



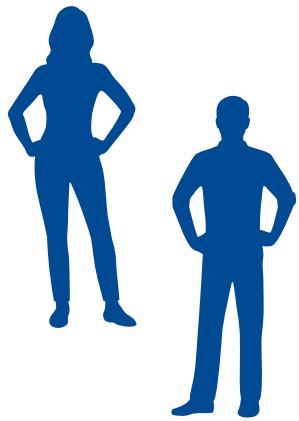
angewandte **w**irtschafts- und **m**edienpsychologie

Interventionelle Versuchspläne

Thema 04

Effekte finden

Intervention



Untersuchungsobjekte



Intervention
(Treatment)



Messung der Effekte



Wir wollen viel „Signal“, wenig „Rauschen“



Kevin Doncaster, Flickr, 2016, [URL](#)

- ▶ Experimentelle Kontrolle des „Rauschens“
- ▶ Das zentrale Bemühen des Untersuchers gilt der experimentellen Kontrolle der sog. Datenfluktuation (Datenvarianz; Rauschen).
- ▶ Variabilität in Daten ist natürlich und allgegenwärtig.
- ▶ Der Grundgedanke der experimentellen Versuchsplanung besteht darin, unerwünschte Variabilität (Rauschen) von erwünschter (Signal) zu trennen.
- ▶ Rauschen ist Variabilität in den Messergebnissen, die nicht durch die Theorie erklärt werden, sondern z. B. von Messfehlern stammen.
- ▶ Signal ist die Variabilität, die durch die Theorie erklärt wird. So sollte z. B. in einer Gruppe, die das neue Medikament bekommt, eine höhere Heilungsrate bestehen, als in der Kontrollgruppe. Es gibt also Variabilität zwischen den Gruppen, die erwünscht ist.

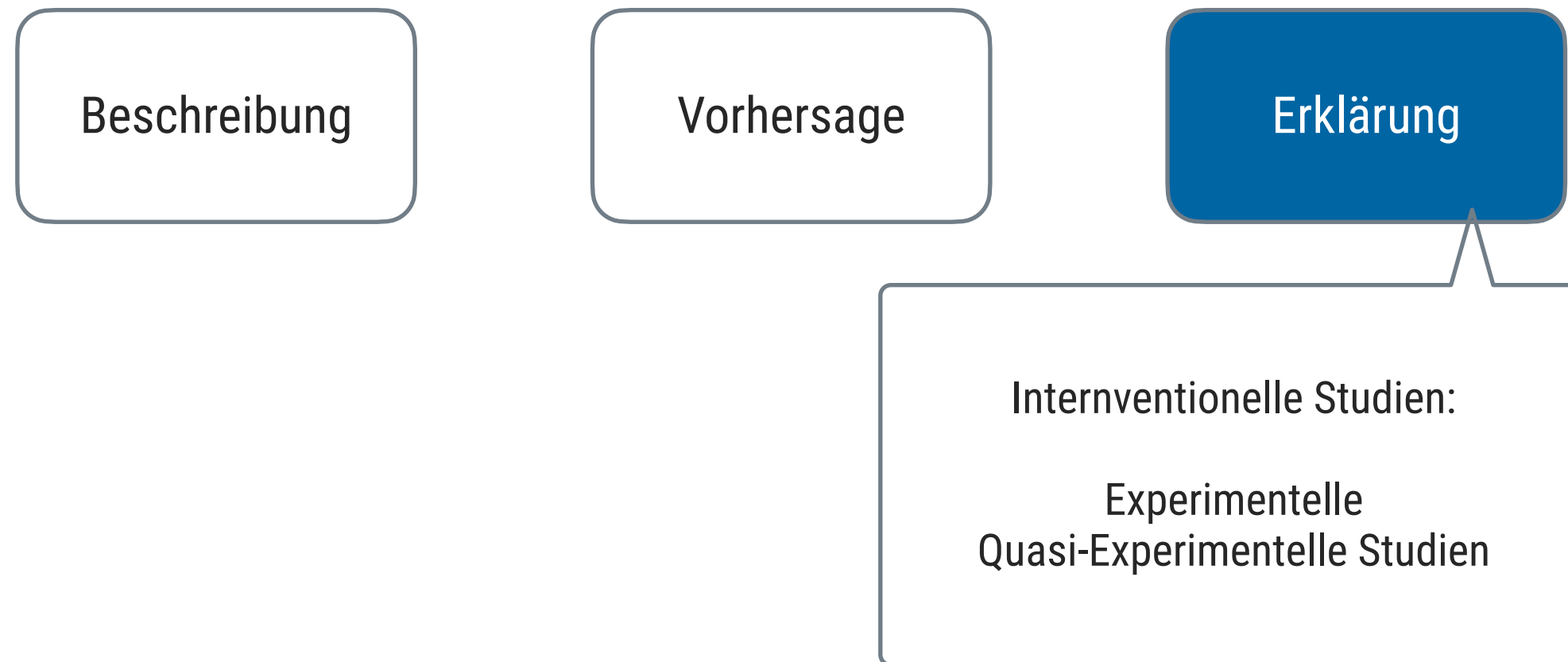
Drei Arten epistemologischer Studien

Beschreibung

Vorhersage

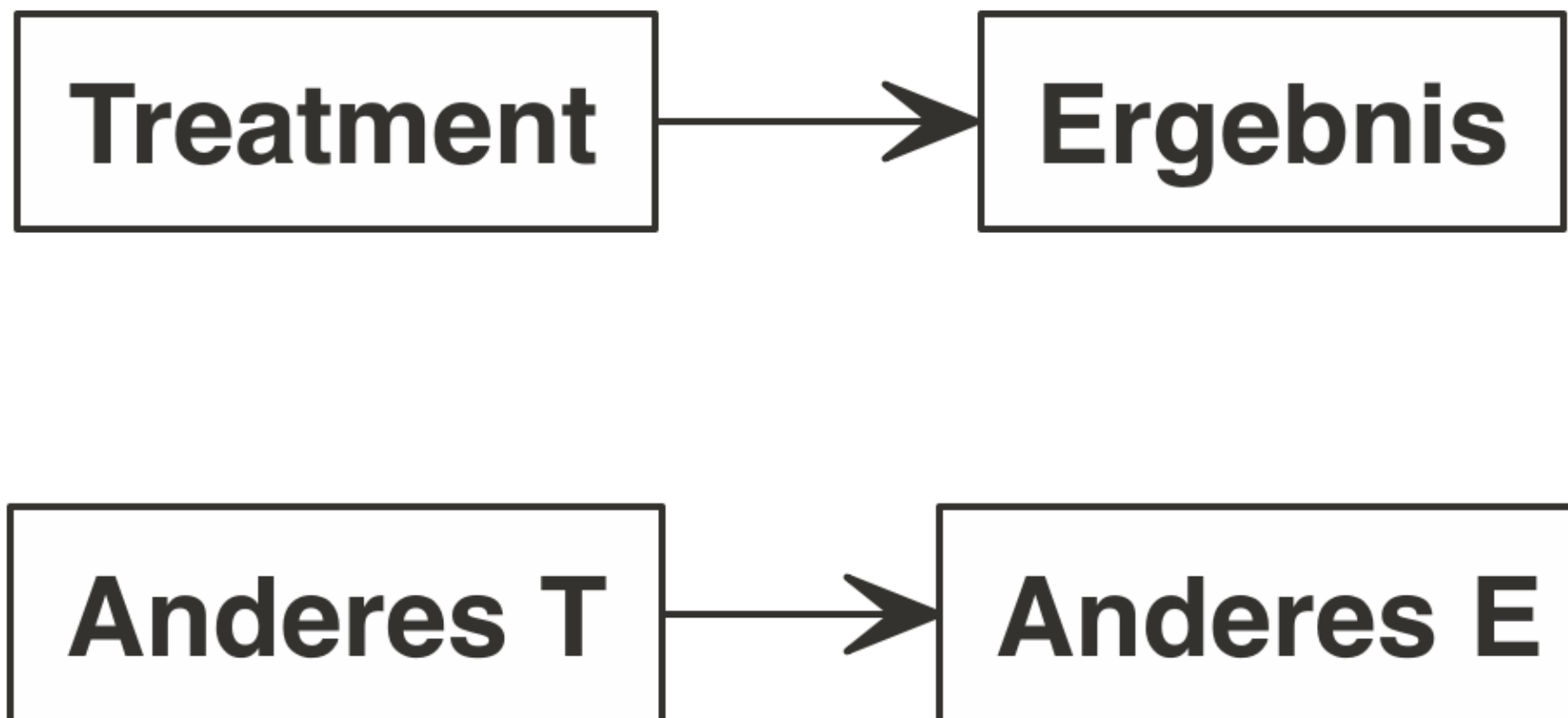
Erklärung

Drei Arten epistemologischer Studien (2)



Was ist ein Effekt?

- ▶ Ein Effekt (eines Treatments T/einer Intervention) ist hier definiert als der Unterschied im Ergebnis zwischen was passiert ist nachdem das Treatment verabreicht wurde und was passiert wäre, wenn stattdessen eine andere Intervention verabreicht worden wäre, unter der Annahme, dass sonst alles andere gleich gewesen wäre.



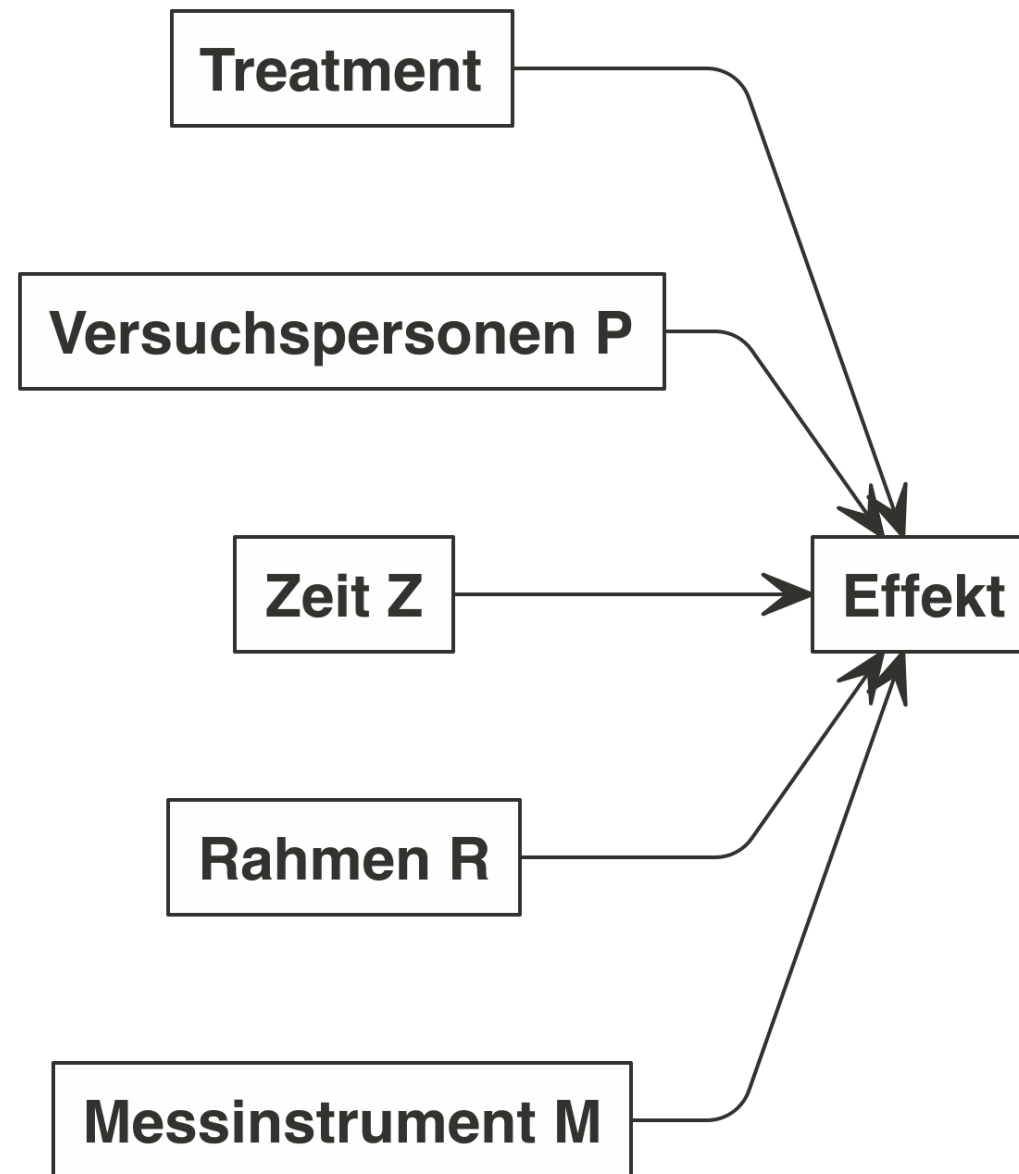
Leider ist es unmöglich, einen Effekt zu bestimmen

- ▶ Leider ist es per definitionem unmöglich, zu messen, was der Fall wäre, wenn die Welt anders wäre bzw. wir eine andere Intervention angewendet hätten.
- ▶ Man spricht daher von einer *kontrafaktischen* Definition.
- ▶ Auf dieser Basis definiert man den mittleren Treatmenteffekt (average treatment effect, AVE) wie folgt (bei einer Stichprobe oder Population der Größe N und für die AV Y).

$$AVE = \sum_i^N [Y_i(1) - Y_i(0)]$$

- ▶ Nota bene: Man kann bei einer Person (i) nur entweder $Y(0)$ oder $Y(1)$ beobachten.

Der Effekt ist eine Funktion von fünf Größen



Die fünf Determinanten eines Effekts

- ▶ Ursache (U) – wie?
 - ▶ Damit ist das Treatment gemeint, soweit es die Ursache für den Effekt ist.
 - ▶ Art und Menge des Treatment spielt natürlich eine Rolle (50 mg Ibuprofen wirken anders als 500 mg).
- ▶ Versuchsobjekte (O) – wer?
 - ▶ Häufig Personen in der Psychologie, aber auch Firmen, Teams etc.
 - ▶ Auf dem Versuchsobjekt werden das Treatment angewendet und die Effekte gemessen.
 - ▶ Die Effekte können als zwischen den Versuchsobjekten variieren.
- ▶ Zeit (Z) – wann?
 - ▶ Der Effekt hängt vom Zeit des Treatments ab und von der Periode zwischen Applikation und Messung des Effekts.
- ▶ Rahmen (R) – wo?
 - ▶ Eine Kopfschmerztablette wirkt vielleicht besser, wenn sie während eines entspannenden Bades an einem ruhigen Abend zuhause eingenommen wird (als während eines stressigen Projektmeetings).
- ▶ Messinstrument (M) – was?
 - ▶ Je nach verwendeten Messinstrument oder Zielvariablen können die Effekte des Treatments deutlicher oder weniger deutlich zum Tragen kommen.

Effektstärke als Funktion der fünf Determinanten

$$ES = f(UOZRM)$$

Konstruktvalidität als Funktion der fünf Determinanten

- ▶ Fehlzuordnungen in den fünf Determinanten begrenzen die Konstruktvalidität eines Effekts.
- ▶ Ursache (U) – wie?
 - ▶ Eine neue Coachingmethode (T) hat keinen Effekt, aber die Freundlichkeit des Coaches wurde fälschlich als Effekt des Treatments T interpretiert.
- ▶ Versuchsobjekte (O) – wer?
 - ▶ Die Forscher gaben an, die Teilnehmer:innen waren „Health care professionals“, dabei handelte es sich um ungelernte Hilfskräfte.
 - ▶ Die Teilnehmer waren nicht kooperativ und haben sich nicht an die Instruktionen gehalten.
- ▶ Zeit (Z) – wann?
 - ▶ Die Forscherin nahm an, das Treatment habe keinen Effekt, aber sie hat nur zu früh gemessen.
- ▶ Rahmen (R) – wo?
 - ▶ Vielleicht funktioniert Coaching nur im reichen, demokratischen Westen bei gebildeten Menschen, hängt also vom Rahmen (Ort, Kontext) ab? Die Forscher sind sich über diese Beschränkung aber nicht im klaren.
- ▶ Messinstrument (M) – was?
 - ▶ Der Statistiktest des Professors hat nicht das Statistikwissen gemessen, sondern die Spick-Kompetenz (aber die dafür sehr genau). Leider fiel das dem Prof nicht auf (leider für den Prof, die Studis fanden es super).

Interne Validität als Spezialfall der Konstruktvalidität

- ▶ Die interne Validität kann als Spezialfall der Konstruktvalidität gesehen werden.
- ▶ Bedrohungen bzw. Einschränkungen der internen Validität beziehen sich nur bestimmte Fehlbeschreibungen beim Determinanten der Ursache (U): Störvariablen die auch ohne das Treatment T zu einen Effekt führen würden.
- ▶ Versuchsobjekte (O) – wer?
 - ▶ In einer medizinischen Studie werden die gesünderen Patienten in die Gruppe mit dem neuen Medikament gebracht, aber die kränkeren in die Gruppe mit dem bisherigen Standard-Medikament.
- ▶ Zeit (Z) – wann?
 - ▶ Zur Messung von Stress wird Speichel-Kortisol entnommen. Die Proben der Experimentalgruppe bleiben aber zulange der Raumtemperatur und Sauerstoff ausgesetzt (im Gegensatz zu den Proben der Kontrollgruppe).
- ▶ Rahmen (R) – wo?
 - ▶ Die Verabreichung des neuen Medikaments fand in einer Wellness-Klinik statt. Das alte Medikament im Keller des alten Krankenhauses.
- ▶ Messinstrument (M) – was?
 - ▶ In der Experimentalgruppe wurde ein erfahrener Beobachter zur Analyse des Assessment-Centers eingesetzt, in der Kontrollgruppe der Praktikant, der von Tuten und Blasen keine Ahnung hatte.

Externe Validität

- ▶ Die externe Validität fragt, wie gut ein Effekt verallgemeinert werden kann.
- ▶ Ursache (U) – wie?
 - ▶ Eine Forscherin hat nur einen Teil eines neuen Coachingskonzepts in einem Treatment umgesetzt. Sind die Ergebnisse jetzt auf das ganze Coachingkonzept verallgemeinerter?
 - ▶ Gibt die Forscherin korrekt an, dass nur ein Teil im Treatment umgesetzt wird, liegt hingegen keine Schwäche in der Konstruktvalidität vor.
- ▶ Versuchsobjekte (O) – wer?
 - ▶ Ein Krebsmedikament wurde nur an älteren Frauen evaluiert (und dies hat das Forscherteam korrekt angegeben). Ob der Effekt wohl auch für andere Altersgruppen und Geschlechter gilt?
- ▶ Zeit (Z) – wann?
 - ▶ Ein Allergiemittel reduziert kurzfristig die relevanten Symptome. Aber ob es wohl auch langfristig Effekte hat?
- ▶ Rahmen (R) – wo?
 - ▶ Ob psychologische Erkenntnisse auch in nicht-studentischen Populationen gelten?
- ▶ Messinstrument (M) – was?
 - ▶ Finden sich die Effekte des Vorschultrainings wohl nicht nur in einem bestimmten Intelligenztest, sondern auch in anderen? Haben sie vielleicht auch Effekte auf die emotionale (nicht nur kognitive) Entwicklung des Kindes?

Störfaktoren der internen Validität 1/2

Störfaktor	Beeinträchtigung
Zeitgeschehen (history)	Die beobachteten Effekte gehen nicht allein auf die Untersuchungsbedingungen, sondern auf unkontrollierte zwischenzeitliche Ereignisse zurück.
Reifung (maturation)	Wenn sich der zu untersuchende Sachverhalt bezüglich biologischer (und/oder psychosozialer Reifungsmerkmale verändert, ist mit reifungsbedingten Effekten zu rechnen, die den eigentlichen Befund überlagern.
Mehrfache Testung (test sophistication)	Bei mehrfacher Erhebung derselben Messdaten an ein und demselben Individuum können die während des zweite (dritten ...) Messzeitpunkts erhobenen Daten aufgrund vorangegangener Testung beeinflusst sein.
Instrumentierung (instrumentation)	Die gemessenen Werte gehen z. T. auf die (zwischenzeitlich erfolgte) Veränderung der Messinstrumente zurück (z. B. aufgrund mangelnder Objektivität und Reliabilität eines Tests.)
Statistische Regression (regression)	Werden mehr oder weniger extrem verschiedene Leistungsgruppen z. B. mit Hilfe eines Vortests gebildet, dann kann die mangelnde Reliabilität des Instruments zu einer statistischen Regression zur Mitte bei der zweiten Testung führen.

Störfaktoren der internen Validität 2/2

Störfaktor	Beeinträchtigung
Auswahlverzerrung (selection)	Bei nicht-zufälliger Bildung von Versuchsgruppen können die damit von Anfang an bestehenden systematischen Ausgangsunterschiede zwischen den Gruppen den eigentlichen Effekt überlagern.
Ausfalleffekte (experimental mortality)	Fallen im Untersuchungsverlauf Versuchspersonen von verschiedenen Versuchsgruppen aus, so kann das die eigentlichen Effekte beeinflussen, wenn die Ausfallquote für die Gruppen systematisch verschieden ist (z. B. mehr „Dropouts“ in Kontrollgruppe).
Versuchsleitereffekte (experimenter-bias effects)	Bleiben die Eigenschaften, Verhaltensweisen und/oder Versuchserwartungen des Untersuchers unkontrolliert, kann das eine systematische Beeinträchtigung der eigentlichen Befunde auslösen.
Interaktive Effekte (carry-over effects)	Wird ein Individuum unter verschiedenen Untersuchungsbedingungen untersucht und bleiben dabei Übertragungseffekte unkontrolliert, können dadurch die Untersuchungsbefunde verfälscht werden.

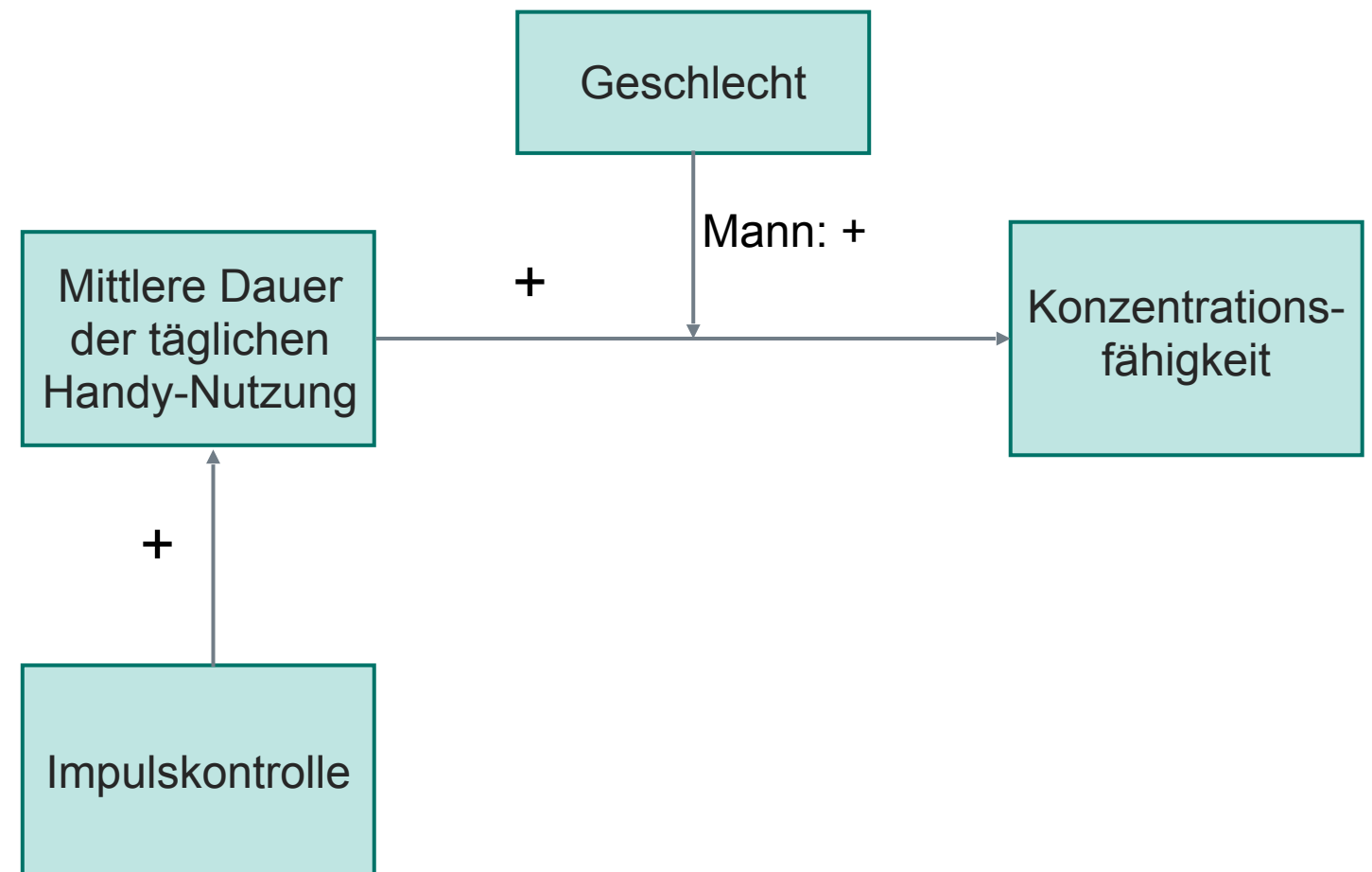
Übersetzen Sie Ihre Forschungsfrage in ein Modell

- ▶ Suchen Sie in der Literatur nach theoretischen Modellen bzw. Theorien, die die von Ihnen untersuchten Konstrukte *gemeinsam* betrachten.
- ▶ Man kann sich auch selber Modelle ausdenken; diese sollten aber möglichst stark in existierenden Theorien abgebildet sein und durch empirische Belege gestützt sein.

Laienformulierung

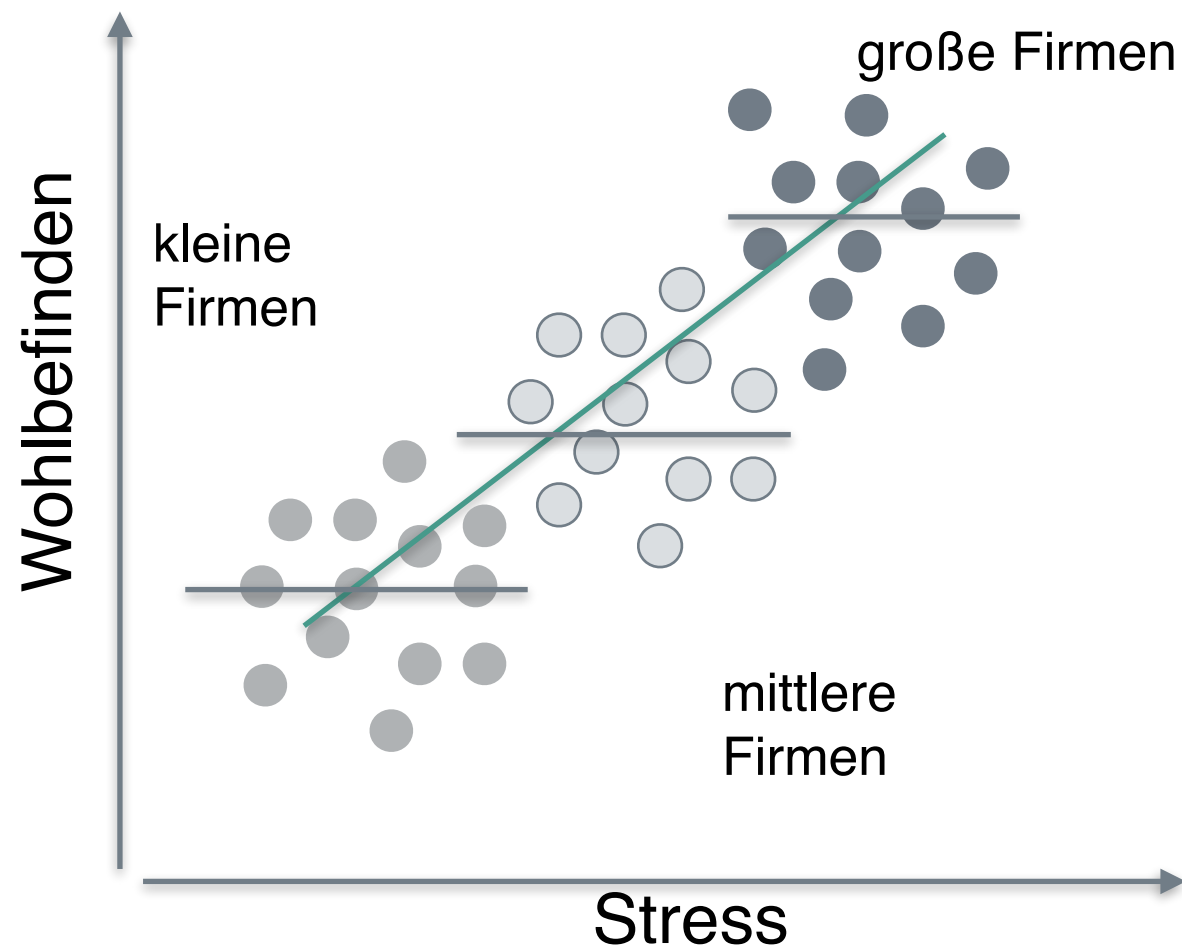
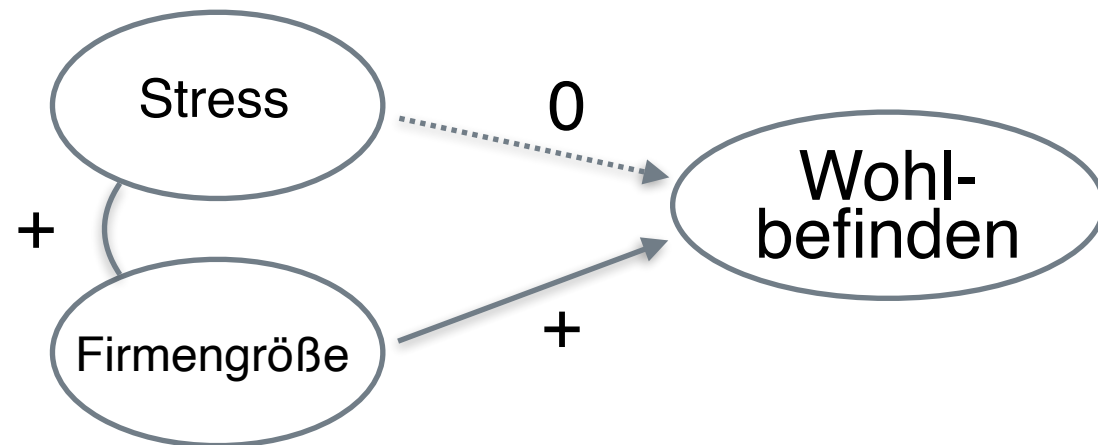
Wissenschaftliches Modell

„Wer viel am Handy rumdaddelt, der ist halt nicht so auf Zack, im Hirn, und so, vor allem bei Jungs, übrigens, liegt an den Genen. Ach ja, wer sich halt nicht im Griff hat, der daddelt halt mehr.“

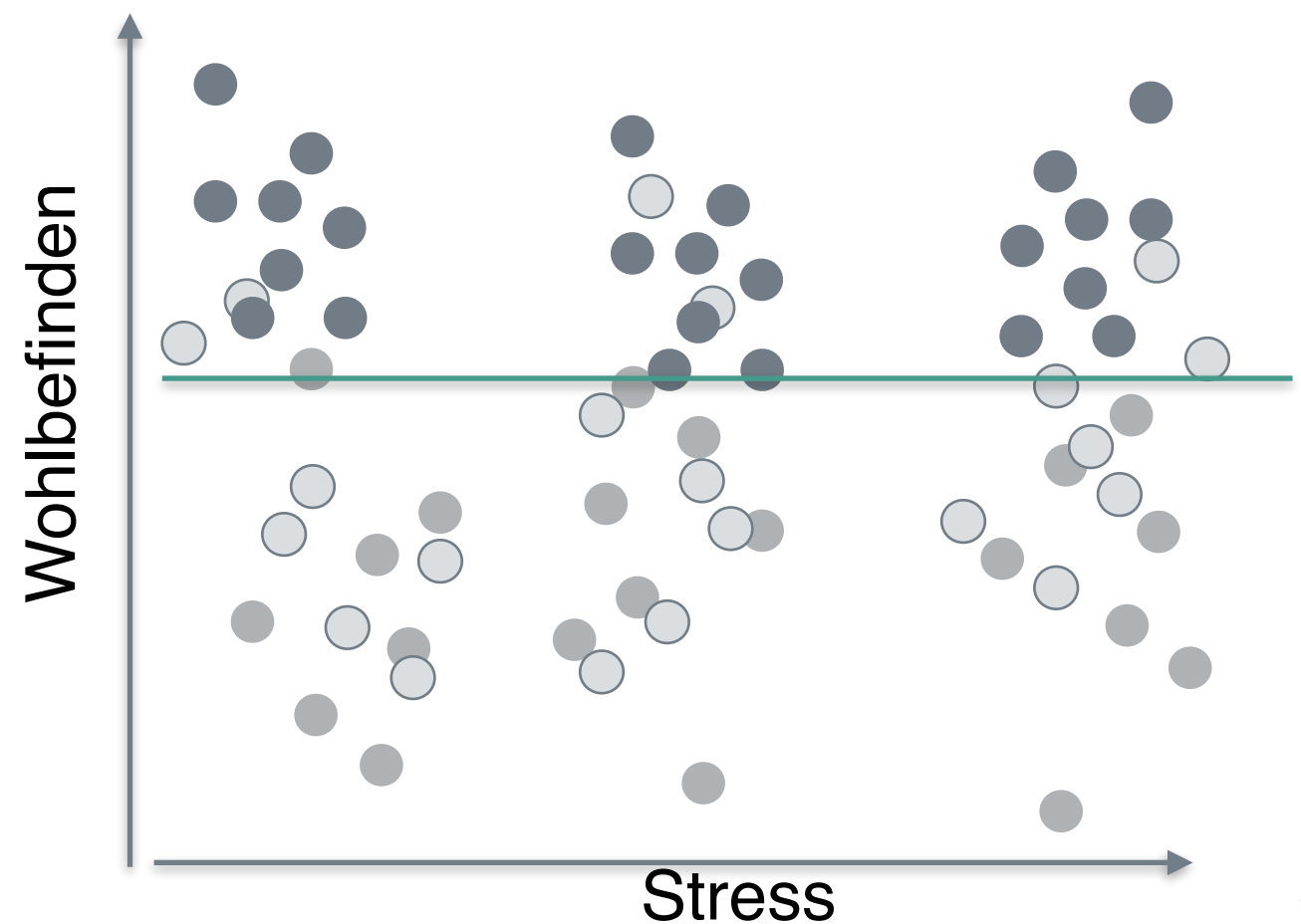
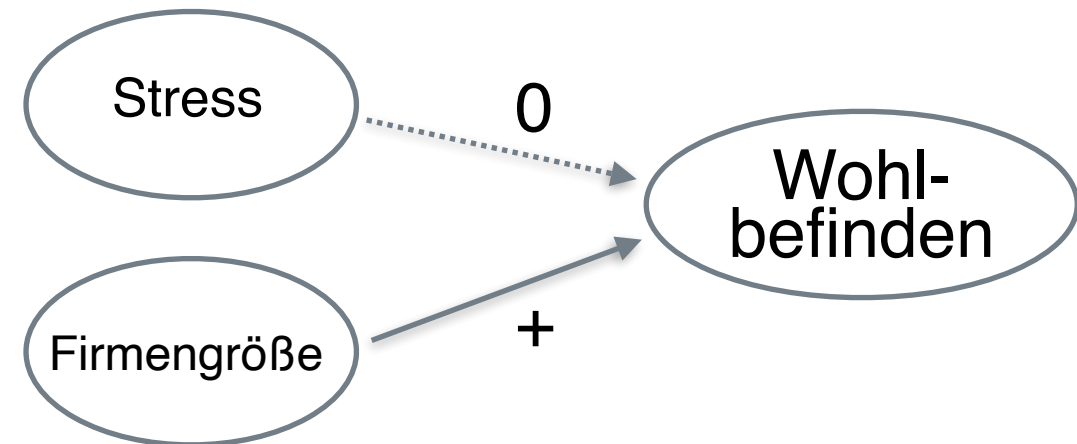


Aufdecken verborgener Zusammenhänge

Multivariate Beobachtungsstudie



Randomisiertes Experiment



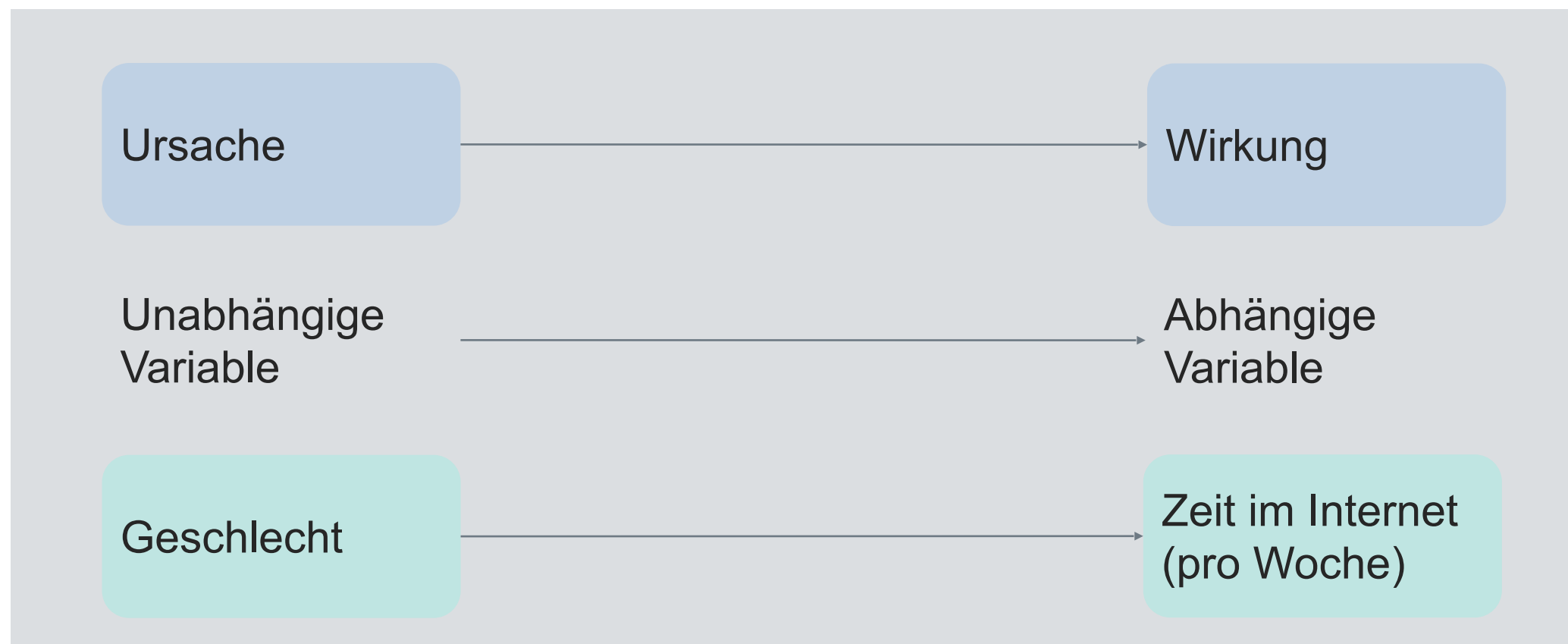
Variablen in experimentellen Studien (Kausalhypothesen)

Unabhängige Variable (UV)

- unabhängige Variable; „Ursache“
- Ihr Wert hängt von **keiner anderen Variable** (in der Studie) ab

Abhängige Variable (AV)

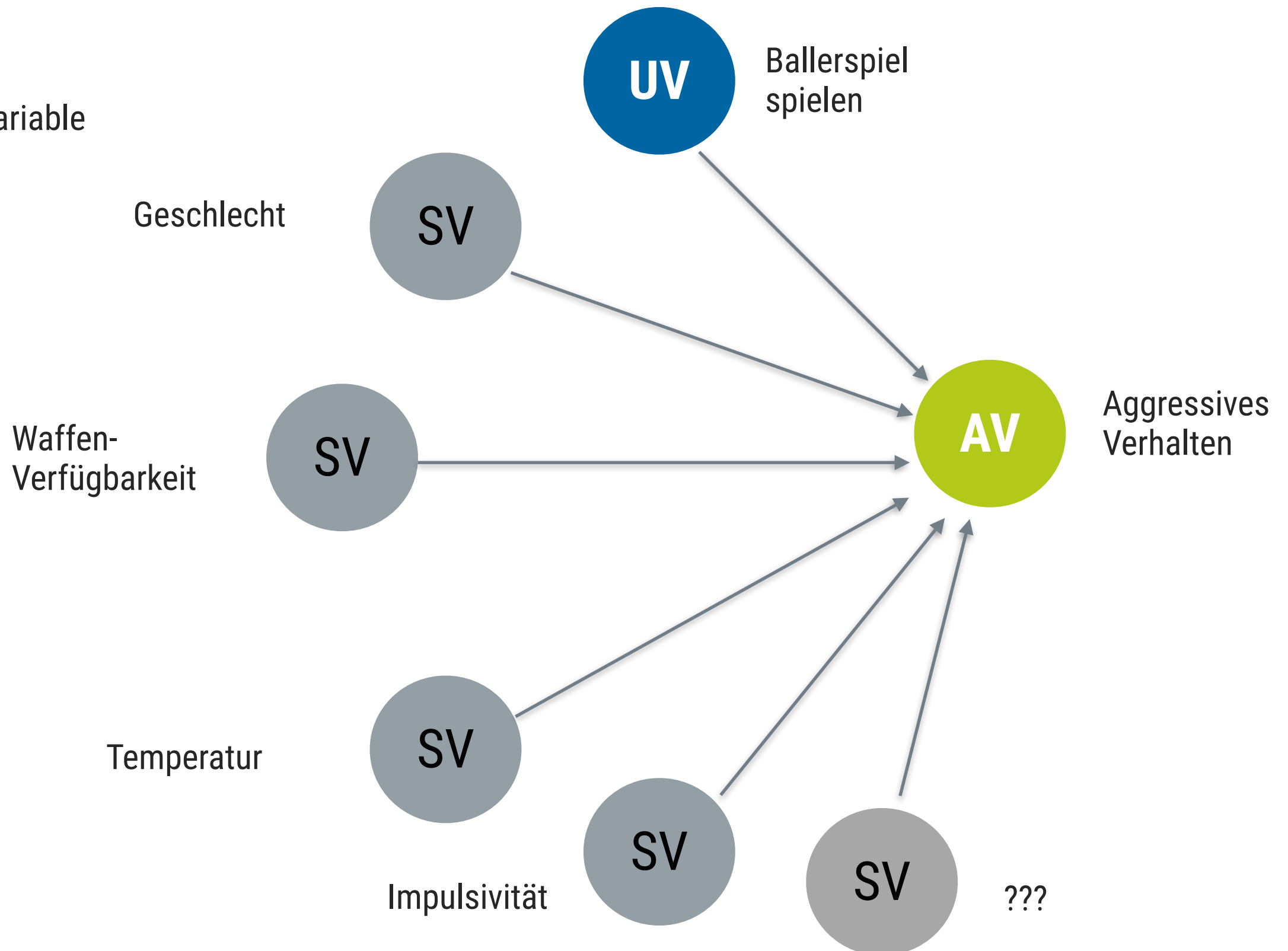
- abhängige Variable; „Wirkung“
- Ihr Wert hängt von der **Variation der unabhängigen Variable** ab



Die Begriffe UV und AV machen nur Sinn, wenn es eine Intervention in der Studie gibt.

Beispiel für eine Kausalhypothese: „Ballerspiele machen aggressiv“

*SV: Störvariable



Wann ist eine Ursache wirklich eine Ursache?

Theorie nach John Stuart Mill:

Die Stufen der UV gehen systematisch mit den Stufen der AV einher (Korrelation)

KOVARIATION

**Kausal-
schluss**

(Ursache
als Ursache bestätigt)

Veränderungen in der UV
müssen zeitlich vor den
Änderungen der AV
auftreten

Alternativerklärungen für
den Zusammenhang von UV
und AV müssen
ausgeschlossen sein

UV ZEITLICH VOR AV

KEINE ANDEREN URSACHEN
MÖGLICH

Störvariablen im Experiment

- ▶ Meist gibt es viele kausale Variablen, die eine AV beeinflussen (können)
- ▶ In einer Studie oder einem Experiment interessiert häufig nur der Einfluss einer Variablen, um eine spezifische Hypothese zu prüfen
- ▶ Alle in dieser Untersuchung nicht interessierenden Variablen, die die AV beeinflussen, heißen Störvariablen
- ▶ Störvariablen mögen selbst theoretisch interessant sein, in einem Experiment versucht man jedoch, sie unter Kontrolle zu bringen, um den Einfluss der interessierenden UV untersuchen zu können
- ▶ Kontrolliert man Störvariablen nicht, kann eine Veränderung in AV nicht eindeutig auf UV zurückgeführt werden, ein Kausalschluss ist dann nicht möglich.
- ▶ Beobachtungsstudien haben meist weniger Kontrolle über Störvariablen als Studien mit Intervention.
- ▶ Studien in kontrolliertem Umfeld sind meist in geringerem Maße Störvariablen ausgesetzt als Studien in „freier Wildbahn“.
- ▶ Studien mit Randomisierung umgehen das Problem der personengebundenen Störvariablen (wenn die Fallzahl groß genug ist).

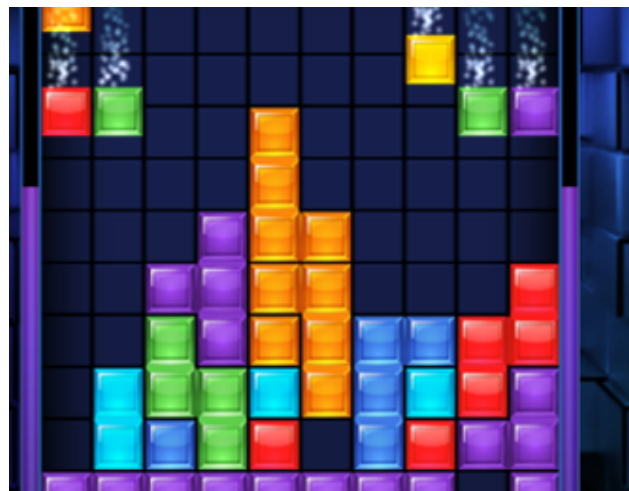
Systematische Störvariable

UV

Art des Computerspiel



Gruppe 1:
Computerspiel
mit **viel** Gewalt
Temperatur:
heißer Raum



Gruppe 2:
Computerspiel
mit **wenig** Gewalt
Temperatur:
kalter Raum

AV

Aggressives
Verhalten

- ▶ Hier lässt sich der Einfluss der Raumtemperatur **nicht** vom Einfluss der Computerspiels trennen.
- ▶ Der Einfluss der Raumtemperatur variiert systematisch mit der UV.
- ▶ Man spricht von einer systematischen Störvariablen, wenn die Variable auch einen Einfluss auf die AV hat.
- ▶ Die beiden Variablen (Art des Computerspiels und Raumtemperatur) sind dann konfundiert.
- ▶ Die interne Validität ist hier gefährdet!

UN-Systematische Störvariable

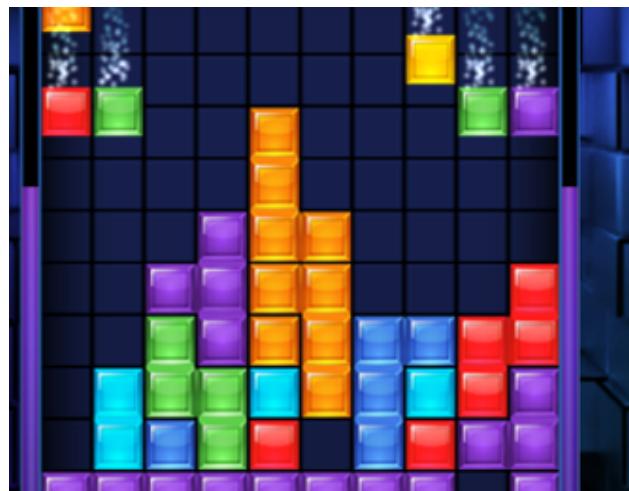
UV

Art des Computerspiel



Gruppe 1:
Computerspiel
mit **viel** Gewalt

Die Heizung war kaputt!
Es war mal kalt und mal
warm, ganz gemischt und
ohne System!



Gruppe 2:
Computerspiel
mit **wenig** Gewalt

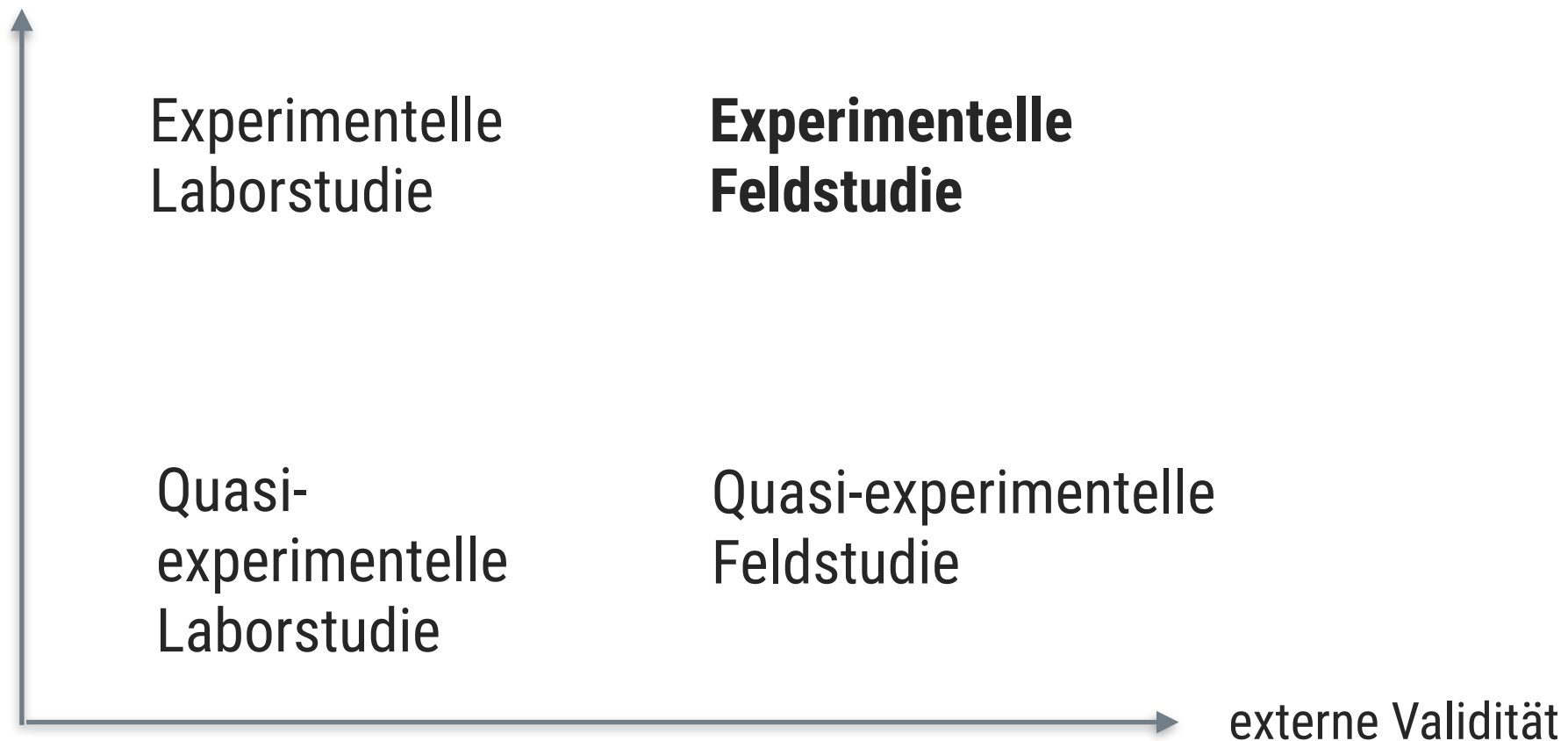
AV

Aggressives
Verhalten

- ▶ Schwankt eine Störvariable nicht systematisch mit der UV, so liegt keine (massive) Einschränkung der internen Validität vor.
- ▶ Unter einer nicht-systematischen Störvariablen versteht man eine Variable, die einen Einfluss auf die AV hat (mit ihr kovariiert), aber nicht systematisch mit der UV kovariiert
- ▶ Durch das nicht-systematische Variieren der Raumtemperatur „mittelt“ sich der Einfluss der Raumtemperatur wieder heraus.
- ▶ Es liegt dann keine Konfundierung vor.

Feldforschung vs. Laborforschung

interne Validität



Wann verzichtet man auf ein Experiment?

- ▶ Zumeist sind Experimente im Vergleich zu Beobachtungsstudien vorzuziehen, da Studien mit hoher interner Validität sind besser als Studien mit geringer interner Validität (*ceteris paribus*), wenn man an Ursachen interessiert ist.
- ▶ Es gibt aber Situationen, in denen man auf eine Beobachtungsstudie ausweichen wird:

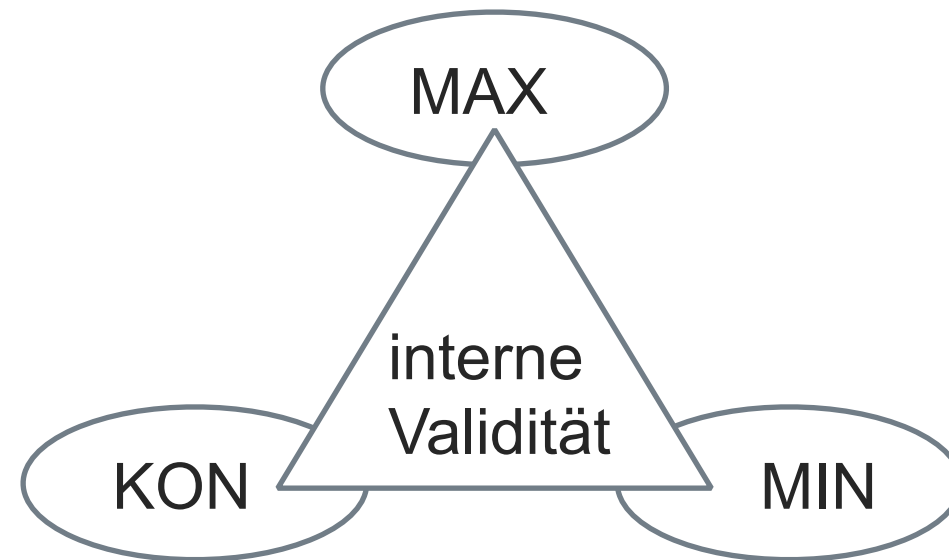
Problem	Beispielexperiment
Ethische Bedenken	Effekt von 20 Jahre Kette rauchen auf Krebsentwicklung
Unpraktisch	Effekt von langfristigem Kaffeekonsum auf Blutdruck
Unmöglich	Effekt von Geschlecht oder Persönlichkeit auf Intelligenz
Aufwändig	Effekt von 10 Jahren Dauermeditation auf Stressempfinden

Sicherung der internen Validität: Das Max-Kon-Min-Prinzip

Bedeutung der Varianzarten

- ▶ Logik der Bewertung von Versuchsplänen: Wenn die Primärvarianz (Behandlungsvarianz) größer ist als die Fehlervarianz (Systematischer Fehler + Zufallsfehler), dann hat die UV einen Einfluss auf die AV.
- ▶ Auswirkungen dieser Logik auf
 - ▶ **Planung von Versuchen:** Max-Kon-Min-Prinzip und Kontrolltechniken („Wie mache ich Wirkungen der UV sichtbar?“)
 - ▶ **Überprüfung der Daten:** Inferenzstatistische Prüfung zur Absicherung gegen den Zufall („Sind Wirkungen der UV größer als „zufällige“ Schwankungen?“)

Das Max-Kon-Min-Prinzip



- ▶ **MAX**imiere die *Primärvarianz (Behandlungsvarianz)*: Wähle die Stufen der UV so, dass möglichst große Unterschiede in der AV zwischen den Gruppen entstehen.
- ▶ **KON**trolliere die *Sekundärvarianz (Systematischer Fehler)*: Sorge dafür, dass systematische Fehler (bekannte Störvariablen) in allen Gruppen gleich wirken und bestimme deren Einfluss.
- ▶ **MIN**imiere die *Fehlervarianz (Zufallsfehler)*: Vermeide unsystematische (zufällige) Merkmalsfluktuationen auf Seiten der Versuchssituation, der Datenerfassung und der Datenverarbeitung.

Sicherung der internen Validität: Das Max-Kon-Min-Prinzip

- **Maximiere die Primärvarianz**

- Wahl von Extremgruppen
- Intensivierung des Stimulusmaterials
- Manipulationskontrolle

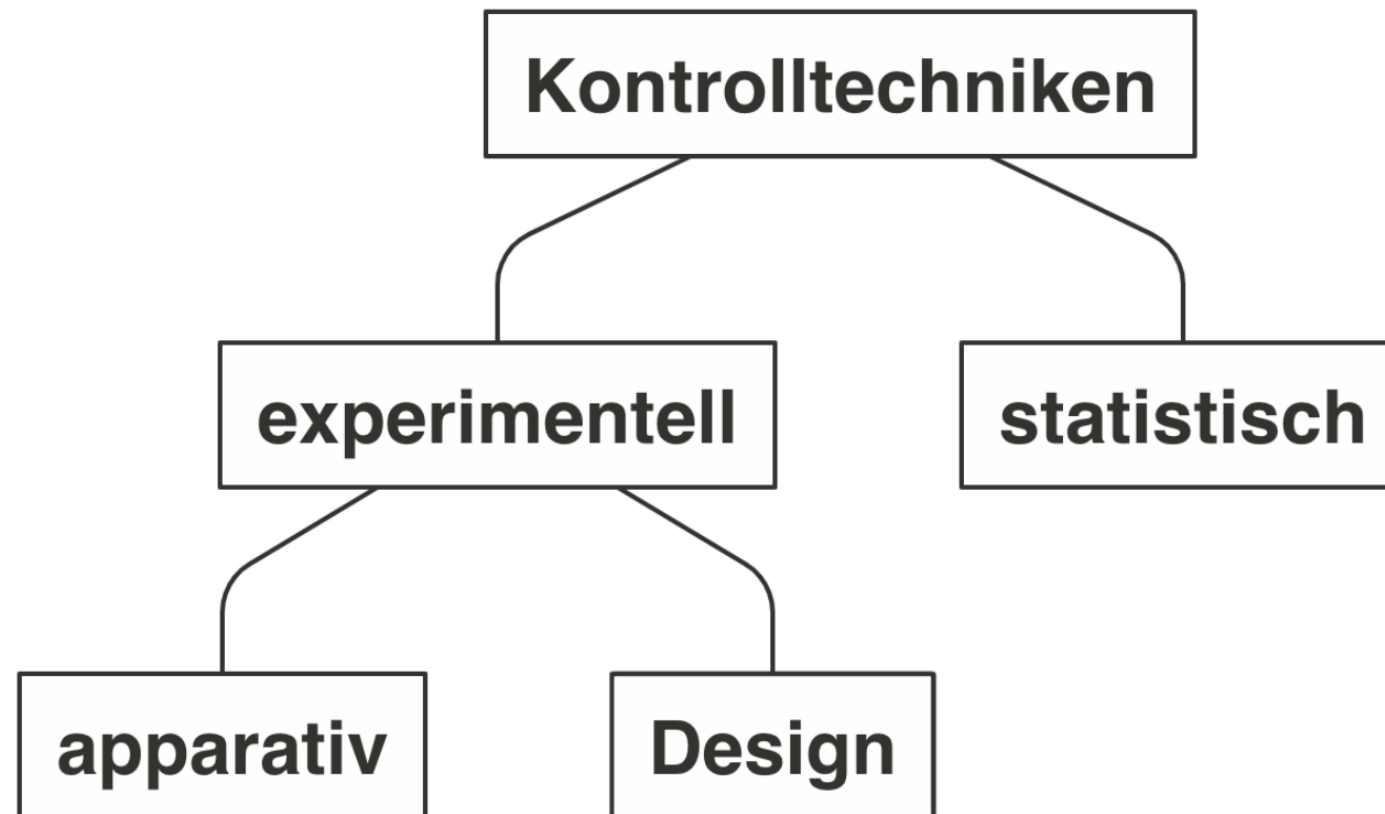
- **Kontrolliere die Sekundärvarianz (systematischer Fehler)**

- Coverstory und Verblindung
- Konstanthaltung eines Störfaktors für alle Versuchsgruppen und experimentellen Bedingungen
- Randomisierung der Probanden und der Bedingungen
- Umwandlung eines Störfaktors in eine weitere experimentelle UV („Kontrollvariable“)
- Elimination des Störfaktors (z. B. schalldichter Raum)
- Ausbalancieren (in jeder Gruppe z. B. die Hälfte in den lauten und die Hälfte in leisen Raum)
- Nachträgliche statistische Kontrolle: Kovarianzanalyse

- **Minimiere die Fehlervarianz (unsystematischer Fehler)**

- Wahl eines Wiederholungs-Versuchsplans
- Anheben der Standardisierung der Untersuchungssituation
- Erhöhung der Zuverlässigkeit und Gültigkeit des Messinstruments

Taxonomie an Kontrolltechniken



MAXimierung der Primärvarianz

- ▶ Wähle die Stufen der UV so, dass möglichst große Unterschiede in der AV zwischen den Gruppen entstehen.
- ▶ Kontrolltechniken:
- ▶ Wahl von mehreren Bedingungen (> 2 Stufen), „optimale Stufen“
- ▶ Wahl von mehrfaktoriellen Designs (Umwandlung eines Störfaktors in eine weitere Variable („Kontrollvariable“))
Achtung: Bei mehr als 2 Faktoren wird die Interpretation von Wechselwirkungen problematisch!
- ▶ Wahl von extremen Bedingungen („Extremgruppen“)
Achtung: Gefahr von Trivialergebnissen!
- ▶ Ziel: Effekte der UV durch die Versuchsplanung möglichst „maximal“ zum Vorschein bringen

KONtrolle der Sekundärvarianz - Überblick

- ▶ Sorge dafür, dass systematische Fehler (bekannte Störvariablen) in allen Gruppen gleich wirken und/oder bestimme deren Einfluss.
- ▶ Kontrolltechniken:
 - ▶ vor der Datenerhebung:
 - ▶ Umwandlung von Sekundärvarianz bzw. Störvariable in UV
 - ▶ Randomisierung oder Parallelisierung (Blockbildung/Schichtung)
 - ▶ Messwiederholung
 - ▶ während der Datenerhebung: Konstanthaltung
 - ▶ nach der Datenerhebung: Kovarianzanalytische Kontrolle
- ▶ Ziel: Effekte von „Nicht-UVn“, die als Störvariablen einen systematischen Einfluss haben können, bestmöglich unter Kontrolle halten.

KONtrolle der Sekundärvarianz vor der Datenerhebung I

Umwandlung von Störvariablen in zusätzliche UVs

- ▶ Vorgehen:
 - ▶ Umwandlung einer Störvariablen in zusätzliche UV
 - ▶ Beispiele: Alter, Geschlecht, Intelligenz, Schulbildung, Vorwissen
 - ▶ Ziel: Kontrolle von Störvariablen (z. B. interindividuelle Varianz)
- ▶ Vorteile:
 - ▶ Zusätzliche Informationen über Störvariablen
 - ▶ Aus umgewandelter UV resultierende Varianz kann aus Gesamtvarianz herausgerechnet werden (z. B. Kovarianzanalyse in der Datenauswertung)
- ▶ Nachteil:
 - ▶ Kenntnis über Störvariablen, die mit AV korrelieren, wird benötigt

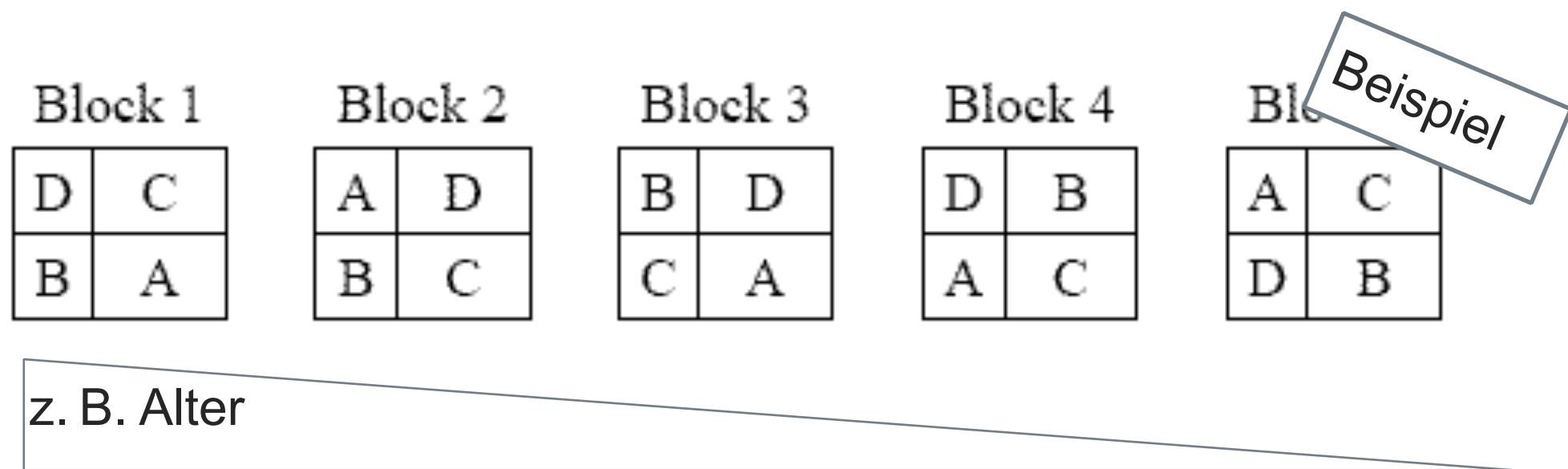
KONtrolle der Sekundärvarianz vor der Datenerhebung II

Randomisierung

- ▶ Annahme: Gebildete Zufallsstichproben, die derselben Population entstammen, gleichen einander weitgehend
- ▶ Vorgehen:
 - ▶ Zufällige Zuweisung der Probanden zu den Versuchsbedingungen
 - ▶ Beispiele: Münzwurf, Loseziehen, Zufallszahlentabelle
- ▶ Vorteil:
 - ▶ Anwendung, wenn Vielzahl möglicher Störvariablen kontrolliert werden soll, über deren Effekt nichts Genaueres bekannt ist
 - ▶ Unbekannte Unbekannte werden automatisch mit kontrolliert (wenn man Glück hat!)
- ▶ Nachteile:
 - ▶ Nicht effektiv, wenn Stichproben klein sind
 - ▶ Vergleichbarkeit ist nie sicher, daher sollten die relevanten Statistiken der Gruppen jeweils berichtet werden (z. B. Alter, Bildung, Persönlichkeit...)
 - ▶ Insgesamt ist Parallelisierung wahrscheinlich effektiver als Randomisierung*

*Saint-Mont, U. (2015). Randomization Does Not Help Much, Comparability Does. *PloS one*, 10(7), e0132102.

Blocking (Stratifizierung)



KONtrolle der Sekundärvarianz vor der Datenerhebung III

Parallelisierung

- ▶ Spezialfall einer Umwandlung von Störvariablen in UV (z. B. Personenmerkmale wie Alter, Intelligenz)
- ▶ Zuordnung der Vpn zu den Versuchsbedingungen aufgrund der Merkmale, in denen man eine Einflussgröße auf die AV erwartet

Blockbildung („Mischung“ aus Randomisierung und Parallelisierung)

- ▶ Auswahl von Vpn, die sich hinsichtlich Parallelisierungsmerkmal gleichen
- ▶ Aufstellen einer Rangreihe (bezogen auf Ausprägung des Parallelisierungsmerkmals)
- ▶ Bildung von “Blöcken” von Vpn mit jeweils benachbarten Rangplätzen
(Gedanke: Vpn eines Blocks sind sich hinsichtlich Parallelisierungsmerkmal ähnlicher als Vpn aus unterschiedlichen Blöcken)
- ▶ Zuordnung der Vpn eines “Blocks” zu Versuchsbedingung erfolgt dann per Zufall (“Randomisierung”): “Statistische Zwillinge”

KONtrolle der Sekundärvarianz vor der Datenerhebung IV

Messwiederholung

- ▶ Vorgehen: Alle Vpn werden unter allen Versuchsbedingungen untersucht
- ▶ Ziel: Kontrolle der interindividuellen Varianz (keine interindividuellen Unterschiede zwischen Bedingungen)
- ▶ Vorteile:
 - ▶ Kenntnis über Personenvariablen, die mit AV korrelieren, nicht nötig
 - ▶ Versuchsdurchführung ist ökonomisch
- ▶ Nachteile:
 - ▶ Vpn sind keine “statische” Einheiten, die von Messung zu Messung konstant bleiben (z. B. Lernfähigkeit, Ermüdung)
 - ▶ erhöhter Aufwand (für Vpn)
 - ▶ Probleme bei abhängigen Untersuchungsplänen
 - ▶ Konfundierung zwischen den UVn und der gewählten Darbietungsabfolge der Bedingungen
 - ▶ Übertragungseffekte („Carry-over Effekte“)
- ▶ Methode: Konstanthaltung von Zeiteffekten
 - ▶ Vollständige Permutation oder unvollständige Permutation aller möglichen Behandlungskombinationen

Minimierung der Fehlervarianz I

- ▶ Vermeide unsystematische (zufällige) Merkmalsfluktuationen auf Seiten der Versuchssituation, der Datenerfassung und der Datenverarbeitung.
- ▶ Kontrolltechniken:
- ▶ Kontrollierte Untersuchungsbedingungen (Abschirmung, Eliminierung, Instruktion)
- ▶ Einsatz reliabler, valider und objektiver Erhebungsinstrumente (Beobachter, Messinstrumente)
- ▶ Doppelte Dateneingabe
- ▶ Ziel: Auswirkungen von unbekannten Störvariablen und (Daten-) Fehlern so klein wie möglich halten.

Minimierung der Fehlervarianz II – Typische Störvariablen

- ▶ Situation
 - ▶ Untersuchungsort und Untersuchungszeit
 - ▶ Atmosphäre (Technik, weißer Kittel, ...)
- ▶ Versuchsperson
 - ▶ Motivation: „Intelligente Vp“, soziale Erwünschtheit, Bewertungsangst
 - ▶ Erwartung: Placebo (daher: mind. Einfachblindversuch)
 - ▶ Prozesse in der Vp: Aktivierung – Ermüdung, Lernen – Übung
- ▶ Instruktion
 - ▶ Eine zentrale Rolle, um diese Störfaktoren zu eliminieren oder konstant zu halten, spielt die Instruktion der Versuchsteilnehmer.
 - ▶ Die Instruktion umfasst nicht nur den verbalen Anweisungsteil, sondern alle Versuchsumstände: die Umgebung, das Verhalten des Versuchsleiters, usw.
 - ▶ Die Instruktion sollte daher möglichst schriftlich ausformuliert und standardisiert sein.

Minimierung der Fehlervarianz III – Typische Störvariablen

Verlässlichkeit der Messwerte

- ▶ Bewertungsstandard: Gütekriterien der Messen
 - ▶ Validität: Grad der Genauigkeit, das zu messen oder vorherzusagen, was gemessen oder vorhergesagt werden soll
 - ▶ Reliabilität: Grad der Genauigkeit, mit dem etwas gemessen wird (unabhängig davon, ob dies auch gemessen werden soll)
 - ▶ Objektivität: Grad der Unabhängigkeit der Ergebnisse vom Untersucher
- ▶ Datenerhebung: Beispiel Fragebogen
 - ▶ Besser Onlinebefragung als schriftliche Befragung (keine Übertragungsfehler; Minimierung von Missing Values; ...)
- ▶ Datenerhebung: Beispiel Verhaltensbeobachtung
 - ▶ Systematische statt zufälliger Beobachtung
 - ▶ Geschulte Beobachter, immer zwei Beobachter



Erarbeiten Sie ein experimentelles Design

- ▶ Bitte entwickeln Sie jeweils ein eigenes Beispiel für ein
 - ▶ Einfaktorielles experimentelles Forschungsdesign
 - ▶ Zweifaktorielles experimentelles Forschungsdesign

- ▶ Sie entwickeln eine Studie zur Untersuchung des Einflusses von Arbeitsbelastung auf die Arbeitszufriedenheit. Dazu können Sie in drei Abteilungen eines Unternehmens jeweils 30 Mitarbeiterinnen und Mitarbeiter beobachten und befragen. Bitte entwickeln Sie jeweils konkrete Vorschläge in der angegebenen Zahl zur
 - ▶ Maximierung der Primärvarianz
 - ▶ Kontrolle der Sekundärvarianz
 - ▶ Minimierung der Fehlervarianz

Wie ist die interne Validität der „Ballerstudie“ zu retten?

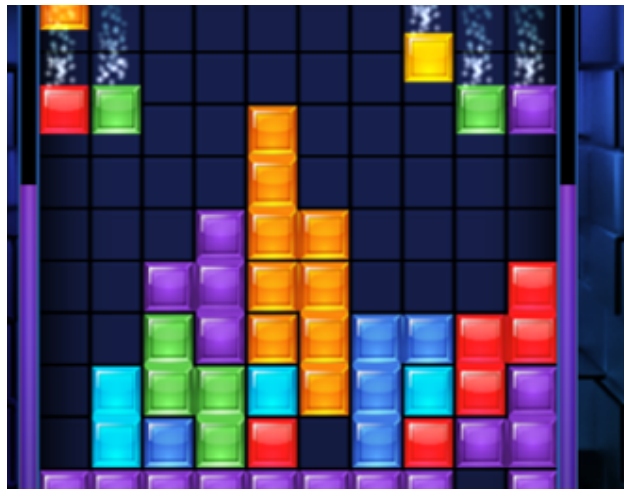
UV

Art des Computerspiel



Gruppe 1:
Computerspiel
mit **viel** Gewalt

Die Heizung war kaputt!
Es war mal kalt und mal
warm, ganz gemischt und
ohne System!



Gruppe 2:
Computerspiel
mit **wenig** Gewalt

AV

Aggressives
Verhalten

Versuchsvorbereitung

Denken Sie an diese Punkte:

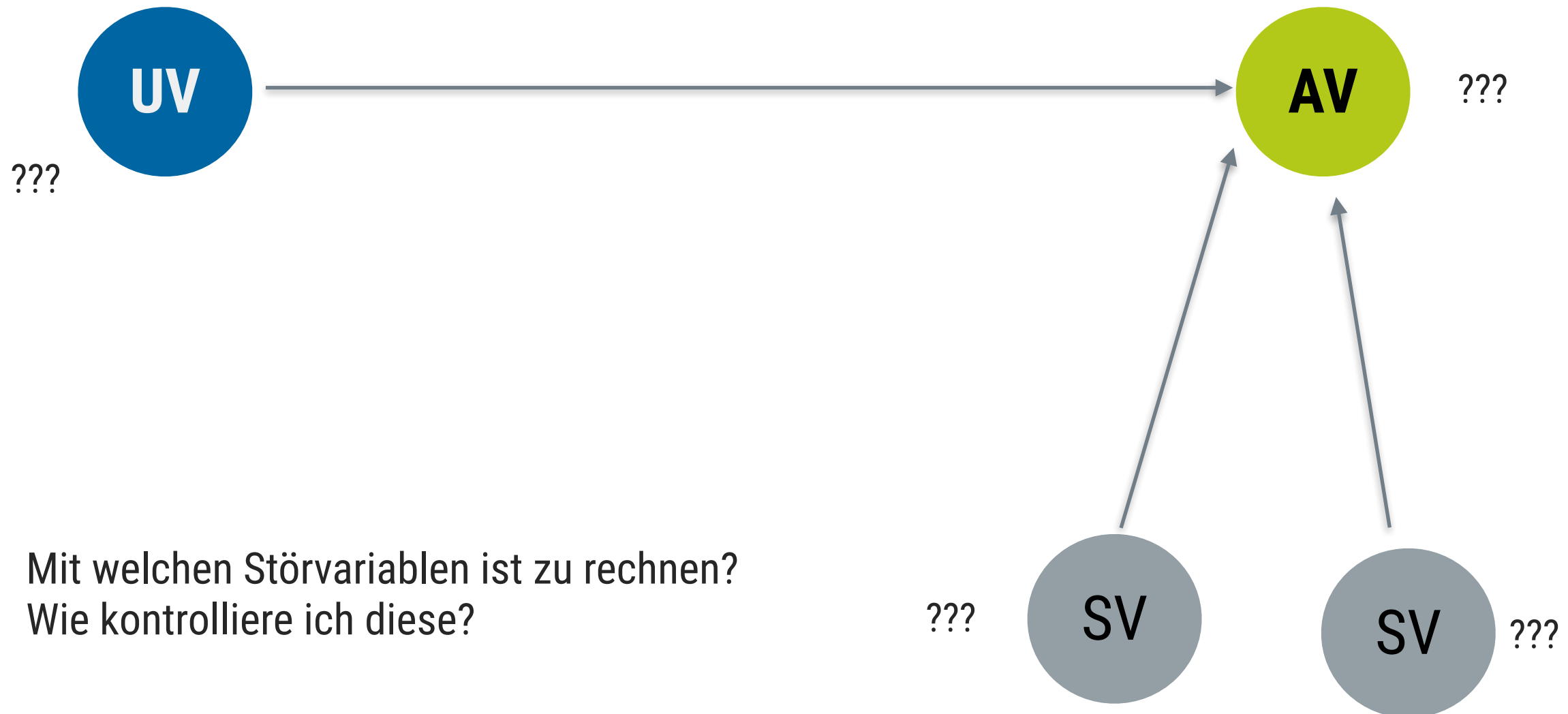
- ▶ Prüfen der ethischen Unbedenklichkeit der Versuchsdurchführung
- ▶ Erstellung bzw. Vorbereitung des nötigen Materials (Instruktionen, Fragebögen, Apparate)
- ▶ Evtl. Vortests und Probeläufe planen
- ▶ Qualität der Messverfahren und Stimuli prüfen

Typische Fehler:

- ▶ Motivation der Teilnehmenden: Wer überredet/gezwungen wird mitzumachen, könnte nur halbherzig mitmachen, was die Ergebnisse verfälschen würde
- ▶ Stör-und Fehlereinflüsse: z. B. selektive Stichprobe; Versuchsleiterartefakt, Hawthorne-Effekt („boa, ich bin in der Wissenschaft!“)
- ▶ Umstände der experimentellen Situation
- ▶ „Drop-out“ – Versuchspersonen gehen während der Studie verloren
- ▶ Qualität der Datenerhebung



Was sind UV, AV und SV meiner Studie?



UV und AV sollten in Ihrem Bericht klar benannt sein. Maßnahmen zur Kontrolle von Störvariablen sollten angesprochen werden. Eine diagrammatische Darstellung dieser Art bezeichnet man auch als "Pfaddiagramm", wobei die Pfeile die "Pfade" darstellen.



Stellen Sie Ihr Modell als Pfaddiagramm dar!

UV(s) (Ursache)

AV (Wirkung)

Beispiel

Markendarbietung
(offen anstatt verblindet)

Bekanntheitsgrad
(der Marken)

Einstellung
(gegenüber Marken)

Präferenz
(für Produkt A anstatt B)

Produkt A

Produkt B



Erstellen Sie das Pfaddiagramm für Ihre Studie!



Wie groß ist unsere Gruppe im Schnitt?

- ▶ Geben Sie Ihre Körpergröße (und einige Korrelate) [in dieses Formular](#) ein (anonym)!
- ▶ Betrachten Sie die Daten Ihrer Gruppe (heutiges Datum) [in diesem Datenblatt](#)!
- ▶ Ziehen Sie aus diesen Daten eine Stichprobe der Größe $n=5$ und eine mit $n=20$ (Körpergröße).
- ▶ Berechnen Sie jeweils Mittelwert und Standardabweichung!
- ▶ Geben Sie dann die Kennwerte der zwei Stichproben wiederum [in dieses Formular](#) ein!
- ▶ Die Daten zu den Kennwerten der Stichproben können Sie [hier](#) abrufen (entweder per CSV-Download oder per Import von GoogleSheets in R).
- ▶ Beantworten Sie auf Basis der Daten zu den Stichprobenkennwerten folgende Fragen:
 - ▶ Streuen die Mittelwerte der Stichprobe?
 - ▶ Streuen die Standardabweichungen der Stichprobe?
 - ▶ Hängt die Streuung der Mittelwerte mit der Stichprobengröße zusammen?

Klassifikation quantitativer Versuchsplänen

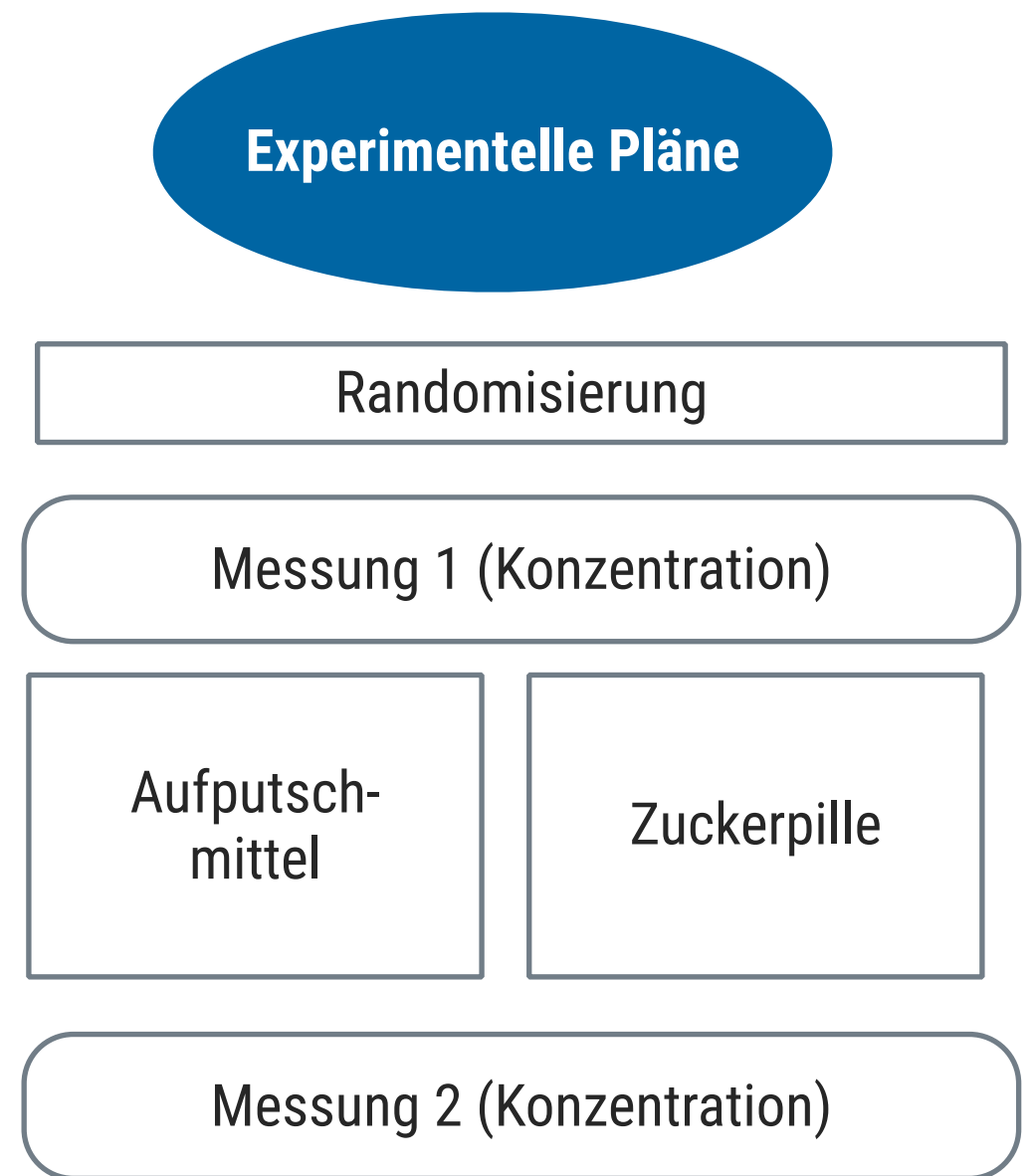
Klassifikationskriterien für Untersuchungsdesigns

Kriterium	Varianten von Designs
Wissenschaftstheoretischer Ansatz	qualitativ, quantitativ, Mixed-Methods
Erkenntnisziel	grundlagenwissenschaftlich, anwendungswissenschaftlich (unabhängige vs. Auftragsstudie)
Gegenstand	Empirische Studie (Original- vs. Replikationsstudie), Methodenstudie, Theoriestudie (Review vs. Metaanalyse)
Datengrundlage (empirisch)	Primäranalyse, Sekundäranalyse, Metaanalyse
Erkenntnisinteresse (empirisch)	explorativ (gegenstandsbeschreibend, theoriebildend), deskriptiv (populationsbeschreibend), explanativ (hypothesenprüfend)
Bildung & Behandlung von Untersuchungsgruppen (explanativ)	Experimentelle, quasi-experimentelle, nicht-experimentelle Studie (Korrelationsstudie vs. Ex-post-facto-Studie)
Untersuchungsort	Labor, Feld
Anzahl Untersuchungszeitpunkte (empirisch)	(Quasi-)experimentell mit/ohne Messwiederholung (between-subjects vs. within-subjects), nicht-experimentelle mit/ohne Messwiederholung (Querschnitt vs. Trend vs. Längsschnitt)
Anzahl Untersuchungsobjekte (empirisch)	Gruppenstudie (Stichprobenstudie vs. Vollerhebung), Einzelfallstudie

Im Überblick: Zwei Arten von Versuchsplänen in der Forschung



Geringe interne Validität



hohe interne Validität

Wo möglich, sollten experimentelle Pläne bevorzugt werden.

Within- vs. Between-Designs

within subject

Messung 1 (Konzentration)

Reihenfolgen-Randomisierung

Intervention 1

Intervention 2

Messung 2 (Konzentration)

Intervention 2

Intervention 1

Messung 3 (Konzentration)

höhere Power

Wo möglich, sollten within-Subject-Pläne bevorzugt werden.

between subject

Messung 1 (Konzentration)

Randomisierung

Experimental-gruppe
(Gruppe 1)
Intervention 1

Kontrollgruppe
(Gruppe 2)
Intervention 2

Messung 2 (Konzentration)

geringere Power

Vor- und Nachteile von Within-Designs

**within
subject**

Messung 1 (Konzentration)

Reihenfolgen-Randomisierung

Intervention 1

Intervention 2

Messung 2 (Konzentration)

Intervention 2

Intervention 1

Messung 3 (Konzentration)

Vorteile

- ▶ Personen werden mit sich selbst verglichen, dadurch werden personen-gebundene Störvariablen neutralisiert
- ▶ Weniger „Rauschen“ als bei beim between-subject-Design, da nur intraindividuelle Unterschiede eingehen, die frei sind von interindividuellen Unterschieden
- ▶ Daher weniger Versuchspersonen nötig, um Effekt zu entdecken

Nachteile

- ▶ Carry-Over-Effekte müssen beachtet werden (Kontrolle von Reihenfolge-Effekten nötig)
- ▶ Within-Designs sind nicht immer einsetzbar
- ▶ Zeitaufwand pro Versuchsperson höher



Cooler Studies von PLOS One – Lassen Sie sich inspirieren!

[Meta-Milgram: An Empirical Synthesis of the Obedience Experiments](#)

[Loss of Control Increases Belief in Precognition and Belief in Precognition Increases Control](#)

[Temptation at Work](#)

[Behavioral Priming: It's All in the Mind, but Whose Mind?](#)

[The Distance Between Mars and Venus: Measuring Global Sex Differences in Personality](#)

[The Eyes Don't Have It: Lie Detection and Neuro-Linguistic Programming](#)

[Always Gamble on an Empty Stomach: Hunger Is Associated with Advantageous Decision Making](#)

[A Virtual Reprise of the Stanley Milgram Obedience Experiments](#)

[The Power of Kawaii: Viewing Cute Images Promotes a Careful Behavior and Narrows Attentional Focus](#)

[Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach](#)

[A Propaganda Index for Reviewing Problem Framing in Articles and Manuscripts: An Exploratory Study](#)

[Non-Disruptive Tactics of Suppression Are Superior in Countering Terrorism, Insurgency, and Financial Panics](#)

Das Experiment – Der „Königsweg“ der empirischen Forschung

Das Experiment ist eine Sonderform der wissenschaftlichen Beobachtung, die folgende Kriterien erfüllt:

- ▶ Trennung mindestens einer unabhängigen Variablen (UV) von einer abhängigen Variablen (AV), wobei die AV gemäß der zu prüfenden Hypothese von der UV (kausal) beeinflusst wird.
- ▶ Systematische Variation bzw. Manipulation der UV, um ihre Auswirkung auf die AV zu beobachten;
- ▶ Kontrolle anderer Einflussfaktoren („Störvariablen“) auf die AV
- ▶ Randomisierung, d. h. die zufällige Zuordnung von Versuchspersonen zu Versuchsbedingungen oder der Reihenfolge von Versuchsbedingungen

Wann immer Experimente ethisch unbedenklich bzw. vertretbar und praktisch durchführbar sind, stellen sie das Mittel der Wahl dar, um Kausalzusammenhänge zu prüfen.

Manipulation der UV

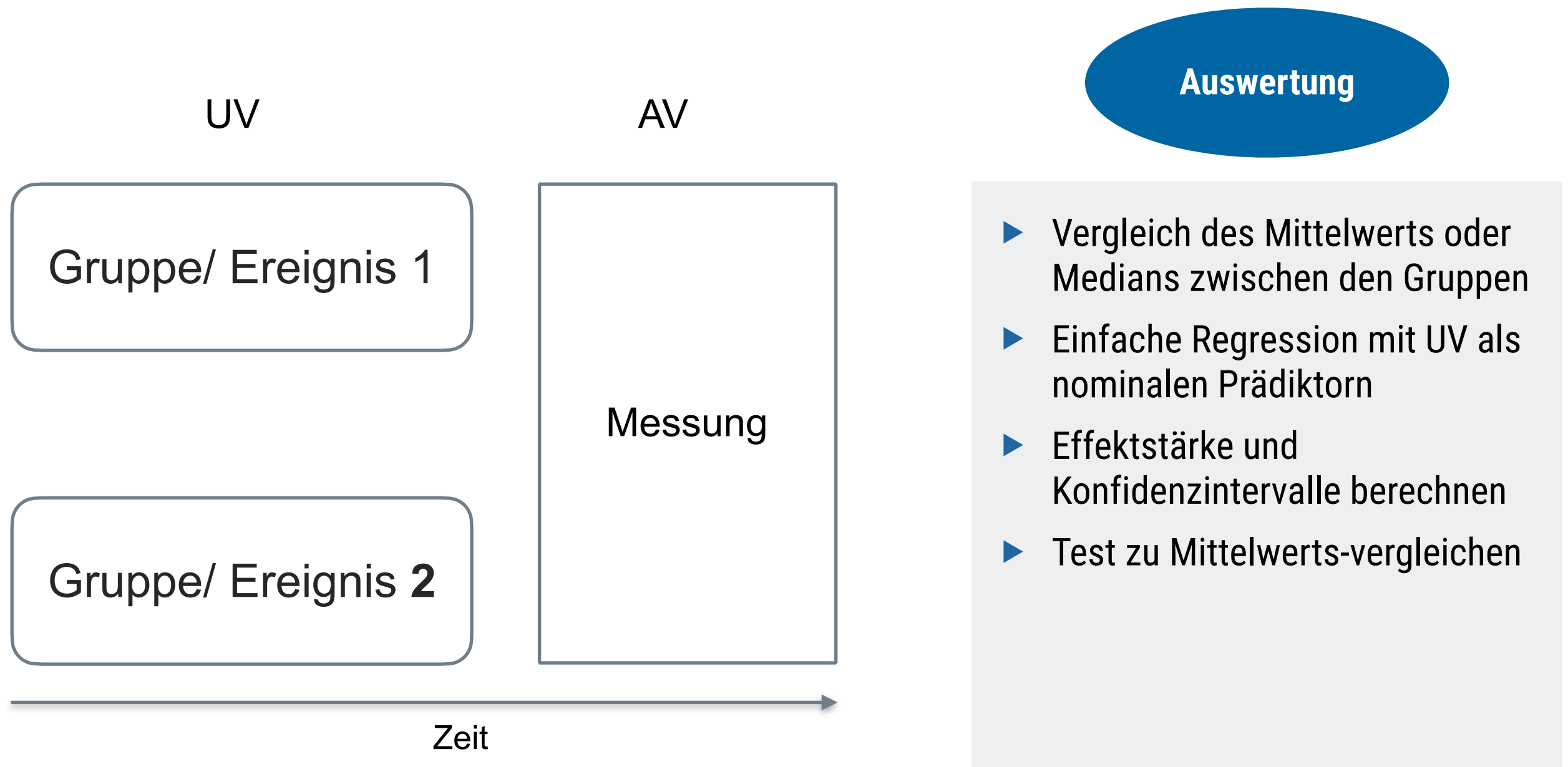
- ▶ Es werden verschiedene Realisierungen der UV hergestellt und die Probanden unter den verschiedenen Bedingungen beobachtet
- ▶ Einfachste Variante: Experimentalgruppe versus Kontrollgruppe (z. B. Medikament vs. Placebo)

Kontrolle von Störvariablen und Randomisierung

- ▶ Störvariablen werden so weit wie möglich in ihrem Einfluss auf die AV ausgeschaltet bzw. deren Korrelation mit der AV verhindert
- ▶ Variieren die Werte der AV wie vorhergesagt mit der UV, so ist die Hypothese eines kausalen Einflusses der UV auf AV gestützt worden (aber nicht bewiesen!)

Typische quantitative Versuchspläne

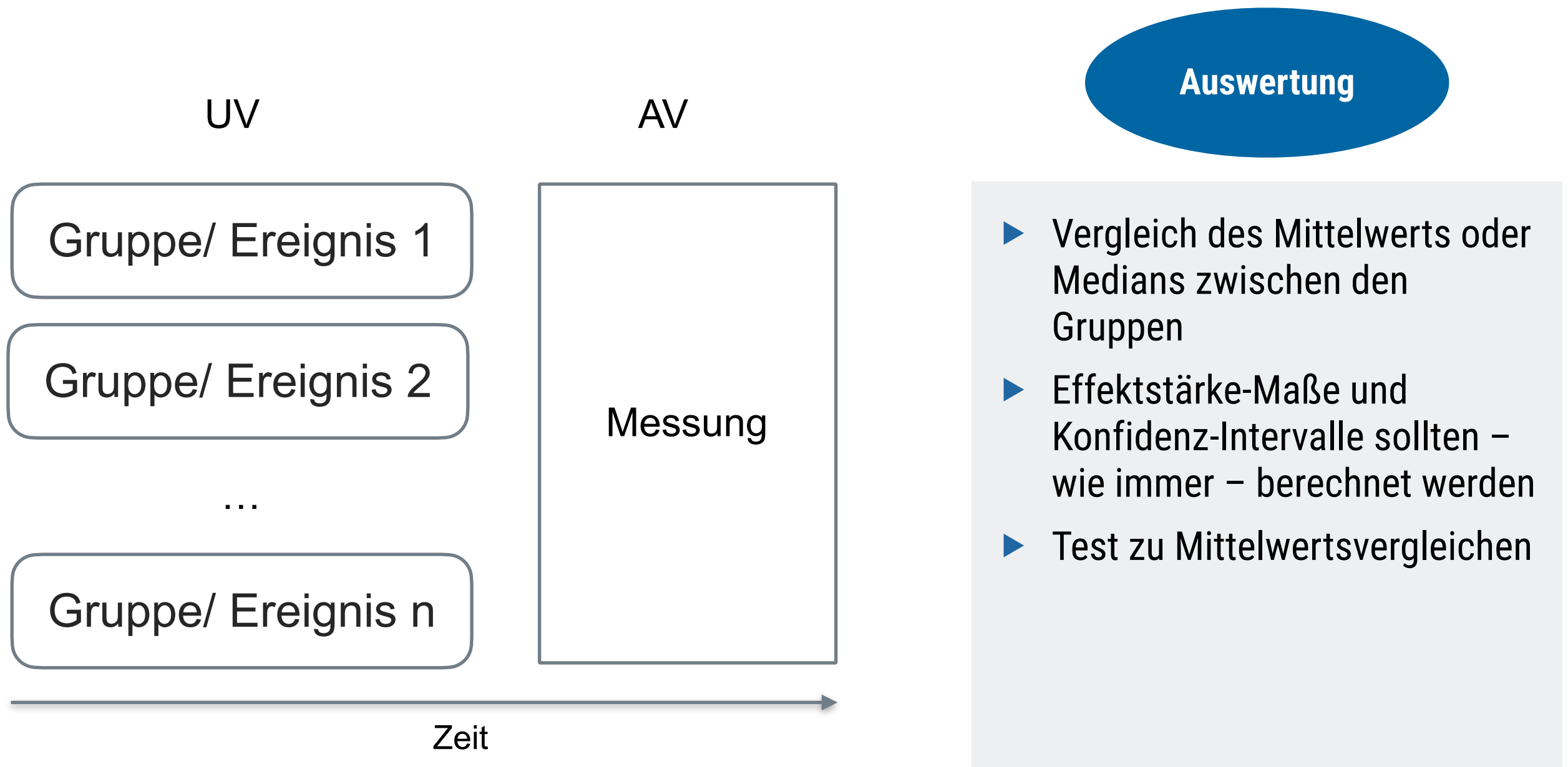
Experimentell: Zwei Gruppen querschnittlich



Beispiel

Ist die Arbeitszufriedenheit (AV) höher in der Entwicklungsabteilung oder im Vertrieb? Führt Koffein zu höherer Aufmerksamkeit als ein Placebo?

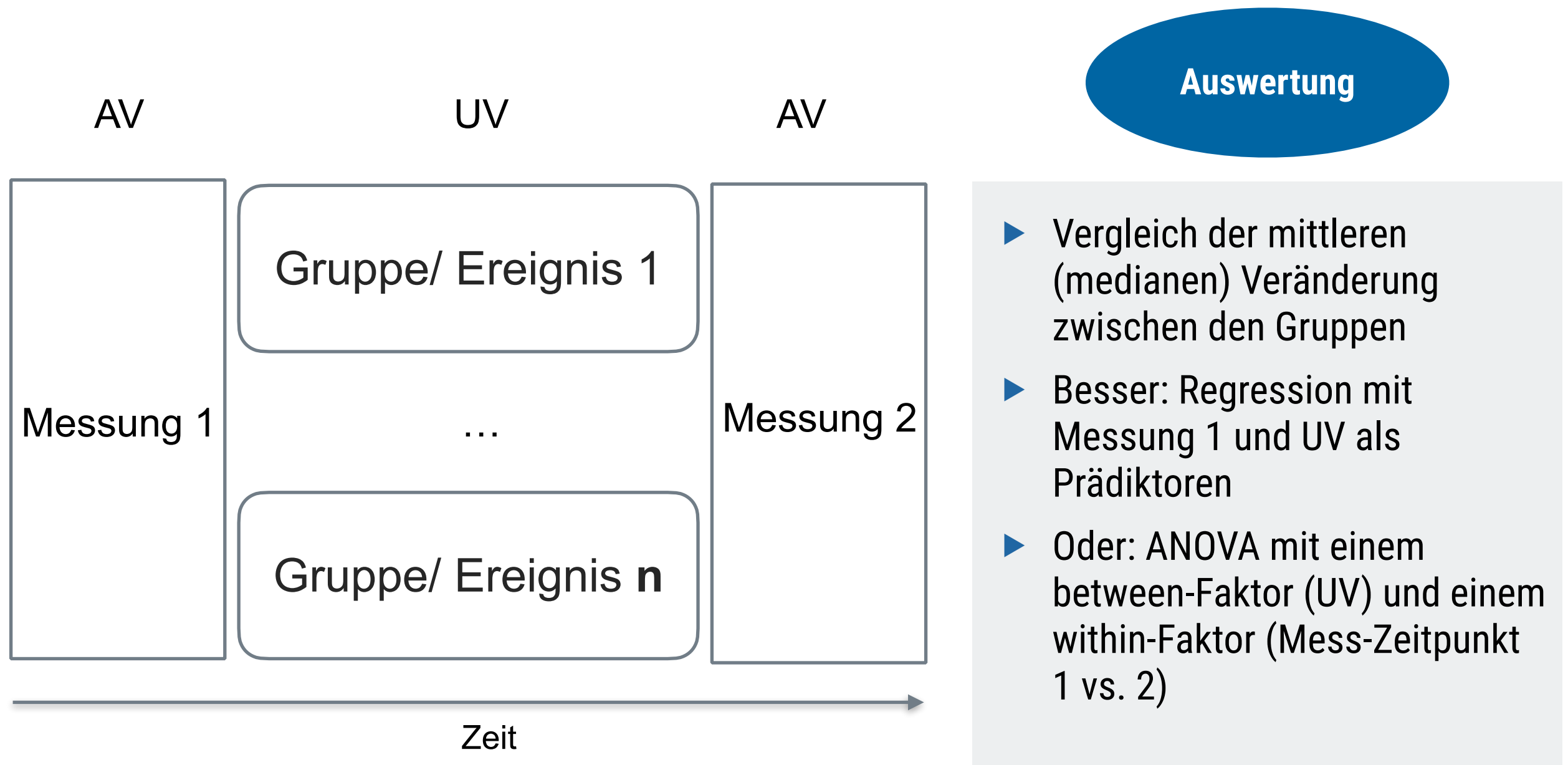
Experimentell: Mehr als zwei Gruppen querschnittlich



Beispiel

Ist die Arbeitszufriedenheit (AV) höher in der Entwicklungsabteilung, in der Produktion oder im Vertrieb? Führt Koffein zu höherer Aufmerksamkeit als ein Placebo oder als eine Kontrollgruppe?

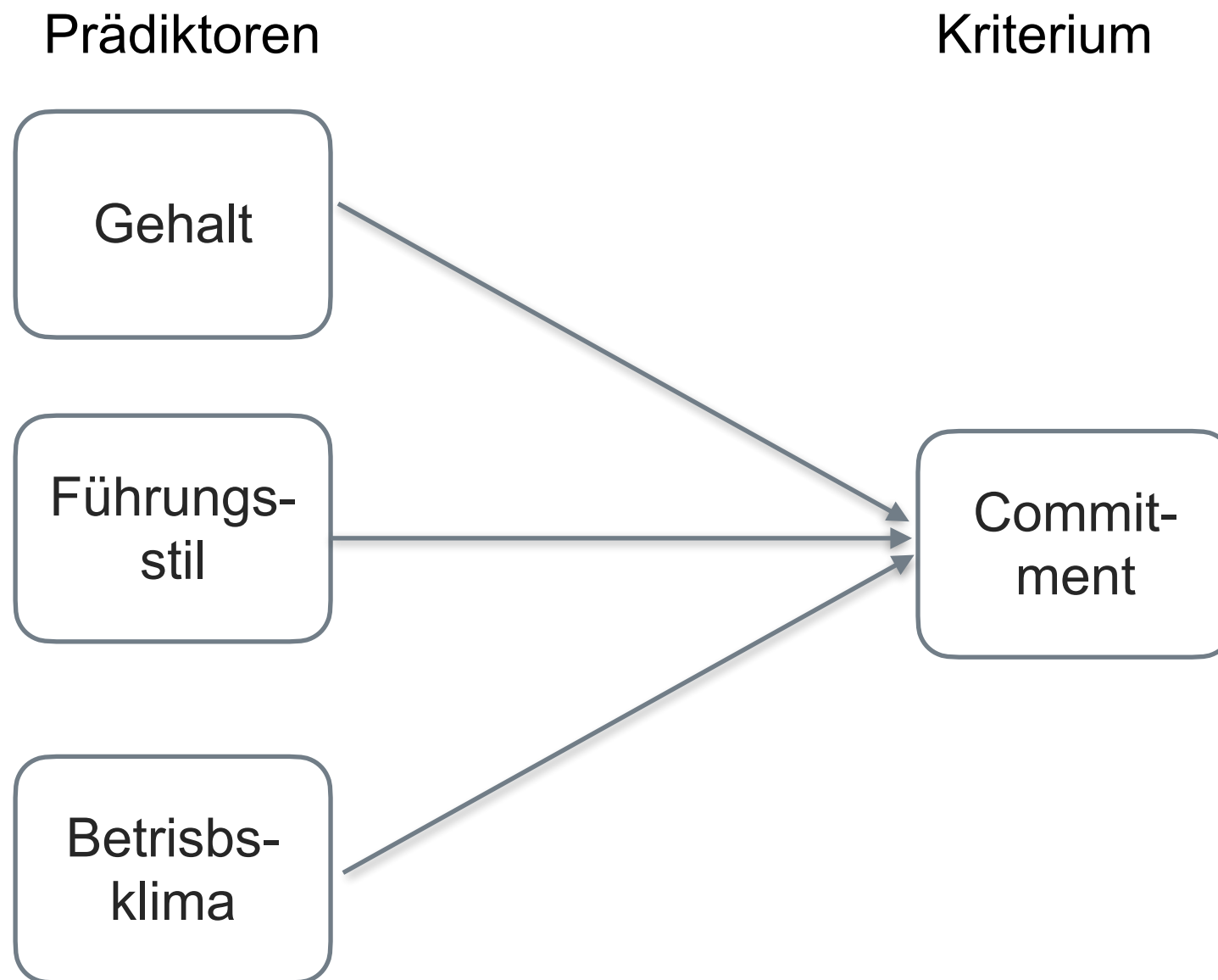
Experimentell: Zwei (oder mehr) Gruppen, vorher-nachher



Beispiel

Erhöht Koffein die Aufmerksamkeit (gemessen an "Baseline" zu Messzeitpunkt 1) stärker als ein Placebo?

Beobachtend: Ein oder mehr Prädiktoren



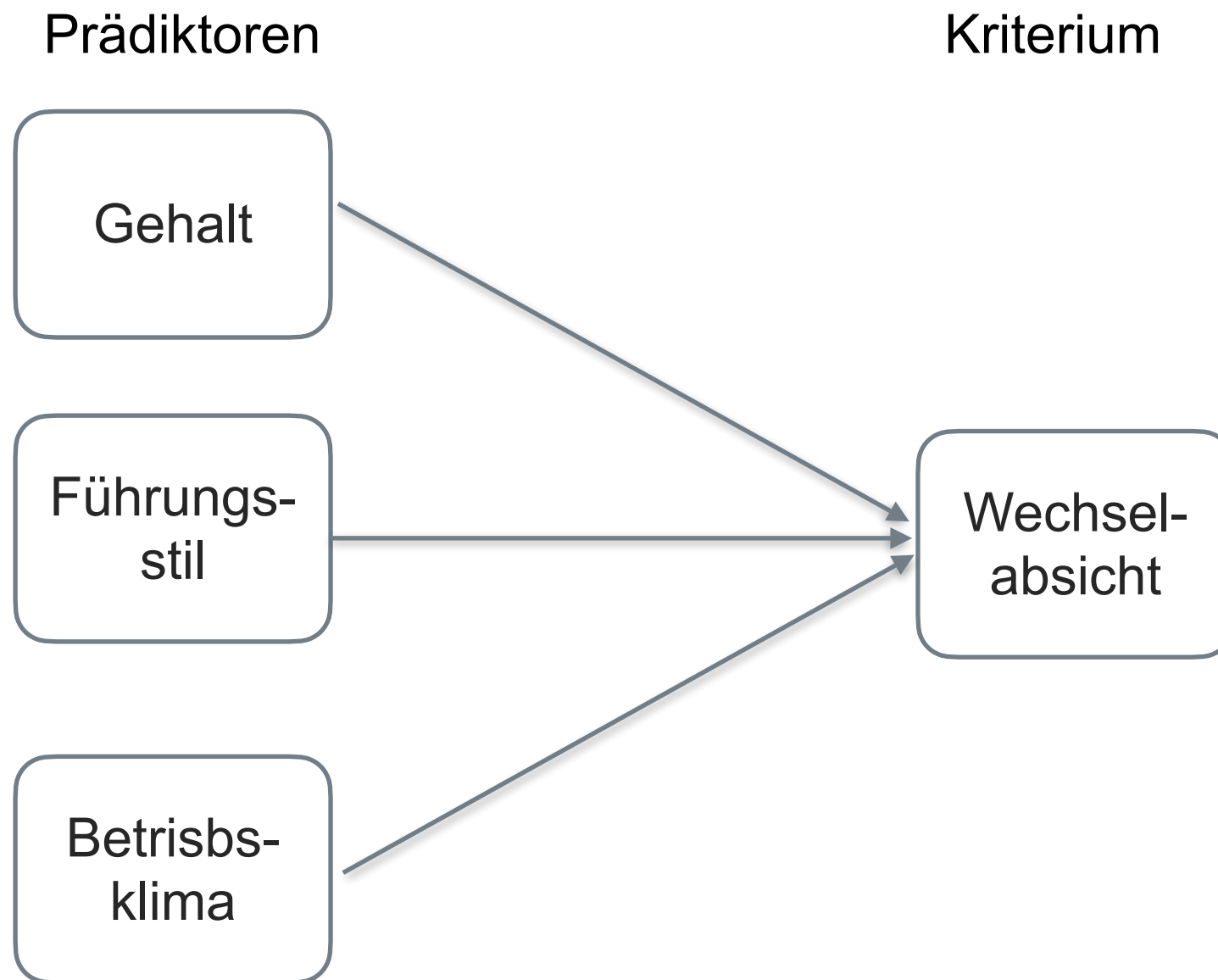
Auswertung

- ▶ Regression
- ▶ Einfache Regression bei einem Prädiktor
- ▶ Multiple Regression bei mehreren Prädiktoren
- ▶ Die Prädiktoren können beliebig skaliert sein; das Kriterium ist bei der "normalen" (multiplen) Regression metrisch skaliert

Beispiel

Welchen Einfluss haben die Prädiktoren Gehalt, Führungsstil und Betriebsklima auf das Commitment der Mitarbeiter?

Beobachtend: Mehrere Prädiktoren, Kriterium dichotom



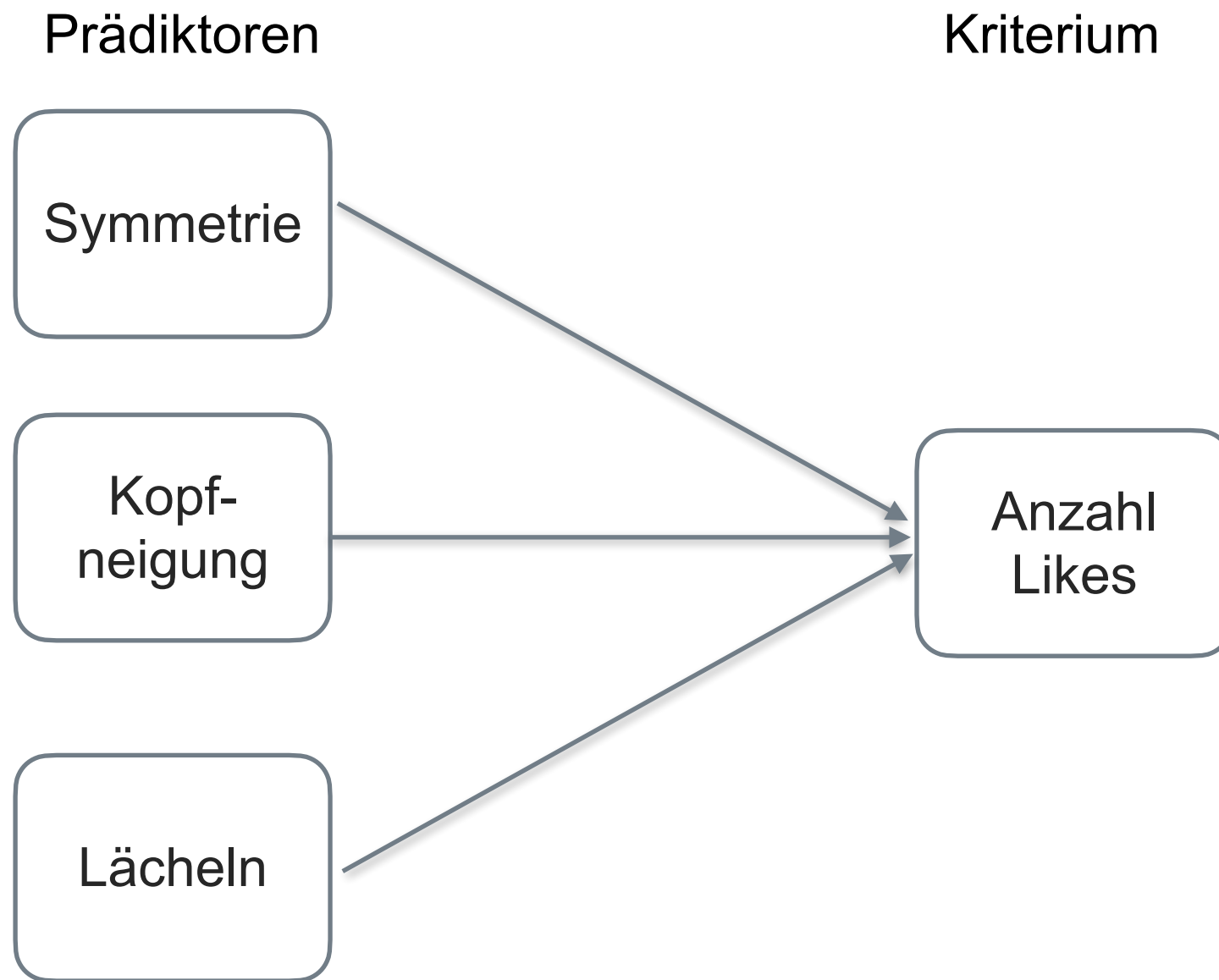
Auswertung

- ▶ *logistische* Regression bei dichotomem Kriterium
- ▶ Die Prädiktoren können beliebig skaliert sein; das Kriterium ist bei der "normalen" (multiplen) Regression metrisch skaliert

Beispiel

Welchen Einfluss haben die Prädiktoren Gehalt, Führungsstil und Betriebsklima auf die Wechselabsicht (ja vs. nein) der Mitarbeiter?

Beobachtend: Mehrere Prädiktoren, Kriterium ist Häufigkeit



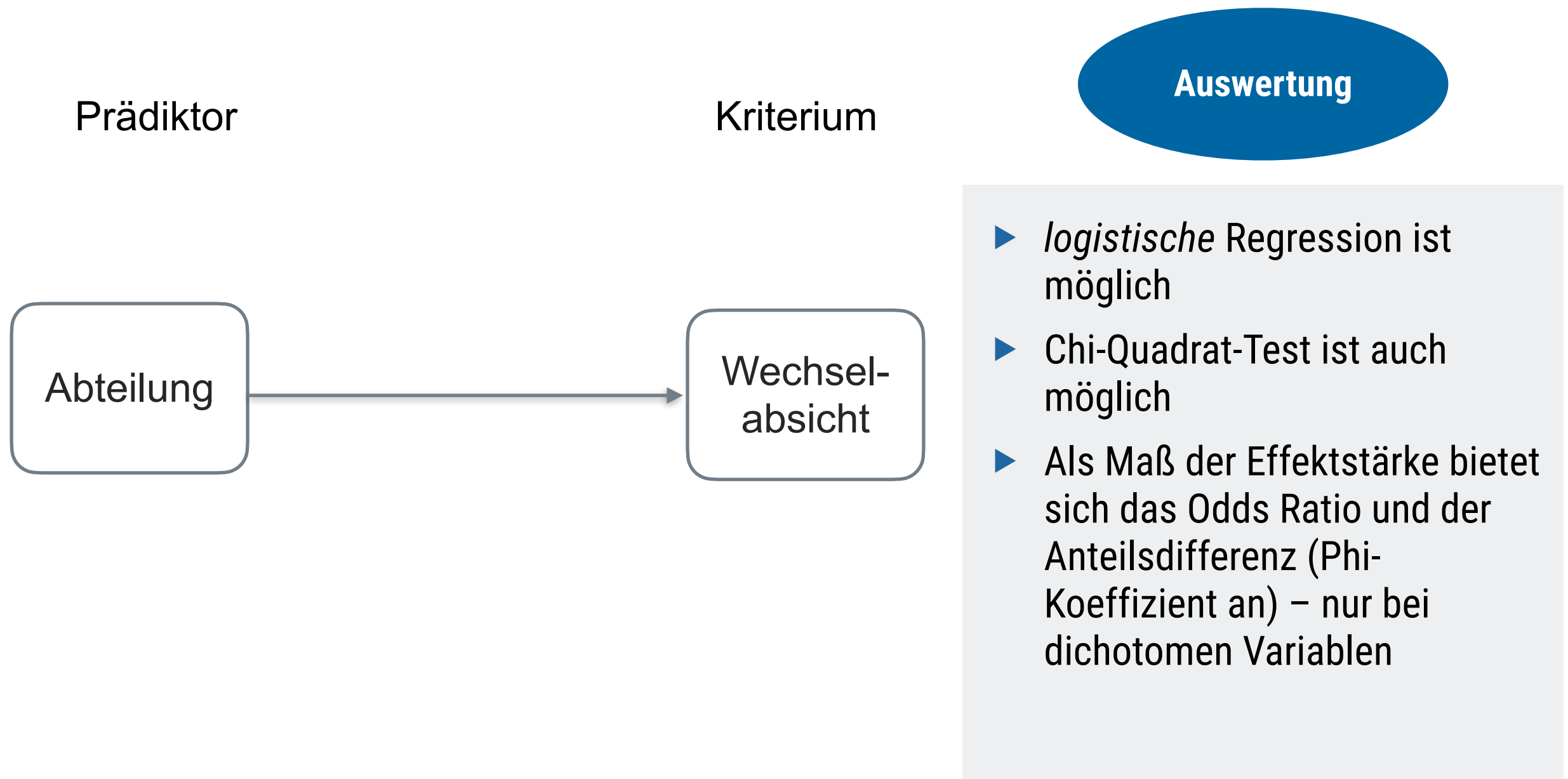
Auswertung

- ▶ *Poisson*-Regression bei Häufigkeits-Kriterium
- ▶ Alternativ: Normale Regression
- ▶ Die Prädiktoren können beliebig skaliert sein; das Kriterium ist eine Häufigkeit
- ▶ Zur Not machen Sie die normale Regression

Beispiel

Sie wollen wissen, wie sich einige Merkmale des Gesichts auf Facebook-Likes auswirken.

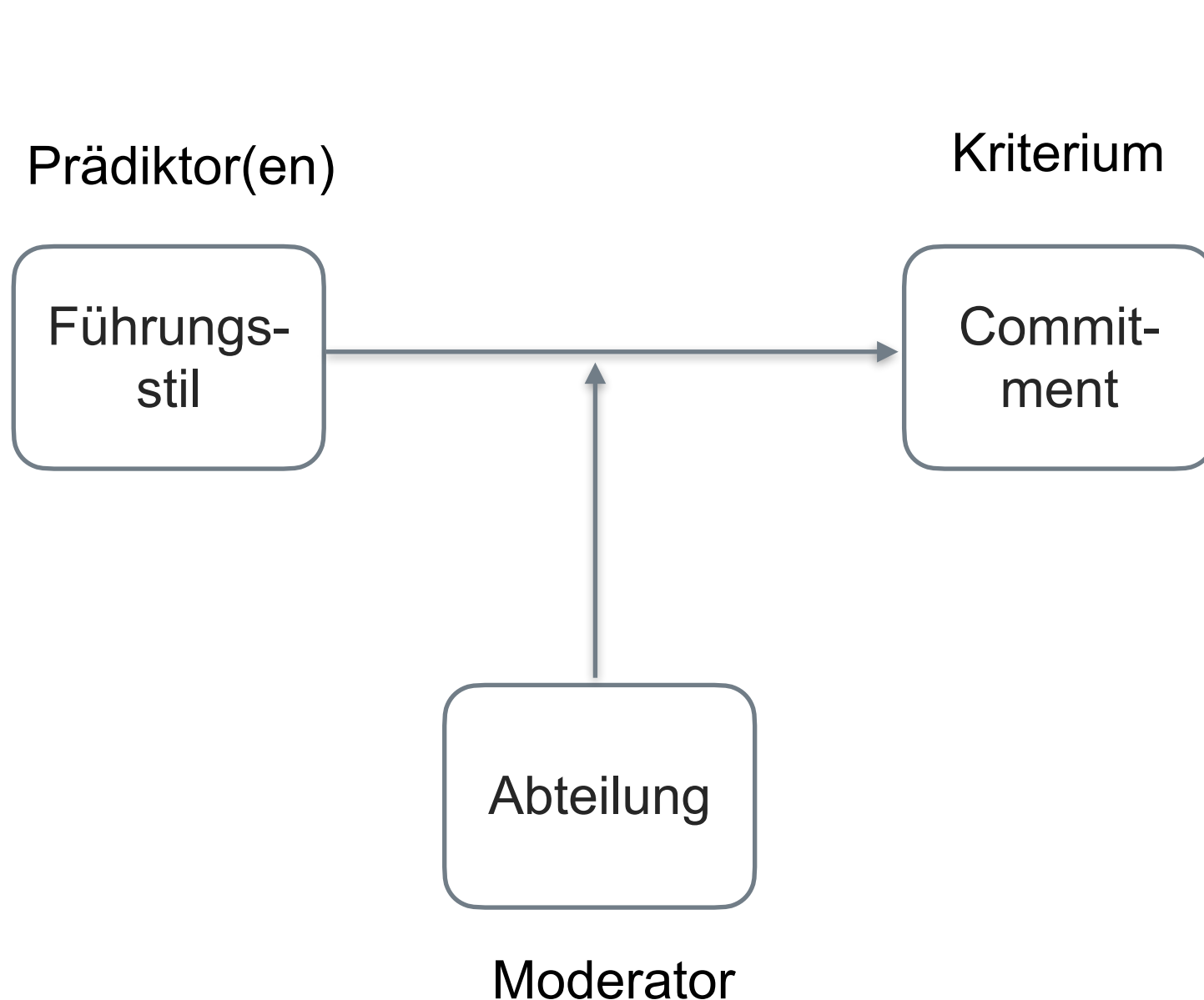
Beobachtend: Ein nominaler Prädiktor, Kriterium nominal



Beispiel

Gibt es einen Zusammenhang (Unterschied) zwischen Abteilung(en) und Wechselabsicht (ja vs. nein)?

Beobachtend: Moderatoreffekt



Auswertung

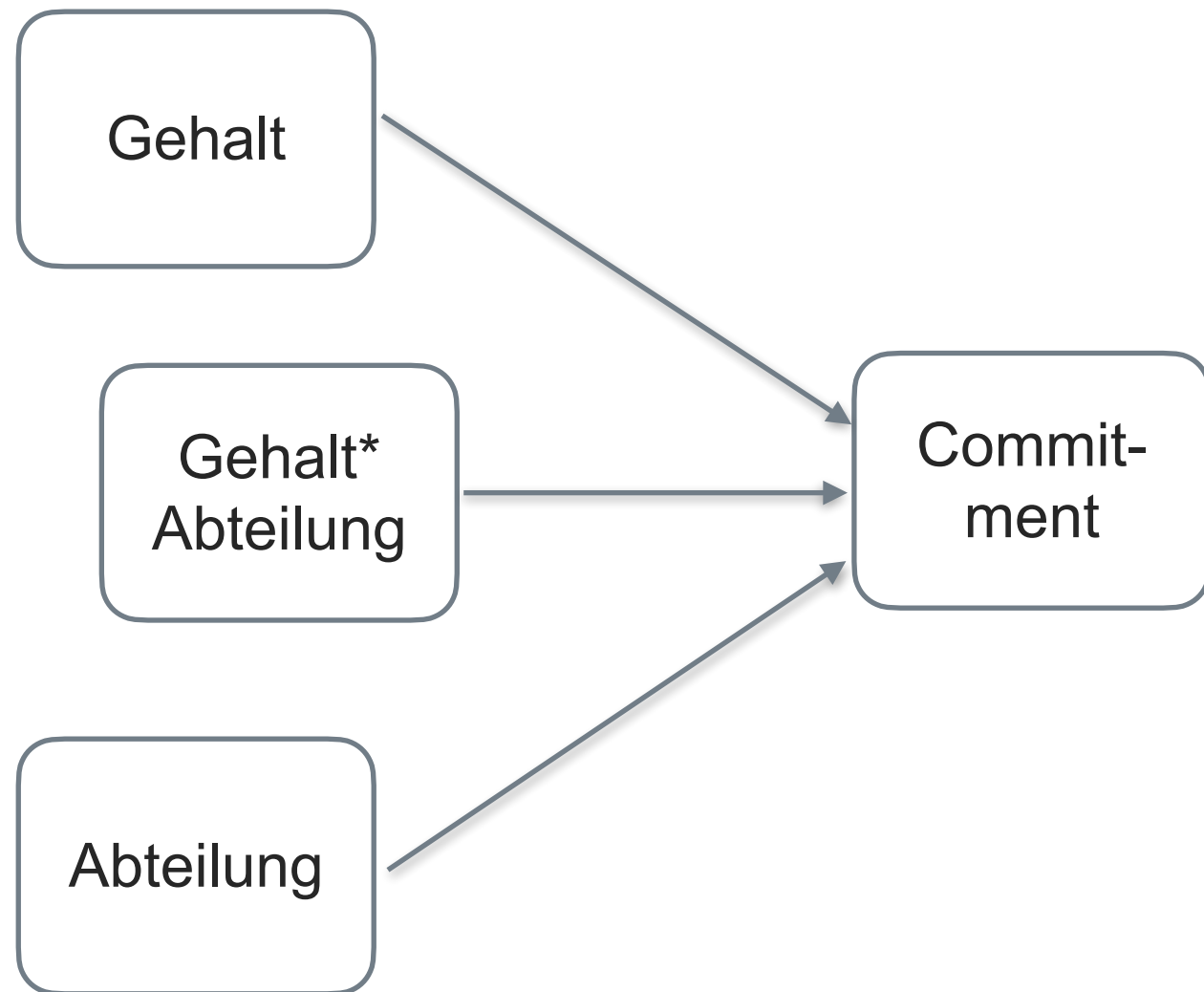
- ▶ Regression
- ▶ Es gehen sowohl Führungsstil als auch Abteilung als Prädiktoren ein
- ▶ Aber zusätzlich geht noch der *Interaktionsterm* ein: das Produkt der beiden Variablen, von denen wir eine Interaktion erwarten
- ▶ Moderatoreffekte werden häufig untersucht

Beispiel

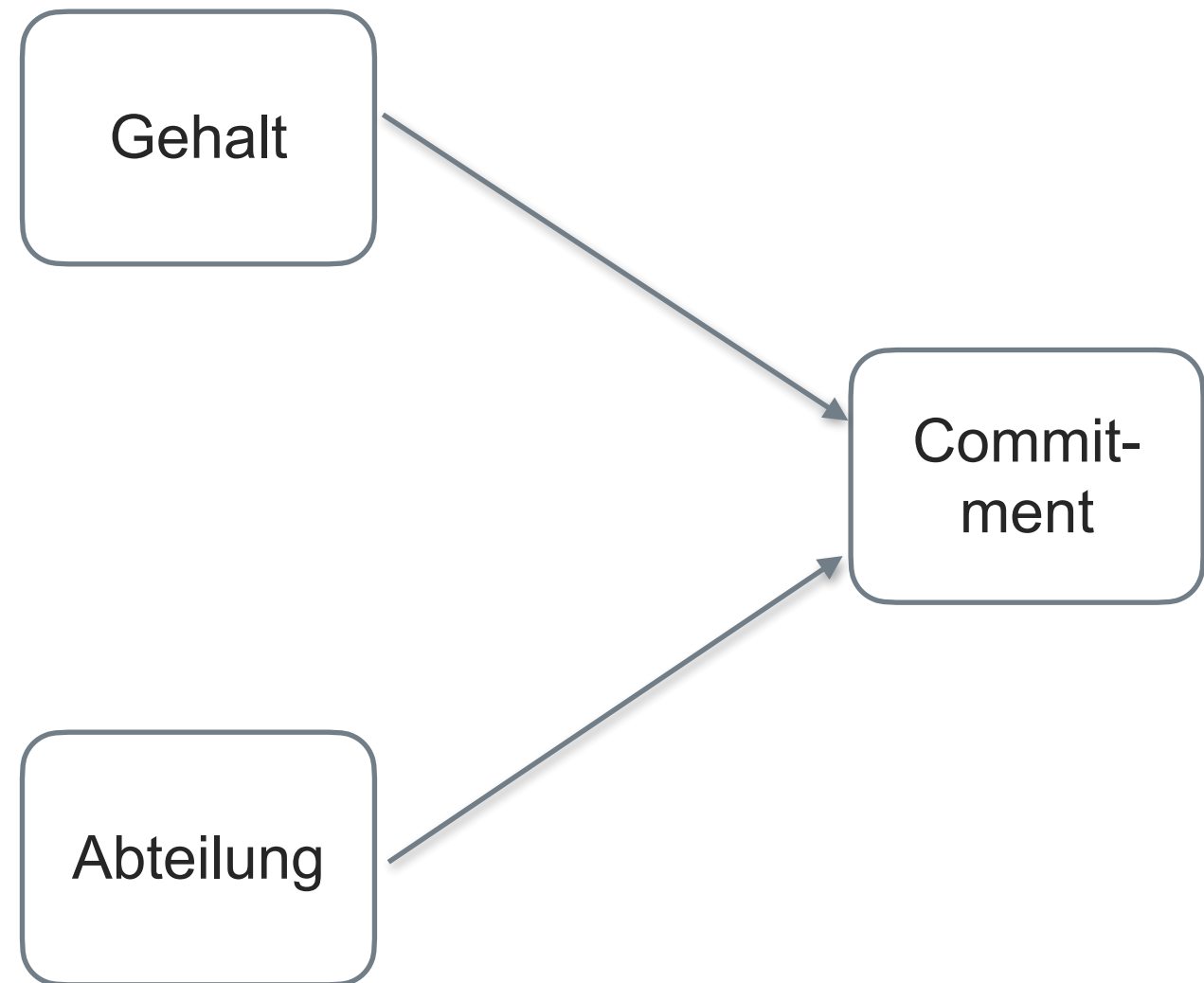
Ist der Einfluss des Führungsstil *unterschiedlich* je nach Abteilung (hinsichtlich Commitment)?

Was ist ein Moderatormodell und was nicht?

Moderatormodell



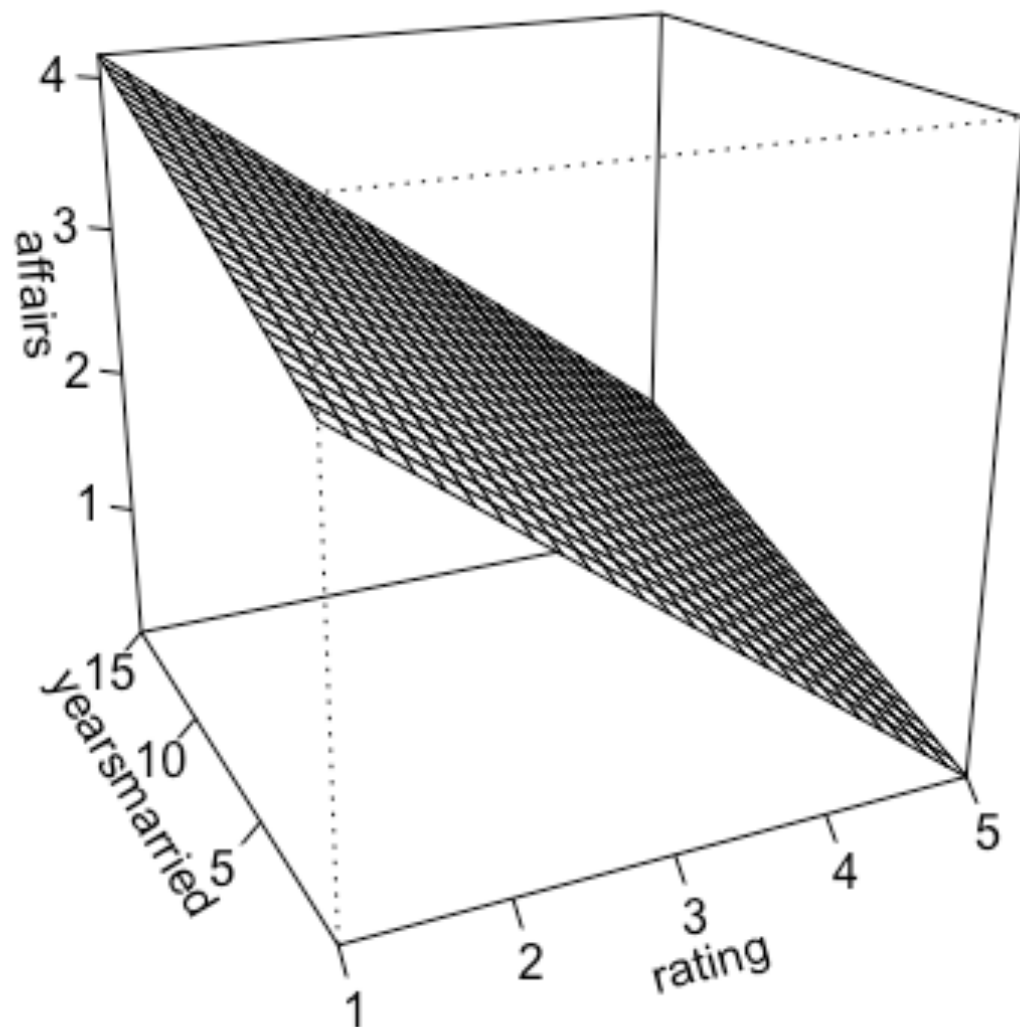
reines Prädiktorenmodell



Veranschaulichung eines Interaktions-/Moderatoreffekts

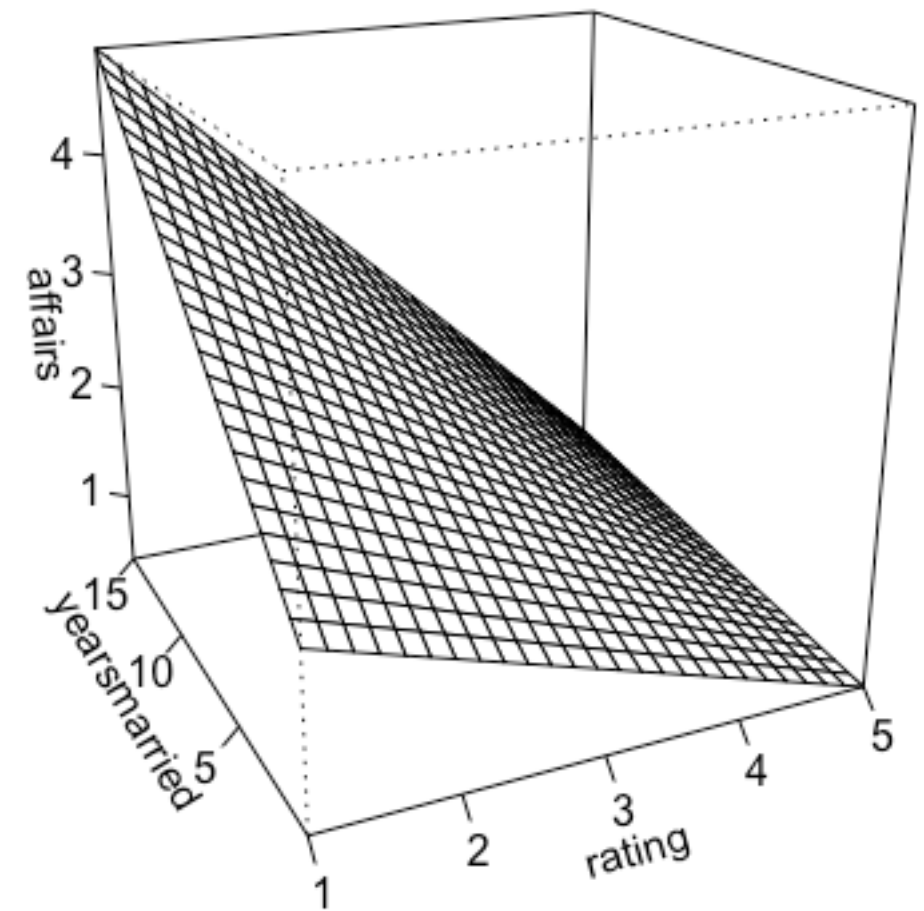
Regressionsmodell: `affairs ~ yearsmarried + rating (+ yearsmarried*rating)`

ohne Interaktionseffekt: $y \sim x_1 + x_2$



Die Stufen von x_1 wirken für alle Stufen von x_2 gleich auf y

mit Interaktionseffekt: $y \sim x_1 + x_2 + x_1*x_2$



Die Stufen von x_1 wirken unterschiedlich auf y je nach den Stufen von x_2

Visualisieren Sie einen Interaktionseffekt in 3D!

Laden Sie den Datensatz "affairs"; erstellen Sie Diagramme mit "Regressionsebenen"!

```
lm1 <- lm(affairs ~ yearsmarried + rating, data = affair)
summary(lm1)
```

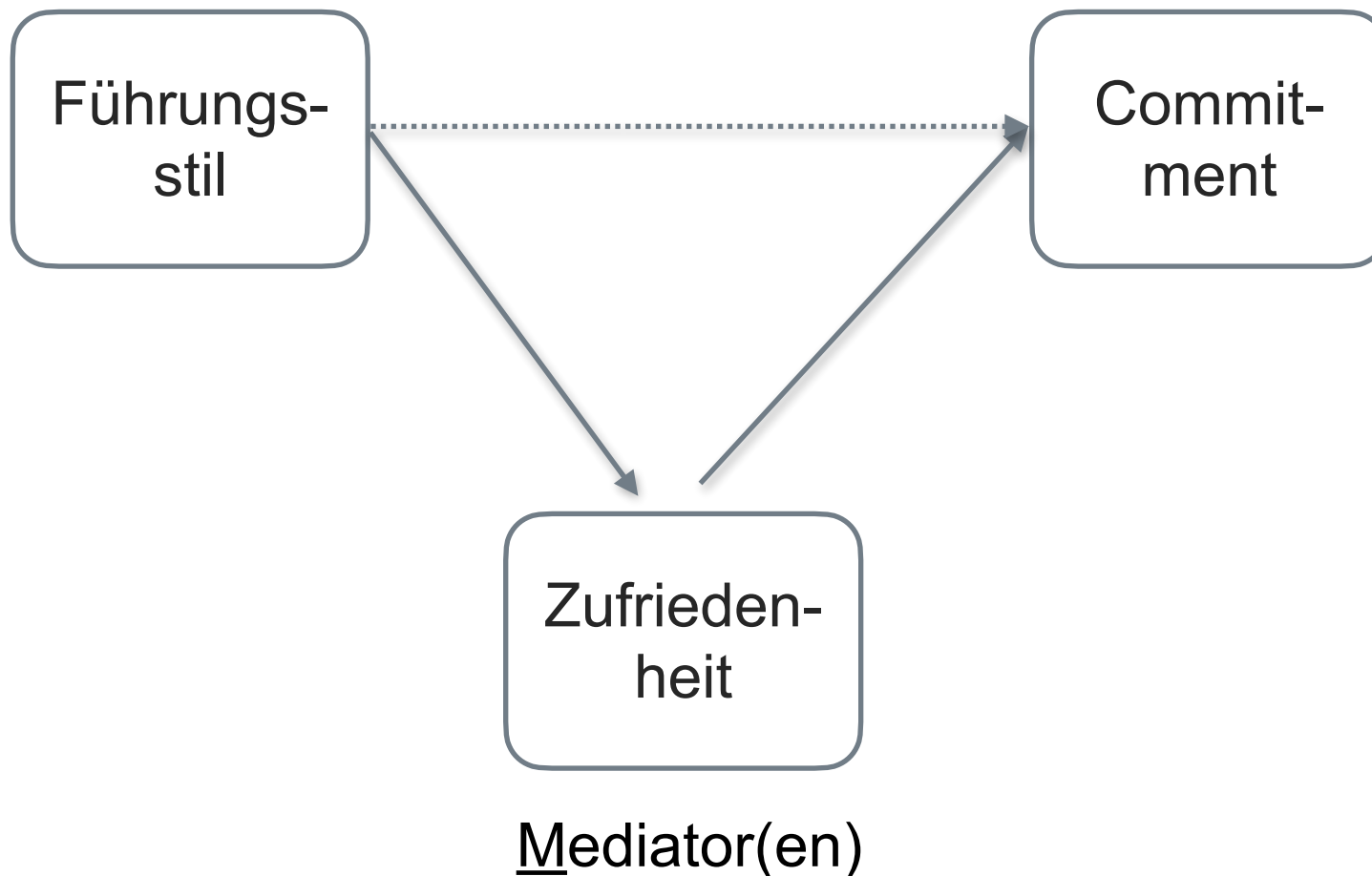
```
lm2 <- lm(affairs ~ yearsmarried + rating + yearsmarried*rating,
data = affair)
summary(lm2)
```

```
persp(lm1, yearsmarried ~ rating, zlab = "affairs")
persp(lm2, yearsmarried ~ rating, zlab = "affairs")
```

Beobachtend: Mediatoreffekt

Prädiktor(en)

Kriterium



Auswertung

- ▶ In Regressionsmodellen M1-M3 müssen Effekte vorhanden sein
- ▶ M1: $P \rightarrow M$
- ▶ M2: $P \rightarrow K$
- ▶ M3: $P + M \rightarrow K$
- ▶ Dabei muss der Einfluss von P in M3 größer sein als in M2 (Mediatoreffekt)

Beispiel

Wird der Einfluss vom Führungsstil auf das Commitment durch die Zufriedenheit der Mitarbeiter erklärt?

Hinweise

- ▶ Dieses Dokument steht unter der Lizenz CC-BY 3.0.
- ▶ Autor: Sebastian Sauer
- ▶ Für externe Links kann keine Haftung übernommen werden.
- ▶ Dieses Dokument entstand mit reichlicher Unterstützung vieler Kolleginnen und Kollegen aus der FOM. Vielen Dank!