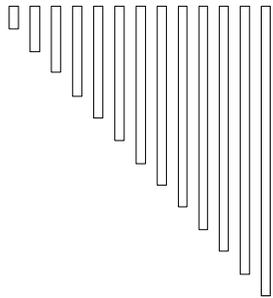
Mit was verbringen Sie ihre Freizeit?

- Musik hören
- Ins Kino gehen
- Sport machen

...

Versus

Schildern Sie Ihre Freizeitaktivitäten.



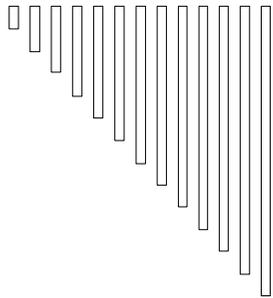
Prüfung der Gesamtheit der Fragen

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- Ist jede Frage erforderlich?
→ Wiederholung ähnlicher Frage vermeiden
 - Könnten Antworten durch die Fragenreihenfolge beeinflusst werden? → Reihenfolgeeffekte vermeiden
 - Ist die Polung der Fragen ausgeglichen?
→ Personen mit einer bestimmten Einstellung sollten nicht durchweg mit «ja» bzw. «nein» antworten müssen.
- Fragebogenentwurf in einem Probedurchlauf testen (Quellen für Missverständnisse? Verbesserungen?)

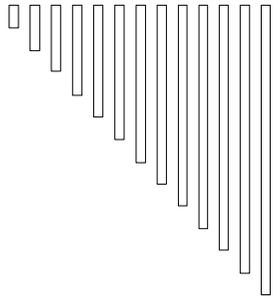


Begleitinformation

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Ziel: Informierte Teilnahme; möglichst starke Teilnahme

- Wer ist verantwortlich? → Anschrift, Telefon, Email ...
- Anrede
- Ziel der Untersuchung, Verwendung der Information
- Angaben zur Auswahl der Person → „Sie sind wichtig!“
- Antwortappell
- Anleitung zum Ausfüllen; geschätzte Dauer
- Rücklauftermin (*deadline*)
- Zusicherung der Anonymität
- Zusicherung von Vertraulichkeit der Daten
- Dank für die Mitarbeit



Postalische Befragung

- Planung und Rücklauf

- Kategorien und Ratings

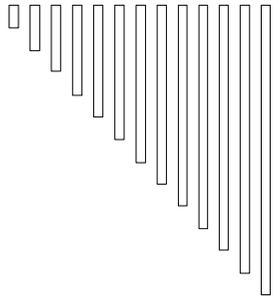
- Güte der Daten

- Versand des Fragebogens per Post

Vorteile?

- Planung der Stichprobe möglich

- Teilnahme zu einem passenden Zeitpunkt



Postalische Befragung

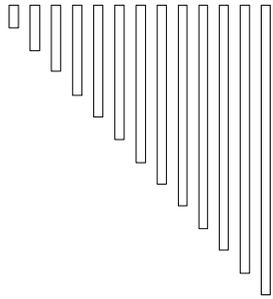
- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

Hauptnachteile?

- Aufwand: teuer bei großen Studien; Dateneingabe
- Geringe Kontrolle: Angaben der Personen kaum prüfbar, Motivation der Teilnehmer oft unklar, Bearbeitungssituation unkontrolliert (Ablenkung? Hilfe? ...)
→ Gefährdung der Qualität der Daten
- Selbstselektion I: Nicht alle Personen der geplanten Stichprobe antworten auch → Rücklauf



Rücklauf

- Planung und Rücklauf

- Kategorien und Ratings

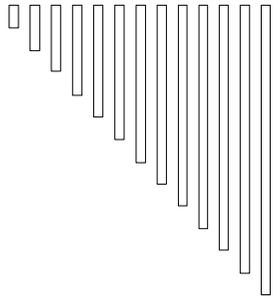
- Güte der Daten

Drei Maße geben Auskunft

- Rücklaufquote

- Rücklaufcharakteristik

- Rücklaufstatistik



Rücklaufquote

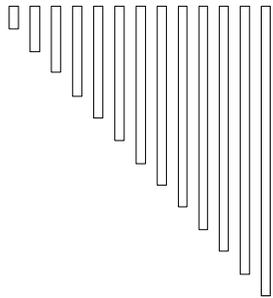
- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

Anteil der zurückgeschickten Fragebogen an der Gesamtzahl der ausgesandten Bögen.

- Schwankt stark (zwischen 10% und 90%)
- Problem: Geringer Rücklauf gefährdet die → Repräsentativität



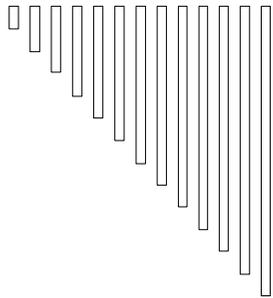
Was den Rücklauf verbessert

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- Aktuelle, interessante Inhalte
- Eine homogene Zielgruppe
- Erfahrung mit schriftlichen Texten
- Eine Ankündigung der Befragung
- Ein Brief, der sich von Werbesendungen unterscheidet
- Eine nicht zu starke Personalisierung
- Eine *deadline*
- Ein frankierter Rückumschlag
- Ein Anreiz (→ Geld, Gewinn)
- Ein Erinnerungsschreiben

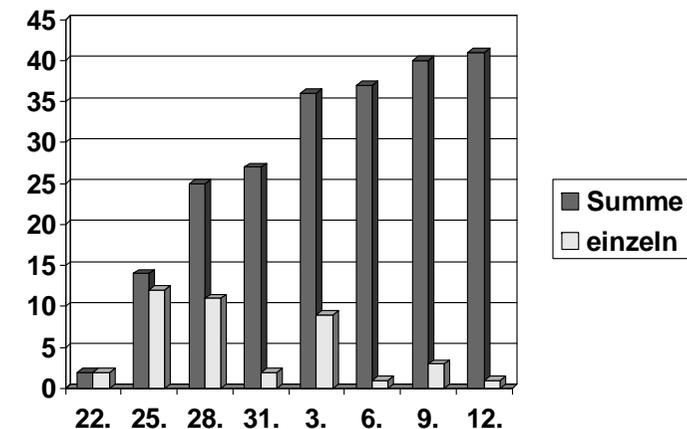


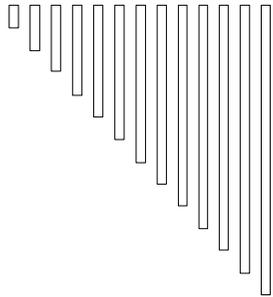
Rücklaufcharakteristik

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Statistik des zeitlichen Rücklaufs der Fragebögen → Häufigkeit pro Tag

- Gibt Auskunft darüber, wie die Stichprobe ausgeschöpft wird,
- ob eine Erinnerung nötig ist.
- Negativ beschleunigte Kurve
- Das Maximum kann nach ca. 7 Tagen extrapoliert werden
- 70-80% davon sind i.d.R. nach 10 Tagen erreicht



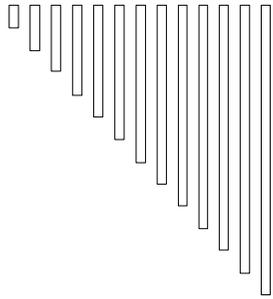


Rücklaufstatistik

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Analyse der biografischen Merkmale der Antworter. Ziel: Hinweise für Probleme mit der → Repräsentativität finden

- Antwortet nur eine bestimmte Teilmenge der Zielgruppe?
- Gib es Personengruppen, die systematisch fehlen?
- Antworten Früh-Antworter anders als Spät-Antworter?
 - Gibt u.U. Hinweise auf ein anderes Antwortverhalten der Nicht-Antworter



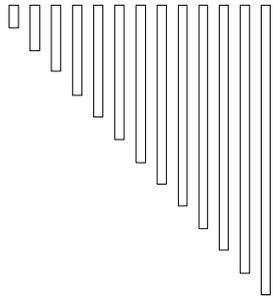
Rücklaufstatistik

- Planung und Rücklauf

- Kategorien und Ratings
- Güte der Daten

Häufige Merkmale von Nicht-Antwortern (Binder, 1979)

- Geringere Bildung, geringere Intelligenz
- Geringeres Interesse am Thema
- Geringere Beziehung zur Untersuchung



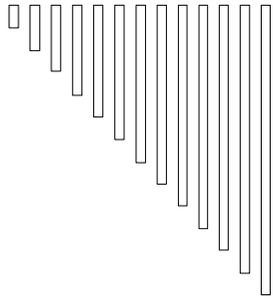
Internetbefragung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

- Per email → Fragebogen als Attachment
- Über eine Internetseite → direktes Ausfüllen des Bogens
- Mehrfachteilnahme ist auszuschließen
- Anonymität ist zu gewährleisten

Vorteile?

- Ablauf noch stärker standardisierbar
- Ökonomisch (Datenerfassung und Auswertung)
- Erreichbarkeit großer Personenkreise



Internetbefragung

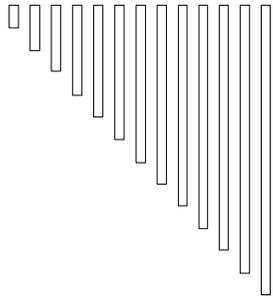
- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

Hauptnachteile?

- Selbstselektion II: Man erreicht überhaupt nur einen bestimmten Personenkreis. Unterscheiden sich Internet-Nutzer von Nicht-Nutzern?
- Selbstselektion I: Antworter / nicht-Antworter → Rücklauf
Wie bei postalischer Befragung
→ Gefährdung der Repräsentativität
- Geringe Kontrolle: Wie bei postalischer Befragung
→ Gefährdung der Qualität der Daten



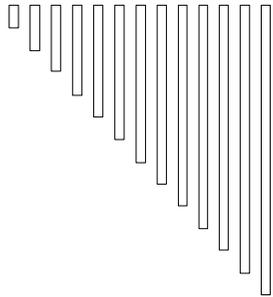
Antwortfehler I

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- Selbstdarstellung: Personen wollen ein Bild von sich vermitteln (v.a. wenn sie negative Folgen befürchten).
- *Self disclosure*: Geringe Bereitschaft zur Selbstenthüllung
- Erinnerungsfehler: v.a. bei Ereignissen, die länger zurückliegen. Fehlende Information wird rekonstruiert und manchmal (fehlerhaft) ergänzt.
 - *priming* und Kontexteffekte
 - Thema Gedächtnis



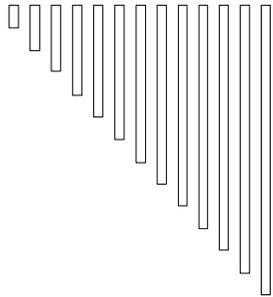
Antwortfehler II

- Planung und Rücklauf

- Kategorien und Ratings

- Güte der Daten

- *Sponsorship bias*: Vermutungen über den Auftraggeber der Befragung (und dessen Ziele) beeinflussen die Antwort
- Fehlende Antworten: Keine Meinung; keine Information über den Sachverhalt; Unentschlossenheit; Absicht; Versehen?
- Gezielte Falschangaben



Erinnerungsfehler

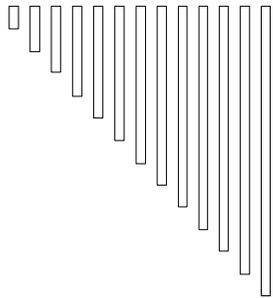
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Nach einem Unfallvideo wurde gefragt: Wie schnell fuhr das Auto als es ... [Verb]

- ... smashed
- ... collided
- ... bumped
- ... hit
- ... contacted



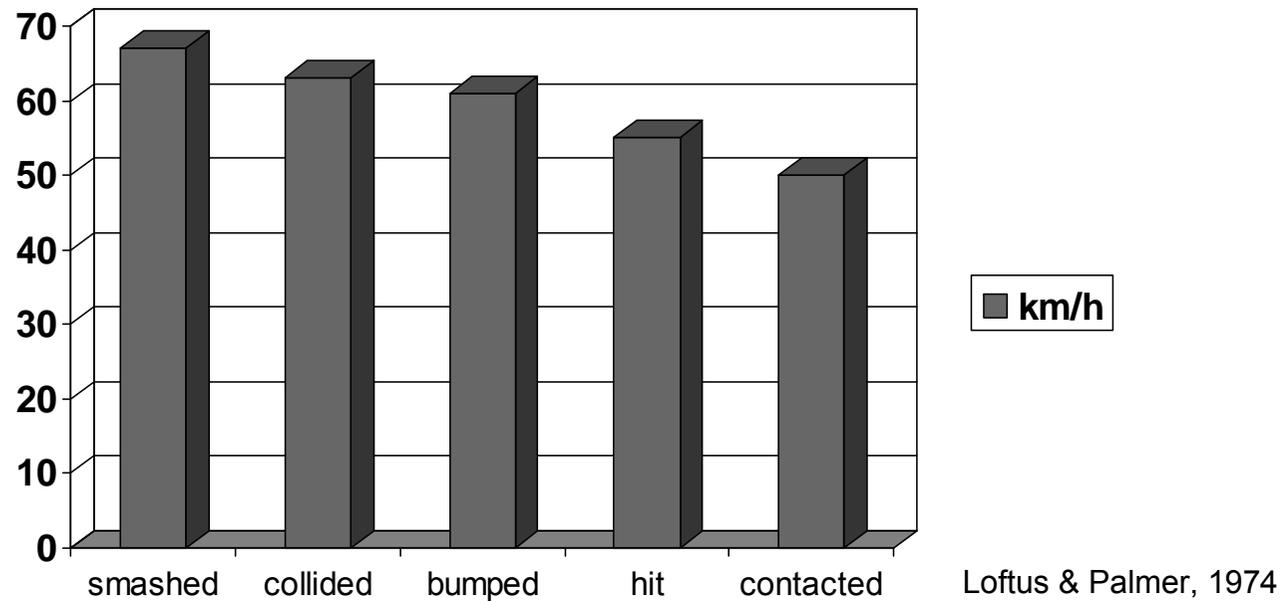
Loftus & Palmer, 1974



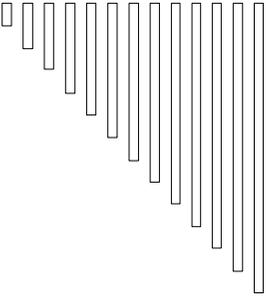
Erinnerungsfehler

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Schätzung der Geschwindigkeit

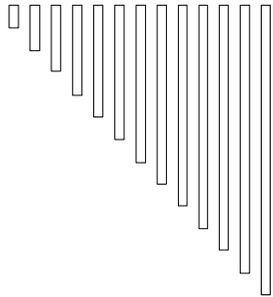


Auf den Wortlauf kommt es an!



Übersicht

- Von der Planung bis zum Rücklauf
- Kategorien und Ratings als Antwortformat
 - Categoriesysteme und ihre Eigenschaften
 - Ratingskalen in verschiedenen Formen
 - Urteilsfehler, die man kennen sollte
 - Ratingskalen versus (Mess-) Skalen
- Güte von Fragebogendaten

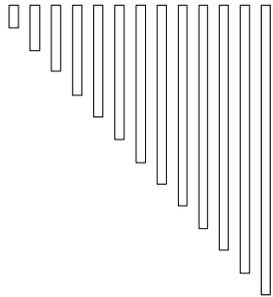


Kategorisierung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Auszählen der Häufigkeit verschiedener Antwortkategorien bei offenen oder multiple-choice Antworten.

- Natürliche Kategorien qualitativer Merkmale
 - Geschlecht (dichotom): Männer vs. Frauen
 - Studienfächer (polytom): Mathematik, Chemie, Physik, Psychologie, Sinologie, Geschichte, VWL, ...
- Künstliche Kategorien aus quantitativen Merkmalen
 - Verschiedene Ausprägungen werden zu einer Kategorie zusammengefasst.
 - Alter: Kinder, Jugendliche, Erwachsene



„Gute“ Categoriesysteme

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Drei Merkmale müssen erfüllt sein

- **Genauigkeit:** Exakte Definition der Kategorien. → Klare Entscheidung, in welche Kategorie eine Antwort fällt.
- **Exklusivität:** Kategorien schließen sich gegenseitig aus. → Jede Antwort wird nur einmal gezählt.
- **Exhaustivität:** Alle Kategorien zusammen sind erschöpfend. → Alle Objekte können gezählt werden.
 - Lässt sich durch eine «Restkategorie» erreichen für alle nicht weiter kategorisierbaren Objekte.

Gute Categoriesysteme?

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Worin besteht das Problem?

Berufe I
Schreiner
Arzt
Dachdecker
Lehrer
Handwerker

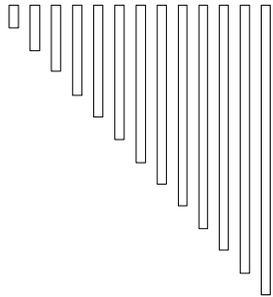
Handwerker =
Oberkategorie

Berufe II
Schlosser
Arzt
Angestellter
Lehrer
Bäcker

Angestellter =
keine Berufsart

Berufe III
Schlachter
Arzt
Bäcker
Lehrer
Metzger

Schlachter =
Metzger



Künstliche Kategorien

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

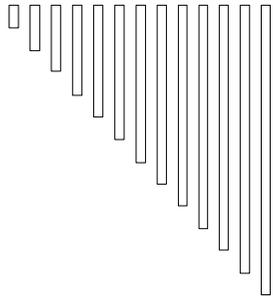
Kategorien aus quantitativen Merkmalen fassen Merkmalsausprägungen zu Gruppen zusammen.

□ Beispiel: Alter in Jahren

- Werte: 19, 22, 23, 24, 24, 24, 25, 28, 28, 30, 32, 38, 42, 49, 50, 51, 53, 53, 57, 57, 59

□ Zu entscheiden:

- Wie viele Kategorien?
- Von welcher Breite?



Vorgehen

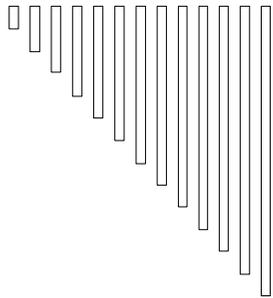
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Werte: 19, 22, 23, 24, 24, 24, 25, 28, 28, 30, 32, 38, 42, 49, 50, 51, 53, 53, 57, 57, 59

- Range der Werte bestimmen:
Kleinsten Wert (19),
größten Wert (59); range = 40
- Unterteilung in gleich breite Teilintervalle, z.B. 4 Kategorien der Breite 10; 2 Randkategorien
- Auszählen der Häufigkeiten

Jahre	Häufigkeit N (%)
< 20	1 (04.8)
20-29	8 (38.1)
30-39	3 (14.3)
40-49	2 (09.5)
50-59	7 (33.3)
≥ 60	0 (0)

Verteilung hat 2 Gipfel



Mögliche Probleme

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

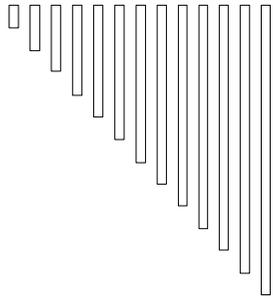
- Kategorien zu breit
Unterschiede in der Verteilung verdeckt

Jahre	Häufigkeit N (%)
< 40	12 (57.1)
≥ 40	09 (42.9)

Je nachdem gewinnt man Übersicht, aber verliert Information!
→ Die richtige Wahl der Breite ist entscheidend

- Kategorien zu eng
Häufungen weniger gut sichtbar

Jahre	Häufigkeit N (%)
< 20	1 (04.8)
20-24	5 (23.8)
25-29	3 (14.3)
30-34	2 (09.5)
35-39	1 (04.8)
40-44	1 (04.8)
45-49	1 (04.8)
...	...

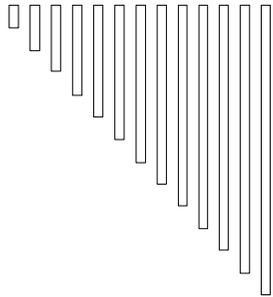


Ergebnisdarstellung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Kreuztabelle der Merkmale (hier: soziale Schicht und Art der Erkrankung), nach Gleis et al., 1973.

	Hohe soziale Schicht (%)	Niedrige soziale Schicht (%)
Psychische Störungen des höheren Lebensalters	44 (35.2)	53 (30.3)
Abnorme Reaktionen	29 (23.2)	48 (27.4)
Alkoholismus	23 (18.4)	45 (25.7)
Schizophrenie	15 (12.0)	23 (13.1)
Manisch depressive Leiden	14 (11.2)	6 (03.4)
	125 (100)	175 (100)



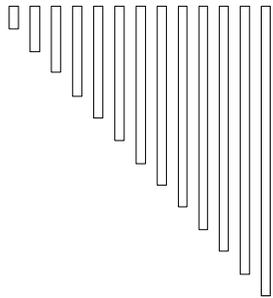
Ergebnisdarstellung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Kreuztabelle der Merkmale (hier: soziale Schicht und Art der Erkrankung), nach Gleis et al., 1973.

Analyse: Prüfung auf Verteilungsmerkmale z.B. mit Hilfe von χ^2 -Tests

Im Beispiel: Gleicht die Verteilung der Krankheiten bei niedriger sozialer Schicht der Verteilung bei hoher sozialer Schicht?



Einschätzen auf Ratingskalen

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Erfassung einer Merkmalsausprägung auf einer Skala mit markierten Abständen

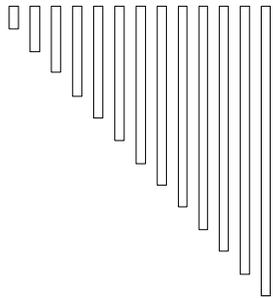
Beispiele: Bitte markieren Sie auf der Skala, was Ihrem Urteil nach zutrifft:

A. Herrn XY finde ich

unsympathisch -3 -2 -1 0 1 2 3 sympathisch

B. Partei XY halte ich in Wirtschaftsfragen für

wenig kompetent -3 -2 -1 0 1 2 3 sehr kompetent



Annahmen

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

A. Herrn XY finde ich

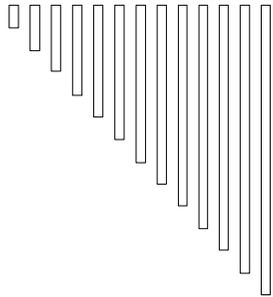
unsympathisch [-3] [-2] [-1] [0] [1] [2] [3] sympathisch

B. Partei XY halte ich in Wirtschaftsfragen für

wenig kompetent [-3] [-2] [-1] [0] [1] [2] [3] sehr kompetent

2 Annahmen

- Die Stufen werden als gleich groß interpretiert (→ Intervallskala)
- Die Skalenverankerung wird als gleich interpretiert

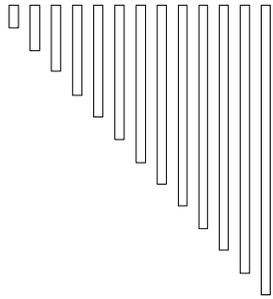


Arten von Ratingskalen

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Unterscheidung nach

- Art der Skalenverankerung
- Art der Skalenpolung
- Zahl der Abstufungen
- Art der Urteilsdimension



Art der Skalenverankerung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Zahlen

- Die Veranstaltung weckt mein Interesse an der Thematik.

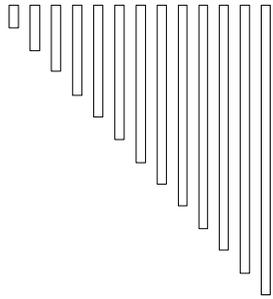
trifft nicht zu -3 -2 -1 0 1 2 3 trifft zu

Sprachliche Bezeichnungen

- Der Dozent leitet die Veranstaltung souverän.
Stimmt ... gar nicht teilweise völlig

Grafische Symbole

- Die Veranstaltung macht mir Spass.   

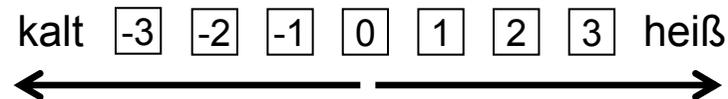


Art der Skalenpolung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

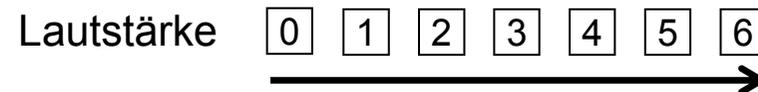
bipolar

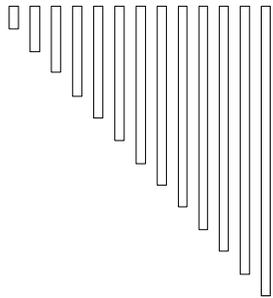
- Mit neutraler Mittelposition
- Betont die Gegensätzlichkeit der Skalenhälften



unipolar

- Keine neutrale Mittelposition





Zahl der Abstufungen

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

- Keine verbindliche Regel, meist 3 bis 7
- Gerade oder ungerade Zahl? Bei bipolaren Skalen macht es einen Unterschied. Bei gerader Zahl: keine neutrale Antwort!

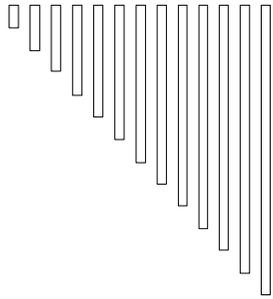
kalt heiß

kalt | heiß

- Wie viele Stufen? Je mehr Stufen, desto größer die verlangte Differenzierungsfähigkeit. Im Extrem: keine Stufen

kalt |-----X-----| heiß

- Fraglich, ob Personen so fein urteilen können.



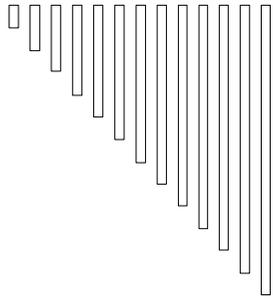
Art der Urteilsdimension

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Je nach Fragestellung

- Häufigkeit: nie, selten, manchmal, häufig, immer
- Intensität: sehr stark, stark, mittel, gering, sehr gering
- Zustimmung: stimme nicht zu ... stimme voll zu
- etc.

→ Die Skalenverankerungen sollten die gewünschte Dimension gut wiedergeben und leicht verständlich sein.



Mögliche Urteilsfehler I

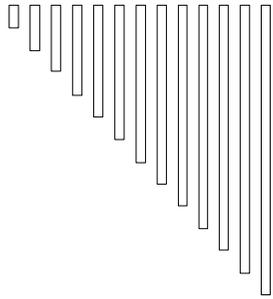
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Fehler, die den Umfang der Skala betreffen

- Boden- oder Deckeneffekt: Wurde die Skala bei den ersten Einschätzungen schon ausgeschöpft, können extremere Ausprägungen nicht mehr angegeben werden.
- Tendenz zur Mitte: Vermeidung extremer Urteile (weil unklar ist, ob man die ganze Skala ausschöpfen kann)

Abhilfe?

- Umfang der Skala durch Beispiele erklären



Mögliche Urteilsfehler II

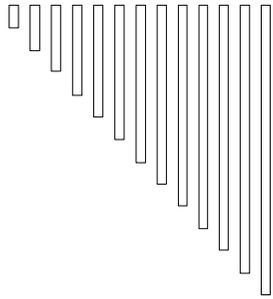
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Fehler aus dem Spektrum der Urteile

- Verzerrung durch ungleiche Merkmalsverteilung: Hat man z.B. viele positive Urteile und wenige negative, differenziert man bei den positiven stärker.
- Reihenfolgeeffekte: Wechsel von einem extremem zu einem gemäßigten Urteil führt zur Überbetonung des Kontrasts.

Abhilfe?

- Reihenfolge der Fragen variieren.



Mögliche Urteilsfehler III

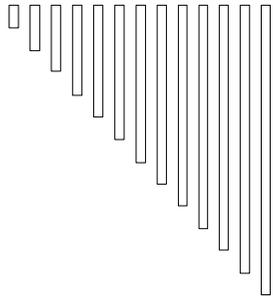
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Fehler bei der Beurteilung von Personen

- Halo-Effekt: Gesamteindruck beeinflusst die Beurteilung eines Einzelmerkmals (oder umgekehrt).
- Milde-Härte-Fehler: Manche Personen beurteilen systematisch zu positiv oder zu negativ.
- Rater-Ratee-Interaktion: Man bezieht in sein Urteil mit ein, wie man sich selbst beurteilen würde.

Vorbeugung?

- Betonung einer unverzerrten Beurteilung
- Hinweis auf die Gefahr von Fehlern
- Mögliche Konsequenzen der Beurteilung erläutern

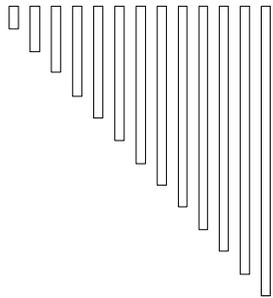


Ratingskala vs. Skala

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Ratingskala	Skala
<input type="checkbox"/> -3 <input type="checkbox"/> -2 <input checked="" type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	Härtegrade 1, 2, ... 9, 10
<input type="checkbox"/> Wird benutzt, um die Antwort für ein Item zu erfassen (Antwortskala) <input type="checkbox"/> Wird für ein einzelnes Item verwendet <input type="checkbox"/> Das Item ist nicht mit einem Skalenwert assoziiert <input type="checkbox"/> Wird passend für eine Frage gewählt	<input type="checkbox"/> Wird benutzt, um z.B. die Ausprägung einer Eigenschaft zu erfassen <input type="checkbox"/> Umfasst mehrere Items <input type="checkbox"/> Jedes Item hat einen Skalenwert <input type="checkbox"/> Resultiert aus einem Prozess der Skalierung

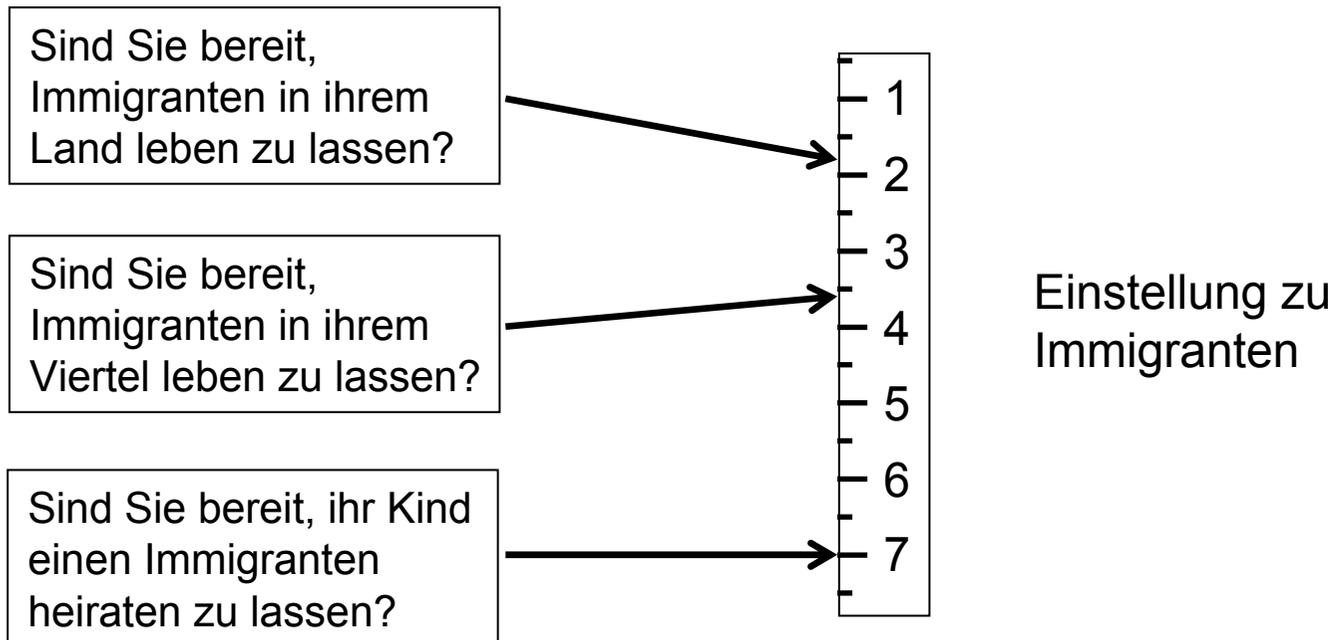
Z.B: Thurstone-, Likertskala

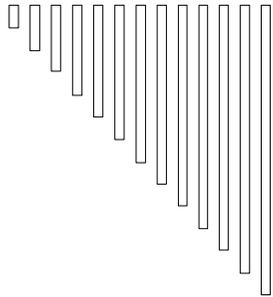


Skalierung

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Konstruktion einer Skala mit definierten Eigenschaften





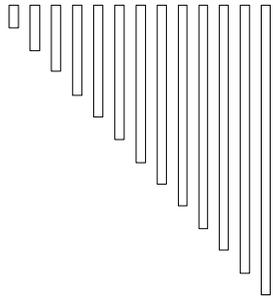
Thurstone-Skala



- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Einstellungsmessung (Thurstone & Cave, 1929)

- Schritt 1: Sammlung einer großen Zahl von Items, deren Bejahung auf eine unterschiedliche Ausprägung des Merkmals schließen lassen.
- Schritt 2: Experten stufen die Ausprägung des Merkmals ein, das mit einer Bejahung eines Item ausgedrückt wird (11-stufiges Rating).
- Schritt 3: Berechnung des Skalenwertes eines Items als Median der Expertenbeurteilungen.
- Wert für die Einstellung einer Person: Summe der Skalenwerte aller mit «ja» beantworteten Items



Thurstone-Skala

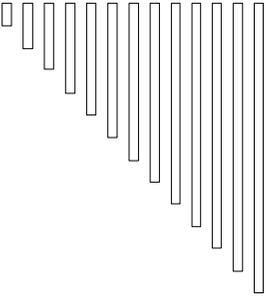


- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Beispiel: Nachbarschaftliche Kontakte in Siedlungen

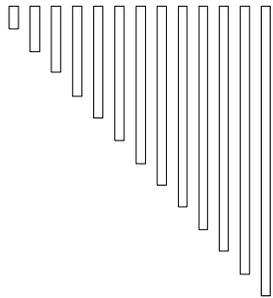
Expertenurteil: schlecht (-5), neutral (0), gut (+5)

- | | | |
|-------------------------------------|--|-------|
| <input checked="" type="checkbox"/> | Keinem Menschen in der Nachbarschaft würde etwas auffallen, wenn mir etwas zustieße. | -3.05 |
| <input type="checkbox"/> | Ich komme mir in dieser Siedlung oft vor wie ein Fremder. | -2.00 |
| <input checked="" type="checkbox"/> | Ich habe oft den Eindruck, dass sich die Menschen in meinem Bezirk nur flüchtig kennen. | -0.50 |
| <input type="checkbox"/> | Ich kenne hier in der Nachbarschaft fast jeden mit Namen. | +0.90 |
| <input type="checkbox"/> | In diesem Wohnbezirk ist es kaum möglich, sich auch nur für kurze Zeit von den anderen zurückzuziehen. | +1.79 |
| <input type="checkbox"/> | Hier in der Siedlung haben Menschen kaum Geheimnisse voreinander. | +3.30 |



Übersicht

- Von der Planung bis zum Rücklauf
- Kategorien und Ratings als Antwortformat
- Güte von Fragebogendaten
 - Objektivität
 - Reliabilität
 - Validität

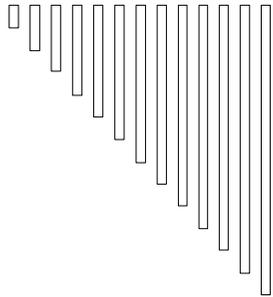


Objektivität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Anwenderunabhängigkeit des Messverfahrens

- Durchführung: Ergebnis unabhängig vom Anleiter
 - Standardisierte Instruktion
- Auswertung: Ergebnis unabhängig vom Auswerter
 - Aufgaben mit geschlossenem Antwortformat
 - Standardisierung der Auswertung (→ Software)
- Interpretation: Ergebnis unabhängig vom Beurteiler
 - Definition, was bestimmte Antworten bedeuten
 - Vorgabe von Normtabellen, z.B. für Altersstufen, Geschlechter oder Bildungsgruppen



Reliabilität

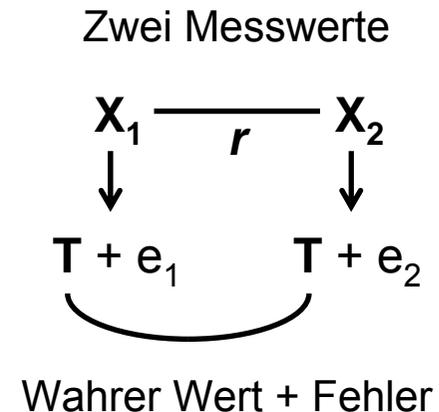
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

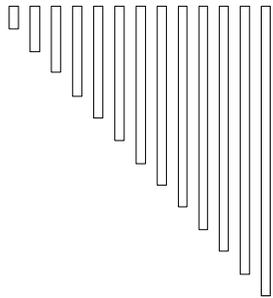
Zuverlässigkeit des Verfahrens

- Wird mittels Korrelation (r) bestimmt
- gut: $r > 0.9$; mittelgut: $r = 0.8-0.9$

Methoden

- Retest-Reliabilität (→ Testentwicklung)
- Paralleltest-Reliabilität (→ Testentwicklung)
- Interne Konsistenz (→ Skalenentwicklung; Test):
Wenn mehrere Items dasselbe messen sollen
 - Homogenität und Cronbach α



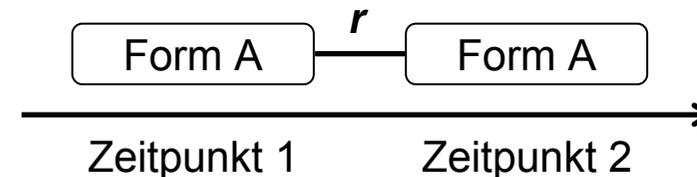


Retest-Reliabilität

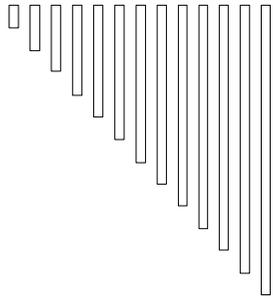
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Stabilität der Messung über die Zeit

- Derselbe FB wird 2mal vorgegeben
- Wie stark korrelieren die Werte?



- Führt eher zu einer *Überschätzung* der Reliabilität (z.B. wenn Antworten erinnert werden)
- Nicht einsetzbar bei Merkmalen, die sich in kürzerer Zeit ändern (Meinungen; Stimmungen)

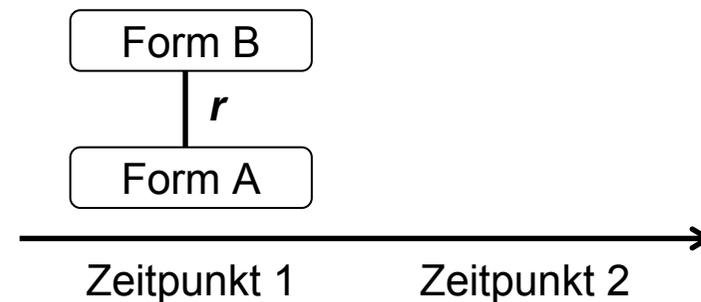


Paralleltest-Reliabilität

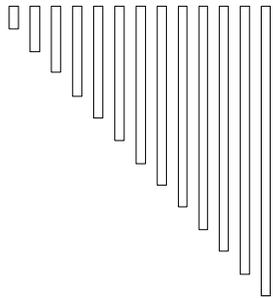
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Äquivalenz zweier Fragebogenversionen

- 2 analoge Versionen werden vorgegeben
- wie stark korrelieren die Werte?



- Zeitliche Veränderung kein Problem
- 2 Versionen = höherer Aufwand
- Es sind immer beide Versionen reliabel oder nicht reliabel



Homogenität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Durchschnittliche Korrelation zwischen allen Aufgabenpaaren

Item 1

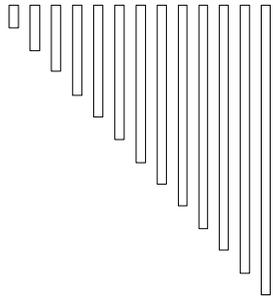
Item 2

Item 3

Item 4

	Item 1	Item 2	Item 3	Item 4
Item 1	1.00			
Item 2	0.89	1.00		
Item 3	0.91	0.92	1.00	
Item 4	0.88	0.93	0.95	1.00

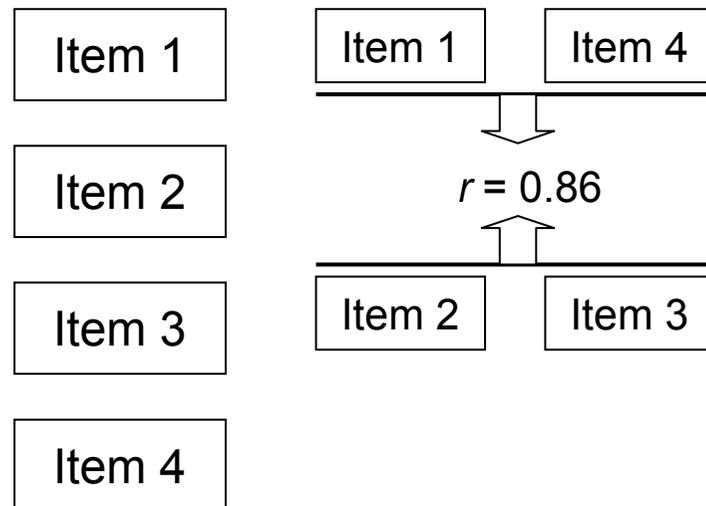
Homogenität $m = 0.91$



Testhalbierungsreliabilität

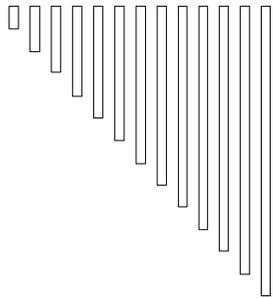
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Korrelation zwischen zwei Itemhälften



Je mehr Items, desto höher die Reliabilität
→ Halbe Testlänge
→ *Unterschätzung* der Reliabilität

Korrektur: Spearman-Brown-Formel



Cronbach alpha

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

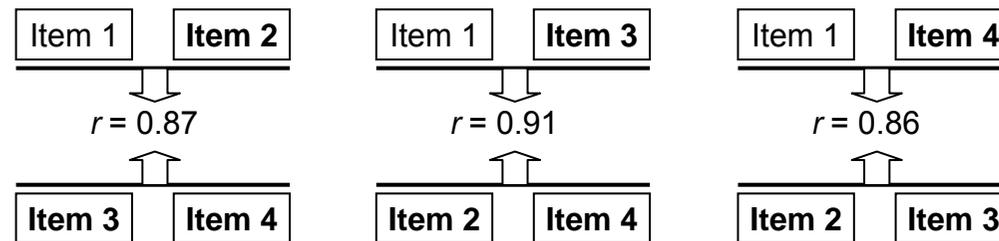
Durchschnittliche Korrelation aller Halbierungen

Item 1

Item 2

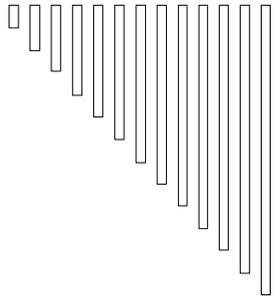
Item 3

Item 4



$$\alpha = (0.87 + 0.91 + 0.86) / 3 = 0.88$$

- *Unterschätzt* die Reliabilität
- Je mehr Items, desto höher α
- Je höher die Iteminterkorrelationen, (Homogenität) desto höher α



Validität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Misst der Fragebogen das, was er messen soll

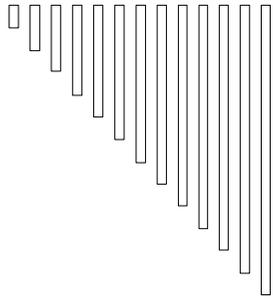
→ Objektivität und Reliabilität sind unwesentlich, wenn ein Messverfahren nicht valide ist

→ Wird meist ebenfalls mittels Korrelation (r) bestimmt

→ gut: $r > 0.6$; mittelgut: $r = 0.4-0.6$

Methoden

- Inhaltsvalidität
- Kriteriumsvalidität
- Konstruktvalidität



Inhaltsvalidität

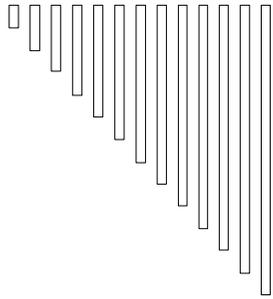
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Logische oder Augenscheinvalidität (*face-validity*)

→ Gegeben, wenn der Inhalt das zu messende Konstrukt in seinen wichtigsten inhaltlichen Aspekten erfasst

Beispiel: Ein Test der Grundrechenarten ist inhaltlich nicht valide, wenn er die Multiplikation nicht umfasst.

- Kann nicht numerisch bestimmt werden
- Ist eher eine Zielvorgabe für die Fragebogenentwicklung als ein Gütekriterium im engeren Sinne



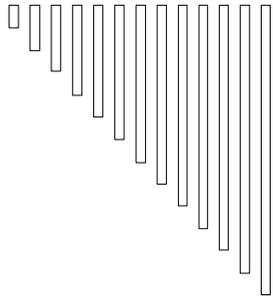
Kriteriumsvalidität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

→ Gegeben, wenn die Ergebnisse mit anderen Erhebungsmethoden für das Merkmal (= Kriterium) korrelieren

Beispiel: Der frühere Eignungstest zum Medizinstudium (auch ein Fragebogen) korrelierte positiv mit dem späteren Studienerfolg.

□ Problem: Es gibt nicht für jedes latente Merkmal ein gutes Außenkriterium (z.B. Intelligenz, Religiosität, etc).



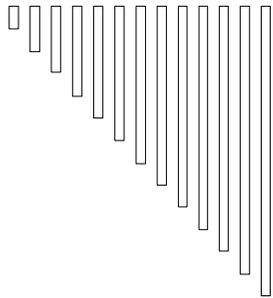
Konstruktvalidität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

→ Gegeben, wenn anhand der Ergebnisse theoretisch abgeleitete Zusammenhänge zu anderen Konstrukten bestätigt werden

Beispiel: Fragebogen zum Attributionsstil korreliert mit Angaben zu Emotionen, die auf entsprechenden Attributionen basieren.

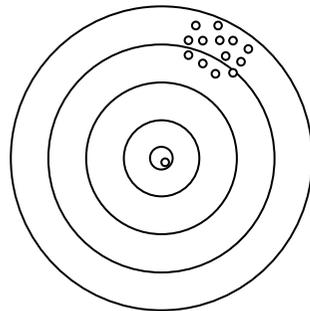
- Konvergente Validität: Wie gut kann ein Konstrukt durch mehrere Methoden gemessen werden?
- Diskriminante Validität: Wie gut können verschiedene Konstrukte durch eine Messung differenziert werden?



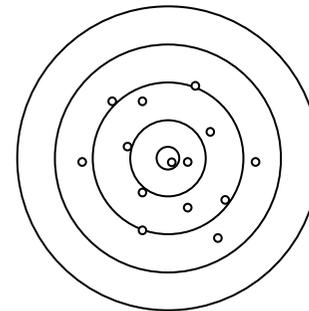
Reliabilität vs. Validität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

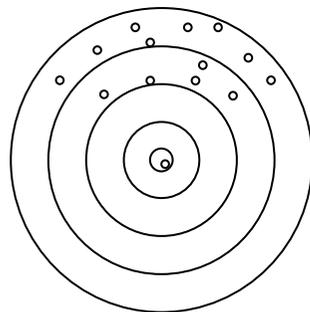
Ziel eines Sportschützen: Treffer im Zentrum



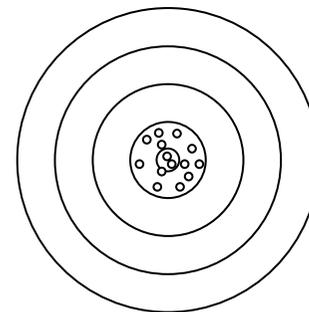
reliabel aber
nicht valide



valide aber
nicht reliabel

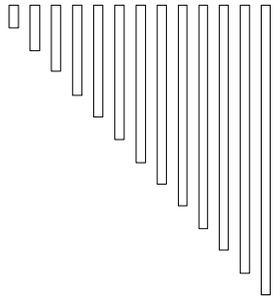


weder reliabel
noch valide



reliabel und
valide

Gütekriterien im Überblick



- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Gütekriterien des
Messinstruments

Gütekriterien der
Untersuchung

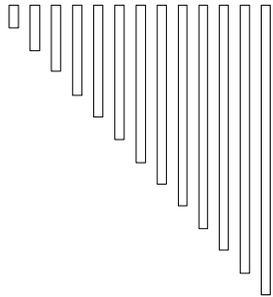
Objektivität des
Messinstruments

Reliabilität des Messinstruments

Validität des Messinstruments

Interne Validität der Untersuchung

Externe Validität der Untersuchung

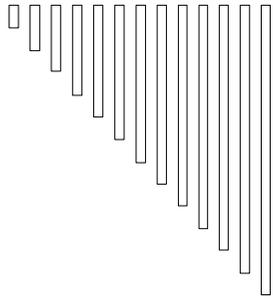


Interne vs. externe Validität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

- **Interne Validität:** → potentielle Störvariablen sind ausgeschlossen, → das Untersuchungsdesign erlaubt die intendierte Schlussfolgerung.
 - Wichtig hierfür: Design der Studie
 - Eingruppendesign (z.B. Umfrage)
 - Mehrgruppen-Design (z.B. Experiment)

- **Externe Validität:** → Ergebnisse sind auf andere Personen (Orte/Zeiten) verallgemeinerbar.
 - Wichtig hierfür: Repräsentativität



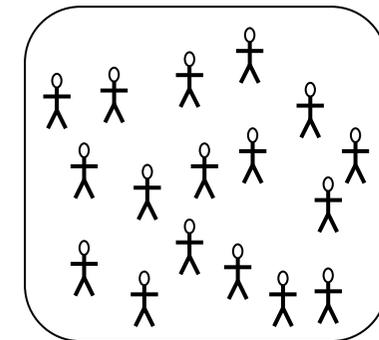
Vollerhebung

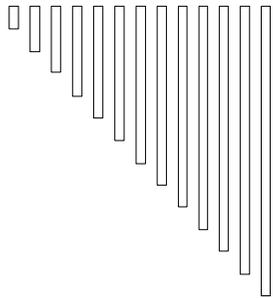
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Population: Gesamtheit aller Personen, über die eine Aussage gemacht werden soll.

- **Vollerhebung:** Untersuchung der gesamten Population (z.B. eine Schulklasse)

→ Die Repräsentativität bzgl. der Personen steht nicht in Frage.



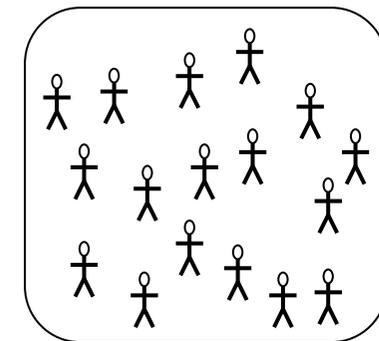


Stichprobe

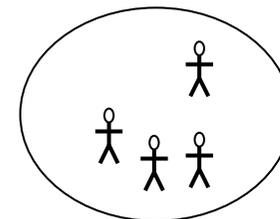
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

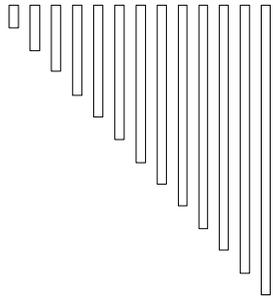
Population

- **Stichprobe:** Untersuchung eines Ausschnitts der Population
 - Ist die Stichprobe repräsentativ?
 - Wie genau lassen sich die Verhältnisse in der Population dabei schätzen?



Rück-
schluss ↑ ↓ Aus-
wahl

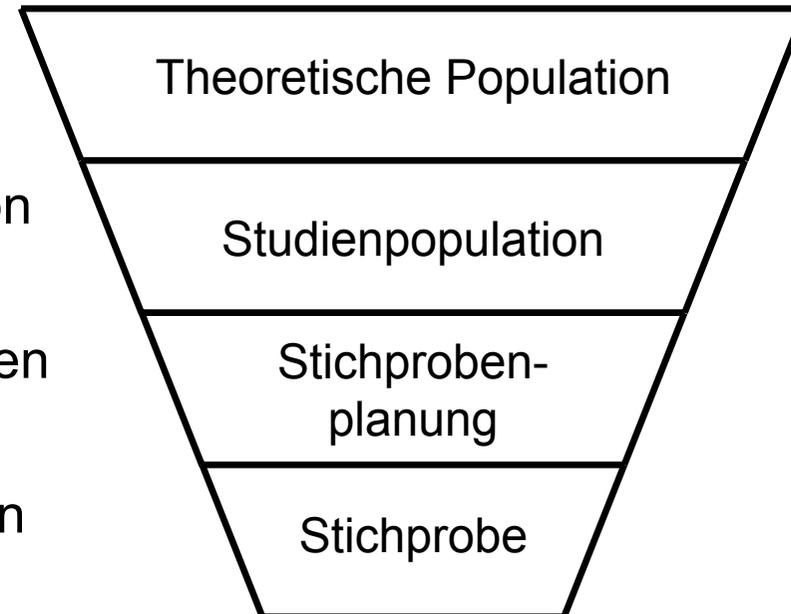


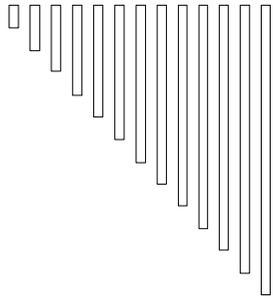


Population und Stichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

- Auf wen wollen Sie generalisieren?
- Auf welche Population haben sie Zugriff?
- Welchen Anteil können Sie erreichen?
- Müssen Sie Personen aussondern?



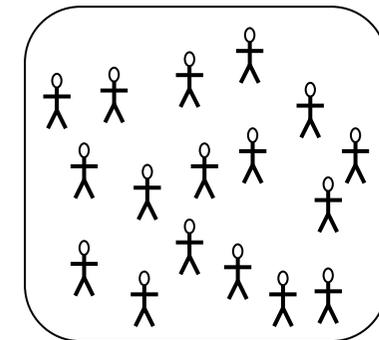


Was ist Repräsentativität?

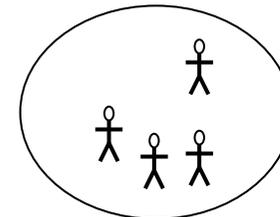
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

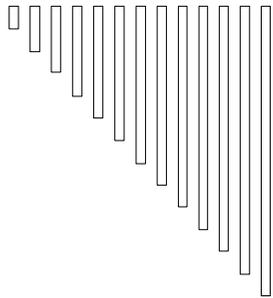
Allgemein: Die Stichprobe gibt die Verhältnisse in der Population im Hinblick auf ein interessierendes Merkmal möglichst unverzerrt wieder.

- Spezifische Repräsentativität: Die Stichprobe entspricht der Population nur in wenigen Merkmalen
- Globale Repräsentativität: Die Stichprobe entspricht der Population in nahezu allen Merkmalen



Rück-
schluss ↑ ↓ Aus-
wahl



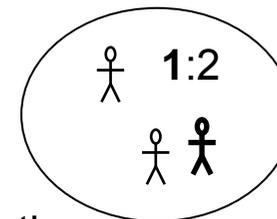
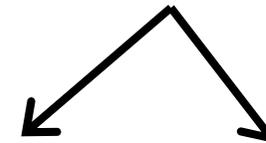
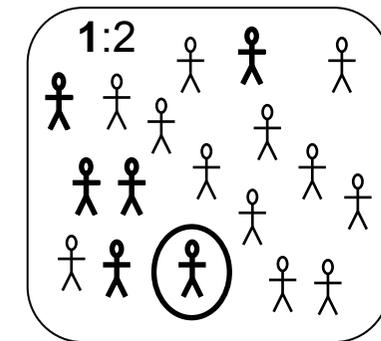


Stichprobengröße

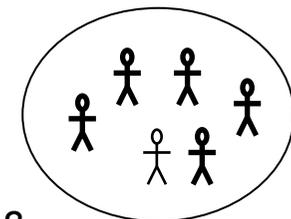
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Häufige Meinung: „Die untersuchte Stichprobe ist so klein, das Ergebnis kann nicht repräsentativ sein!“

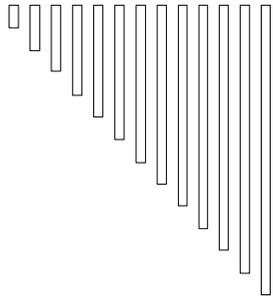
- Verwechslung von Repräsentativität und Stichprobengröße
- Große Stichproben sind nicht generell repräsentativer als kleine
- Problematisch für die Repräsentativität: systematische Auswahlfehler



klein und
repräsentativ



groß
nicht repräsentativ



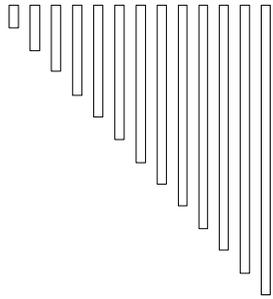
Herstellung von Repräsentativität

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Definition über das Auswahlverfahren

Eine Stichprobe ist dann repräsentativ, wenn das Auswahlverfahren keine Elemente der Population in Bezug auf die interessierenden Merkmale bevorzugt.

- **Wie stellt man dies sicher?** → Durch Zufallsauswahl
Für die probabilistische Stichprobe gilt
 - Alle Personen haben dieselbe Chance, in die Stichprobe zu kommen
 - Es gibt keine systematischen Auswahlfehler
 - Die Ergebnisse lassen sich generalisieren



Einfache Zufallsstichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Charakterisierung: Jedes Element der Population kommt mit gleicher Wahrscheinlichkeit in die Stichprobe

Realisierung

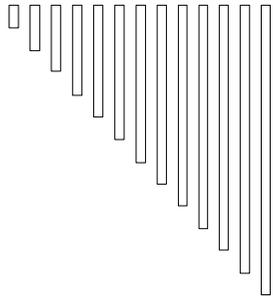
- (1) Herstellung einer vollständigen Liste aller Elemente
- (2) Auswahl: Los / Zufallszahlen



Auswahl

Zufallszahlen		
884	860	263
006	429	845
252	512	148
153	682	214





Einfache Zufallsstichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Voraussetzungen

- Alle Elemente der Population müssen für die Auswahl zugänglich sein und teilnehmen

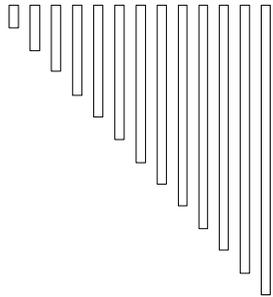
Eigenschaften

- Global repräsentativ
- Genauigkeit der Stichprobenwerte lässt sich mathematisch berechnen (→ Vertrauensintervalle z.B. für Mittelwerte), hierfür ist die Stichprobengröße wichtig



Auswahl





Geschichtete Zufallsstichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

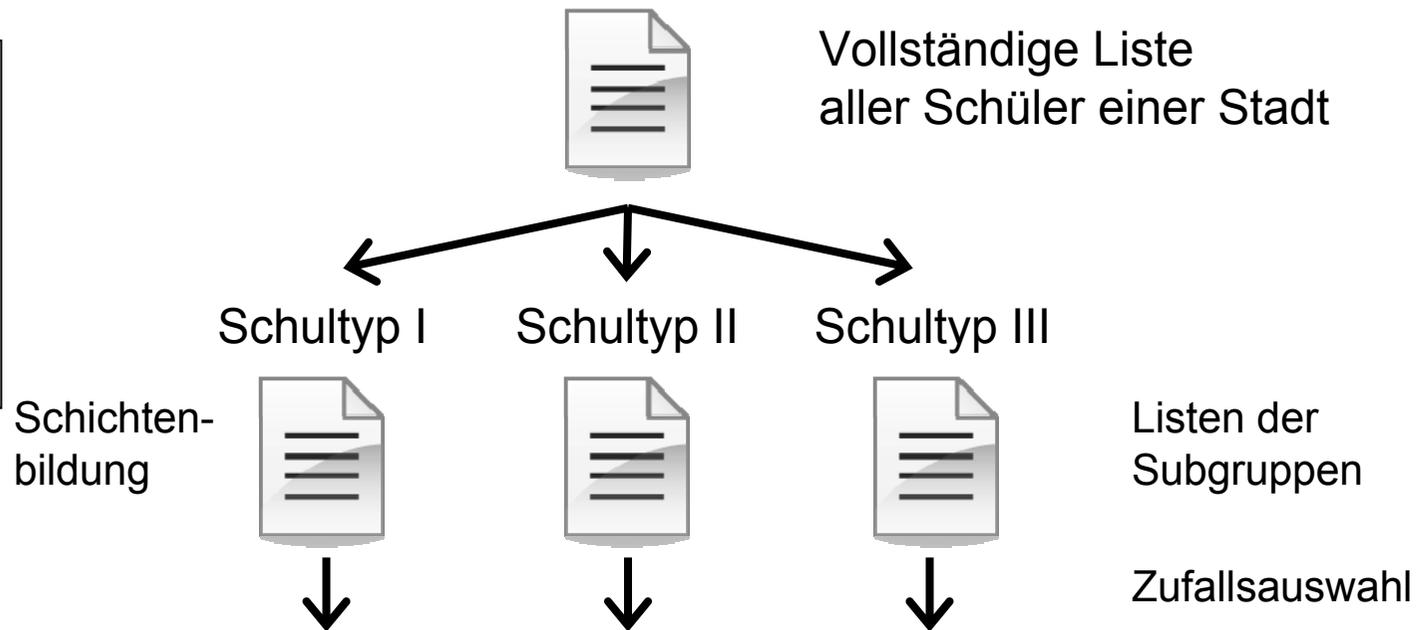
Charakterisierung: Die Population wird anhand relevanter Merkmale in Gruppen (Schichten) eingeteilt. Aus jeder Schicht wird eine Zufallsstichprobe proportional zur Gruppengröße gezogen.

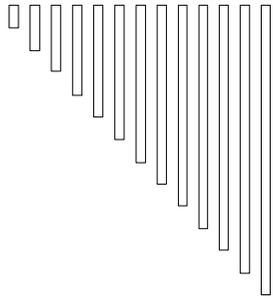
Realisierung

- (1) Herstellung einer vollständigen Liste aller Elemente
- (2) Aufteilung der Elemente in Schichten
- (3) Zufallsauswahl pro Schicht

Geschichtete Zufallsstichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten





Geschichtete Zufallsstichprobe

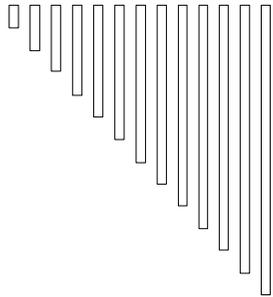
- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Voraussetzungen

- Alle Elemente der Population müssen für die Auswahl zugänglich sein und teilnehmen
- Schichtungsverteilung muss bekannt sein

Eigenschaften

- Spezifisch repräsentativ (im Hinblick auf das Schichtungsmerkmal)
- Genauere Schätzungen der Populationswerte möglich



Zufällige Klumpenstichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Charakterisierung: Von natürlichen Gruppen (Klumpen) wird eine Zufallsauswahl von Klumpen getroffen; diese werden vollständig untersucht

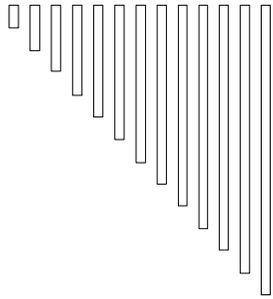
Beispiel: Population: Schüler einer bestimmten Schule;
Klumpen: Einzelne Schulklassen

Voraussetzungen

- Alle Klumpen für die Auswahl zugänglich; alle Personen nehmen teil

Eigenschaften

- Ökonomisch! Aber: nur eingeschränkt repräsentativ



Ad hoc Stichprobe

- Planung und Rücklauf
- Kategorien und Ratings
- Güte der Daten

Charakterisierung: Willkürliche Aufnahme von Personen nach Verfügbarkeit und freiwilliger Teilnahme

Voraussetzungen

Keine

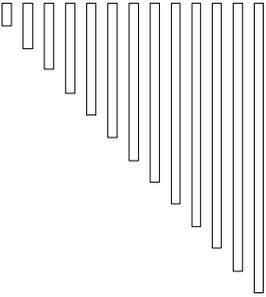
Eigenschaften

Sehr ökonomisch!

Nicht repräsentativ

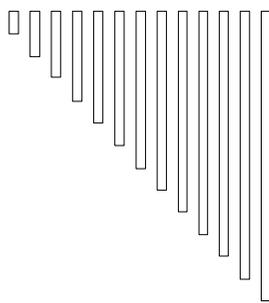
→ Leider Grundlage vieler Untersuchungen

→ Gefährdung der externen Validität einer Untersuchung



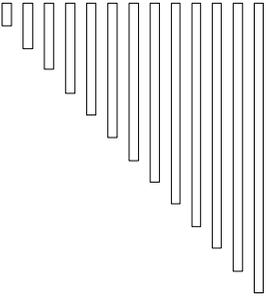
Übersicht

- Von der Planung bis zum Rücklauf
- Kategorien und Ratings als Antwortformat
- Güte von Fragebogendaten
- Fazit



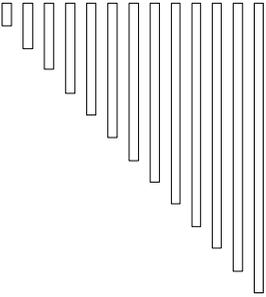
Fazit I

- Bei der Konstruktion von Fragebögen sind vielfältige Entscheidungen zu treffen
→ Konsequenzen für das Ergebnis
- Kategoriensysteme müssen sorgfältig gestaltet sein.
- Subjektive Einschätzungen werden mit (Rating-/Antwort-) Skalen erhoben.
- Skalierungen sind hingegen nach festen Regeln konstruierte Messinstrumente.



Fazit II

- Die Güte von Fragebögen bemisst sich an ihrer Objektivität, Reliabilität und Validität.
- Repräsentativität ist entscheidend für stichprobenbasierte Studien.
- Für repräsentative Stichproben lässt sich die Genauigkeit von Parameterschätzungen berechnen (→ Vertrauensintervalle).
- Anhand des Rücklaufs versucht man Schlüsse über die Repräsentativität zu ziehen.



Literatur

- Beller, S. (2008). *Empirisch forschen lernen*. Bern: Huber. (Neuaufgabe in Vorbereitung)

