

howtodo.at

Einleitung

Seit Jahren kommen wiederholt Studierende auf mich zu, haben Fragen zu ihren empirischen Erhebungen, zu Details statistischer Datenanalyse oder zum Einsatz von SPSS. Über mehrere Jahre hinweg entstanden ist das Buch „*How to do Empirie, how to do SPSS*“, das 2016 erschienen ist. Das Buch hat sich gut verkauft, war letztendlich ausverkauft.

Eine zweite (überarbeitete) Auflage wäre leicht, aber nicht besonders herausfordernd gewesen. Viel spannender war es, How to do weiterzuentwickeln – mit allen Erfahrungen und Rückmeldungen, die seit dem Erscheinen gesammelt werden konnten.

Gemeinsam mit facultas entstand ein neues Konzept: Wir teilen das „Doppel-Buch“ der Erstauflage und lassen daraus zwei eigenständige Publikationen entstehen, eine für Empirie und eine für Statistik und Datenanalyse (mit SPSS).

Warum das?

Die Nutzung des Buchs hat sich oft auf einen der beiden Teile konzentriert. Viele müssen empirische Sozialforschung zunächst KONZIPIEREN. Datenanalyse liegt dabei noch in der Ferne.

Erst dann, wenn die AUSWERTUNG ansteht, rücken Statistik und Tools wie SPSS näher.

Einleitung /.2

Deshalb versorgen wir unsere Zielgruppe(n) jetzt „phasengerecht“, in „dünneren Häppchen“. Das verbindende Element bleibt natürlich bestehen! Beide Bücher sind nach wie vor so gestaltet, dass sie ein großes Ganzes ergeben – einzeln oder kombiniert verwendbar.

Es gibt unterschiedlichste Literatur, die Detail-Fragen beantwortet. Es bedeutet aber meist viel Recherche, Antworten zu finden und zu interpretieren. Meine beiden Bücher haben den Anspruch, das bereits gemacht zu haben, die wichtigen Fragen zu beantworten. Die Bücher haben viele Abbildungen, Querverweise, ein gutes Stichwortverzeichnis und begleitende Übungsdaten im Netz.

Bereits nach zwei Jahren war es an der Zeit, neue Zweitauflagen zu entwickeln.

Die Bände wurden dort und da – Anregungen der Leserschaft folgend – geschärft, noch besser strukturiert und durchgehend farbig gedruckt. Alle Links sind nun über QR-Codes direkt abrufbar, die Zitierweise folgt der aktuellen APA 7-Norm.

Empirie-, Statistik- und SPSS-Slides

Die folgenden Slides sind eine Ergänzung zu den Inhalten meiner beiden Bücher

- Braunecker, C. (2023). *How to do empirische Sozialforschung. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. ISBN 978-3-8252-6160-3

und

- Braunecker, C. (2023). *How to do Statistik und SPSS. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. ISBN 978-3-8252-6161-0

Warum gibt's diese Slides online?

Die Folien sind gedacht als praktisch verwendbarer Themenkurzüberblick. Oder als animierte Haupt-, Übergangs- oder Zwischenslides für eigene Darstellungen – probieren Sie es ab Seite 4 einmal im Präsentationsmodus aus ;-). Auf wiederholte Anregung der interessierten Kollegenschaft stelle ich sie als „offene“, frei weiterverwend- und -gestaltbare Powerpoint-Datei (sowie als pdf) zur Verfügung.

Sehr freuen würde mich, wenn meine beiden Bücher **Einzug in Ihre Literaturliste** (im Idealfall als **Pflichtliteratur**) bzw. **in die Buchsammlung studentischer Leserschaft** finden.

Daneben wäre auch eine **gute Rezension** zu dem einen oder anderen Band eine nette Sache 😊!

Forschen – einfach wie ein Kochrezept

Ich möchte **Schnitzel** zubereiten ...

... **aus Soja** mit
Semmelbrösel und **Butterschmalz** ...

... **heute Abend**, bei mir **zu Hause** ...

... **drei** Schnitzel ...

... **in** meiner extra dafür gekauften **neuen Bratpfanne** ...

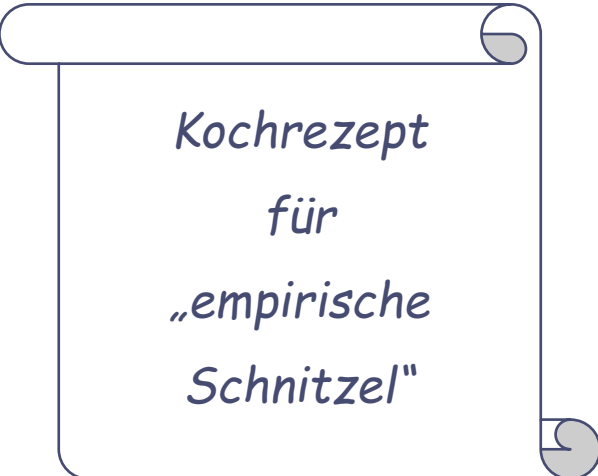
Damit nichts schief läuft, **probiere ich die neue Bratpfanne** bereits **morgens aus**.

Am Nachmittag **überprüfe ich die Zutaten**.

Am Abend **koche ich**,

lade dazu meine beiden besten Freunde ein ...

... und **decke den Tisch** festlich.

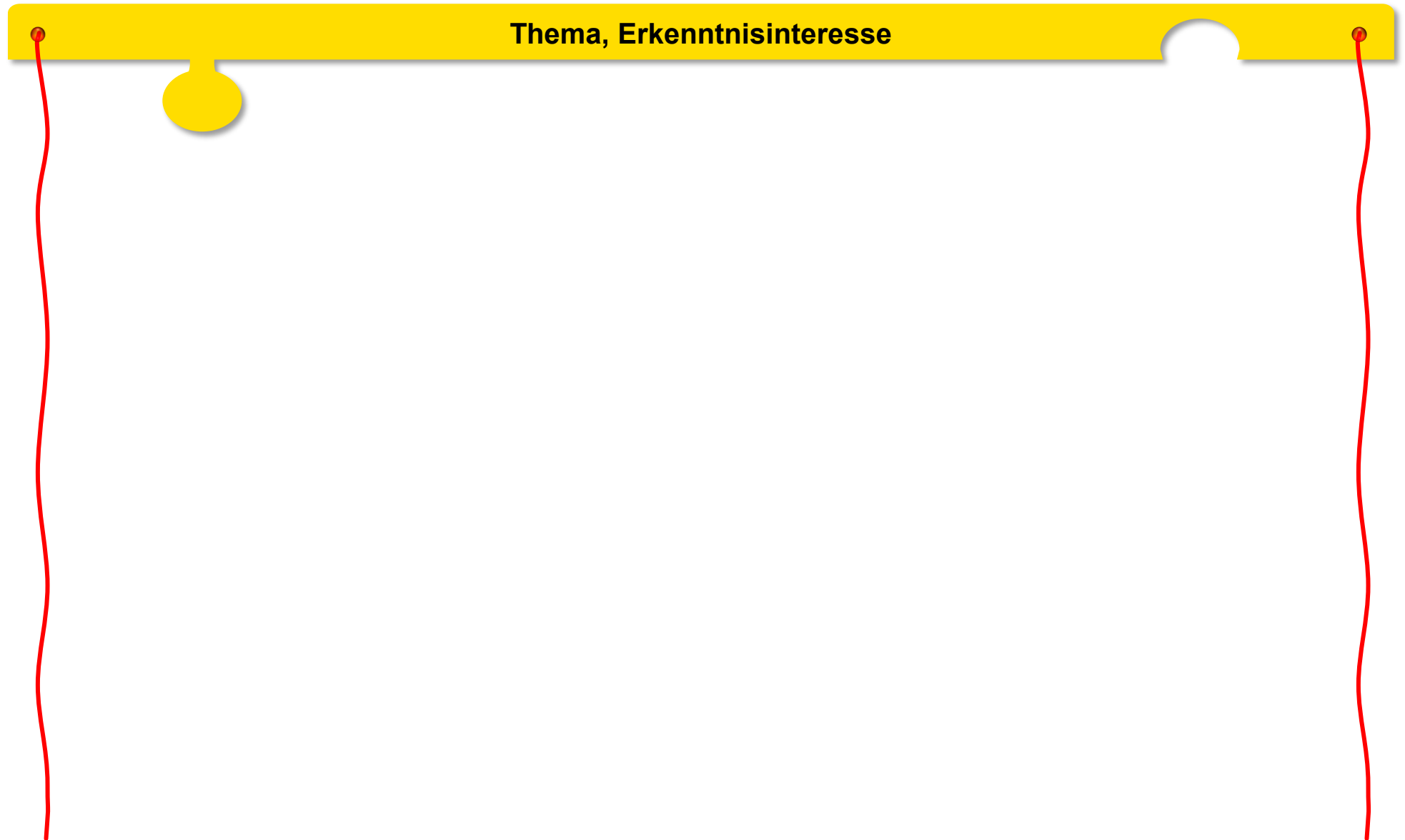


*Kochrezept
für
„empirische
Schnitzel“*

Forschen – Zutaten der Reihe nach und ineinander verwoben!



Empirischer Forschungsprozess



Keine Empirie ohne präzises Erkenntnisinteresse (Thema)

Ein klar und exakt formuliertes Erkenntnisinteresse (Thema)
ist die Ausgangsbasis jeder empirischen Erhebung!

- Nicht zu breit formuliert:
„Online-Befragung zu gebackenem Essen“

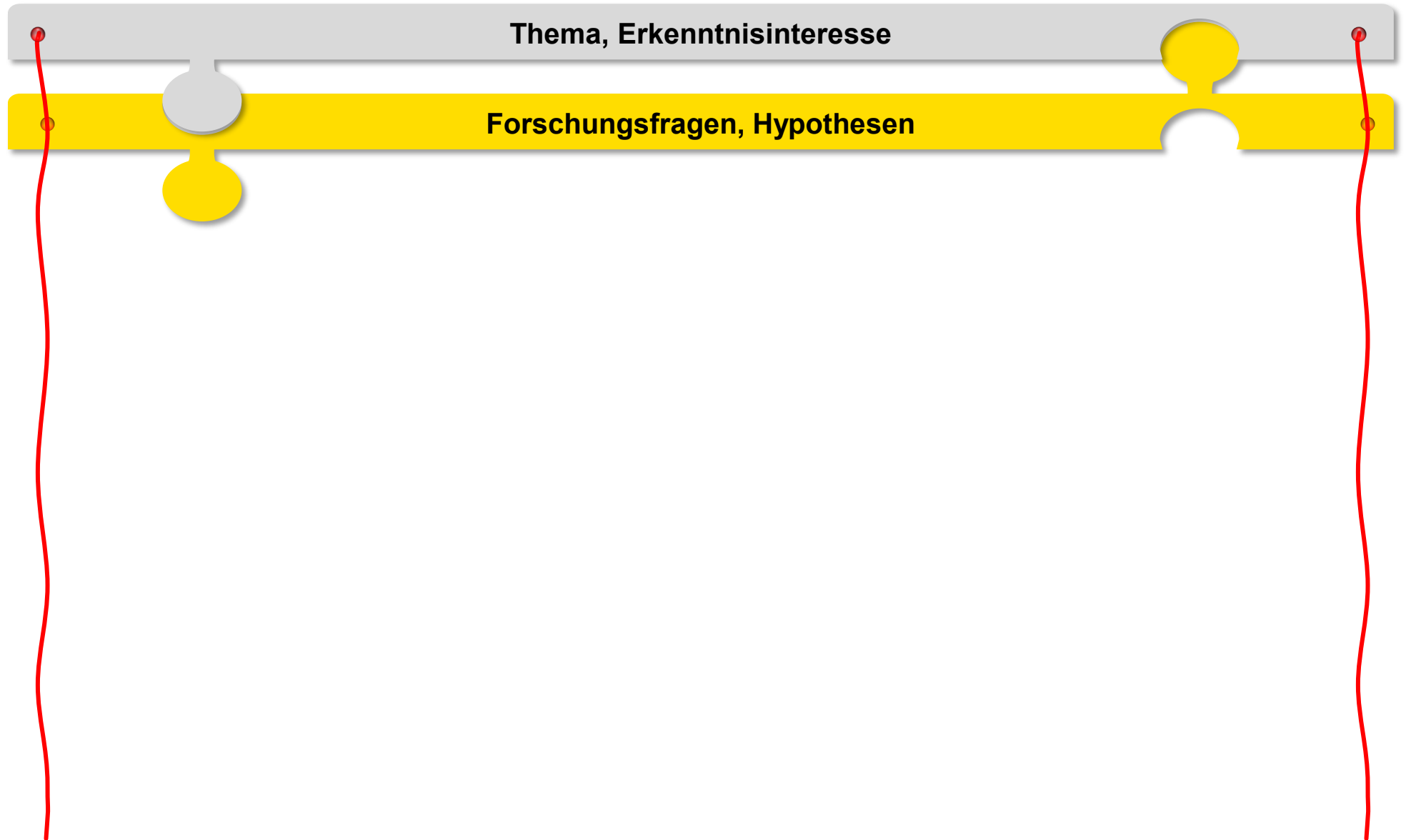
Besser konkreter:

*„Konsumgewohnheiten der deutschen Wohnbevölkerung
in Bezug auf Schnitzel“*

- ... die inhaltliche Klammer,
an der die Forschungsfragen und Hypothesen andocken.



Empirischer Forschungsprozess



Forschungsfragen und Hypothesen sind unverzichtbar

Forschungsfragen und/oder Hypothesen stehen mit der Erhebung und deren Instrument(en) in Einklang und Wechselwirkung.

- Erst aus den Forschungsfragen bzw. Hypothesen heraus ergeben sich die konkreten Fragen im Fragebogen: Diese werden „ab und zu“ etwas „daran vorbei“ formuliert ;-)) – oder sind überhaupt „zuerst da“ ...
- Ein Fragebogen benötigt **nicht WENIGER**, auch **nicht MEHR**, **sondern GENAU SO VIELE** Fragen wie zur Abdeckung der Forschungsfragen und/oder Hypothesen erforderlich sind.
- Nicht jede Forschungsfrage benötigt eine Hypothese, nicht jede Hypothese eine Forschungsfrage.
- Formulierung:
„Welche(r)“ oder „Wie“ bei Forschungsfragen,
„Wenn, dann“ oder „Je, desto“ bei Hypothesen.



Forschungsfragen und Hypothesen

Forschungsfragen

- drücken ein neutrales Erkenntnisinteresse in Frageform aus
- definieren die genauen Inhalte und Formulierungen im Erhebungsinstrument (Fragebogen, Leitfaden, Codierschema, Beobachtungsprotokoll)
- ▶ **Empirie beantwortet** Forschungsfragen.

Hypothesen

- stellen Annahmen bzw. Behauptungen auf, die auf Basiswissen (aus Vorerhebungen, Literatur usw.) beruhen
- sind „vermutete Antworten auf Forschungsfragen“ *)
- sollten nicht bloße Aussagen sein
 - empfohlen: Konditional- („Wenn, dann“) und Vergleichssätze („Je, desto“)
- ▶ **Empirie prüft** vorab formulierte Hypothesen (quantitativ) oder aus den **empirischen Erkenntnissen** werden Hypothesen formuliert (qualitativ).

*) vgl. Herczeg, P. & Wippersberg, J. (2021). *Kommunikationswissenschaftliches Arbeiten. Eine Einführung* (2. Auflage). facultas/utb. S. 93.



Hypothesen (inhaltlich, nicht statistisch)

- ▶ **Hypothesen sind Vermutungen über Ergebnisse von Datenerhebungen.**
- ▶ Eine Hypothese formuliert das, was zu untersuchen ist, als (wissenschaftlich) theoriegestützt überprüfbare Aussage.
 - Eine Hypothese kann gerichtet (= präziser) und ungerichtet sein.
 - Eine Hypothese sollte möglichst kurz und leicht fassbar sein.
 - Eine (SOWI-)Hypothese ist eine Wahrscheinlichkeitsaussage.
 - Eine Hypothese zielt auf neuartige Erkenntnisse ab.
 - Eine Hypothese kann nie „vollständig“ verifiziert, höchstens momentan „gestützt“ werden. Niemals können alle denkbaren Möglichkeiten der Überprüfung ausgeschöpft werden.



Inhaltliche und statistische Hypothesen

Inhaltliche Hypothesen

- ▶ formulieren **inhaltliche Zusammenhänge**:
 - Wenn Personen älter als 40 sind, dann essen sie eher Schnitzel, als wenn sie jünger sind.

Statistische Hypothesen

- ▶ formulieren **Operationalisierung** und **statistische Kennwerte**:
 - **Nullhypothese**: Der **Prozentanteil** von Personen, die **mehrmals wöchentlich Schnitzel** essen, **ist** unter Personen, die älter als 40 sind, **gleich (ähnlich)** wie unter jüngeren Personen.
 - **Alternativhypothese**: Der **Prozentanteil** von Personen, die **mehrmals wöchentlich Schnitzel** essen, **ist** unter Personen, die älter als 40 sind, **höher** als unter jüngeren Personen.
- ▶ **In der Regel** werden **inhaltliche UND statistische** Hypothesen aufgestellt.



Präzisierung, Hierarchisierung, Detaillierung

- 1. Welche altersspezifischen Unterschiede treten beim Schnitzelkonsum **generell** auf?
 - 1.1. Welche altersspezifischen Unterschiede treten beim Konsum von **Gemüseschnitzeln** auf?
 - 1.2. Welche altersspezifischen Unterschiede treten beim Konsum von **Sojaschnitzeln** auf?
 - 1.3. Welche altersspezifischen Unterschiede treten beim Konsum von **Hühnerschnitzeln** auf?
 - 1.4. Welche altersspezifischen Unterschiede treten beim Konsum von **Schweinsschnitzeln** auf?
 - 1.5. usw.

	Altersunterschiede			
Gemüse	%	%	Ø	Ø
Soja	%	%	Ø	Ø
Huhn	%	%	Ø	Ø
Schwein	%	%	Ø	Ø

- **Erhebungsvariablen:**
 - **Alter** – z.B. „bis 40“ Jahre | „älter“ oder offen gefragt |__|__|
 - **Konsumation generell** von Schnitzel aus
 - Gemüse ☐ Soja ☐ Huhn ☐ Schwein ☐ usw.
 - oder **Anzahl** (in den letzten ...) gegessener **Schnitzel** aus
 - Gemüse |__|__| Soja |__|__| Huhn |__|__| usw.



Präzisierung, Hierarchisierung, Detaillierung

- 1. Wenn Personen älter als 40 sind, essen sie **EHER Schnitzel**, als wenn sie jünger sind.
 - 1.1. Wenn Personen älter als 40 sind, essen sie eher **Gemüseschnitzel**, als wenn sie jünger sind.
 - 1.2. Wenn Personen älter als 40 sind, essen sie eher Schnitzel **aus Soja**, als wenn sie jünger sind.
 - 1.3. Wenn Personen älter als 40 sind, essen sie eher Schnitzel **vom Huhn**, als wenn sie jünger sind.
 - 1.4. Wenn Personen älter als 40 sind, essen sie eher Schnitzel **vom Schwein**, als wenn sie jünger sind.
 - 1.5. usw.

	bis 40	41 plus
Gemüse	%	%
Soja	%	%
Huhn	%	%
Schwein	%	%

- **Erhebungsvariablen:**
 - **Alter** – bis 40 Jahre – älter als 40 Jahre
 - **Konsumation (generell, z. B. in den letzten 12 Monaten)**
 - von Schnitzel aus – Gemüse ☐ – Soja ☐
 - Huhn ☐ – Schwein ☐ usw.



Präzisierung, Hierarchisierung, Detaillierung

- 1. Wenn Personen älter als 40 sind, essen sie monatlich **DURCHSCHNITTlich MEHR** Schnitzel, als wenn sie jünger sind.
 - 1.1. Wenn Personen älter als 40 sind, essen sie monatlich Ø mehr **Gemüseschnitzel**, als wenn sie jünger sind.
 - 1.2. Wenn Personen älter als 40 sind, essen sie monatlich Ø mehr Schnitzel **aus Soja**, als wenn sie jünger sind.
 - 1.3. Wenn Personen älter als 40 sind, essen sie monatlich Ø mehr Schnitzel **vom Huhn**, als wenn sie jünger sind.
 - 1.4. Wenn Personen älter als 40 sind, essen sie monatlich Ø mehr Schnitzel **vom Schwein**, als wenn sie jünger sind.
 - 1.5. usw.

	bis 40	41 plus
Gemüse	Ø	Ø
Soja	Ø	Ø
Huhn	Ø	Ø
Schwein	Ø	Ø

- **Erhebungsvariablen:**
 - **Alter** – bis 40 Jahre – älter als 40 Jahre
 - **Anzahl** in den letzten **30 Tagen** gegessener **Schnitzel**
 - aus – Gemüse |__|__| – Soja |__|__|
 - Huhn |__|__| – Schwein |__|__| usw



Präzisierung, Hierarchisierung, Detaillierung

1. Je älter Personen sind,
DESTO MEHR Schnitzel essen sie.

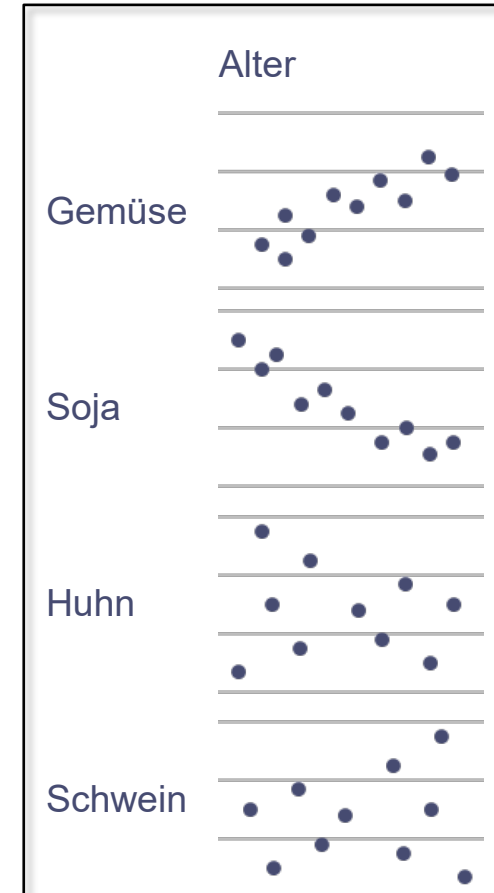
1.1. Je älter Personen sind, desto mehr essen sie
Gemüseschnitzel.

1.2. Je älter Personen sind, desto mehr essen sie
Schnitzel **aus Soja**.

1.3. Je älter Personen sind, desto mehr essen sie
Schnitzel **vom Huhn**.

1.4. Je älter Personen sind, desto mehr essen sie
Schnitzel **vom Schwein**.

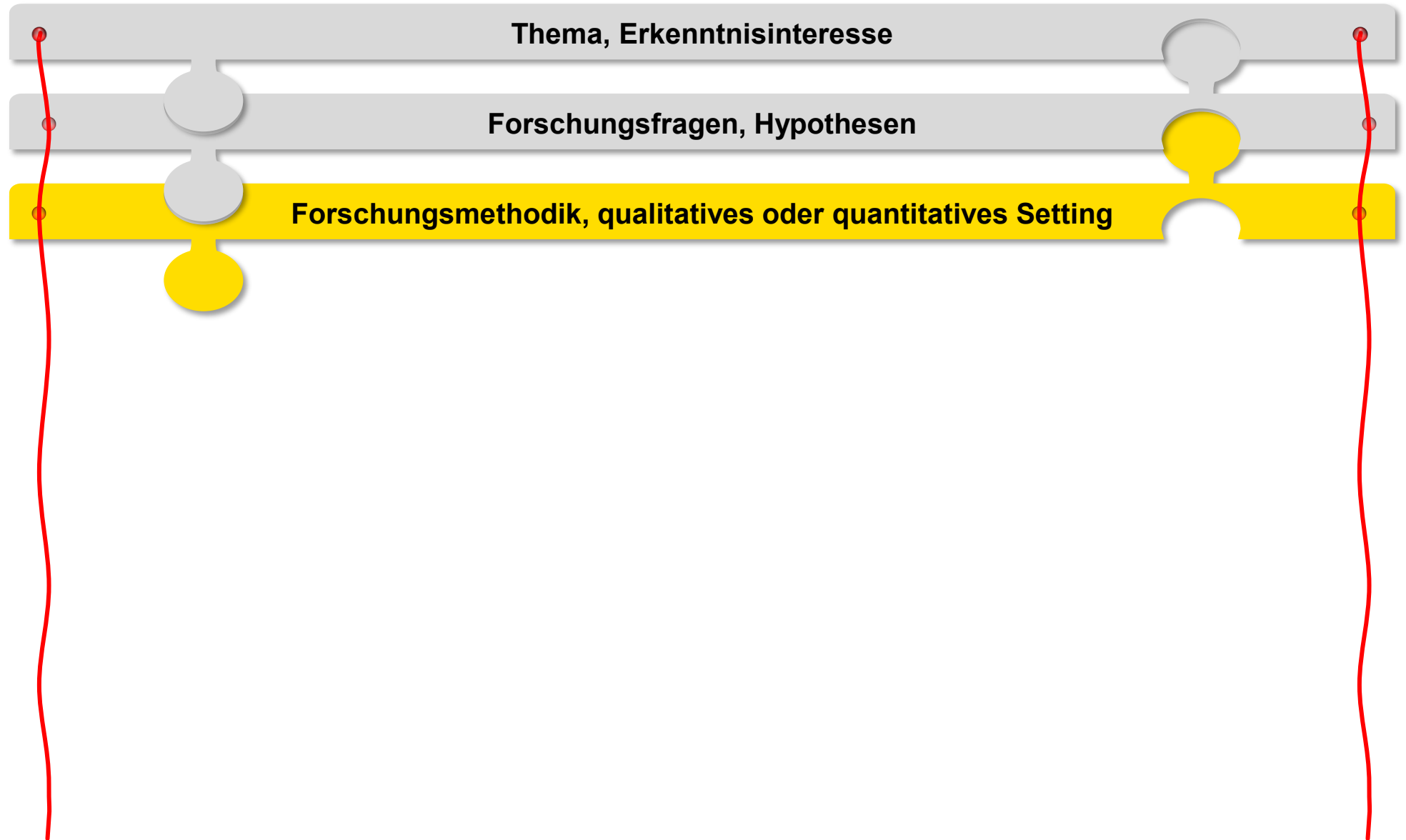
1.5. usw.



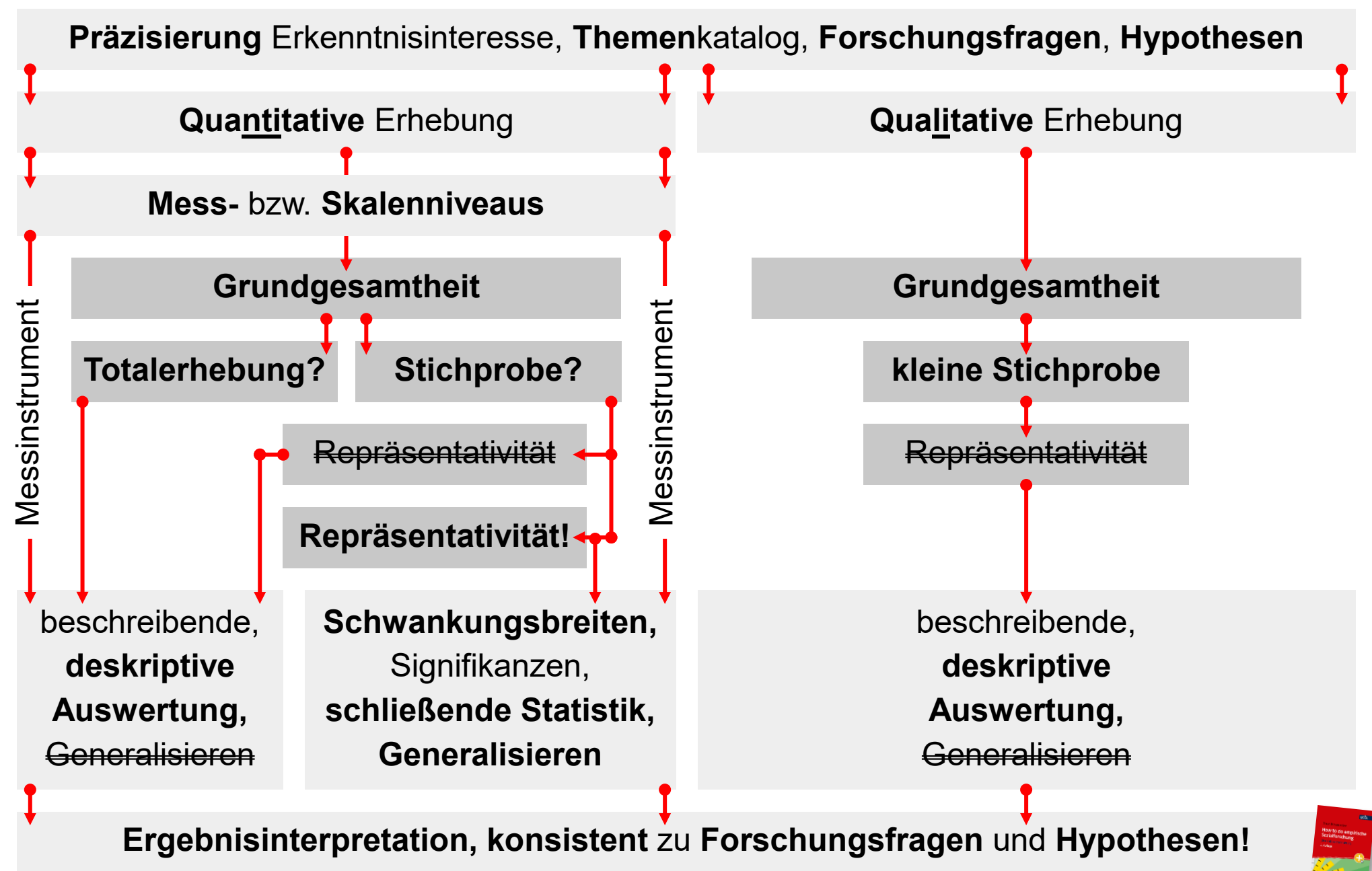
- **Erhebungsvariablen:**
- **Alter** offen gefragt |__|__|
 - **Anzahl** in den letzten **30 Tagen** gegessener **Schnitzel**
 - aus – Gemüse |__|__| – Soja |__|__|
 - Huhn |__|__| – Schwein |__|__| usw.



Empirischer Forschungsprozess



Gesamtzusammenhänge empirischer Forschung



Vgl.: Braunecker, C. (2023). *How to do empirische Sozialforschung. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb.



Qualitativ ≠ Quantitativ

QUALITATIV

- Beschreiben, Motive, inhaltliche Tiefe
- bei neuen Themen, zur Erforschung von Dimensionen
- vor allem individuelle und psychologische Aspekte
- überschaubare Erhebungsanzahl
- kein einheitlicher Fragebogen (dafür Leitfaden)
- verbalisierte Ergebnisse

Statistik

QUANTITATIV

- Ergebnisse auf breiter Basis
- zahlenmäßige Interpretationen
- Abtestung „bekannter“ Zustände
- große Menge an Erhebungen
- standardisierter Fragebogen oder Erhebungsprotokoll
- Prozentzahlen, Mittelwerte, Kennzahlen

Statistik



Inhaltsanalyse

- **Zählen oder Bewerten von** Aussagen oder anderen klar definierten **Inhalten**
- in Print-, elektronischen und Online-Medien
- Codebuch (Analyseeinheiten, -inhalte, Codierungsregeln) und Codierschema

! Pretest !

! INTER- und INTRA-Coder-Reliabilität !

Codierschema (beispielhafter Auszug)								
Codierer_ID	Medium	Datum	Uhrzeit	Seite	Platzierung	Anzahl_Wort _ 'Schnitzel'	Kontext	usw.
4711	Zeitung A	01.01.		1	Titel	5	positiv	
4711	Zeitung B	01.01.		10	Chronik	1	negativ	
0815	abc.de	02.01.	12:01		Startseite	7	neutral	
0815	xyz.at	02.01.	18:20		Startseite	12	positiv	
0815	Zeitung A	03.01.		5	Politik	3	neutral	
usw.								



Beobachtung

- **objektiviertes Erfassen** von Situationen, Handlungen und Verhaltensweisen
- Beobachtungsbogen • „teilnehmend“ oder „von außen“
- Eyetracking • Mystery-Tests • Laborbeobachtung • Feldbeobachtung

• oft auch **in Kombination mit Befragungen**

Beobachtungsbogen für ein Einkaufszentrum (beispielhafter Auszug)

Immer zur vollen Stunde ausfüllen! Nur den Schnitzelstand beobachten!

Beobachter:in
Datum
Uhrzeit
Standort Portal
Bereich Süd <input type="checkbox"/>
Bereich West <input type="checkbox"/>

Anzahl Personen beim Schnitzelstand

Männer, erwachsen	<input type="text"/> <input type="text"/> <input type="text"/>
Frauen, erwachsen	<input type="text"/> <input type="text"/> <input type="text"/>
Jugendliche, etwa 12 bis 18	<input type="text"/> <input type="text"/> <input type="text"/>
Kinder, unter 12	<input type="text"/> <input type="text"/> <input type="text"/>

Anzahl Personen mit ...

keiner Einkaufstasche	<input type="text"/> <input type="text"/> <input type="text"/>
einer Einkaufstasche	<input type="text"/> <input type="text"/> <input type="text"/>
zwei bis drei Einkaufstaschen	<input type="text"/> <input type="text"/> <input type="text"/>
mehr als drei Einkaufstaschen	<input type="text"/> <input type="text"/> <input type="text"/>

Anzahl Personen, die im Moment ...

langsam am Gang gehen, ohne beim Schnitzelstand stehenzubleiben	<input type="text"/> <input type="text"/> <input type="text"/>
langsam am Gang gehen, beim Schnitzelstand stehenbleiben und schauen	<input type="text"/> <input type="text"/> <input type="text"/>
schnell am Gang gehen, ohne beim Schnitzelstand stehenzubleiben	<input type="text"/> <input type="text"/> <input type="text"/>

usw.



Gruppendiskussion (Fokusgruppe)

- **Diskussionsrunde** („runder Tisch“) von rund zehn Personen
- Dauer ein bis zwei Stunden, gelenkt durch Moderation
- Leitfaden

- homogene oder heterogene Teilnehmerschaft ► je nach Themenstellung
- gute Möglichkeit, Gegenstände oder Infomaterial direkt miteinzubeziehen
- Gruppensituation: wechselseitige Anregungen im Gespräch, Abbau von Hemmungen
- Aufzeichnung elektronisch ► dann Transkription ► dann Auswertung

Diskussionsleitfaden für eine Fokusgruppe (beispielhafter Auszug)

Warm-Up-Phase (~ 10 min.)

- Begrüßung, Vorstellung, für die Teilnahme bedanken
- Informationen über den Ablauf der Diskussion: keine richtigen oder falschen Antworten, jede Meinung willkommen, alles anonym, rund 90 Minuten Dauer
- Kurze Vorstellungsrunde der Teilnehmer:innen (Vorname, Alter, Bezug zum Erhebungsgegenstand)

Teil 1: Schnitzel-Konsumgewohnheiten unter der Woche (~ 15 min.)

Wenn Sie am Abend ein Wiener Schnitzel essen: Wie fühlen Sie sich da? Hat das Auswirkungen auf Ihren Tagesablauf? Welche? Läuft ein Wochentag anders ab, wenn Sie am Abend Schnitzel essen?

usw.



Qualitative Einzelbefragung

- **Einzelgespräche** aus Befragtenperspektive
- völlig freier Gesprächsverlauf oder zumindest frei formulierbare Fragen
- ausschließlich oder überwiegend offene Fragen

Bandbreite von völlig narrativ:

„Denken Sie an gestern. Wie war der Tag? Erzählen Sie mir, was Sie vom Aufstehen an alles gemacht haben, bis Sie am Abend das Schnitzel gegessen haben.“



... bis zu strukturiertem Fragebogen, qualitativ:

„Wenn Sie am Abend ein Schnitzel essen: Wie geht es Ihnen dabei?“



Ergebnis = Inhalte:

„Gestern habe ich mich seit dem Aufwachen auf das Schnitzel am Abend gefreut, musste den ganzen Tag daran denken“ • „Mit einem Schnitzel verbinde ich Freizeit“ • „Schnitzel ist Genuss“ • „Zuerst wird gevöllert, danach kommt die Reue“ • „viel zu fett“



Umfrage, Befragung (quantitativ)

- meist **strukturiertes**, (voll) **standardisiertes Interview**
- Formulierungen/Fragereihenfolge vorgegeben, KEIN Spielraum für Befragende
- PAPI • CATI • CAWI • Mehrthemenumfrage MTU • Panel • Tracking

(fast) ausschließlich geschlossene Fragen:

„Essen Sie zumindest gelegentlich Schnitzel?“

☐ JA ☐ NEIN

„WENN JA: Wie schmeckt Ihnen Schnitzel?“

Vergeben Sie bitte eine Schulnote:“

① ② ③ ④ ⑤ ⑥

offene Fragen meist nur als

„Sonstiges, und zwar: _____“ oder „Warum: _____“ 

► Bereits VOR der Erhebung ist viel an inhaltlichem Wissen notwendig

Ergebnis = Zahlen:

„Unabhängig vom Wohnort gaben in Österreich zwei Drittel an, zumindest ab und zu gerne Schnitzel zu essen. In Deutschland traf dies insgesamt nur auf jede(n) Zweite(n) zu – im Süden allerdings deutlich öfter als im Norden [angenommene Ergebnisse].“



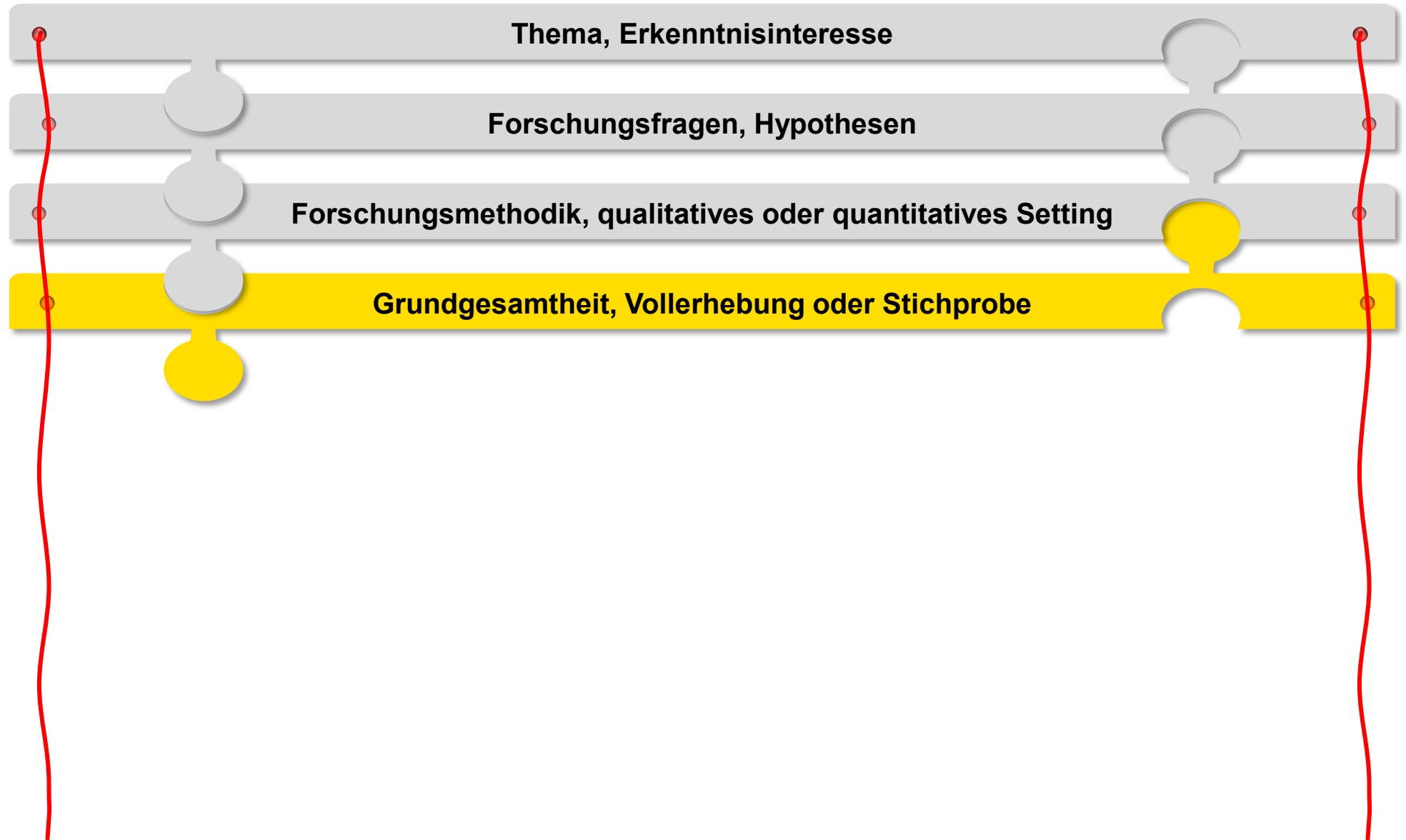
Experimentelles Design

- Analyse von Ursache-Wirkung-Beziehungen (**Kausalzusammenhängen**)
 - Versuchs- und Kontrollgruppe
 - keine Störvariablen (einflusstragenden Nebeneffekte) !
 - unabhängige Variable (Ursache) VOR abhängiger Variable (Wirkung) !
 - Beobachtungs- • Befragungsexperiment • Labor- • Feldexperiment

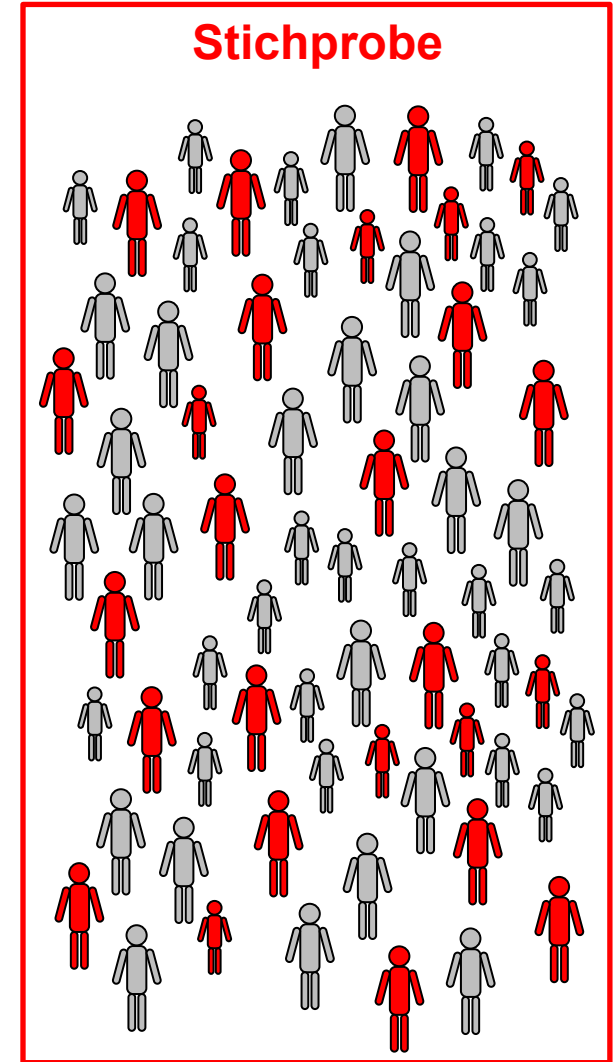
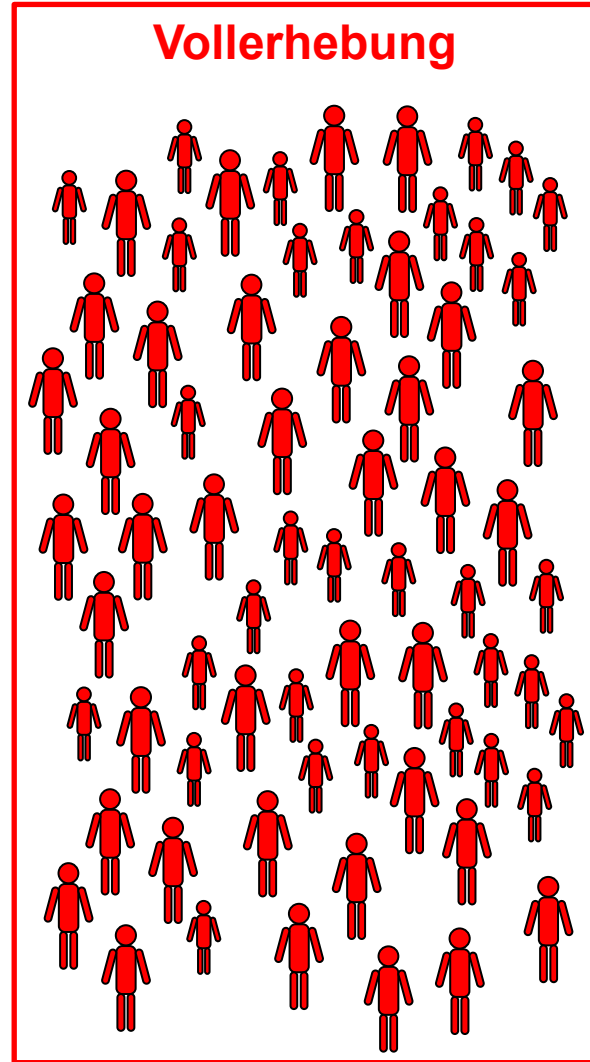
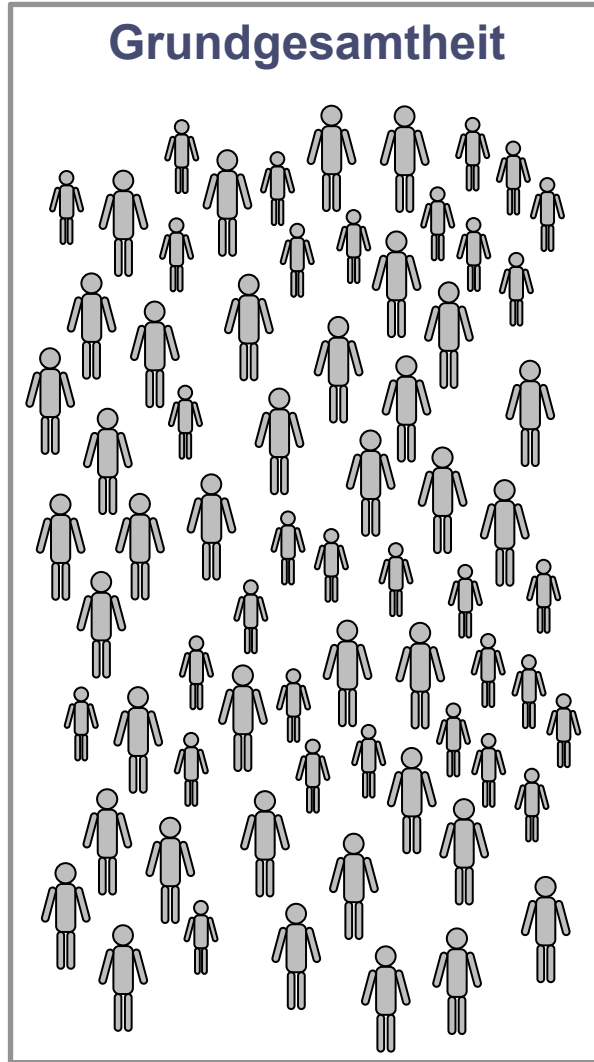
Mögliche Designs			Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇
	Einzelfall-Studie	EG	U	M					
	Pre-Test/Post-Test	EG	M	U	M				
	Zeitreihe	EG	M	M	M	U	M	M	M
	Mehrfall-Studie	EG (rand)	M	U ₁	M	Z _{1 usw.} = Zeitpunkt der Messung M = Messung			
		EG (rand)	M	U ₂	M				
	Post-Test mit Kontrollgruppe	EG (rand)	U	M	EG = Experimentalgruppe KG = Kontrollgruppe				
		KG (rand)	keine U	M					
	Pre- und Post-Test mit Kontrollgruppe	EG (rand)	M	U	M	(rand) = Randomisierung U = Ursache (Stimulus)			
KG (rand)		M	keine U	M					



Empirischer Forschungsprozess



Vollerhebung oder (Zufalls?-)Stichprobe?



!!! genau definieren !!!

► Struktur bekannt?

► adressierbar?

☹ Zeit, Kosten

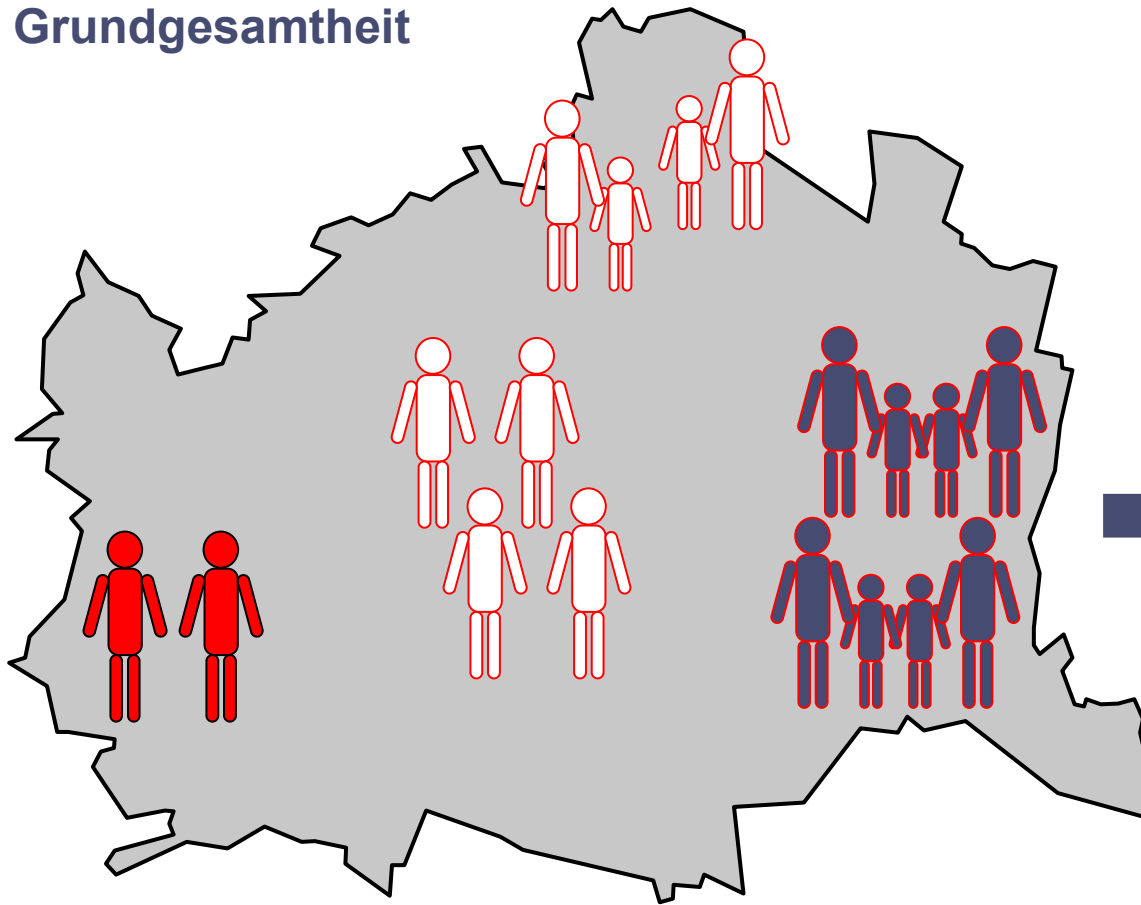
☹ Datenschutz!

☹ Verweigerungen?

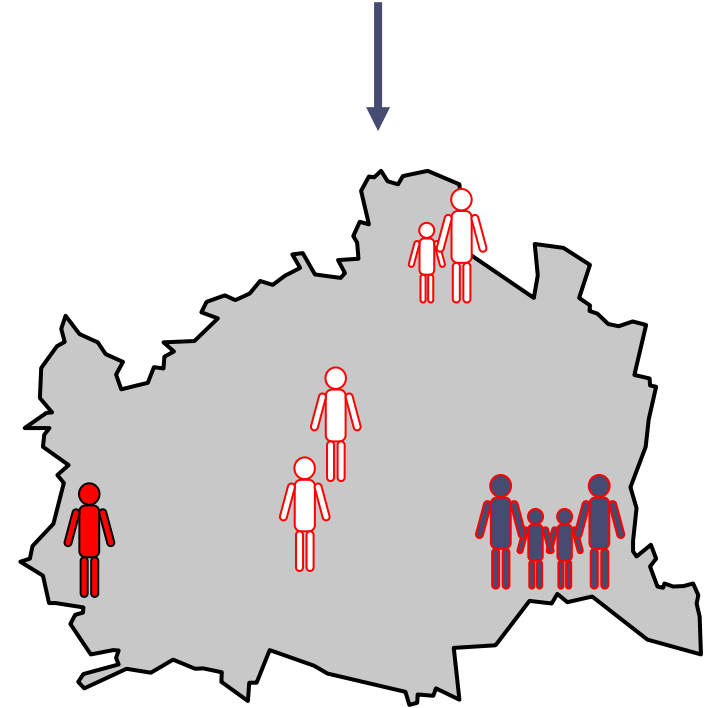


Repräsentativität

Grundgesamtheit



repräsentative
Stichprobe



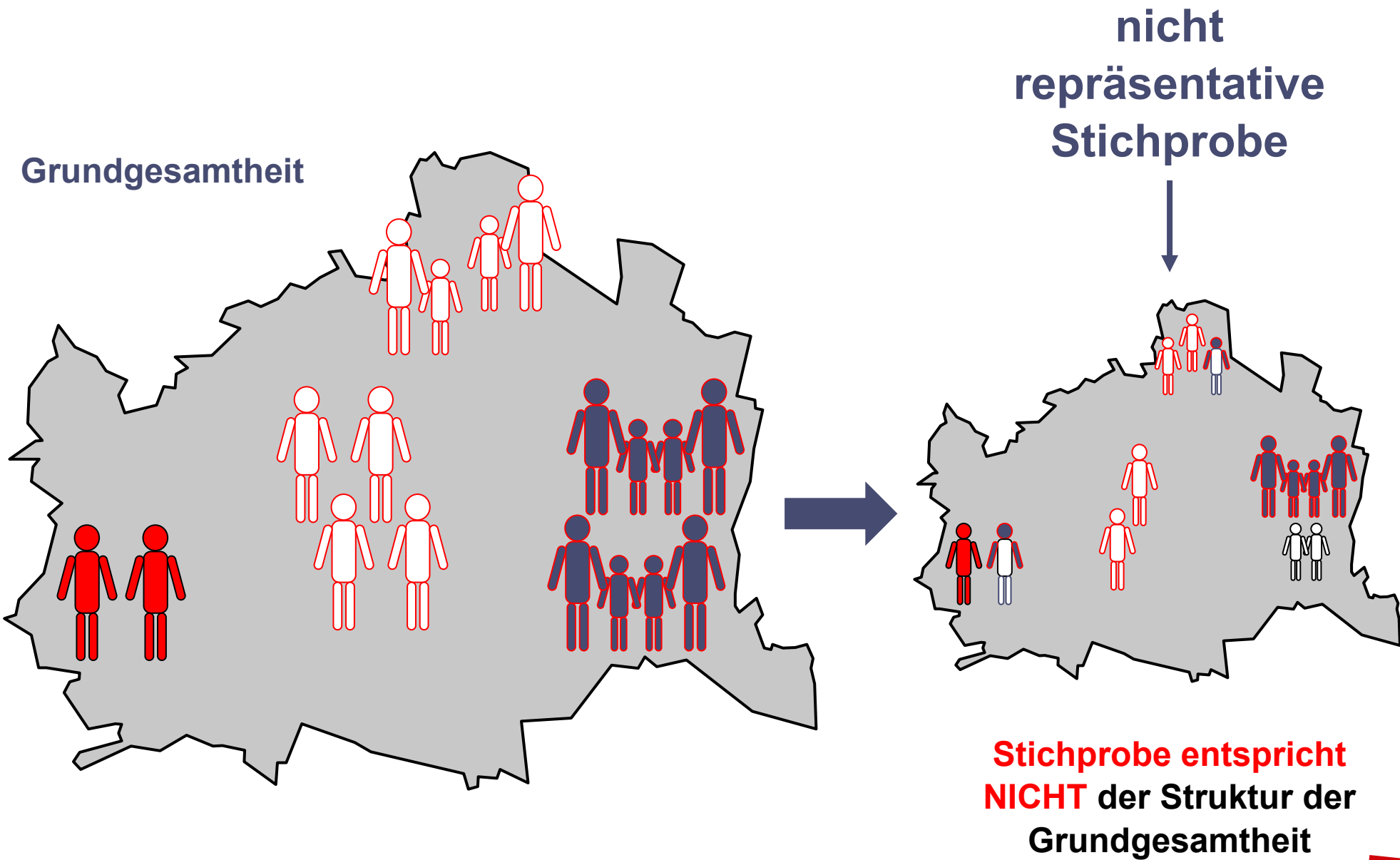
- Repräsentativität \neq Stichprobengröße!
- Hybrid-Samples für „On- und Offliner“ ...

strukturell **exaktes, kleines
Abbild** der Grundgesamtheit!

- nur möglich bei definierter Grundgesamtheit!



Fehlende Repräsentativität



Vgl.: Braunecker, C. (2023). *How to do empirische Sozialforschung. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. S. 61–68.



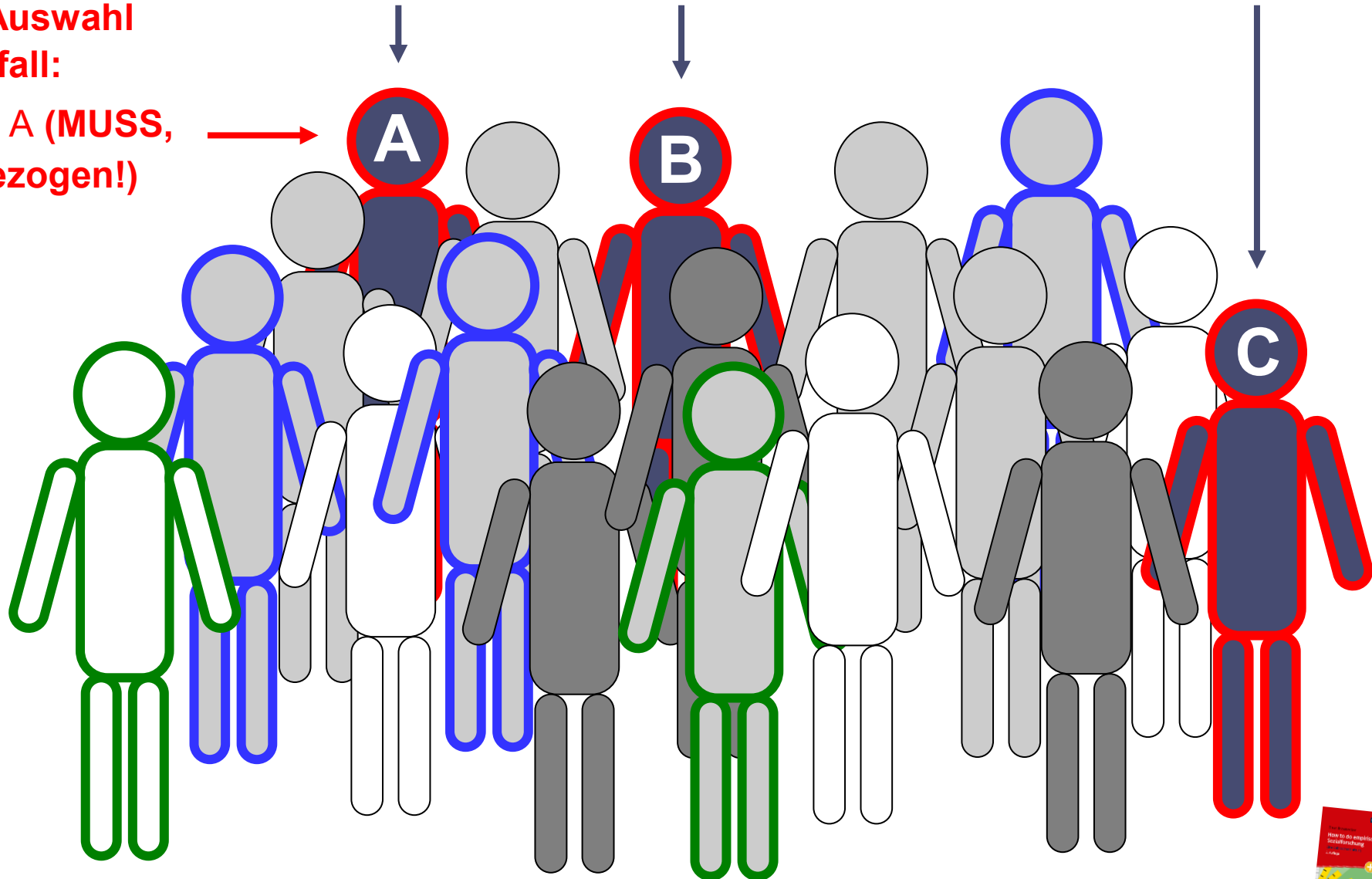
Stichproben-Auswahlverfahren

► Auswahl OHNE Zufall:

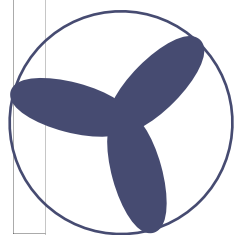
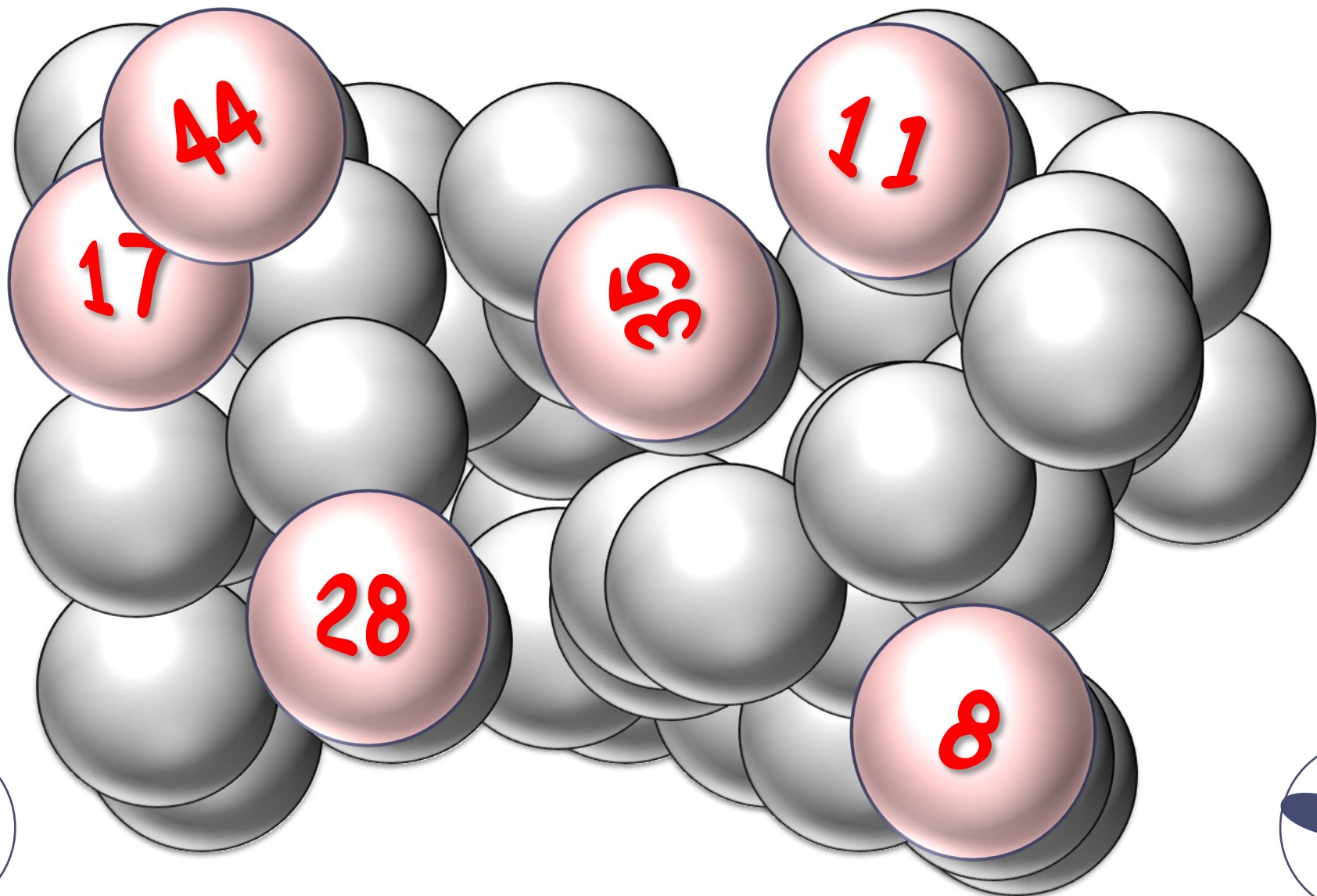
Person A oder B oder C (KANN, weil idente Merkmale)

► EINE Auswahl per Zufall:

Person A (MUSS, weil gezogen!)



Reine Zufallsstichprobe



Vgl.: Braunecker, C. (2023). *How to do empirische Sozialforschung. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. S. 69–70.

Zufallsstichprobe – die (mathematisch) beste Wahl!

- Zuerst nach Zufallszahlen sortieren und dann x-beliebige Datensätze entnehmen ...

ZufallsZahl	Kunden-Nummer	Vorname	Zuname	Geburts-datum	E-Mail-Adresse
=ZUFALLSZAHL()	01	Ernst	Mayer	23.09.1950	ernst.mayer@[...].de
0,05087	15	Helene	Müller	10.02.1960	helene.müller@[...].de
0,12148	03	John	Berger	21.05.1989	john.berger@[...].de
0,18116	13	Emelie	Franz	13.10.1955	emelie.franz@[...].de
0,23312	06	Hermann	Mann	13.09.1982	hermann.mann@[...].de
0,25621	04	Anna	Sommer	20.05.1973	anna.sommer@[...].de
0,32622	08	Wilhelm	Huber	20.11.1957	wilhelm.huber@[...].de
0,51188	12	Rudolf	Herbst	30.04.1990	rudolf.herbst@[...].de
0,70729	01	Linda	Schwarz	12.07.1954	linda.schwarz@[...].de
...
0,99392	5000	Charlotte	Weiss	07.01.1941	charlotte.weiss@[...].de



Proportionale und disproportionale Stichprobe

Kund:innendaten

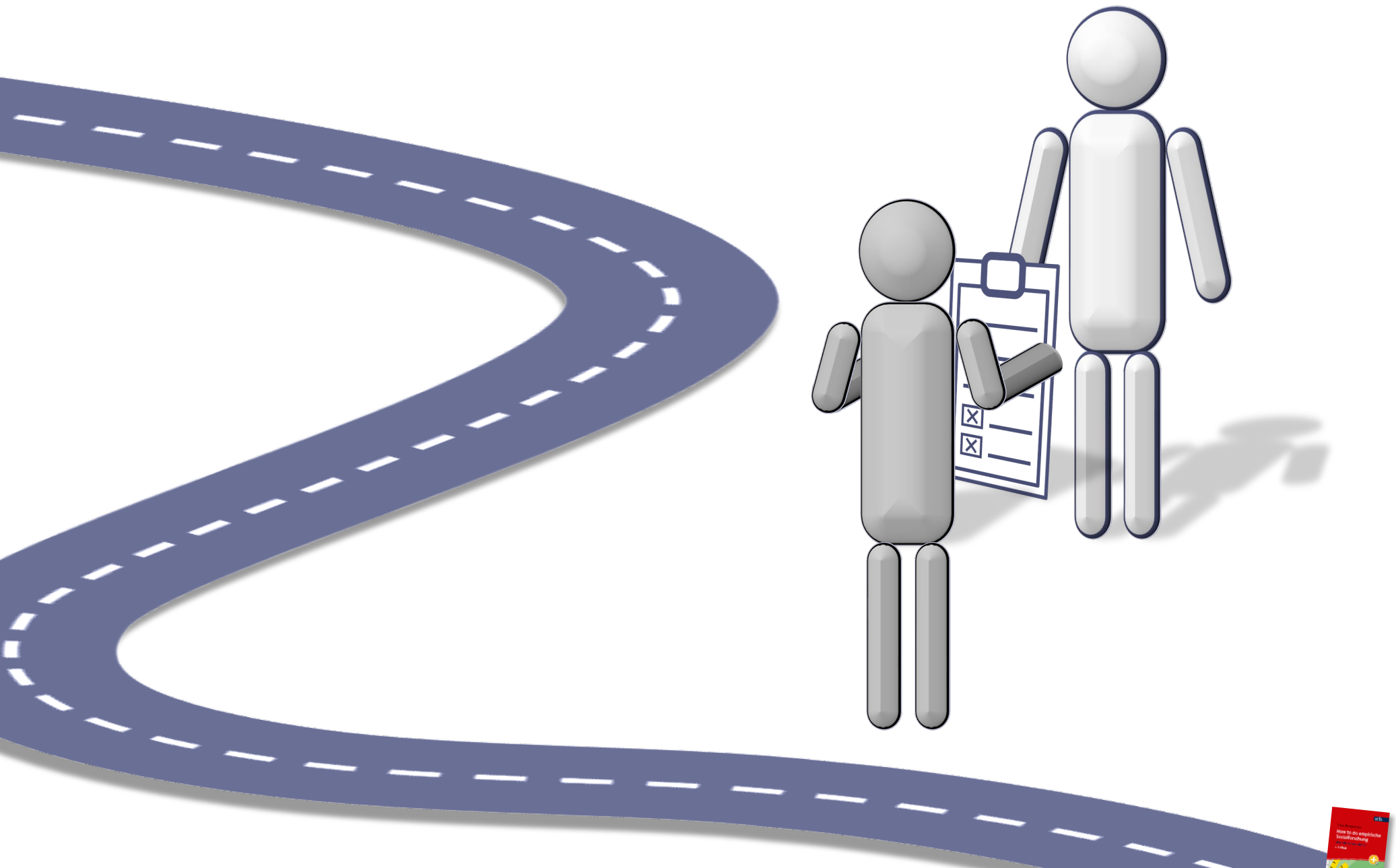
Vertriebsregion	Anzahl der Kund:innen		Stichprobe		Gewichtung	
	im letzten Jahr	in %	proportional	disproportional	Berechnung	Faktor
Nord	15 615	36,1	361	361		1,00
Süd	1 091	2,5	25	100	= 25 / 100	0,25
Ost	14 690	33,9	339	339		1,00
West	11 898	27,5	275	275		1,00
insgesamt	43 294	100,0	1 000	1.075		1 000

► JEDE Region besitzt jetzt eine für sinnvolle Aussagen über die Region vernünftige Größe.

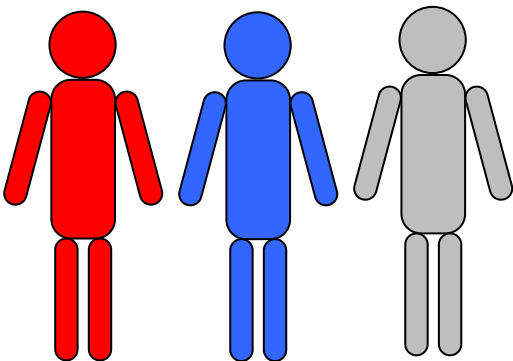
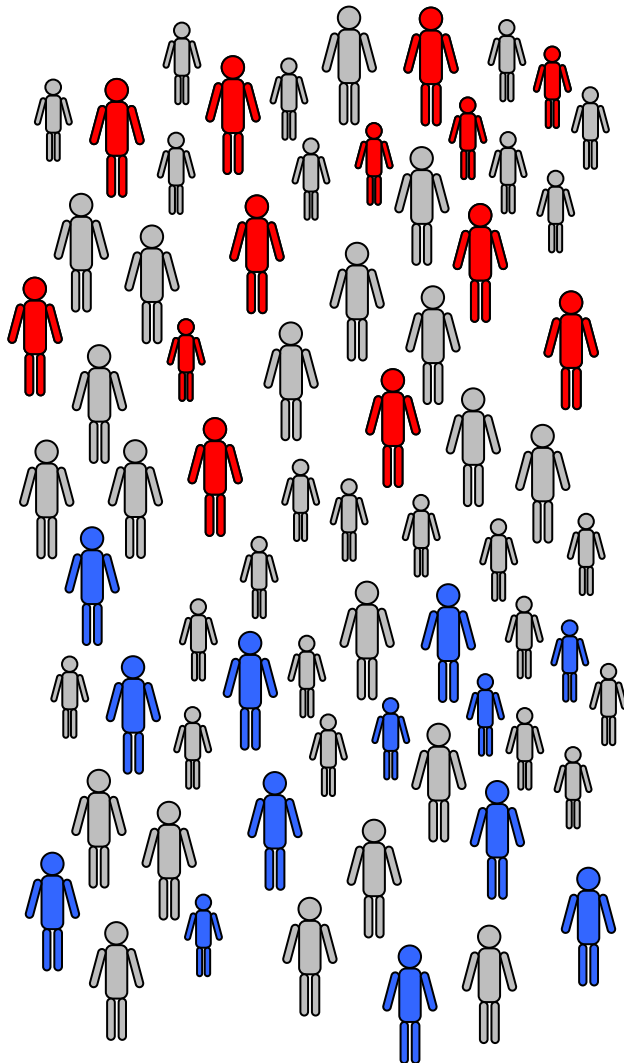
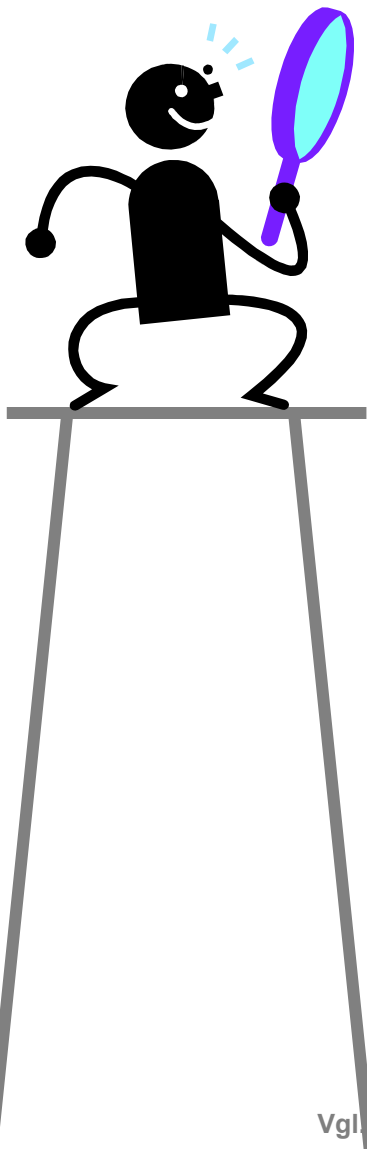
► Für das Gesamtergebnis müssen alle disproportionalen Datensätze jedoch gemäß ihrem tatsächlichen Anteil wieder abgewichtet (reproportionalisiert) werden.



Convenience Sample



Quotenstichprobe



	Anzahl
15 bis 30 Jahre alt	9
31 bis 50 Jahre alt	9
51 Jahre und älter	9
mag Schnitzel	14
mag kein Schnitzel	13

► usw.



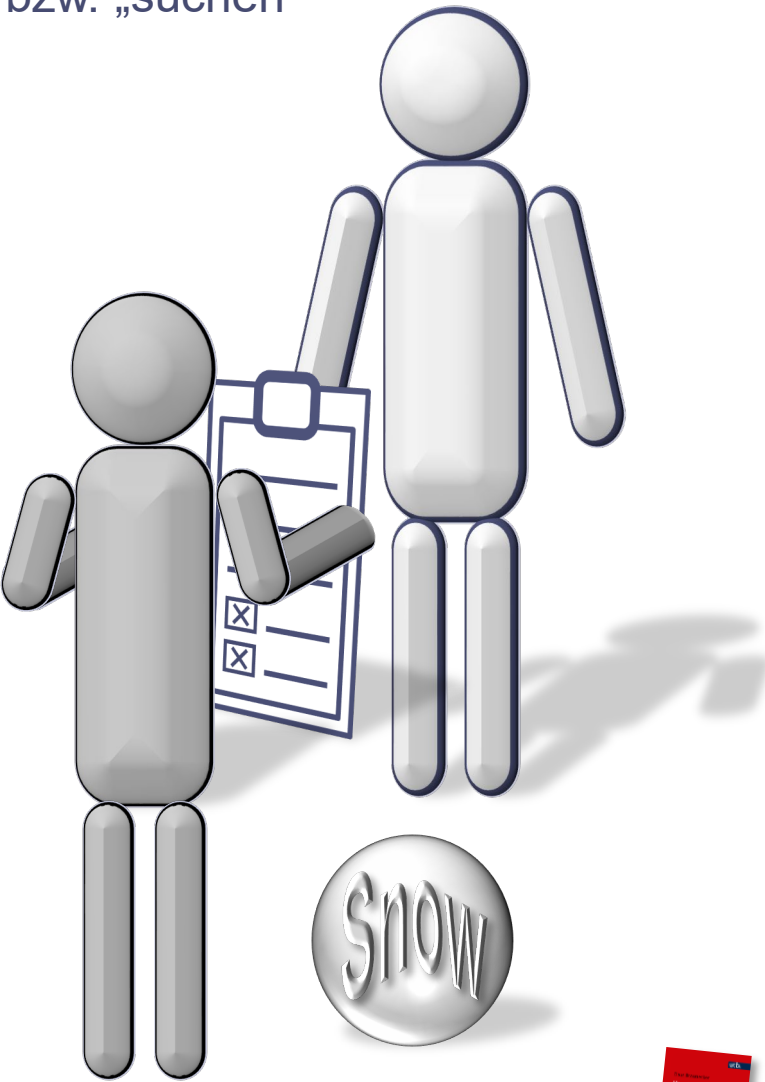
Bewusste Quotenstichprobe

► Befragte nach Kriterien auswählen

Wohnort	gelegentlich Schnitzel?	Anzahl
Deutschland	ja	250
	nein	250
Österreich	ja	250
	nein	250
Anzahl wenn bekannt nach der realen Verteilung, sonst (willkürlich) fixiert		1.000

Convenience Sample

► „Irgend jemanden“ befragen bzw. „suchen“



oft als Mix



- ▶ **Auch Convenience- oder Quotensamples** können mehrstufig ablaufen (vorgeschichtet). Sie **sollten** aber zumindest in Teilbereichen **auch Zufallselemente** beinhalten.
 - „Zufälligkeit“ kann z.B. bereits bei der Vorstrukturierung der Erhebungsorte ein „besseres“ Sample ermöglichen.
 - Jeden fünften Menschen zu befragen ist besser als nur die „Sympathischen“ ...
 - ...

Zufall schadet nie und sollte immer „irgendwo“ mit dabei sein!

- ▶ **„Je Zufall, desto besser“ !!!**



Zufall – auch bei willkürlichen Stichproben „ein wenig“ möglich

Einkaufszentrum 1		Einkaufszentrum 2		Einkaufszentrum 3	
Montag	Dienstag	Mittwoch	Donnerstag	Freitag	Samstag
8 bis 12 Uhr		12 bis 16 Uhr		16 bis 20 Uhr	

EKZ 1 Mo 08-12	EKZ 2 Mo 08-12	EKZ 3 Mo 08-12
EKZ 1 Mo 12-16	EKZ 2 Mo 12-16	EKZ 3 Mo 12-16
EKZ 1 Mo 16-20	EKZ 2 Mo 16-20	EKZ 3 Mo 16-20
EKZ 1 Di 08-12	EKZ 2 Di 08-12	EKZ 3 Di 08-12
EKZ 1 Di 12-16	EKZ 2 Di 12-16	EKZ 3 Di 12-16
EKZ 1 Di 16-20	EKZ 2 Di 16-20	EKZ 3 Di 16-20
EKZ 1 Mi 08-12	EKZ 2 Mi 08-12	EKZ 3 Mi 08-12
EKZ 1 Mi 12-16	EKZ 2 Mi 12-16	EKZ 3 Mi 12-16
EKZ 1 Mi 16-20	EKZ 2 Mi 16-20	EKZ 3 Mi 16-20
EKZ 1 Do 08-12	EKZ 2 Do 08-12	EKZ 3 Do 08-12
EKZ 1 Do 12-16	EKZ 2 Do 12-16	EKZ 3 Do 12-16
EKZ 1 Do 16-20	EKZ 2 Do 16-20	EKZ 3 Do 16-20
EKZ 1 Fr 08-12	EKZ 2 Fr 08-12	EKZ 3 Fr 08-12
EKZ 1 Fr 12-16	EKZ 2 Fr 12-16	EKZ 3 Fr 12-16
EKZ 1 Fr 16-20	EKZ 2 Fr 16-20	EKZ 3 Fr 16-20
EKZ 1 Sa 08-12	EKZ 2 Sa 08-12	EKZ 3 Sa 08-12
EKZ 1 Sa 12-16	EKZ 2 Sa 12-16	EKZ 3 Sa 12-16
EKZ 1 Sa 16-20	EKZ 2 Sa 16-20	EKZ 3 Sa 16-20



Kalenderwoche a

EKZ 1 Do 08-12	EKZ 2 Mo 08-12	EKZ 3 Mi 08-12
EKZ 1 Do 16-20	EKZ 2 Mo 16-20	EKZ 3 Mi 12-16
EKZ 1 Fr 16-20	EKZ 2 Mi 08-12	EKZ 3 Fr 12-16
EKZ 1 Sa 12-16	EKZ 2 Mi 16-20	EKZ 3 Sa 12-16
EKZ 1 Sa 16-20	EKZ 2 Fr 08-12	EKZ 3 Sa 16-20

Kalenderwoche b

EKZ 1 Mo 08-12	EKZ 2 Mi 12-16	EKZ 3 Mo 08-12
EKZ 1 Di 08-12	EKZ 2 Fr 16-20	EKZ 3 Mo 12-16
EKZ 1 Mi 12-16	EKZ 2 Sa 08-12	EKZ 3 Mo 16-20
EKZ 1 Do 12-16	EKZ 2 Sa 12-16	EKZ 3 Di 12-16
EKZ 1 Fr 12-16	EKZ 2 Sa 16-20	EKZ 3 Mi 16-20

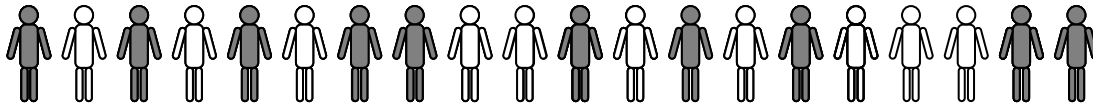
Kalenderwoche c

EKZ 1 Mo 12-16	EKZ 2 Mo 12-16	EKZ 3 Di 16-20
EKZ 1 Mo 16-20	EKZ 2 Di 08-12	EKZ 3 Do 12-16
EKZ 1 Di 16-20	EKZ 2 Di 12-16	EKZ 3 Do 16-20
EKZ 1 Mi 08-12	EKZ 2 Do 16-20	EKZ 3 Fr 16-20



Ermittlung der Stichprobengröße über („Quasi“-)Schwankungsbreiten

ALLE, über die etwas herauszufinden ist (= **Grundgesamtheit**):



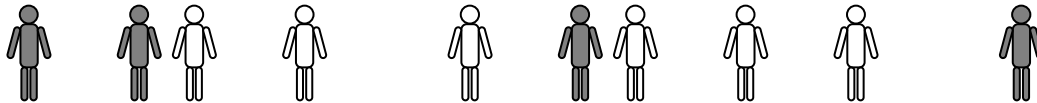
50% der Grundgesamtheit essen gerne Schnitzel.

Weil dieser Prozentsatz ja unbekannt ist, erfolgt eine **empirische Erhebung**:

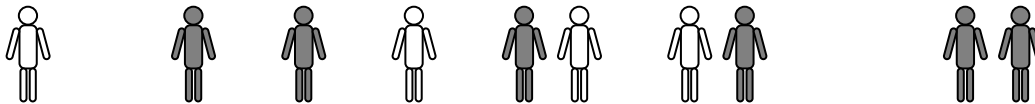
Befragt wird ein **ZUFÄLLIG** ausgewählter Teil (= Stichprobe).



■ Auch dort sind **50%** zu finden, die gerne Schnitzel essen.



■ Der ZUFALL hätte aber genau so gut andere „treffen“ können, wo **40%** Schnitzel gut finden ...



■ ... oder wieder andere, von denen **60%** Schnitzel mögen.

► Also: **Selbst 10% Ergebnisdifferenz** (oder mehr) kann **statistische Unschärfe** sein!

► **Bei erlaubter Schwankung von $\pm 10\%$: Stichprobengröße $n = 100$ reicht aus.**

► Soll die Schwankungsbreite kleiner sein, muss die Stichprobe größer sein!



Die individuell nötige Stichprobengröße ist aus den Standardtabellen für Schwankungsbreiten ablesbar.

Annahmen im Beispiel:

- Als maximale Schwankungsbreite sollen ± 1,8% toleriert werden.
- Das erwartete Ergebnis wird als 50% angenommen (größtmögliche Schwankungsbreite).
- Das Konfidenzintervall soll bei 95,5% liegen.

Tabellenbasis: 2 σ ±		Prozentergebnis einer Zufallsstichprobe >>>						
		5%	10%	15%	20%	30%	40%	50%
		95%	90%	85%	80%	70%	60%	50%
Zahl der Fälle >>>	10 000	0,4	0,6	0,7	0,8	0,9	1,0	1,0
	5 000	0,6	0,8	1,0	1,1	1,3	1,4	1,4
	4 000	0,7	0,9	1,1	1,3	1,4	1,5	1,6
	3 000	0,8	1,1	1,3	1,5	1,7	1,8	1,8
	2 500	0,9	1,2	1,4	1,6	1,8	2,0	2,0
	2 000	1,0	1,3	1,6	1,8	2,0	2,2	2,2
	1 000	1,4	1,9	2,3	2,5	2,9	3,1	3,2
	900	1,5	2,0	2,4	2,7	3,1	3,3	3,3
	800	1,5	2,1	2,5	2,8	3,2	3,5	3,5
	700	1,6	2,3	2,7	3,0	3,5	3,7	3,8
	600	1,8	2,4	2,9	3,3	3,7	4,0	4,1
	500	1,9	2,7	3,2	3,6	4,1	4,4	4,5
	400	2,2	3,0	3,6	4,0	4,6	4,9	5,0
	300	2,5	3,5	4,1	4,6	5,3	5,7	5,8
	200	3,1	4,2	5,0	5,7	6,5	6,9	7,1
	100	4,4	6,0	7,1	8,0	9,2	9,8	10,0
	50	6,2	8,5	10,1	11,3	13,0	13,9	14,7

Schwankungsbreite → Mindeststichprobengröße

- Stichprobengröße (n) = abhängig von max. tolerierter Schwankungsbreite in der kleinsten Subsamplegruppe, die analysiert wird.

$$\sigma = \sqrt{\frac{p \cdot (100 - p)}{n}}$$

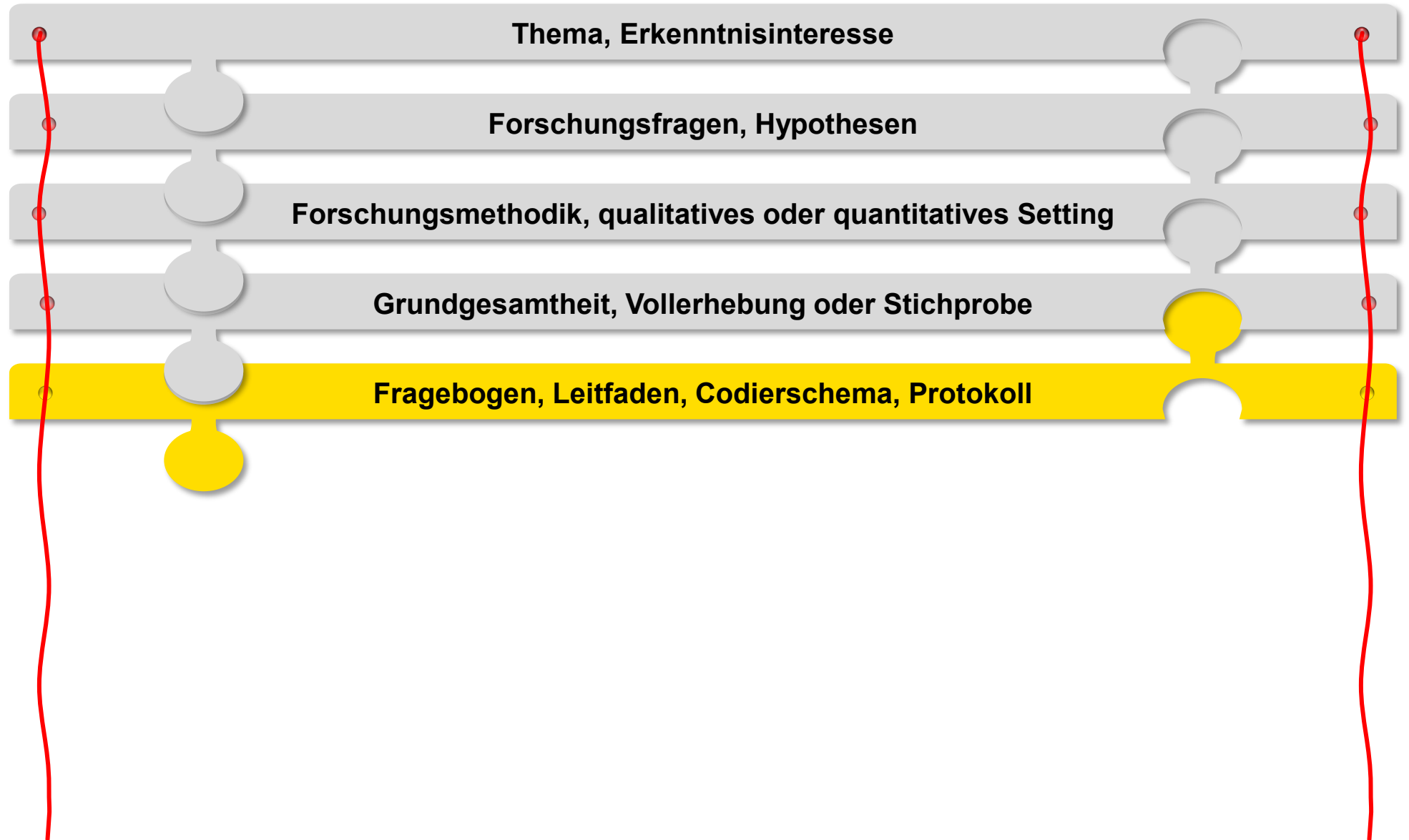
$$\cdot \sqrt{\frac{N - n}{N - 1}}$$

!!! Stichproben**größe** ≠ mehr **Repräsentativität** !!!

wenn Größe von N (Grundgesamtheit) = unbekannt						wenn Größe von N = bekannt und N =			
Sicherheit bei Prozentergebnis von >>>						10.000	5.000	1.000	600
						50 %	50 %	50 %	50 %
Stichprobengröße (n) >>>	95,5%	10 %	20 %	30 %	40 %	50 %	0,0		
	10.000	0,6	0,8	0,9	1,0	1,0	1,0	0,0	
	5.000	0,8	1,1	1,3	1,4	1,4	1,5	1,2	
	3.000	1,1	1,5	1,7	1,8	1,8	1,7	1,4	
	2.500	1,2	1,6	1,8	2,0	2,0	2,0	1,7	
	2.000	1,3	1,8	2,0	2,2	2,2	3,0	2,8	0,0
	1.000	1,9	2,5	2,9	3,1	3,2	3,5	3,4	1,8
	750	2,2	2,9	3,3	3,6	3,7	4,4	4,2	3,2
	500	2,7	3,6	4,1	4,4	4,5	4,9	4,8	2,9
	400	3,0	4,0	4,6	4,9	5,0	5,7	5,6	4,8
	300	3,5	4,6	5,3	5,7	5,8	7,0	6,9	6,3
	200	4,2	5,7	6,5	6,9	7,1	10,0	9,9	9,5
	100	6,0	8,0	9,2	9,8	10,0			9,1



Empirischer Forschungsprozess



Qualitativ? Quantitativ? Kombiniert?

Erkenntnisinteresse, Forschungsfragen bzw. Hypothesen weisen den Weg in Richtung wortreicher Ergebnisse oder knapper(er) Maßzahlen.

- **narrativ:**

„Denken Sie an den gestrigen Tag. Wie war der? Erzählen Sie mir, was Sie vom Aufstehen an alles gemacht haben, bis Sie am Abend das Schnitzel gegessen haben.“

- **Leitfaden:**

„Wenn Sie am Abend ein Schnitzel essen: Wie fühlen Sie sich da? Welche Auswirkungen hat das auf Ihren Tagesablauf?“

- **strukturierter Fragebogen, qualitativ:**

„Wenn Sie am Abend ein Schnitzel essen: Wie geht es Ihnen dabei?“

- **strukturierter Fragebogen, quantitativ:**

„Wie schmeckt Ihnen abends ein Schnitzel? Vergeben Sie bitte eine Schulnote.“

- **strukturierter Fragebogen, quantitativ und qualitativ:**

„Wie schmeckt Ihnen am Abend ein Schnitzel? Vergeben Sie bitte eine Schulnote.“ Wenn Note 3 oder schlechter: *„Warum ist das so?“*



Jemand könnte quantitativ fragen ...

- *„Haben Sie in den letzten 30 Tagen zumindest ein Schnitzel gegessen?“*
– Ja | Nein
- *„Wie viele Schnitzel haben Sie in den letzten 30 Tagen gegessen?“*
– keines | eines | zwei oder drei | vier oder fünf | mehr als fünf
- *„Wie oft essen Sie innerhalb von 30 Tagen im Schnitt Schnitzel?“*
– Antworten Sie bitte auf einer Skala von 0 („nie“) bis 10 („sehr oft“): 0 1 2 3 4 5 6 7 8 9 10
- *„Wie viele Schnitzel haben Sie in den letzten 30 Tagen gegessen?“*
– Geben Sie bitte die Anzahl an: |_|_|

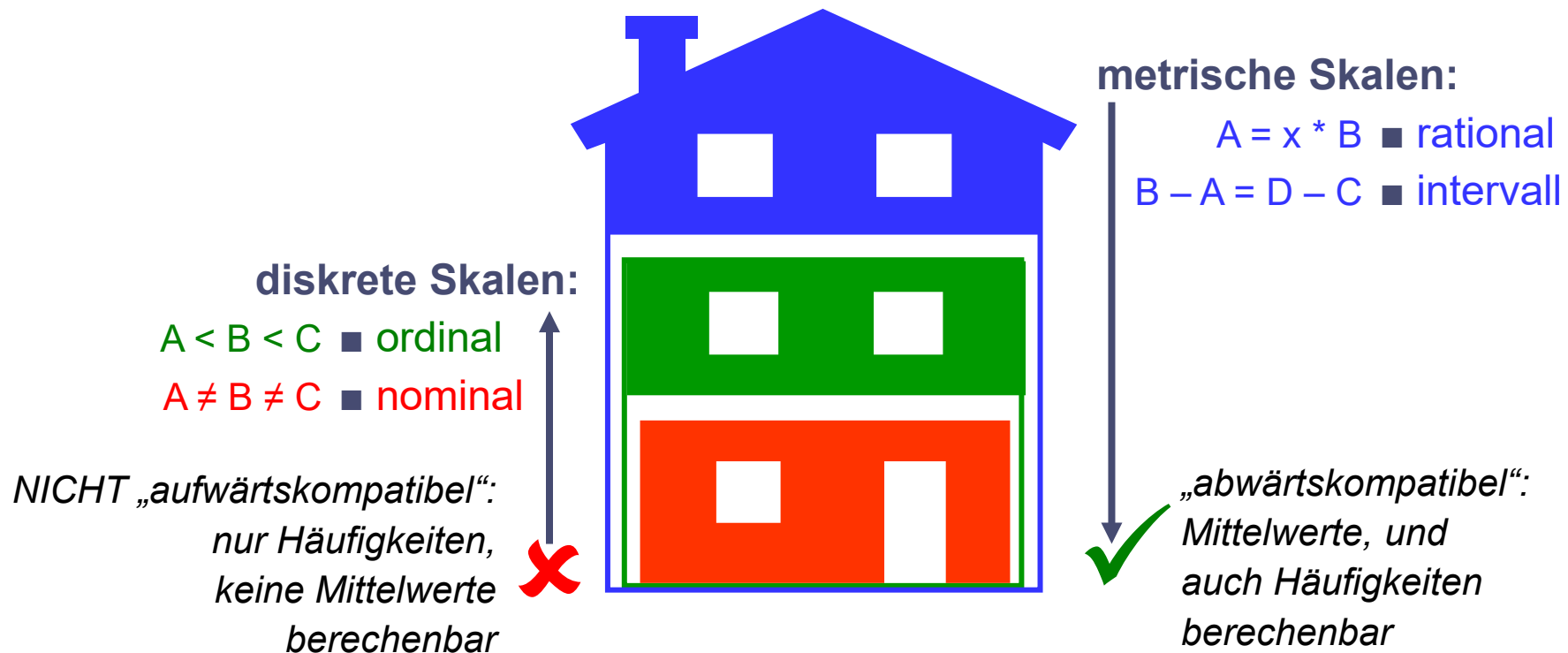
- **An dieser Stelle werden genaue Überlegungen in Richtung „Wie werde ich später auswerten, damit ich die Forschungsfragen beantworten und Hypothesen prüfen kann“ oft zu wenig detailliert durchdacht ...**



Das Messniveau ebnet den Weg zur späteren Auswertung

Je nach Mess- bzw. Skalenniveau ...

- sind **verschiedene Rechenoperationen** bei der Auswertung **sinnvoll** und **zulässig**
- sind die in Zahlen erfassten Antworten unterschiedlich zu interpretieren



- Das höchste Niveau bietet zwar die meisten Auswertungsmöglichkeiten, „zu viel metrisch“ kann aber Befragte ziemlich rasch überfordern.

Arten von Skalen (Messniveaus)

Skalenart, Messniveau	Kennzeichnung		Beispiele		Codierung	mögliche Analyse- Verfahren
(höchstes) rational	Verhältnis- mäßigkeit! EIN realer Nullpunkt: 0 = „nicht ausgeprägt“	A = x * B	<ul style="list-style-type: none">• Einkommen• Länge• Maße• Geschwindigkeit	<i>Alter</i> <i>in Jahren (anzugeben)</i>	15	<ul style="list-style-type: none">• Häufigkeiten• Median• Mittelwert• Varianz• Standard- abweichung• beliebige weitere Verfahren
					16	
					...	
					104	
					105	
(quasi-) intervall	gleiche Distanzen! KEIN (realer) Nullpunkt	B – A = D – C	<ul style="list-style-type: none">• Datum • IQ• Grad Celsius• Schulnoten- Bewertungen, Zustimmungs- Skalen	<i>Alter</i> <i>Jahrgang (anzugeben)</i>	2000	
					1999	
					...	
					1920	
					1919	
ordinal	Rangordnung! mehr – weniger	A < B < C	<ul style="list-style-type: none">• Rangreihen• Schulnoten	<i>Alter</i> <div>15 bis 26 Jahre</div> <div>27 bis 50 Jahre</div> <div>51 Jahre plus</div>	1	<ul style="list-style-type: none">• Häufigkeiten• Median
					4	
					6	
nominal (niedrigstes)	Unterschied! entweder – oder	A ≠ B ≠ C	<ul style="list-style-type: none">• Hauptwohnsitz- Bundesland• zuletzt gewählte Partei	<i>Alter</i> <div>entspricht der Ziel- gruppe 15 bis 26</div> <div>entspricht NICHT der Zielgruppe</div>	12	<ul style="list-style-type: none">• Häufigkeiten
					34	



Diskussionen, geteilte Meinungen bei der Fragebogengenerstellung



- Klassenbreiten: 14–30J • 31–50J • 51J+ ≠ 14–60 J • 61J+
- Zeitlicher Bezugsrahmen: Woche ≠ Monat ≠ Jahr
- Schulnoten ≠ Mittelwert – aber: „quasi-metrische“ Skalen
- Schulnoten – oder andere Form (z.B. internationale Befragung)

■ Items: – Tendenz zur Mitte?



– positiv <> negativ?



– unipolar ≠ bipolar

„Wie gerne mögen Sie Schnitzel? ○ ○ ○“
„Ich ... liebe Schnitzel ○ ○ ○ ○ ○ hasse Schnitzel“

– Richtungstausch?

„Ich ... liebe Schnitzel ○ ○ ○ ○ ○ hasse Schnitzel“
hasse Gemüse ○ ○ ○ ○ ○ liebe Gemüse“

– Skalenbreite?

„Ich ... liebe Schnitzel | △ | hasse Schnitzel“
0 100



Arten von Fragen (Fragetypen)

Fragen, die den Gesprächsverlauf lenken

- Screening- bzw. Sondierungsfragen
- Einleitungsfragen, „Eisbrecherfragen“
- Übergangsfragen
- Filterfragen & Folgefragen

offene und geschlossene Fragen

- halboffene Fragen
- Einfachnennung und Mehrfachnennungen
- Skalenfragen

direkte und indirekte Fragen

- projektive Fragen

spontane und gestützte Abfragen

manipulative Fragestellungen



Tipps für Fragebögen und Online-Formulare

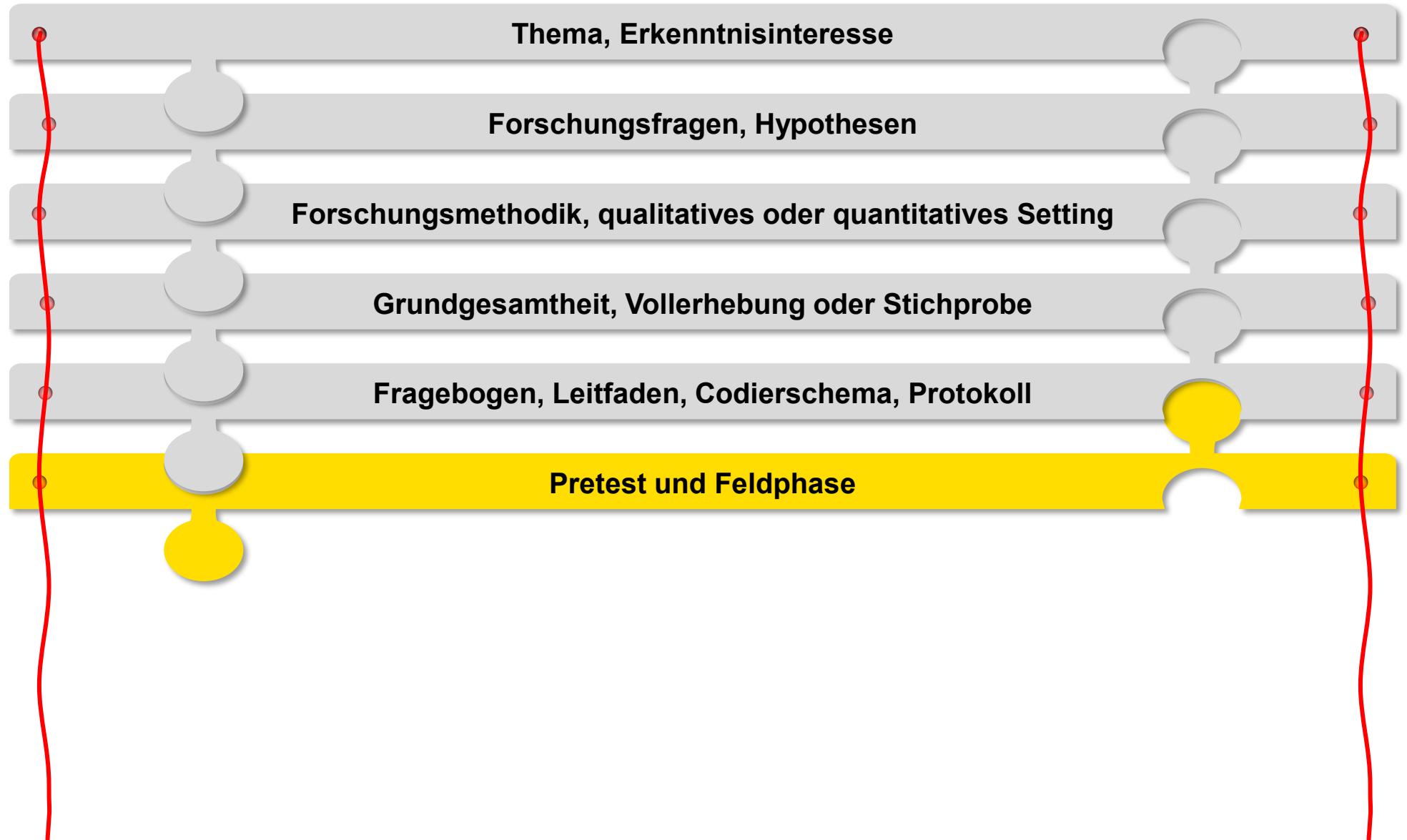
Fragebögen sind immer individuell, vieles ist Geschmackssache.

Aber:

- **Dauer** max. 10 Minuten (20 Minuten gehen gerade noch)
- „**Surveytainment**“
- **zu Beginn:** Begrüßung, Einleitung (worum geht's IN ETWA), durchschnittliche Dauer, Anonymitätszusicherung, DSGVO!
- Orientierung und **Übersicht:** *Struktur*, keine gedankliche „Müllhalde“!
- Heikles und **Sozialstatistik** besser am Ende (ausgenommen Quotenmerkmale)
- **Rotieren** verhindert Platzierungseffekte
- **Ausfüllzwang?** ☹ bei „geraden“ Skalen ohne „weiß nicht“
- ~~Dialekt • Fremdwörter • doppelte Verneinungen • usw.~~
- **am Ende:** „*Danke*“, Abschluss-Statement



Empirischer Forschungsprozess



Pretest – machen Sie ihn unbedingt!

... sonst:

Ein Restaurant macht eine Bedarfserhebung für ein Kinder-Schnitzel auf der Speisekarte. Das Alter potentieller Besteller:innen dieses „Pumuckl-Tellers mit Pommes“ soll erhoben werden.

Ein Link auf ein Online-Formular wird an alle Restaurantgäste verteilt.

Ohne vorherige Pretests.

Die erste (Screening-)Frage: *„Haben Sie Kinder?“*

Wenn JA: *„Wie alt sind Ihre Kinder?“*

Frau Huber antwortet:

„Ja, ich habe zwei Kinder.“

*„**22** und **34** Jahre alt.“* 😊 😊 😊



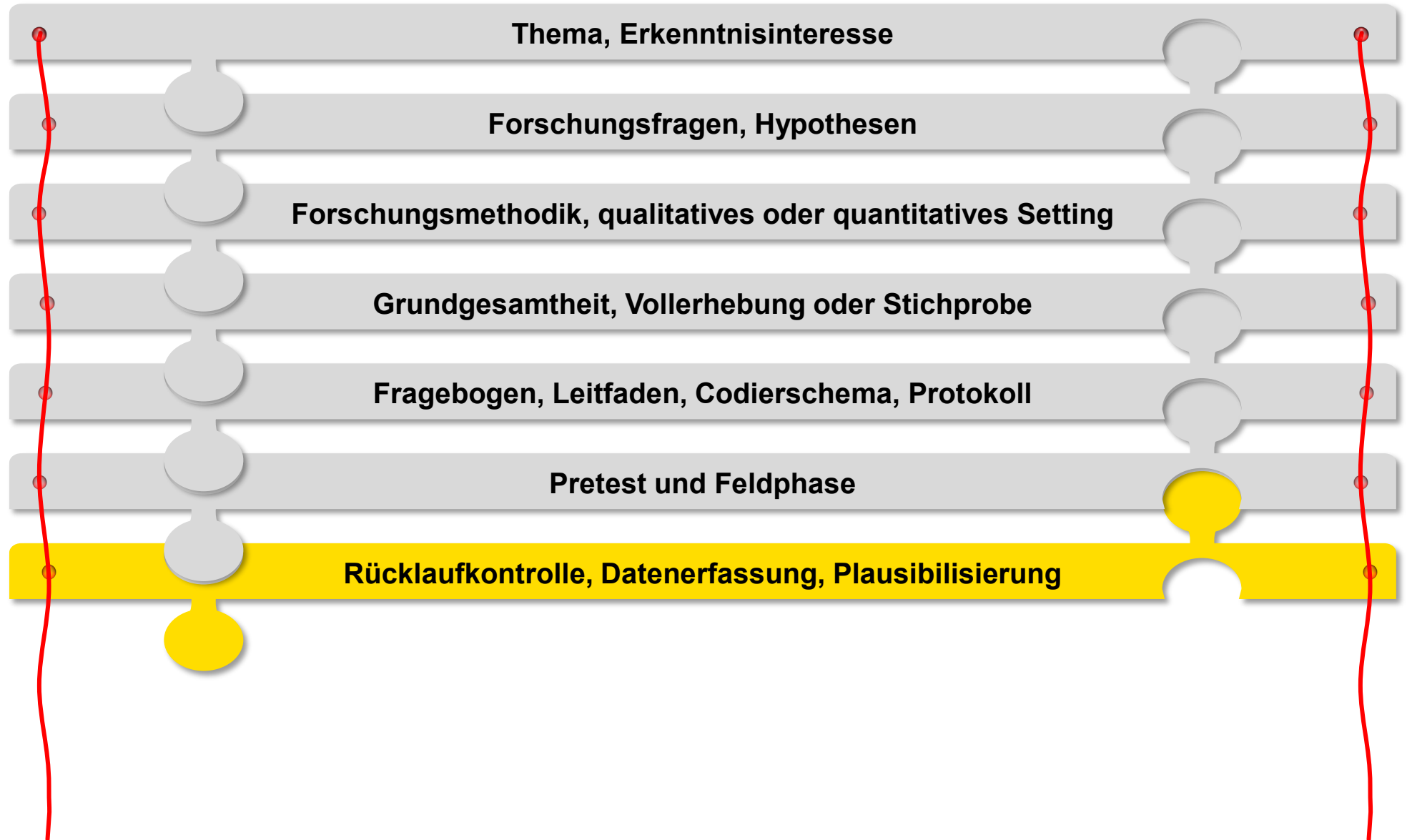
Pretest: Wozu?

Einzelne Testpersonen aus der Zielgruppe (nicht Stichprobe)
beantworten, lesen bzw. kommentieren den Fragebogen/das Online-Formular.

- Funktioniert technisch alles?
- Motivieren die Fragen zum Antworten?
- Werden mehrdeutige Begriffe verwendet?
- Sind die Fragen verständlich und in schlüssiger Reihenfolge?
- Sind die Antwortvorgaben ausreichend, zu knapp oder zu detailliert?
- Ist der Fragebogen/das Online-Formular optisch ansprechend und logisch?
- Sinnvoll ist es, **alles** zu **protokollieren, was nicht O.K. ist**,
dann können keine wichtigen Anpassungen übersehen werden.



Empirischer Forschungsprozess



Strukturvergleich Stichprobe – Grundgesamtheit

Struktur der Befragten

	n	in %	N
Total, alle Befragten	400	100,0	Bevölkerung 14 bis 75 Jahre in % *)
Wohnort liegt im Bundesland	Wien	21,8	21,8
	Niederösterreich	18,8	18,7
	Burgenland	3,5	3,3
	Oberösterreich	16,8	16,7
	Salzburg	6,0	6,3
	Tirol	8,3	8,5
	Vorarlberg	4,8	4,4
	Steiermark	13,8	14,0
	Kärnten	6,5	6,3

1

2

3

*) Quelle: Statistik Austria, 2023a.

Vgl.: Braunecker, C. (2023). *How to do empirische Sozialforschung. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. S. 150–156.



Aufbau eines Datenfiles (Datenauszüge aus BUCHdaten)

► **je Variable eine Spalte, je Datensatz eine Zeile**

► **Variablenamen** stehen in der ersten Zeile **1**

1	FraboNr	Lesen Sie gerne?	Fachbuch gelesen?	Eigenschaften eines idealen Fachbuchs	Empfehlung an ...				Geschlecht
	lfdNr	f_01	f_03	f_05_txt	f_11_1	f_11_2	f_11_3	f_11_4	f_12
	3	3	3	4	3	3	3	3	3
	1	0	0		0	0	0	0	2
	2	0	0		0	0	0	0	1
	3	0	0		0	1	0	0	1
	5	1	0		0	0	0	0	1
	6	0	0		1	1	1	0	1
	7	0	0		0	0	0	0	1
	15	0	1	6 viele Informationen, die sehr	0	1	0	1	1
	16	0	1	6 kompakt, einfache Formulierung	1	1	0	1	1
	17	1	1	6 Praxis mit Beispielen	0	0	0	0	1
	5				5	5	5	5	
	usw.								

► **numerische** **3** und **alphanumerische** (Text, String) **4** Variablen**ausprägungen**

► **Mehrfachangaben** **5** benötigen pro Angabemöglichkeit eine eigene Spalte

■ dichotome Codierung: Antwort gegeben Code 1, sonst Code 0

► **6** = **Filterfrage**: nur bei Code 1 (f_03) ein Eintrag (f_05_txt), sonst bleibt f_05_txt leer

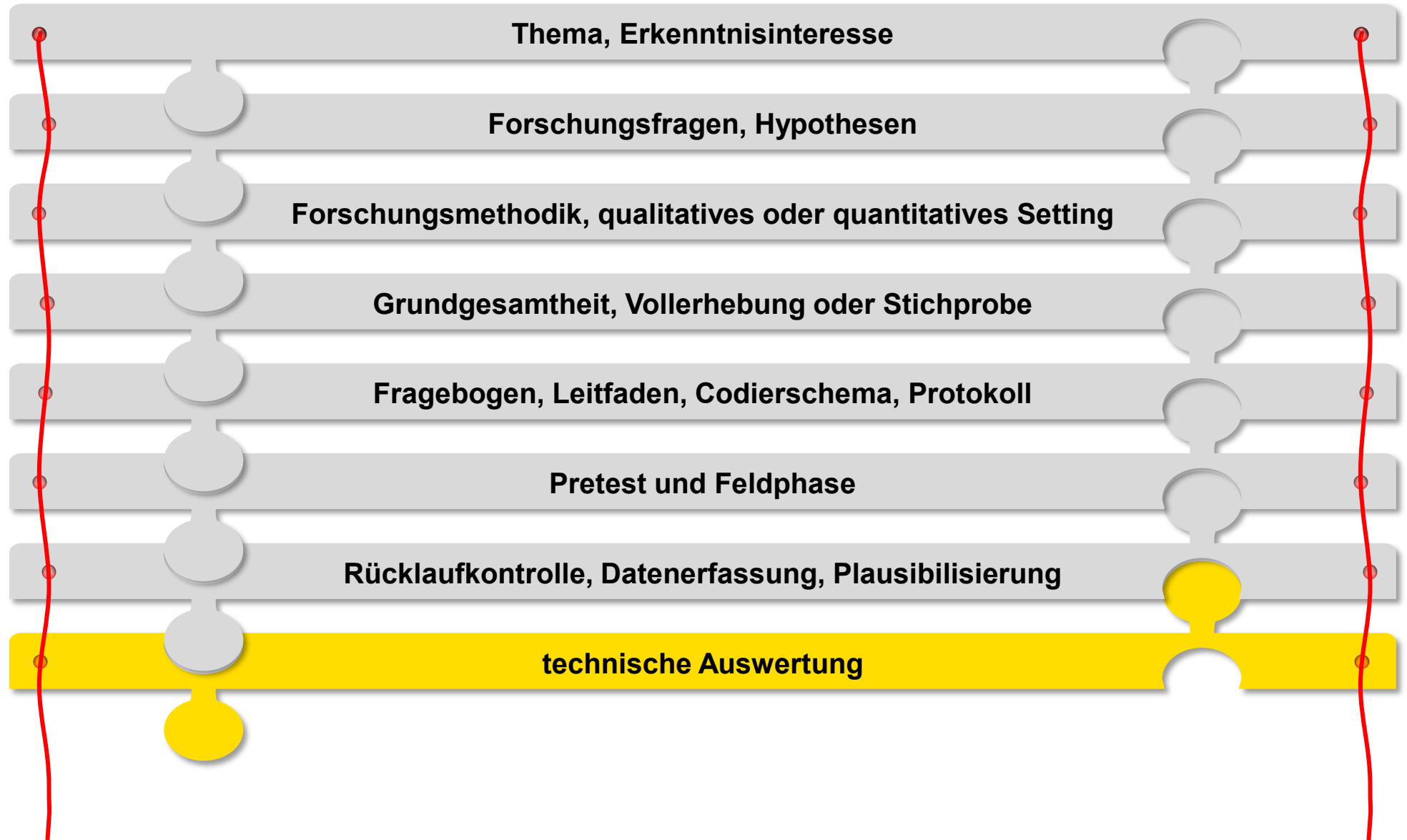


„Was ist und nicht sein darf“ – Datenscreening vor Auswertung!

- zu viele/wenige/falsche **Codes** Code 1 = bis 49 Jahre, Code 2 = 50 Jahre +, **3 = ?**
- **fehlende** Werte, wenn unerlaubt 1 = mag Schnitzel, 2 = mag kein Schnitzel, **0 = ?**
- **Extremwerte** **180 Minuten Wartezeit auf ein Schnitzel**
- **„Text“** bei numerischen Variablen 1,1,2,2,1,2,1,**a**,1,2,2,3,2,_,2,2,1,1,1,1
- **uneinheitliche** Angaben (Schreibweisen) **Schnitzel ≠ Wiener ≠ Wiener Schnitzel**
- **Prozentangaben**, die sich nicht „ausgehen“ **60% Schnitzel auswärts, 45% zu Hause**
- fehlende **Filter** zerstören Basisangaben **„Schnitzel schmecken super“ sagen jene, die noch nie eines gegessen haben**
- alles **contra Hausverstand** **90% der Über-90-jährigen trinken täglich Red Bull; 25 Stunden am Tag usw.**



Empirischer Forschungsprozess



Antwortzuordnung zu Kategorien

		Assoziation		
Daten-satz	„Was sagen Sie zum Schnitzel, das Sie zuletzt gegessen haben?“	positiv	neutral	negativ
1	gut war's, aber zu klein	✓		✓
2	hat schrecklich geschmeckt			✓
3	war diesmal sehr gut, werde bald wieder eins essen	✓	✓	
4	hatte es liefern lassen, es war sehr knusprig	✓	✓	
5	wir mussten zwar ewig darauf warten, war aber super	✓		✓
6	nächstes Mal probiere ich einmal etwas Anderes		✓	
7	ich war satt, es war sehr groß, aber leicht verbrannt	✓	✓	✓
8	usw.			



(Um-)Codieren, Berechnen auf Einzelfallebene: Beispiele

■ Umcodieren in eine neue Variable:

– Wie viele Schnitzel essen Sie im Jahr: |__|__|__|

▶ **Gruppierung:** 0 | 1 bis 4 | 5 bis 9 | 10 bis 19 | 20 plus

– Wie gut hat Ihnen das Schnitzel von zuletzt geschmeckt? ① ② ③ ④ ⑤

▶ **Top Box, Bottom Box:** 1,2 → 1 | 3 → 3 | 4,5 → 5

■ Umcodieren in dieselbe oder neue Variablen:

– Ich ... *liebe Schnitzel* ① ② ③ ④ ⑤ *hasse Schnitzel*

hasse Gemüse ① ② ③ ④ ⑤ *liebe Gemüse*

▶ **Richtungsbereinigung:** 1 → 5 | 2 → 4 | 3 → 3 | 4 → 2 | 5 → 1

■ Berechnen (eines Index) auf Einzelfallbasis:

– Ich liebe ... *Tomaten* ① ② ③ ④ ⑤

Paprika ① ② ③ ④ ⑤

Gurken ① ② ③ ④ ⑤

▶ **Gesamtliebe_{Gemüse}** = $\text{mean}(\text{Note}_{\text{Tomaten}}, \text{Note}_{\text{Paprika}}, \text{Note}_{\text{Gurken}})$



Messniveaus und quantitative Analyseverfahren

Merkmal 1	Merkmal 2	Verfahren
nominal (ordinal)		HÄUFIGKEITSAUSZÄHLUNG ▶ Kapitel 5.1.1, 5.2.2 und SPSS 6.5.1
metrisch (Skala)		MITTELWERTSBERECHNUNG (Zentral-, Lage- und Streuungsmaße) ▶ Kapitel 5.1.2–5.1.4, 5.2.3 und SPSS 6.5.2
nominal (ordinal)	nominal (ordinal)	KREUZTABELLE ▶ Kapitel 5.1.5, 5.2.4 und SPSS 6.5.3
metrisch (Skala)	neben metrisch (Skala) nach nominal (ordinal)	MITTELWERTSVERGLEICH ▶ Kapitel 5.1.6, 5.2.5 und SPSS 6.5.4
metrisch (Skala)	metrisch (Skala)	KORRELATION ▶ Kapitel 5.1.7, 5.2.6 und SPSS 6.5.5



Das Auswertungsverfahren wird durch das Messniveau bestimmt

Variable 1		Variable 2	Verfahren	Fragestellung(en)
nominal oder ordinal	Wie alt sind Sie? – bis 40 J – über 40 J		Häufigkeitsverteilung	Wieviele % sind bis 40 Jahre alt, wieviele % sind älter? ► Kapitel 5.1.1, 5.2.2 und SPSS 6.5.1
Skala oder metrisch	Benoten Sie bitte X: 1 2 3 4 5 Wie alt sind Sie? <input type="text"/> <input type="text"/> <input type="text"/>		deskriptive Statistik oder explorative Datenanalyse	Welche Benotung wird im Schnitt vergeben? Wie ist das Durchschnittsalter? ► Kapitel 5.1.2 – 5.1.4, 5.2.3; SPSS 6.5.2
nominal oder ordinal	Wie alt sind Sie? – bis 40 J – über 40 J	nominal oder ordinal	Wie groß sind Sie? – bis 175 cm – 176 bis 190 cm – über 190 cm	Kreuztabelle Sind jüngere Personen bis 40 eher über 175 cm groß als ältere? ► „Wenn bis 40, dann eher über 175 groß?“ ► Kapitel 5.1.5, 5.2.4; SPSS 6.5.3
Skala oder metrisch	Benoten Sie bitte X: 1 2 3 4 5 Wie groß sind Sie? <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Skala oder metrisch nominal oder ordinal	Benoten Sie bitte Y: 1 2 3 4 5 Wie alt sind Sie? – bis 40 J – über 40 J	Mittelwertsvergleich oder explorative Datenanalyse Welcher mittlere Beurteilungs- unterschied besteht zw. X und Y? Sind Jüngere im Schnitt größer? ► „Wenn bis 40, dann im Schnitt größer?“ ► Kapitel 5.1.6, 5.2.5; SPSS 6.5.4
Skala oder metrisch	Benoten Sie bitte X: 1 2 3 4 5 Wie alt sind Sie? <input type="text"/> <input type="text"/> <input type="text"/>	Skala oder metrisch	Benoten Sie bitte Y: 1 2 3 4 5 Wie groß sind Sie? <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	Korrelation Welchen Zusammenhang gibt es – zwischen Beurteilung X und Y? – zwischen Alter und Größe? ► „Je jünger, desto größer?“ ► Kapitel 5.1.7, 5.2.6; SPSS 6.5.5

Was wie quantitativ auswerten (SPSS) und darstellen?

Essen Sie gerne Schnitzel? – ja nein	Häufigkeitsauswertung • wenn vorhanden auch Analyse von Mehrfachantworten • Kreis-, Linien-, Balken- oder Säulendiagramm
Wie gerne essen Sie Schnitzel? – 1 2 3 4 5	
Wie gerne essen Sie Schnitzel? – 1 2 3 4 5	Deskriptivstatistik • explorative Datenanalyse • Kreis-, Linien-, Balken- oder Säulendiagramm
Wie oft haben Sie in den letzten 30 Tagen Schnitzel gegessen? _ _	
Essen Sie gerne Schnitzel? – ja nein ► Wie alt sind Sie? – bis 30 bis 50 51J +	Kreuztabelle • Signifikanzcheck mittels Chi ² -Test • Säulen- oder Balkendiagramm
Wie gerne essen Sie Schnitzel? – 1 2 3 4 5 Wie gerne essen Sie Gemüse? – 1 2 3 4 5 ► Wie alt sind Sie? – bis 30 bis 50 51J +	deskriptive Mittelwertsvergleiche • Signifikanzcheck mittels Parameter- bzw. parameterfreier Mittelwertsvergleiche • Fehlerbalkendiagramm
Wie oft haben Sie in den letzten 30 Tagen Schnitzel gegessen? _ _ ► Wie alt sind Sie? _ _	bivariate Korrelationen • Pearson- oder Spearman-Koeffizient • Streudiagramm



Arten von Häufigkeiten und Prozentwerten

- 1 Häufigkeit** = Anzahl der Fälle je Merkmalsausprägung
 - ▶ absolute Häufigkeiten
- 2 Prozent** = Anzahl der Fälle je Ausprägung, relativiert zu ALLEN Fällen
 - ▶ relative Häufigkeiten
- 3 Gültige Prozente** = Anzahl der Fälle je Ausprägung, relativiert an nur jenen Fällen, die eine Merkmalsausprägung haben (also ohne „Fehlend“)
- 4 Kumulierte Prozente** = Prozentwerte, summiert in steigender Reihenfolge der Merkmalsausprägungen, beginnend beim kleinsten Wert
 - ▶ $29,2 + 20,7 = 49,9$; $49,9 + 15,0 = 64,9$ usw.

Wie viele Bücher lesen Sie im Schnitt pro Jahr? (in Klassen)

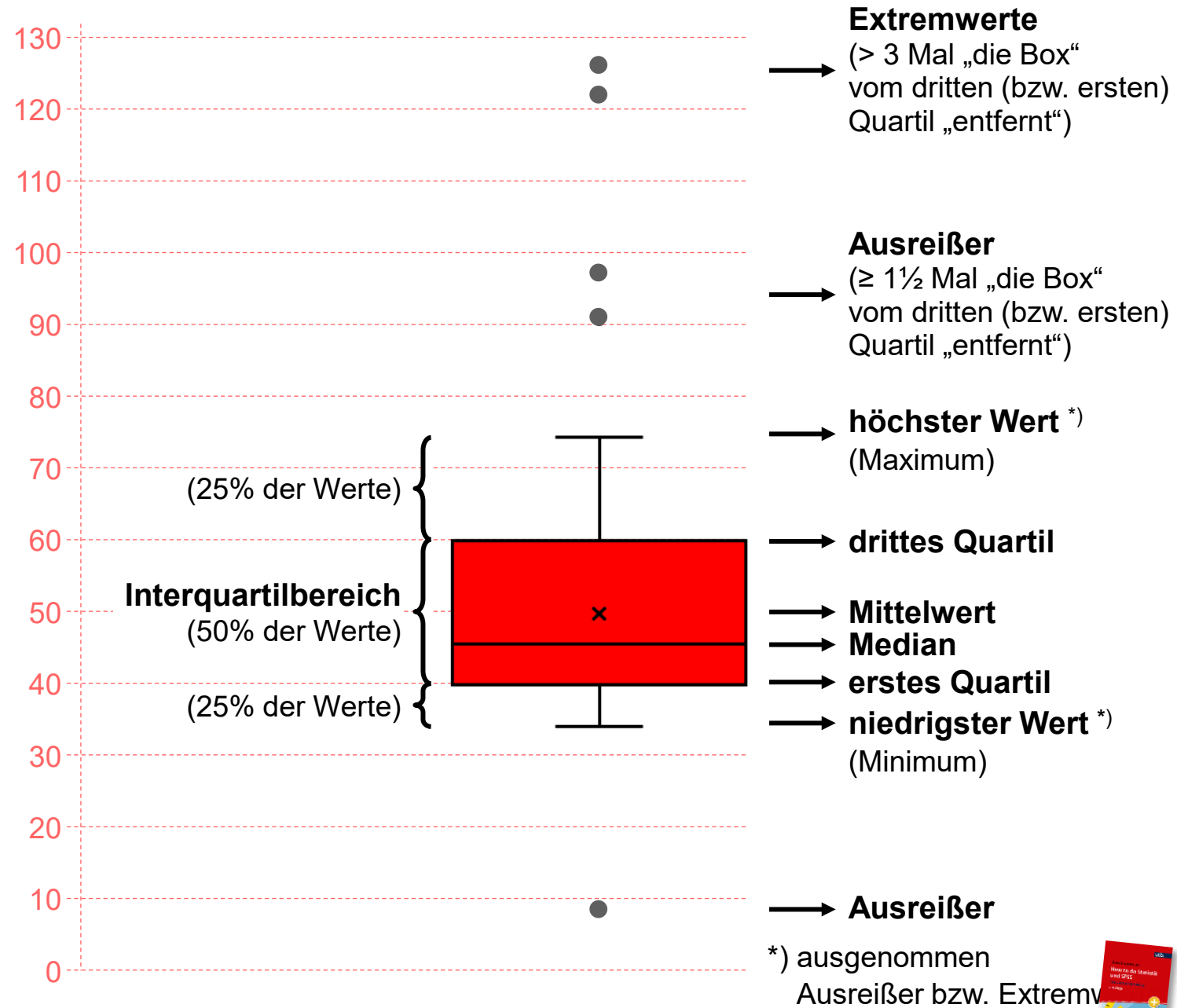
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 Buch	138	27,3	29,2	29,2
	2 Bücher	98	19,4	20,7	49,9
	3 bis 5 Bücher	71	14,1	15,0	64,9
	6 bis 10 Bücher	63	12,5	13,3	78,2
	mehr als 10 Bücher	103	20,4	21,8	100,0
	Gesamt	473	93,7	3 100,0	
Fehlend	System	32	6,3		
Gesamt		1 505	2 100,0		

Vgl.: Braunecker, C. (2023). *How to do Statistik und SPSS. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. S. 53–55.



Boxplot für metrische Verteilungen

- **Boxplots** dienen zur Visualisierung und schnellen Beurteilung der Verteilung und Streuung von Variablen.
- Auch Gruppenvergleiche bzw. das Gegenüberstellen verschiedener Verteilungen sind damit sehr plakativ möglich.



Mittelwert, Varianz und Standardabweichung

1 Varianz

= durchschnittliche quadrierte Abweichung aller Messwerte vom Mittelwert

@ Grundgesamtheit:

▼
$$\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - \mu)^2$$

@ Stichprobe:

▼
$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

2 Standardabweichung

= Wurzel der Varianz

@ Grundgesamtheit:

▼
$$\sigma = \sqrt{\sigma^2}$$

@ Stichprobe:

▼
$$s = \sqrt{s^2}$$

► Je kleiner die **Standardabweichung**, desto besser beschreibt der **Mittelwert** die Verteilung.

Anzahl gelesener Bücher

befragte Person	heuer	im_Vorjahr
1	1	1
2	1	1
3	1	2
4	2	2
5	2	2
6	2	2
7	3	3
8	3	3
9	3	20
Mittelwert	2,0	4,0
Varianz (hier: einer Stichprobe)	0,75	36,50
Standardabweichung (hier: einer Stichprobe)	0,87	6,04



Kreuztabelle

Σ = in Summe

①

beobachtete
Werte

Alter	derzeit_studierend		Σ
	ja	nein	
bis 40 J	② 150	50	200
über 40 J	30	⑤ 70	100
Σ	180	120	300

③

Zeilen-
prozent

Alter	derzeit_studierend		Σ
	ja	nein	
bis 40 J	75,0	25,0	100,0
über 40 J	30,0	70,0	100,0
Σ	① 60,0	40,0	100,0

②

erwartete
Werte

Alter	derzeit_studierend		Σ
	ja	nein	
bis 40 J	② 120,0	③ 80,0	200
über 40 J	④ 60,0	⑤ 40,0	100
Σ	180	120	300

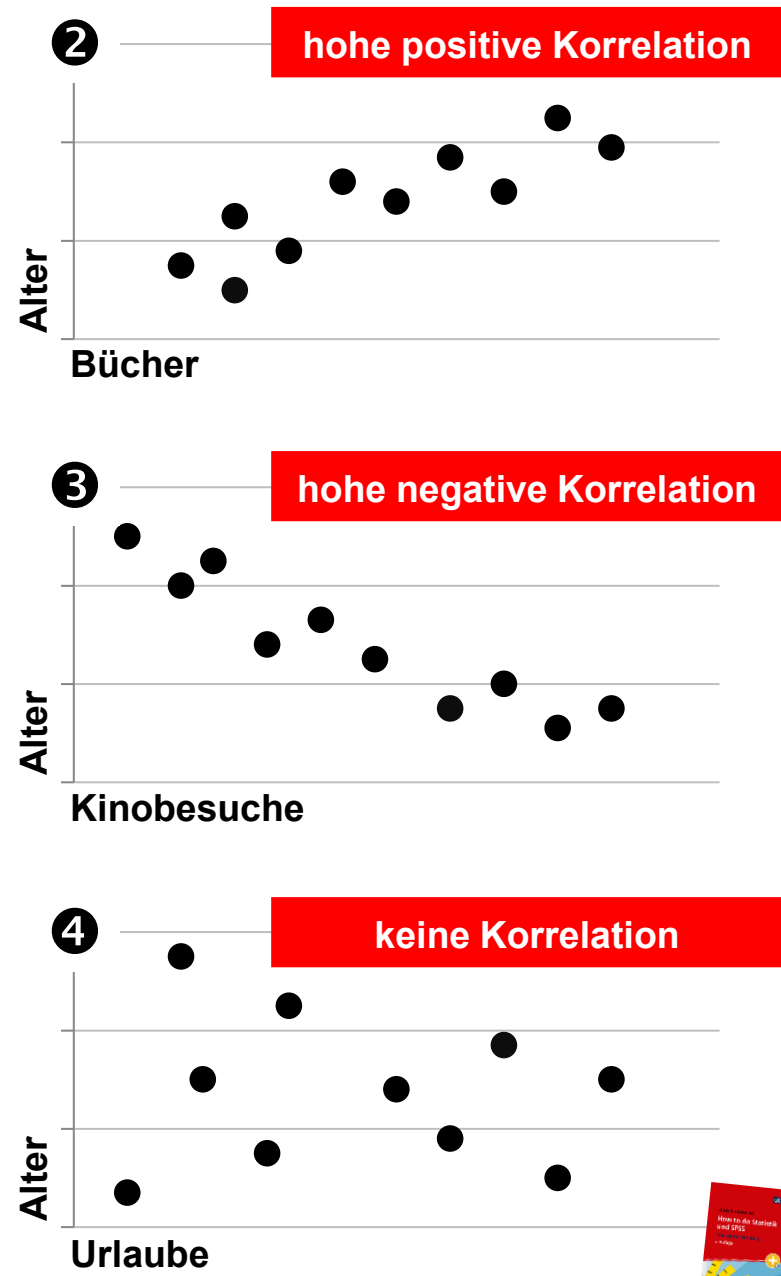
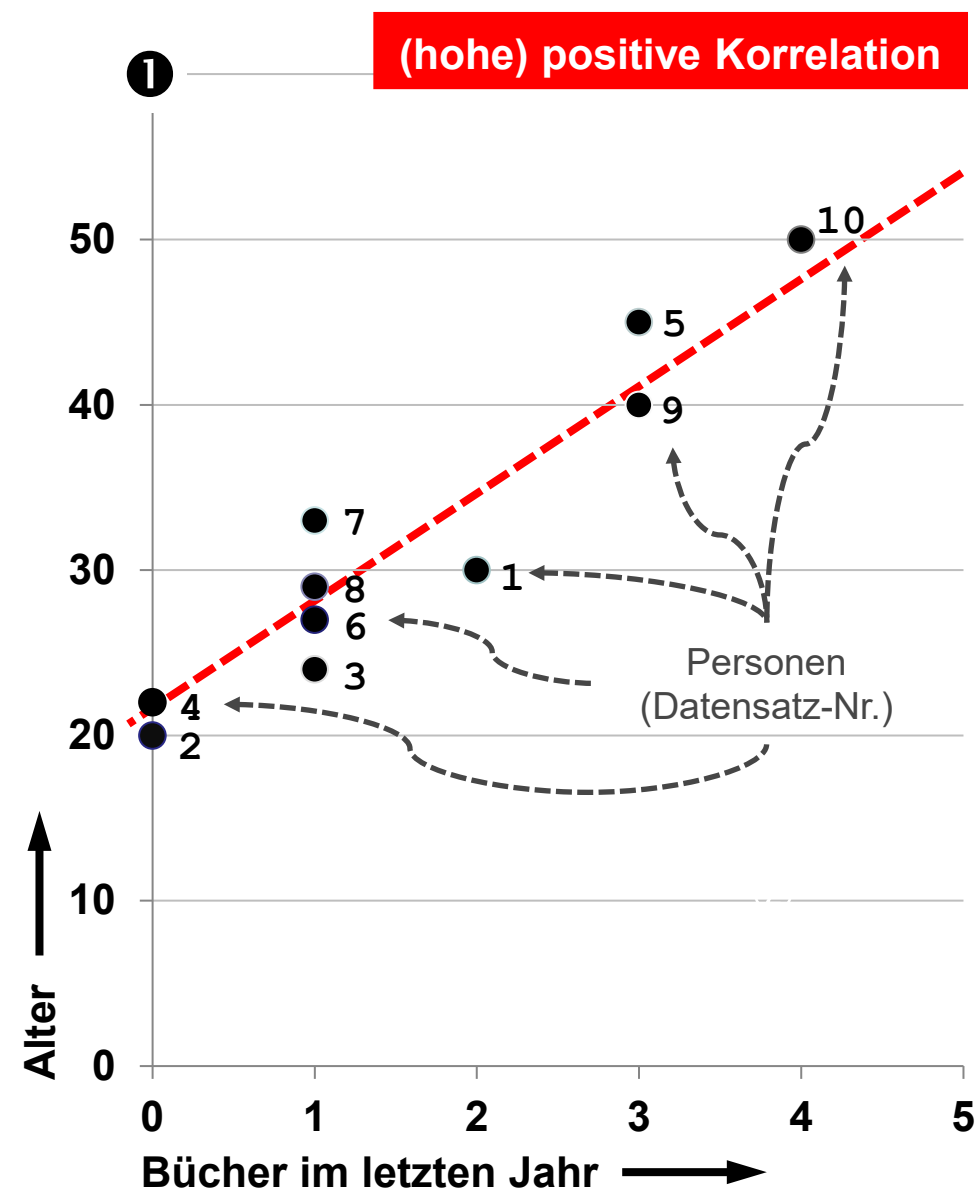
④

Spalten-
prozent

Alter	derzeit_studierend		Σ
	ja	nein	
bis 40 J	83,3	41,7	66,7
über 40 J	16,7	58,3	33,3
Σ	100,0	100,0	100,0



Korrelation



Übersicht gebräuchlicher SPSS-Befehle

Datenhandling:

Datei ↪ **Neu** ▶ Kapitel 6.3.1

↪ **Öffnen** ▶ Kapitel 6.3.1

↪ **Daten importieren** ▶ Kapitel 6.3.1

Daten ↪ **Dateien zusammenfügen** ▶ Kapitel 6.3.2

↪ **Datei aufteilen...** ▶ Kapitel 6.3.2

↪ **Fälle auswählen...** ▶ Kapitel 6.3.2

↪ **Fälle gewichten...** ▶ Kapitel 6.3.3

Datenaufbereitung:

Transformieren ↪ **Variable berechnen...** ▶ Kapitel 6.4.2

↪ **Umcodieren in dieselben Variablen...** ▶ Kapitel 6.4.1

↪ **Umcodieren in andere Variablen...** ▶ Kapitel 6.4.1

Auswertung:

Analysieren ↪ **Deskriptive Statistiken** ▶ Kapitel 6.5

↪ **Mittelwerte und Proportionen vergleichen** ▶ Kapitel 6.5

↪ **Allgemeines lineares Modell** ▶ Kapitel 6.5

↪ **Korrelation** ▶ Kapitel 6.5

↪ **Nicht parametrische Tests** ▶ Kapitel 6.5

↪ **Mehrfachantworten** ▶ Kapitel 6.5.1.1

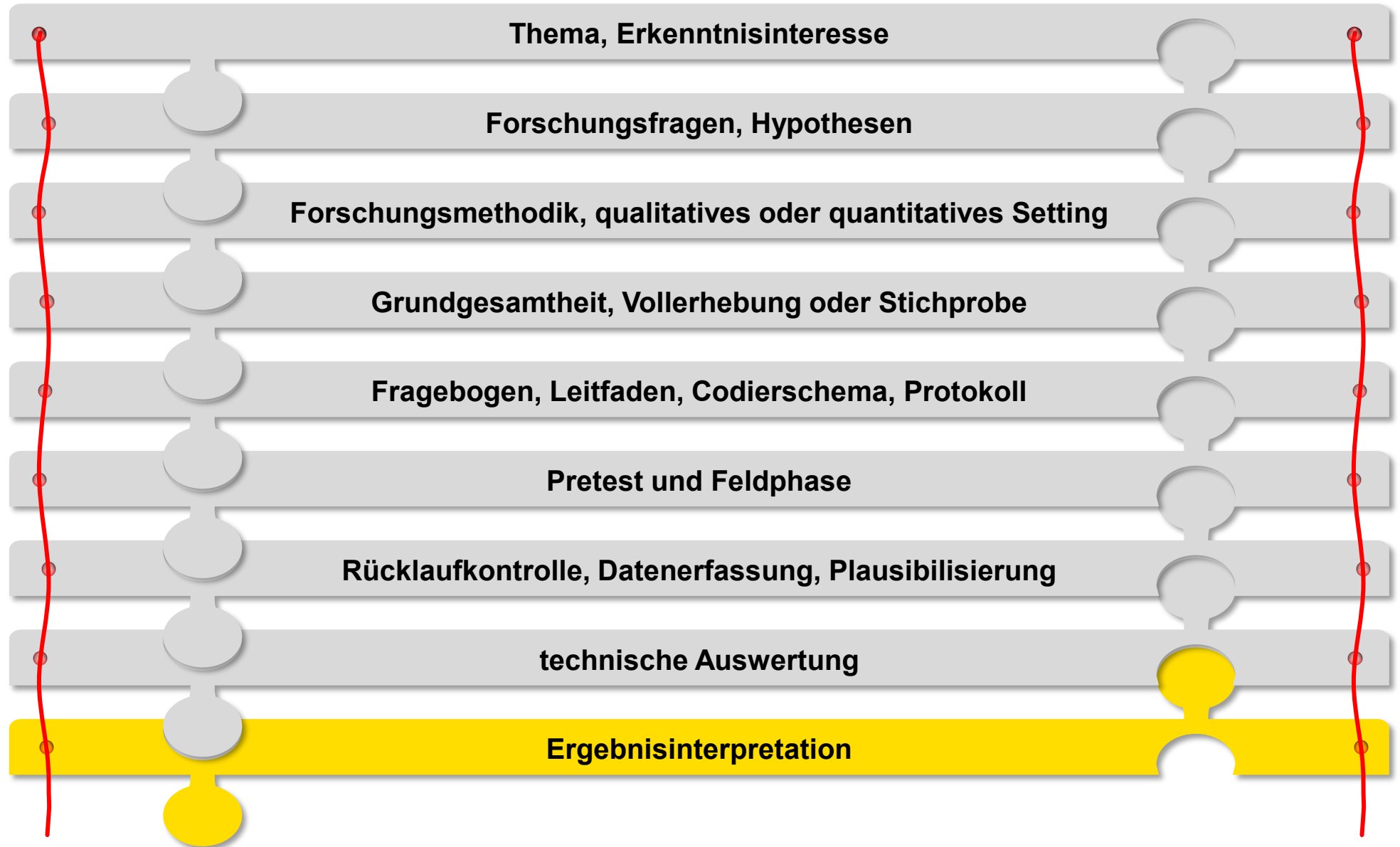
Grafiken:

Grafik ↪ **Diagrammerstellung...** ▶ Kapitel 6.5

Programmeinstellungen: Bearbeiten ↪ Optionen... ▶ Kapitel 6.2.4

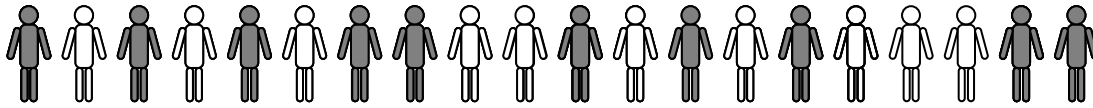


Empirischer Forschungsprozess



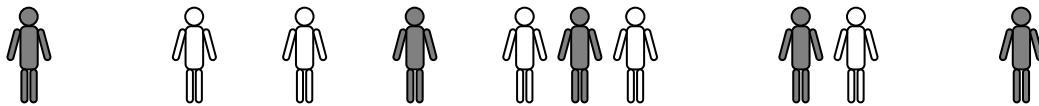
Schwankungsbreiten (ohne Zufallsstichprobe: „Quasi-Schwankungsbreiten“)

ALLE, über die geforscht wird (= **Grundgesamtheit**):

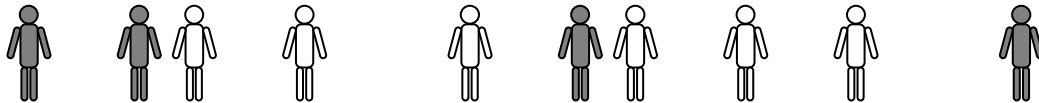


50% der Grundgesamtheit essen gerne Schnitzel.

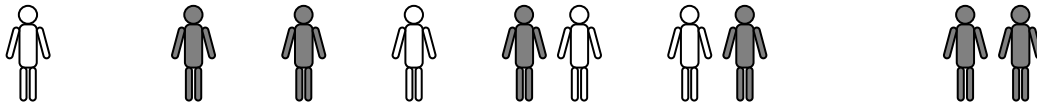
Weil dieser Prozentsatz ja unbekannt ist, findet eine **empirische Erhebung** statt:
Befragt wird ein **ZUFÄLLIG ausgewählter Teil** (= Stichprobe).



■ Auch dort sind **50%** zu finden, die gerne Schnitzel essen.



■ Der ZUFALL hätte aber genau so gut andere „treffen“ können, wo **40%** Schnitzel gut finden ...



■ ... oder wieder andere, von denen **60%** Schnitzel mögen.

► Also: **Selbst 10% Ergebnisdifferenz (oder mehr) kann statistische Unschärfe sein!**

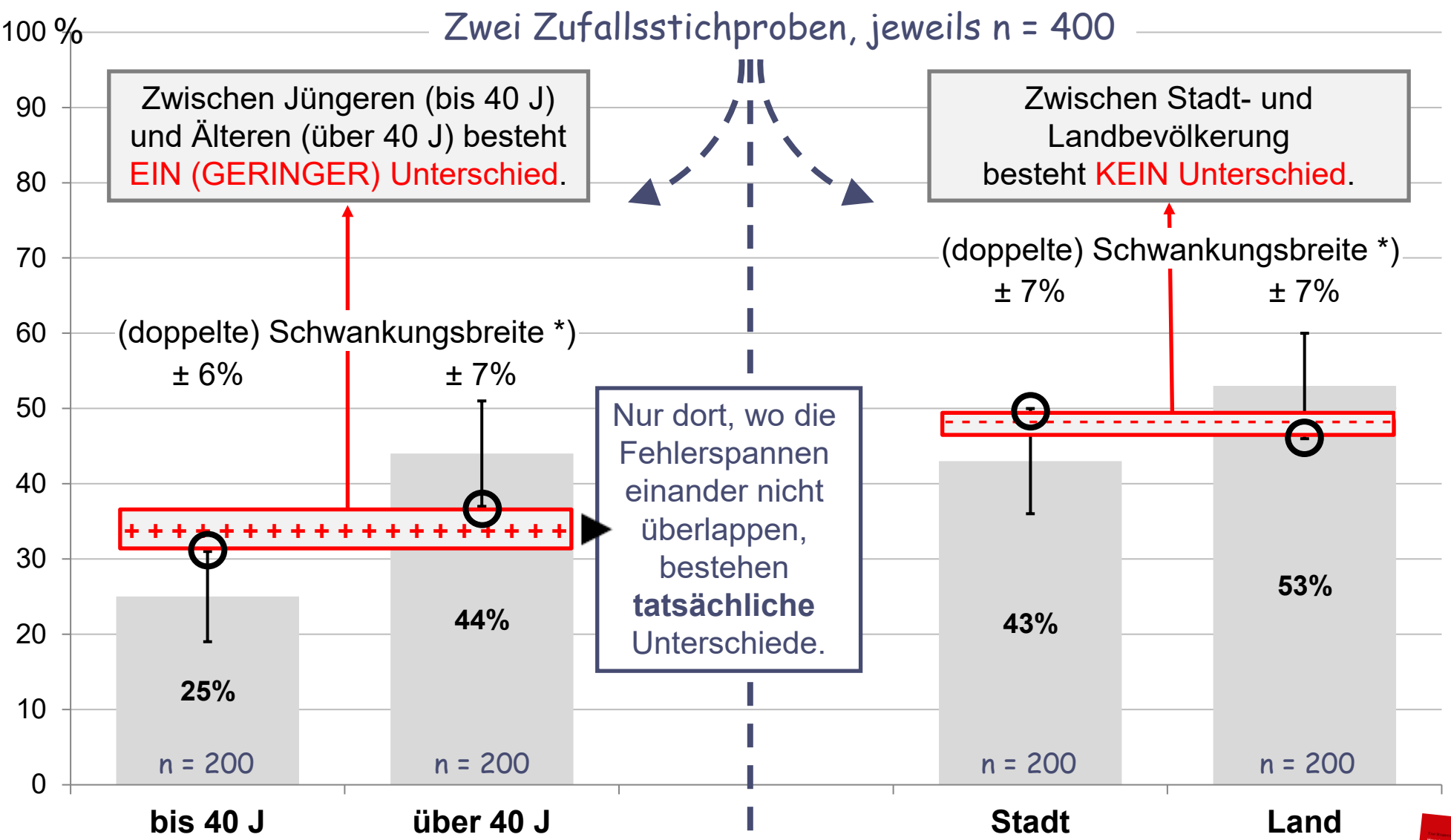
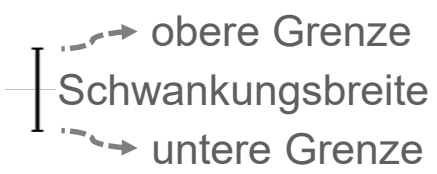
► Je kleiner die Stichprobe, desto größer die Schwankungsbreite.

► Je näher das Ergebnis (Anzahl Schnitzelesser) bei 50%, desto größer die Schwankungsbreite.



± Schwankungsbreiten von Prozentwerten

„Essen Sie gerne Schnitzel?“ – Antwort: „JA“

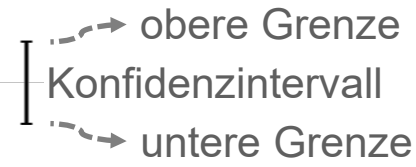


*) : jeweils auf ganze Zahlen gerundet

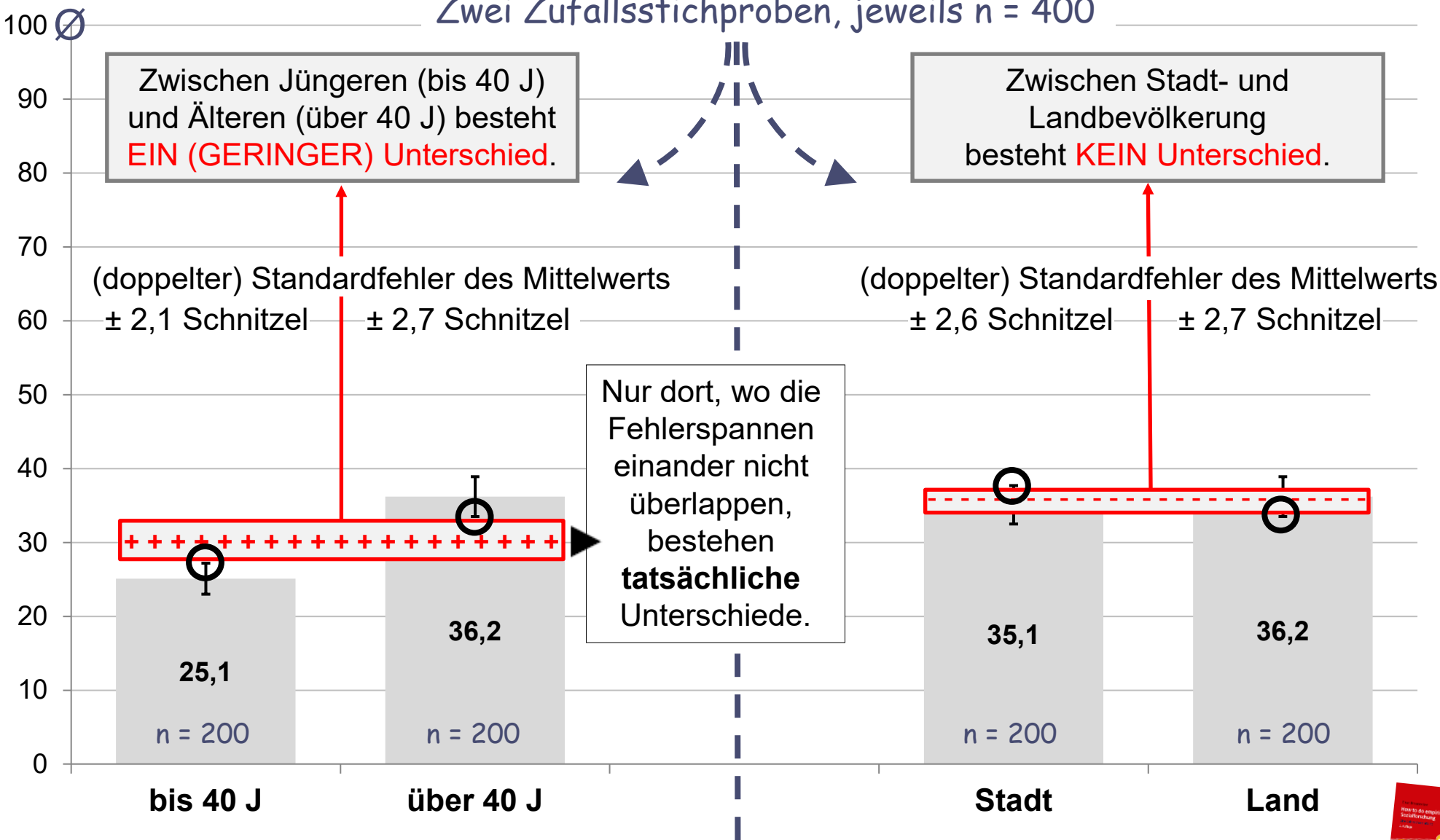


Konfidenzintervalle von Mittelwerten

„Wie viele Schnitzel essen Sie im Jahr? |__|__|__|“



Zwei Zufallsstichproben, jeweils $n = 400$



Signifikanzprüfung – auf die Grundgesamtheit schließen

- ▶ Signifikanzprüfungen drehen sich um folgende Fragen:
 - Besteht ein Ergebnisunterschied zwischen einzelnen Merkmalen?
 - Bewirkt ein Merkmal einen Ergebnisunterschied bei einem anderen?
 - Sind diese Unterschiede (= „Zusammenhänge“) **signifikant**?

Signifikante Ergebnisse:

- **In einer (Zufalls!-)Stichprobe** ermittelte Ergebnisunterschiede
- sind **auf die Grundgesamtheit** real **übertragbar**.
- Sie treten **nicht nur in EINER Stichprobe** (die gerade aktuell analysiert wird) auf,
- sondern kommen **wahrscheinlich in (fast) ALLEN Stichproben**, die zufällig aus der Grundgesamtheit gezogen werden könnten, so oder ähnlich vor.



Null- und Alternativhypothese

■ Nullhypothese H_0 :

1. Zwischen zwei Werten in einer Grundgesamtheit besteht real keine Differenz.
2. In einer ihr entstammenden Stichprobe gefundene Ergebnisse (Ergebnisunterschiede) sind bloß Zufall.
3. Zwischen unabhängiger und abhängiger Variable besteht kein systematischer Zusammenhang.
Die unabhängige Variable hat keinen Einfluss auf die abhängige.

■ Alternativhypothese H_1 :

1. Zwischen zwei Werten in einer Grundgesamtheit besteht real EINE Differenz.
2. In einer ihr entstammenden Stichprobe gefundene Ergebnisse (Ergebnisunterschiede) sind KEIN Zufall.
3. Zwischen unabhängiger und abhängiger Variable besteht EIN systematischer Zusammenhang.
Die unabhängige Variable hat EINEN Einfluss auf die abhängige.

Quelle: Formuliert nach Stangl, W. (2011). *Nullhypothese*. *Online Lexikon für Psychologie und Pädagogik*. <https://lexikon.stangl.eu/62/nullhypothese> [15.12.2019]



Forschungsfragen, Hypothesen, Signifikanz in der Praxis

■ INHALTLICHE Forschungsfrage:

Welchen Zusammenhang gibt es zwischen Alter und Affinität zu Schnitzeln?

■ INHALTLICHE Hypothese:

(ungerichtet:) Wenn Personen jünger sind (bis 40),
dann unterscheiden sie sich im durchschnittlichen Schnitzelkonsum von älteren.

[(gerichtet:) Je älter Personen sind, desto öfter essen sie Schnitzel.]

■ STATISTISCHE Nullhypothese H_0 :

Alter und Schnitzelkonsumationsmenge weisen keinen Zusammenhang auf.
Junge und ältere Personen haben idente (ähnliche) Schnitzelkonsumationsmittelwerte.

■ STATISTISCHE Alternativhypothese H_1 (– hier ungerichtet):

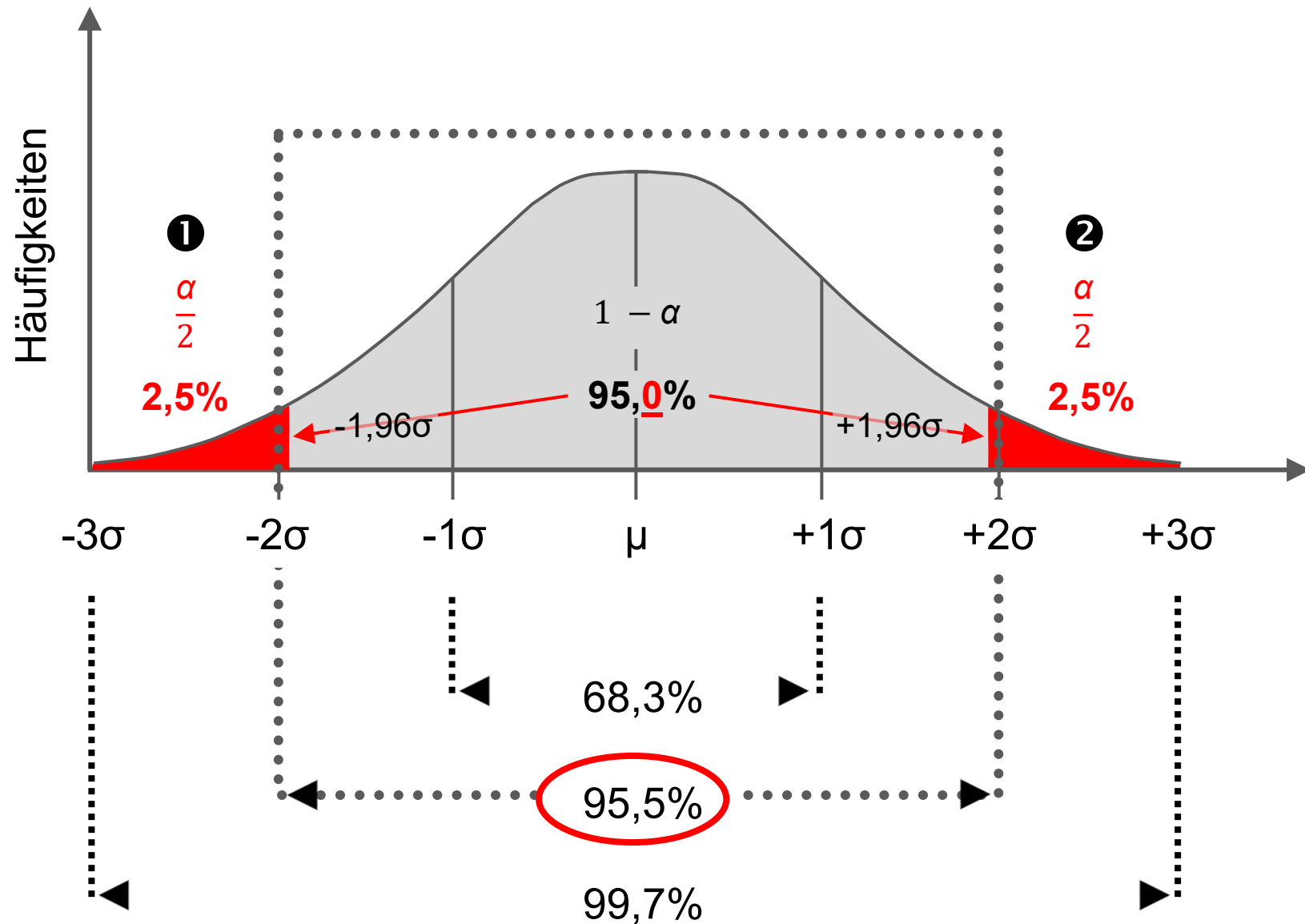
Alter und Schnitzelkonsumationsmenge weisen EINEN Zusammenhang auf.
Junge und ältere Personen haben unterschiedliche Schnitzelkonsumationsmittelwerte.

■ Signifikanz! (– wenn Wahrscheinlichkeit für das Ergebnis $< 5\%$):

*„Die Nullhypothese wird verworfen, die Alternativhypothese vorläufig akzeptiert:
Es besteht ein signifikanter Unterschied zwischen jüngeren und älteren Personen
in Bezug auf Affinität zu Schnitzeln.“*

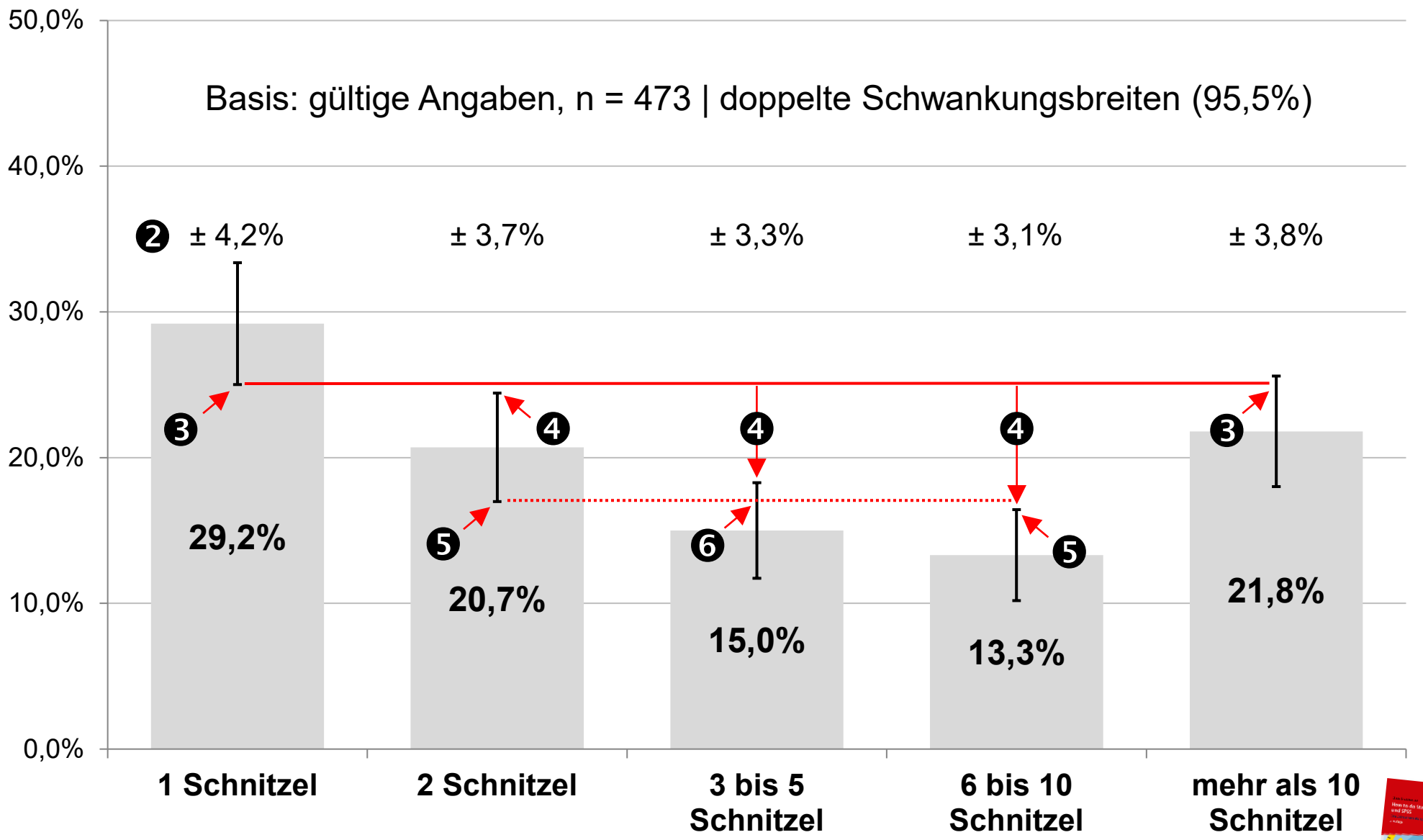
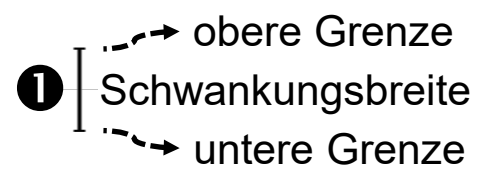


Signifikanzprüfungen: Der Hintergrund

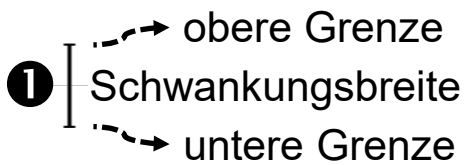


Signifikanz bei Häufigkeiten

„Wie viele Schnitzel essen Sie im Schnitt pro Jahr?“



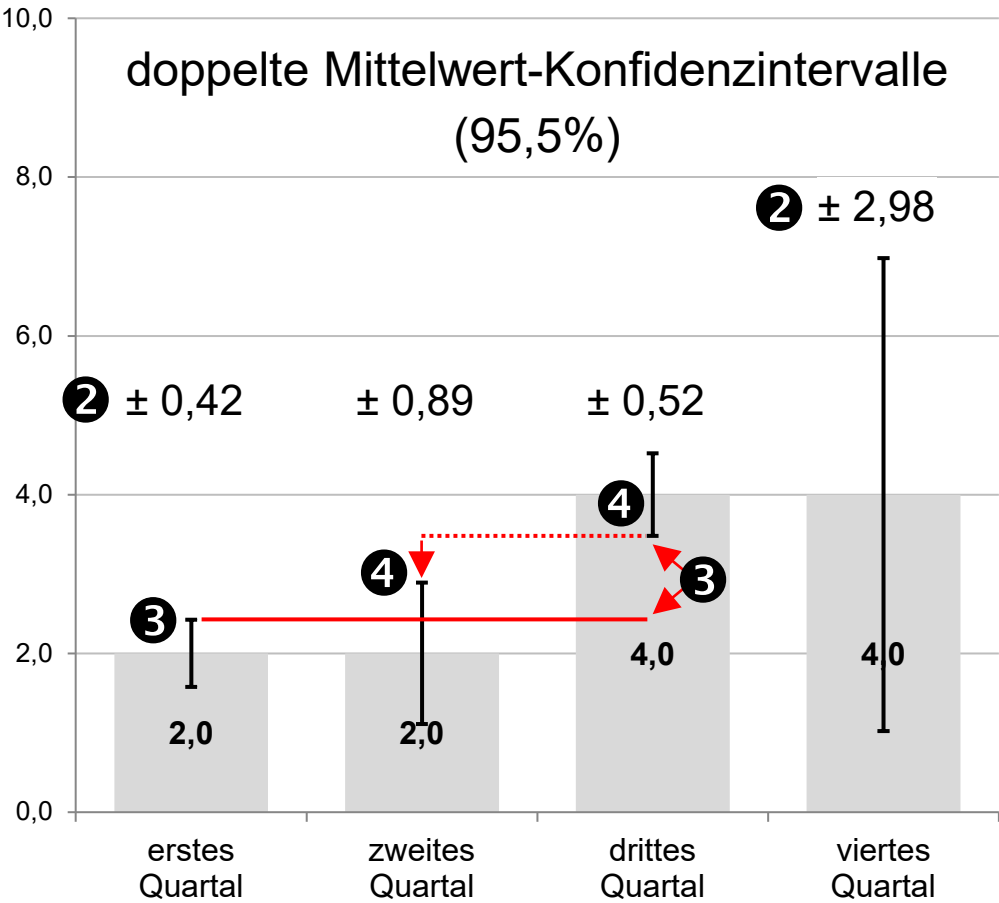
Signifikanz bei Mittelwerten



Anzahl gegessener Schnitzel

befragte Person	erstes Quartal	zweites Quartal	drittes Quartal	viertes Quartal
1	2	1	3	2
2	3	1	4	1
3	2	4	5	1
4	2	1	5	2
5	1	4	5	1
6	2	1	4	2
7	3	2	4	2
8	2	4	3	4
9	2	1	3	10
10	1	1	4	15
Mittelwert	2,0	2,0	4,0	4,0
Standardabweichung (hier: einer Stichprobe)	0,67	1,41	0,82	4,71

Mittelwerte im Quartal



Signifikanz bei Kreuztabellen

	beobachtete Werte		erwartete Werte		Zeilenprozent		Spaltenprozent	
	derzeit studierend		derzeit studierend		derzeit studierend		derzeit studierend	
Alter	ja	nein	ja	nein	ja	nein	ja	nein
bis 40 J	150	50	120,0	80,0	75,0	25,0	83,3	41,7
über 40 J	30	70	60,0	40,0	30,0	70,0	16,7	58,3

$$\chi^2 = \sum_{\text{aller Zellen}} \frac{(\text{beobachtete Werte} - \text{erwartete Werte})^2}{\text{erwartete Werte}} = \frac{(150 - 120)^2}{120} + \frac{(50 - 80)^2}{80} + \frac{(30 - 60)^2}{60} + \frac{(70 - 40)^2}{40}$$

$\chi^2 = 7,5 + 11,25 + 15 + 22,5 = 56,25$ ③ ④ Teststatistik Chi²: kritische Werte

Wenn $\chi^2_{\text{empirisch}} > \chi^2_{\text{kritisch}} \rightarrow \text{H}_0$

$df = (\text{Spalten} - 1) \cdot (\text{Zeilen} - 1) = 1$ ⑤

$\chi^2_{\text{empirisch}} > \chi^2_{\text{kritisch}} (\alpha_{0,05}) \rightarrow \text{H}_0$ ⑥ *

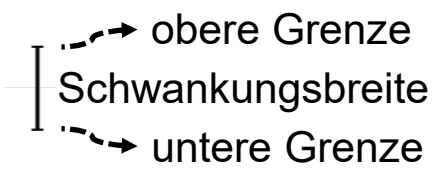
$\chi^2_{\text{empirisch}} > \chi^2_{\text{kritisch}} (\alpha_{0,01}) \rightarrow \text{H}_0$ ⑦ **

$\chi^2_{\text{empirisch}} > \chi^2_{\text{kritisch}} (\alpha_{0,001}) \rightarrow \text{H}_0$ ⑧ ***

Anzahl Freiheitsgrade		Konfidenzniveau (1 - α)			
df		0,90	0,95	0,99	0,999
⑤ 1		2,71	⑥ 3,84	⑦ 6,63	⑧ 10,83
2		4,61	5,99	9,21	13,82
3		6,25	7,81	11,34	16,27
4		7,78	9,49	13,28	18,47
5		9,24	11,07	15,09	20,52
usw.		10,64	12,59	16,81	22,78



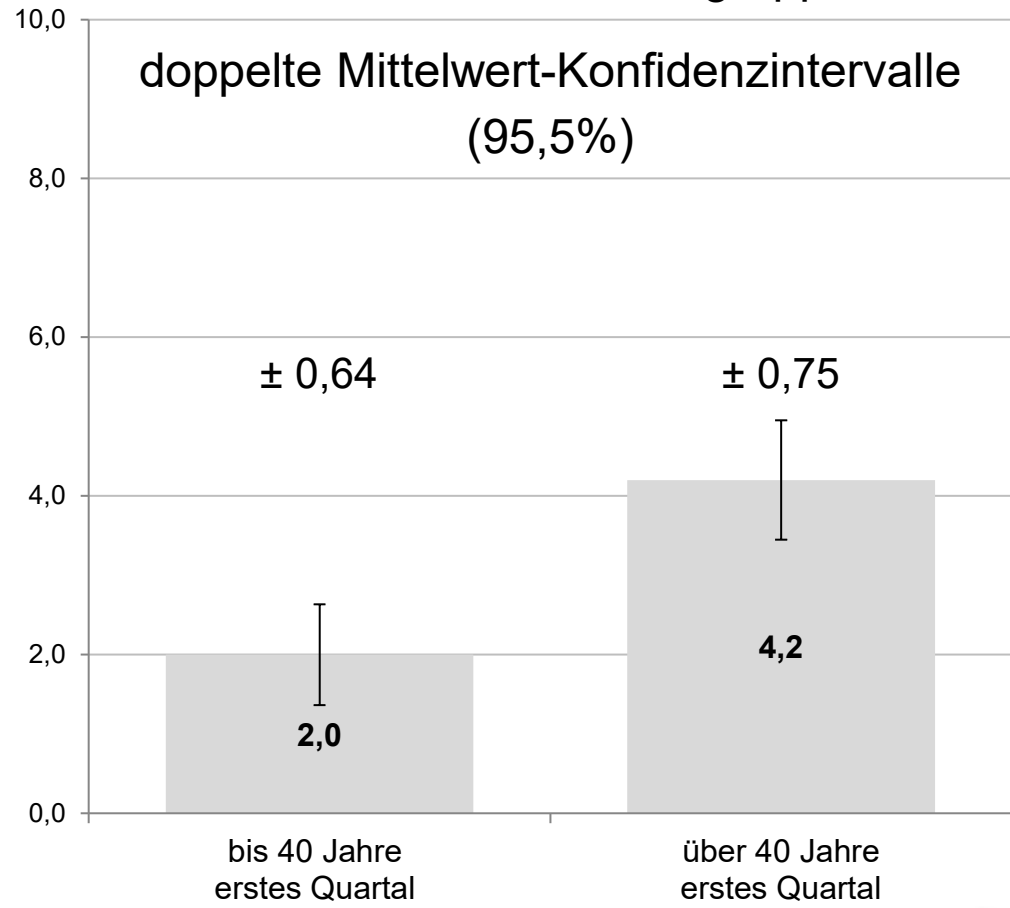
UNabhängiger Mittelwertsvergleich metrisch > nominal



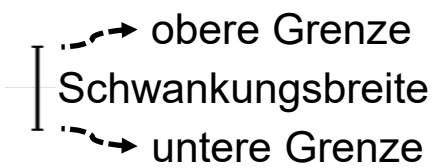
Anzahl gegessener Schnitzel

befragte Person	Alter	erstes Quartal	erstes Quartal
1	bis 40 J	2	
2	bis 40 J	3	
3	bis 40 J	2	
4	bis 40 J	2	
5	bis 40 J	1	
6	über 40 J		4
7	über 40 J		3
8	über 40 J		5
9	über 40 J		4
10	über 40 J		5
Mittelwert		2,0	4,2
Standardabweichung (hier: einer Stichprobe)		0,71	0,84

Mittelwerte nach Altersgruppen



Abhängiger Mittelwertsvergleich metrisch, metrisch, metrisch

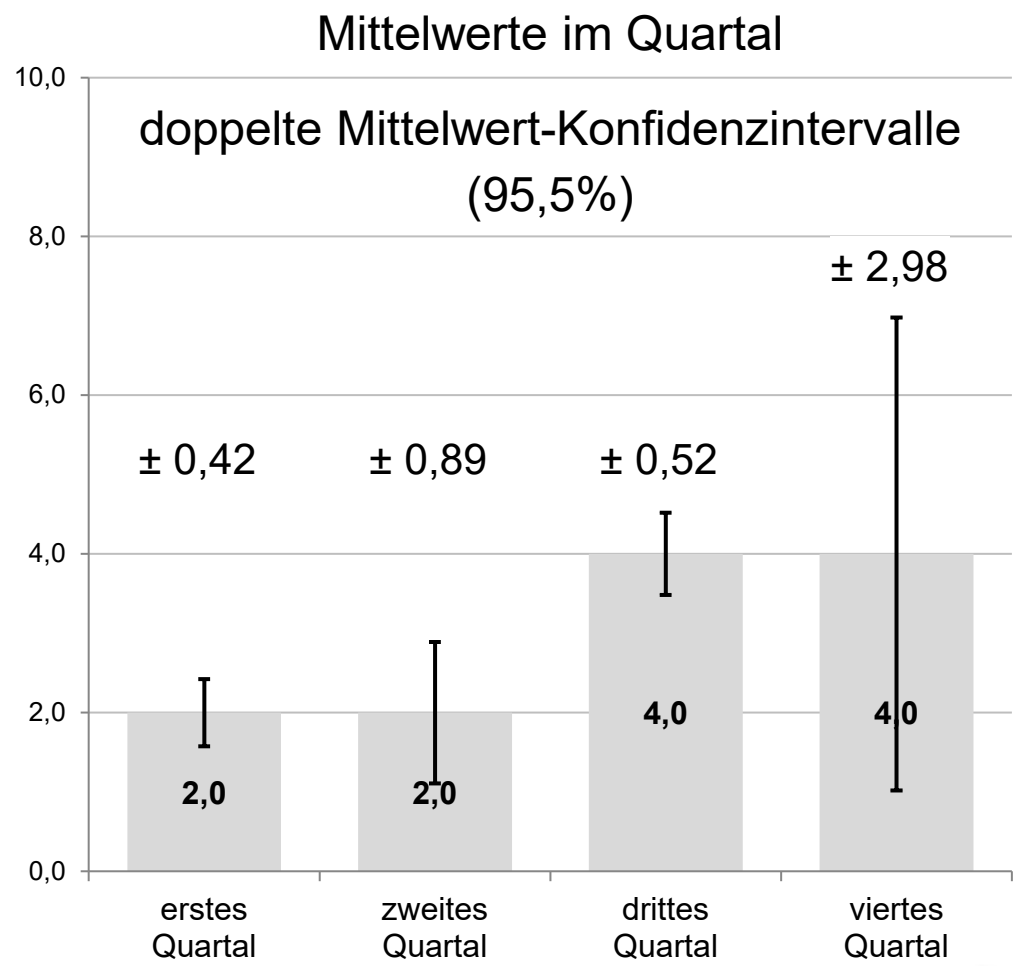


Anzahl gegessener Schnitzel

befragte Person	erstes Quartal	zweites Quartal	drittes Quartal	viertes Quartal
1	2	1	3	2
2	3	1	4	1
3	2	4	5	1
4	2	1	5	2
5	1	4	5	1
6	2	1	4	2
7	3	2	4	2
8	2	4	3	4
9	2	1	3	10
10	1	1	4	15
Mittelwert	2,0	2,0	4,0	4,0

Standardabweichung
(hier: einer Stichprobe)

0,67	1,41	0,82	4,71
------	------	------	------



Signifikanz ein- und zweiseitig prüfen

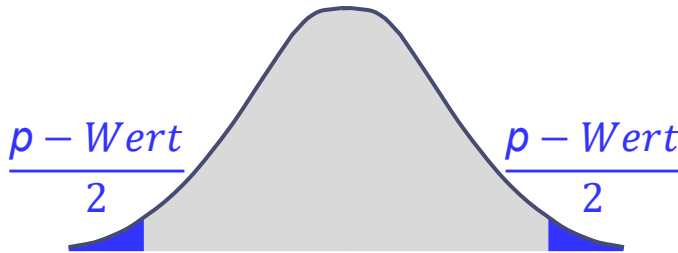
α = Irrtumswahrscheinlichkeit = Signifikanzniveau = 5%

[= „Risiko“ für Fehlentscheidung]

[$1 - \alpha$ = Konfidenzniveau = 95%]

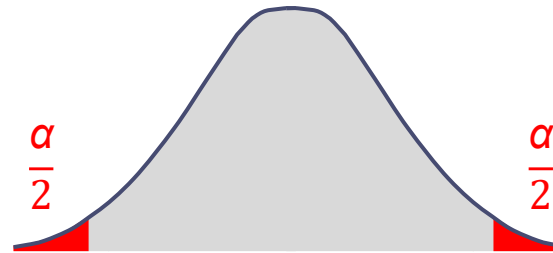
p-Wert = Wahrscheinlichkeit für „dieses“ (oder ein „stärkeres“) Ergebnis, wenn die H_0 gilt

SPSS berechnet
zweiseitigen p-Wert



Ein zweiseitiger p-Wert muss
für einseitige Tests halbiert
werden.

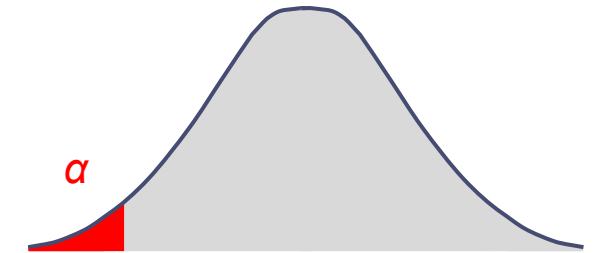
zweiseitiger
Hypothesentest



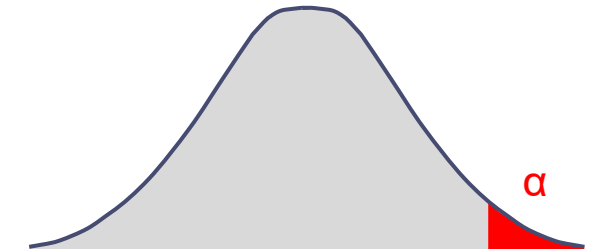
signifikant, wenn: $p - Wert < \alpha$

[Eigentlich, wenn: $2 \cdot \frac{p - Wert}{2} < 2 \cdot \frac{\alpha}{2}$]

einseitiger
Hypothesentest



signifikant, wenn: $\frac{p - Wert}{2} < \alpha$



signifikant, wenn: $\frac{p - Wert}{2} < \alpha$

► Entscheidung für einseitig – ein Muss im VORFELD!

PRO: ist genauer und „schneller“ signifikant

CONTRA: „gefährlich“, wenn das tatsächliche Ergebnis
eine zur H_1 „gegenteilige“ Differenz aufweist,
muss H_0 trotz Unterschieds akzeptiert werden!



Empirischer Forschungsprozess



Tipps für eine erfolgreiche Summary

Jede der folgenden Informationen sorgt für ein Mehr an Transparenz.

- **Exakte Themenstellung, Zweck** – von wem und für wen wurde erhoben?
- Genaue **Wortlaute im Fragebogen!**
- Auf wen (= **Grundgesamtheit**) beziehen sich die Daten?
- **Stichprobenziehung** (Zufallsauswahl, Quotierung oder willkürliche Auswahl)
- **Feldzeit** (wann und wie lange im Feld?)
- Größe und Repräsentativität des **Rücklaufs** (= Nettostichprobe)
- Vorhandensein und Größe von **Gewichtungsfaktoren**



Strukturvergleich Stichprobe – Grundgesamtheit

Struktur der Befragten

		n	in %	N
Total, alle Befragten		400	100,0	Bevölkerung 14 bis 75 Jahre in % *)
Wohnort liegt im Bundesland	Wien	87	21,8	21,8
	Niederösterreich	75	18,8	18,7
	Burgenland	14	3,5	3,3
	Oberösterreich	67	16,8	16,7
	Salzburg	24	6,0	6,3
	Tirol	33	8,3	8,5
	Vorarlberg	19	4,8	4,4
	Steiermark	55	13,8	14,0
	Kärnten	26	6,5	6,3

1

2

3

*) Quelle: Statistik Austria, 2023a.

Vgl.: Braunecker, C. (2023). *How to do empirische Sozialforschung. Eine Gebrauchsanleitung* (2. Auflage). facultas/utb. S. 150–156.

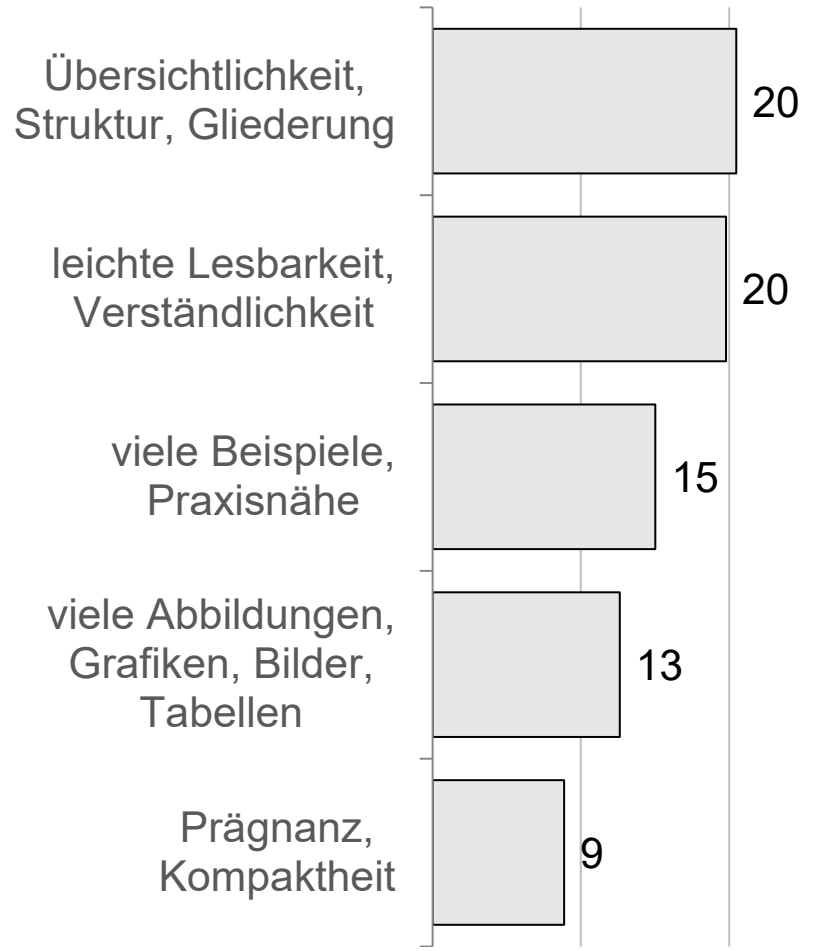


Gute Gliederung und Verständlichkeit sind MUSS-Kriterien!



„Wie muss ein ideales Fachbuch für Sie beschaffen sein?“
Basis: Expert:innengespräch mit 15 Master-Studierenden unterschiedlicher Studienrichtungen an der Universität Wien

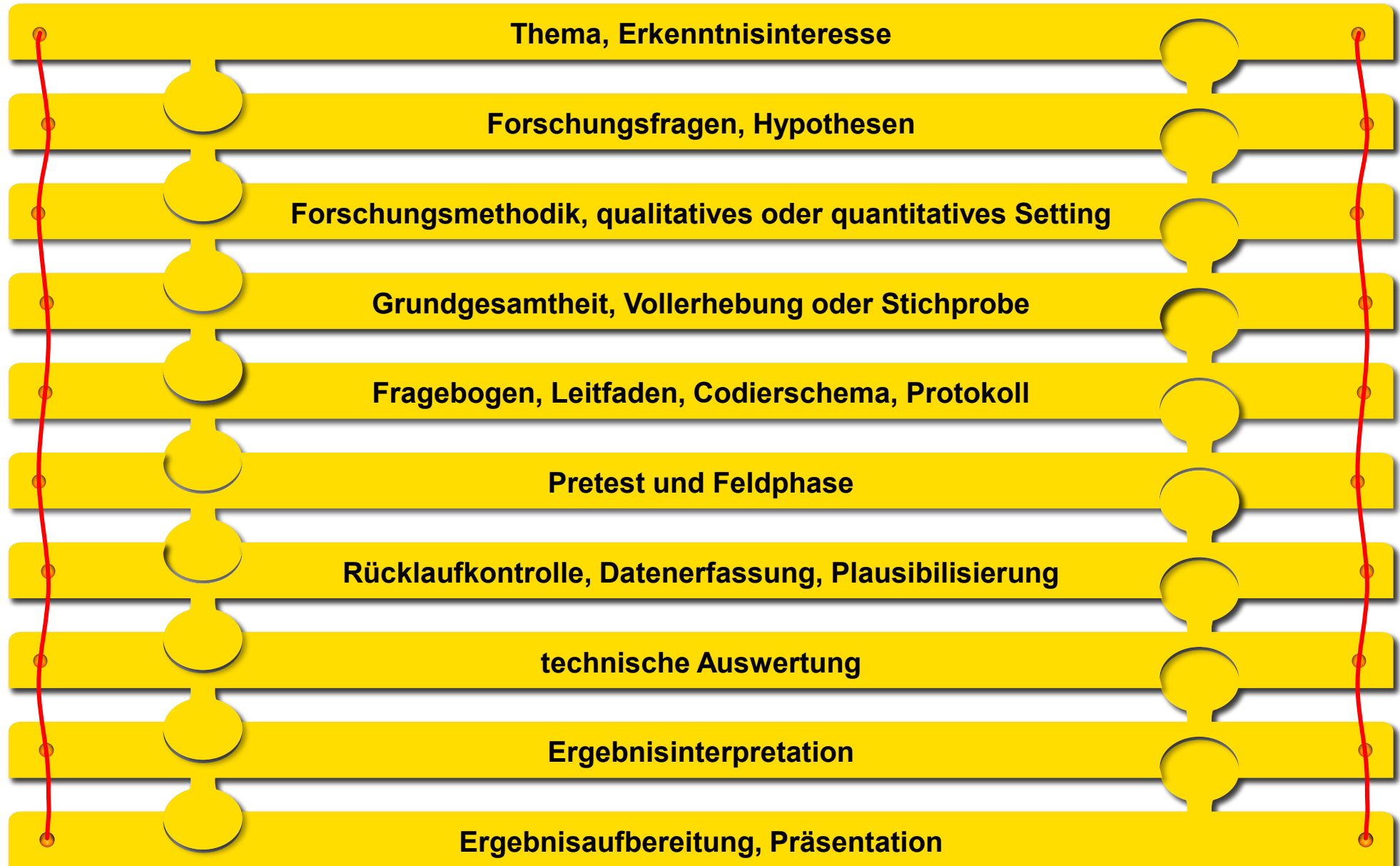
Die wichtigsten fünf Kriterien für Fachbücher:



F 5: „Wie muss ein ideales Fachbuch für Sie beschaffen sein?“
Basis: Haben in den letzten 12 Monaten zumindest ein Fachbuch gelesen | n = 133 | Angabe

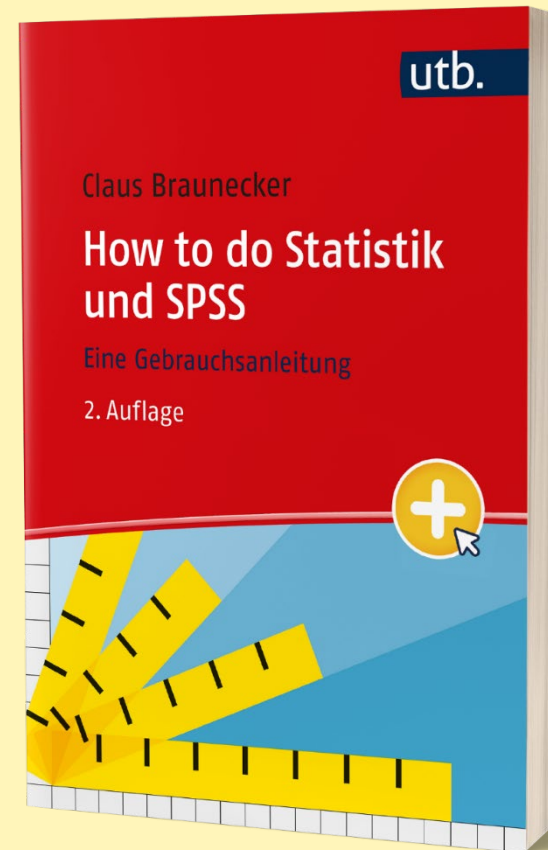
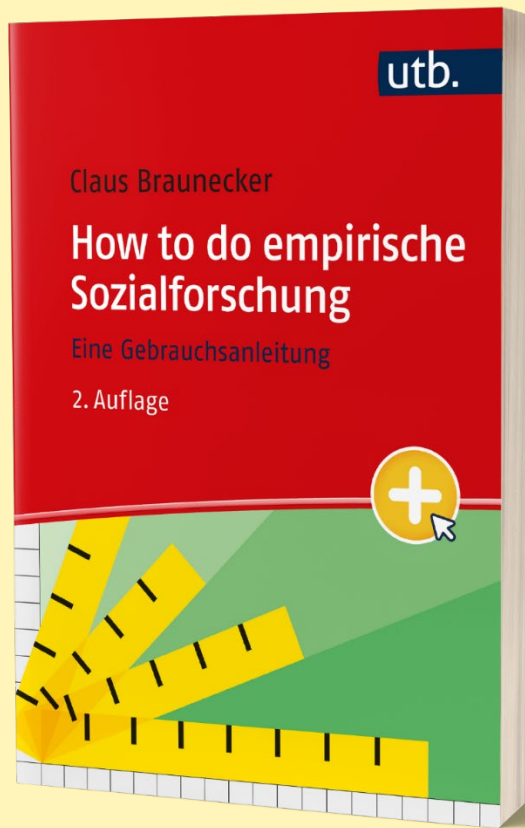


Empirischer Forschungsprozess – zusammenhängendes Ganzes!



Empirie – leicht wie nie.

Statistik und SPSS – ganz ohne Stress.



howtodo.at