

Bortz, Kapitel 8: Hypothesenprüfende Untersuchungen

Hypothesenprüfende Untersuchungen:

- erfordern Vorkenntnisse, die es ermöglichen, vor Durchführung der Untersuchung präzise Hypothesen zu formulieren u. zu begründen
- testen Annahmen über Zusammenhänge, Unterschiede u. Veränderungen ausgewählter Merkmale bei best. Populationen
- Ziele:
 - Prüfung beobachteter Variablenbeziehungen
 - Prognose u. Erklärung von Effekten
- Beitrag zur Stützung kausaler Erklärungen:
 - Identifikation von Ursache-Wirkungs-Beziehungen
 - Testen von Kausalannahmen
- Unterscheidung zw.:
 - log. Stringenz/Schlüssigkeit ihrer Beweisführung
 - Anzahl kausaler Erklärungsalternativen für ihre Ergebnisse
- um anhand der Ergebnisse Hypothesen zu bestätigen/widerlegen, müssen die:
 - Hypothesen präzise formuliert sein (Wirkrichtung, Effektgrößen)
 - Daten kontrolliert erhoben werden (angemessene Operationalisierung/Design)
 - Daten korrekt inferenzstatistisch ausgewertet werden (angemessene Wahl u. Durchführung von Signifikanztests)
- will man Hypothesen über Ursache-Wirkungs-Relationen prüfen
 - liefern Exp. die stringentesten Belege (Exp. sind aber häufig nicht durchführbar)
 - nichtexp. Studien müssen Fehlereffekte u. Störeinflüsse möglichst gering halten

8.1 Grundprinzipien der statist. Hypothesenprüfung

1) Hypothesenarten

Forschungshypothese

- Zusammenhangshypothese (es besteht eine Korrelation zw. zwei/mehreren Merkmalen)
- Unterschiedshypothese (zwei/mehrere Populationen unterscheiden sich in der AV)
- Veränderungshypothese (Ausprägung einer Variable verändert sich im Laufe der Zeit)

Operationale Hypothese

- aus der allg. Forschungshypothese wird eine Vorhersage für ein konkretes Untersuchungsergebnis abgeleitet
- man prognostiziert den Ausgang einer konkreten Untersuchung u. prüft, ob diese auch wirklich zur Klärung der Forschungshypothese beiträgt
- sie resultiert aus der Untersuchungsplanung/Operationalisierung der Variablen

Statistische Hypothesen

- zu der operationalen Hypothese ist eine entsprechende statist. Alternativhypothese zu formulieren u. durch eine komplementäre Nullhypothese zu ergänzen

- jeder Signifikanztest prüft zwei sich ausschließende statist. Hypothesen:
 - H_0 : Nullhypothese
 - H_1 : Alternativhypothese (entspricht i.d.R. der Forschungshypothese)

Beispiele:

Überprüfung einer Zusammenhangshypothese:

Korrelation zw. den Merkmalen ist nicht Null ($H_1: \rho \neq 0$, $H_0: \rho = 0$)

Überprüfung einer Unterschiedshypothese:

es besteht ein Unterschied zw. den Parametern μ_1 und μ_2 ($H_1: \mu_1 \neq \mu_2$, $H_0: \mu_1 = \mu_2$)

Überprüfung einer Veränderungshypothese:

es besteht eine Veränderung zw. den Zeitpunkten t_1 und t_2 ($H_1: \mu_1 \neq \mu_2$, $H_0: \mu_1 = \mu_2$)

Gerichtete/ungerichtete Hypothesen

- statist. Hypothesen können sowohl gerichtet als auch ungerichtet sein
 - ungerichtet: keine Richtung d. Zusammenhangs/Unterschieds/Veränderung
 - gerichtet: Richtung d. Unterschieds/Veränderung bzw. Vorzeichen der Korrelation kann vorhergesagt werden (informationsreicher)

Beispiele:

gerichtete Zusammenhangshypothese:

Korrelation zw. den Merkmalen ist pos. ($H_1: \rho > 0$, $H_0: \rho \leq 0$)

gerichtete Unterschiedshypothese:

polit. Eng. von SoWi-Studis ist größer als das von Mathe-St. ($H_1: \mu_1 > \mu_2$, $H_0: \mu_1 \leq \mu_2$)

gerichtete Veränderungshypothese:

Anteil der Käufer vor der Werbung ist kleiner als der danach ($H_1: \mu_1 < \mu_2$, $H_0: \mu_1 \geq \mu_2$)

- zusammengesetzte Hypothesen: $H_0: \rho \leq 0$ (H_0 wird durch viele Parameter repräsentiert)
- punktuelle/einfache Hypothese: $H_0: \rho = 0$

Spezifische/unspezifische Hypothesen

- statist. Hypothesen können sowohl spezifisch als auch unspezifisch sein
 - unspezifisch: Größe d. Unterschieds/Veränderung u. Höhe d. Korrelation bleibt offen
 - spezifisch: Größe/Höhe wird spezifiziert (dadurch informationsreicher)

Beispiele:

Spezifische gerichtete Zusammenhangshypothese:

Korrelation zw. den Merkmalen liegt nicht unter $\rho = 0,3$ ($H_1: \rho \geq 0,3$, $H_0: \rho < 0,3$)

Spezifische gerichtete Unterschiedshypothese:

Sowi-St. sind um mind. 5 Punkte polit. engagierter ($H_1: \mu_1 \geq \mu_2 + 5$, $H_0: \mu_1 < \mu_2 + 5$)

Spezifische gerichtete Veränderungshypothese:

Werbung erhöht des Käuferanteil um mind. 4 % ($H_1: \mu_2 \geq \mu_1 + 0,04$, $H_0: \mu_2 < \mu_1 + 0,04$)

- spezif. Hypothesen kommen meistens nur in Verbindung mit gerichteten Hypothesen vor (man benötigt dazu entsprechendes präzises Vorwissen)

2) Signifikanztests

Irrtumswahrscheinlichkeit

- Wahrscheinlichkeit, mit der das gefundene emp. Ergebnis auftreten kann, wenn die Populationsverhältnisse der Nullhypothese entsprechen (mit H_1 würde man sich irren)
 - ist sie kleiner als $\alpha\%$, ist das Ergebnis statist. signifikant (α : Signifikanzniveau)
 - kleiner als 5%: signifikant
 - kleiner als 1%: sehr signifikant
- Ergebnis lässt sich mit der Nullhypothese nicht vereinbaren (-> Alternativhypothese)

Stichprobenkennwerteverteilungen

- in jeder hypothesenprüfenden Untersuchung wird ein statist. Kennwert erfasst
 - er soll die gesamte relevante Info einer Untersuchung zusammenfassen
 - Größe des Kennwerts ist stichprobenabhängig
- Stichprobenkennwerteverteilung: Verteilung eines statist. Kennwerts
 - sie ist unbekannt, solange man die wahren Populationsverhältnisse nicht kennt
- → Signifikanztests werden nur eingesetzt, wenn die Ausprägungen der interessierenden Populationsparameter unbekannt sind
- üblich: Behauptung, H_0 wäre richtig (es gelten die damit festgelegten Populationsverhältn.)

Statistische Tabellen

- Frage, wie sich ein Stichprobenkennwert verteilen würde, wenn H_0 gilt
 - mathemat. Problem, das für gebräuchliche statist. Kennwerte gelöst ist
- werden die Kennwerte nach mathemat. Vorschriften transformiert, resultieren statist. Testwerte mit bekannten Verteilungen
 - t-Werte, z-Werte, F-Werte etc.

Signifikante Ergebnisse

- beim Signifikanztest wird der statist. Testwert mit dem Wert verglichen, der von der entsprechenden Testwertverteilung $\alpha\%$ abschneidet
 - ist der Testwert größer als der kritische Tabellenwert -> sign. Ergebnis (H_0 verworfen)
- ein sign. Ergebnis sagt nichts über die Gültigkeit der Hypothesen aus, sondern nur, dass mit 5% oder 1% Irrtumswahrsch. eine tatsächlich richtige H_0 fälschlicherweise verworfen wird
- moderne Statistikprogramme machen statist. Tabellen überflüssig
 - exakte Irrtumswahrsch. wird über Integralrechnung ermittelt (Flächenanteil P)

Beispiel: t-Test

- H_0 : kein Unterschied zw. psych. Belastbarkeit weibl. u. männl. Erwachsener ($\mu_1 = \mu_2$)
- statist. Kennwert für die Überprüfung von Unterschiedshypothesen bei zwei Stichproben: Mittelwertdifferenz (wir durch eine math. Gleichung in einen statist. Testwert transformiert)
- Testwert t folgt bei Gültigkeit der H_0 einer t-Verteilung (bei gl. Varianz u. Normalverteilung)

Ein- und zweiseitiger t-Test

- einseitiger Test: prüft gerichtete Alternativhypothese
 - er lässt für H_0 alle Parameter zu, die $\mu_1 \leq \mu_2$ erfüllen
 - es genügt, die Wahrsch. des Ergebnisses bei Gültigkeit der $H_0: \mu_1 = \mu_2$ zu ermitteln
 - ist der t-Wert größer als der kritische Wert t_α , wird H_0 verworfen
 - kritische t-Wert beim einseitigen Test ist kleiner als der beim zweiseitigen
= daher wird beim einseitigen Test das Ergebnis eher signifikant
 - zweiseitiger Test: prüft ungerichtete Alternativhypothese
 - Werte $t_{(\alpha/2)}$ und $-t_{(\alpha/2)}$ sind die Werte, die von den Extremen der Verteilungsfläche jeweils $\alpha/2\%$ abschneiden
 - ist der t-Wert größer als der kritische Wert $t_{(\alpha/2)}$, wird H_0 verworfen
- nichtsign. Ergebnis: keine Aussage über die Gültigkeit von H_0 u. H_1

Festlegung H_1 und α vor Untersuchungsbeginn

- Hypothesenart (gerichtet etc.) u. Signifikanzniveau (5 oder 1%) müssen vor der Untersuchung festgelegt werden und dürfen im nachhinein nicht mehr geändert werden

3) Probleme des Signifikanztests

- α -Fehler: Fehler 1. Art (man entscheidet sich für H_1 , obwohl H_0 richtig ist)
- β -Fehler: Fehler 2. Art (man entscheidet sich für H_0 , obwohl H_1 richtig ist)

→ darauf resultieren zwei Irrtumswahrscheinlichkeiten

Ermittlung der β -Fehler-Wahrscheinlichkeit

- Wahrsch., mit der man sich irrt, wenn man sich für H_0 entscheidet
- sie kann nur bei spezifischer gerichteter H_1 bestimmt werden ($H_1: \mu_1 \geq \mu_2 + a$)
→ es reicht jedoch aus, wenn $H_1: \mu_1 - \mu_2 = a$ ermittelt wird
- Wahrsch., H_1 fälschlicherweise zu verwerfen sinkt, wenn $X_1 - X_2$ gegen Null geht

Festlegung des β -Fehler-Niveaus

- bislang noch keine Festlegungen wie beim α -Niveau (meist 20%, 10% oder auch 5%)

Teststärke (Power)

- Teststärke: $1 - \beta$
- sie gibt an, mit welcher Wahrsch. man sich für eine richtige H_1 entscheidet
- Untersuchungen mit einer Teststärke unter 50% sollten nicht durchgeführt werden

Praktische Bedeutsamkeit

- Ergebnis, das eine einfache H_1 ($\mu_1 - \mu_2 = a$) verwirft, verwirft auch alle extremeren H_1 ($> a$)
- es sollten zu einer spezif. H_1 nur Parameter zählen, die prakt. bedeutsame Effekte ergeben
→ ein statist. sign. Ergebnis kann in der Praxis völlig bedeutungslos sein

8.2 Varianten hypothesenprüfender Untersuchungen

1) Gefährdung der internen u. externen Validität durch Confounder

Gefährdung der internen Validität

- externe zeitliche Einflüsse
 - andere Einflussgrößen können zeitl. Wandel unterliegen u. Veränderung bewirken
- Reifungsprozesse
 - Vpn selbst können sich verändern
- Testübung
 - Instrument beeinflusst das zu Messende
- mangelnde instrument. Reliabilität
 - Instrument erfasst nur ungenau/fehlerhaft
- statist. Regressionseffekte
 - es kann zu Veränderungen kommen, die artifiziell bedingt sind
- Selektionseffekte
 - ohne Randomisierung können Gruppenunterschiede resultieren, die unabh. von der Intervention sind
- exp. Mortalität
 - Untersuchung wird nicht zu Ende geführt

Weitere Gefährdungen, wenn im Feld die Kontrollgruppe durch die Untersuchung verfälscht wurde:

- empörte Demoralisierung
 - Kontrollgruppe reagiert mit Neid, wenn Treatmentgruppe bessere Bed. hat
- kompensatorischer Wettstreit
 - Ungleichheit könnte bei ihnen Ehrgeiz anstacheln u. sie werden leistungsstärker
- kompensatorischer Ausgleich
 - Leiter bemerkt Ungleichheit u. versucht diese durch Maßnahmen auszugleichen
- Treatmentdiffusion
 - Kontrollgruppe versucht, Reaktionen in der Experimentalgruppe zu imitieren

→ intern valide: Ergebnisse sprechen eindeutig für/gegen die Hypothese u. Alternativerklärungen scheinen unplausibel

Gefährdung der externen Validität

- mangelnde instr. Validität
 - Instrument erfasst nicht das, was es eigentlich soll (oder wird in einem anderen Kontext eingesetzt)
- Stichprobenfehler
 - Verallgemeinerung auf Grundgesamtheit, wenn Stichprobe nicht repräsentativ ist
- exp. Reaktivität
 - Ergebnisse sind nur unter den Laborbed. valide u. können nicht generalisiert werden
- Pretesteffekte
 - Pretests können Sensitivität/Problembewusstsein der Teilnehmer verändern

- Hawthorne-Effekte
 - Bewusstsein, an einer wiss. Untersuchung teilzunehmen, verändert das Verhalten

→ extern valide: Ergebnisse sind auf andere, vergleichbare Personen/Situationen generalisierbar

2) Zusammenhangshypothesen

- werden mittels Interdependenzanalysen überprüft
 - vergleichsweise wenig Aufwand
 - klassisch: einfache Querschnittsuntersuchung (Cross-sectional-Design)
= zu einem best. Zeitpunkt werden mehrere Merkmale an einer Stichprobe erhoben
- Zusammenhang sagt zunächst nichts über die Kausalbeziehung aus
= Ergebnisse beziehen sich primär auf die Art/Intensität des Kovariierens der Merkmale

Bivariate Zusammenhangshypothesen

- formale Struktur: zw. zwei Merkmalen besteht ein Zusammenhang
→ wenn gerichtet: entweder pos., gleichsinniger oder neg., gegensinniger Zusammenhang

Datenerhebung

- für jedes Untersuchungsobjekt werden beide Merkmale erhoben
= Forderung bei einer Korrelation, dass einem Messwert ein anderer eindeutig zugeordnet ist
- Korrelationskoeffizient: quantifiziert die Enge des Zusammenhangs
= seine statist. Bedeutsamkeit überprüft ein Signifikanztest
- Leitlinie zur Überprüfung von Zusammenhangshypothesen:
 - geklärt sein, welche Messwerte der untersuchten Variablen Messwertpaare bilden
 - bei einer Stichprobe u. zwei Merkmalen: es darf pro Untersuchungsobjekt nur ein Messwertpaar in die Korrelationsberechnung eingehen
 - bei zwei Stichproben u. einem Merkmal: Untersuchungsobjekte beider Stichproben müssen Paare bilden, deren Paarlinge nicht austauschbar sein dürfen
 - sind die Paarlinge prinzipiell austauschbar, ist die Zusammenhangshypothese über einen Intraklassenkorrelationskoeffizienten zu prüfen

Bivariate Korrelationen

- bestimmen über einen Korrelationskoeffizienten Enge/Richtung des Zusammenhangs
- für Variablen untersch. Skalenniveaus existieren versch. Korrelationskoeffizienten

y	x	Intervallskala	Ordinalskala	k.Dichotomie	n.Dichotomie	Nominalskala
Intervallskala	Intervallskala	Produkt-M.-Korrelation	Rang-korrelation	Biserale Korrelation	Punkt-biserale Korrelation	Kontingenzkoeffizient
Ordinalskala	Ordinalskala		"	Biserale Rangkorrelat.	Biserale Rangkorrelat.	"
künstl. Dichotomie	Ordinalskala			Tetrachorische Korrelation	Phi-Koeffiz.	"
natürl. Dich.	Ordinalskala				"	"
Nominalskala	Nominalskala					"

- gibt es bei Merkmalen mit untersch. Skalenniveau kein spez. Korrelationsmaß, wird das niedrigere Skalenniveau verwendet
= auch ist es möglich, dies als multivariate Zusammenhangshypothese zu prüfen
- bivariate pos. Korrelation: hohe Ausprägung beider Merkmale
- bivariate neg. Korrelation: hohe Ausprägung des einen, niedrige Ausprägung des anderen M.

Dichotome Merkmale

- zweifach gestufte Merkmale
- künstlich: kontinuierlich verteiltes Merkmal wird auf zwei Stufen reduziert
- natürlich: Merkmal hat tatsächlich nur zwei Ausprägungen

Merkmalsprofile

- es lassen sich auch Zusammenhänge zweier/mehrerer Personen analysieren
- Annahme, dass für jede Person bezügl. mehrerer Variablen Messungen vorliegen, die zusammengenommen ind. Merkmalsprofile ergeben
- Maße: z.B. Q-Korrelation (bei Übereinstimmung im semantischen Differenzial)

Stichprobenfehler

- Stichprobenfehlern können zu starken Verzerrungen führen
 - Gesamtzusammenhang wird unterschätzt/überschätzt (Extremgruppenselektion)
 - Artefakte, die durch mathemat. Abhängigkeit der Merkmale entstehen
- wichtig: Stichprobe muss zufällig gezogen werden (keine systemat. Selektion)

Multivariate Zusammenhangshypothesen

Partielle Zusammenhänge

- häufig wird eine dritte Variable z eingeführt, von der sowohl x als auch y abhängen
- Scheinkorrelation: Zusammenhang zw. x und y verschwindet, wenn man z ausschaltet
→ besteht weiterhin ein Zusammenhang, ist dieser partiell
- Partialkorrelation: überprüft diesen Zusammenhang (setzt Intervallskala voraus)
= bei mehreren z wird eine Partialkorrelation höherer Ordnung verwendet
- Herauspartialisieren von Z: Bereinigung der Merkmale x und y
- Semipartkorrelation: überprüft, wenn z nur aus einer der Variablen herauspartialisiert wird

Multiple Zusammenhänge

- Zusammenhang zw. mehreren Prädiktorvariablen (x) und einer Kriteriumsvariablen (y)
- Multiple Korrelation:
 - überprüft multiple Zusammenhänge
 - wenig Aufwand u. berechenbar, wenn Kriteriumsvariable intervallskaliert ist
- Prädiktoren können auch dichotom oder nominalskaliert sein
 - dichotomes nominales Merkmal (0,1)
 - bei mehr Kategorien wird durch Indikatorvariablen (Dummy-Variablen) kodiert
 - Optimal Scaling: schätzt für die k Kategorien eines nominalen Merkmals metrische Werte, sodass man statt k-1 Dummy-Variablen nur eine Variable benötigt

- multiple Korrelation berücksichtigt auch Kombinationswirkungen der Prädiktoren
 - sagen dadurch mehr aus als bivariate Zusammenhänge
 - bei der Überprüfung multivariater Hypothesen durch mehrere bivariate Korrelationen gehen entsch. Infos verloren (es resultieren Fehlinterpretationen)

Kanonische Zusammenhänge

- Untersuchung zweier Variablenätze (mehrere Prädiktoren/mehrere Kriterien)
- kanonische Korrelation: überprüft kanonische Zusammenhänge
- nominalskalierte Kriterien:
 - Konfigurationsfrequenzanalyse: überprüft Zusammenhangshypothesen mit nominalen Prädiktoren u. nominalen Kriterien
 - Auswertungen erfolgen mit log-linearen Modellen

Faktorielle Zusammenhänge

- wechselseitige Zusammenhänge vieler Merkmale sind durch wenige Faktoren erklärbar
- Faktorstrukturvergleich: überprüft faktorielle Zusammenhänge
- weitere Hypothesen dieser Art beziehen sich auf die Anzahl der Faktoren bzw. darauf, welche Faktoren voneinander abhängig/unabhängig sind
- Konfirmatorische Faktorenanalyse: überprüft diese (ebenso: Strukturgleichungsmodellen)
- Faktorenanalyse: bündelt Variablen gemäß ihrer Interkorrelationen zu Faktoren
 - explorativ: kann ohne Vorannahmen durchgeführt werden
 - konfirmativ: Faktorenladungsmuster wird als Hypothese vorgegeben
- multiple Korrespondenzanalyse: überprüft die faktorielle Struktur von Kontingenztafeln
→ Ziel: Kategorien der Merkmale als Punkte in einem Faktorenraum mit möglichst wenigen Dimensionen abzubilden (demnach: Faktorenanalyse für kategoriale Daten)

Kausale Zusammenhangshypothesen

- Korrelationen geben nur Auskunft über die Enge/Richtung des Zusammenhangs, sagen jedoch nichts über seine Ursachen u. den kausalen Zusammenhang
- Prüfung kausaler Hypothesen kann immer nur annäherungsweise erfolgen
 - es ist möglich, durch Nullkorrelationen Kausalmodelle zu falsifizieren, da diese Korrelationen implizieren
= kausale Hypoth. können durch nicht vorhandene Korrelationen widerlegt werden
 - Kausalinterpretationen von Korrelationen sind höchstens logisch begründbar
- Korrelationsstudien haben so nur eine geringe interne Validität (sind Exp. unterlegen)
- aber: man kann mit ihnen best. Kausalhypothesen von vornh. als unwahrsch. ausschließen

Längsschnittstudien

- Anzahl der konkurrierenden Kausalmodelle wird erheblich eingeschränkt
= das später erhobene Merkmal kann das frühere nicht beeinflussen
- interne Validität von Korrelationsstudien über zeitlich versetzt erhobene Merkmale ist i.d.R. höher als die von Studien, die dieselben Merkmale nur zu einem Zeitpunkt prüfen

Cross-lagged-Panel-Design

- entscheidet über die relative Plausibilität von zwei konkurrierenden Kausalhypothesen
- interne Validität lässt sich über die Einbeziehung von mehr als zwei Messpunkten erhöhen

Pfadanalyse

- prüft Annahmen bezüglich der wechselseitigen Kausalbeziehungen
= kann Kausalhypothesen widerlegen (wenn auch nicht eindeutig bestätigen)
- zwingt den Untersuchenden dazu, sich über Ursache-Wirkungs-Sequenzen Gedanken zu machen (kausale Modelle zu konstruieren)
- wird häufig als Pretest durchgeführt, bevor man gezielt eine Kausalhypothese überprüft

Lineare Strukturgleichungsmodelle (SEM)

- prüfen zusätzlich Hypothesen, die sich auf latente Merkmale u. deren Bezieh. untereinander u. zu den untersuchten Merkmalen beziehen (z.B. LISREL)
- erfordern genaue Überlegung, zw. welchen Variablen kausale Bezieh. bestehen könnten
- können keine Kausalität nachweisen, sondern lediglich Modelle falsifizieren
= Pfadkoeffizienten geben die relative Stärke von Kausaleffekten an, wenn das Modell zutrifft
- bei der Überprüfung der Modellgüte ist H_0 die Wunschhypothese
= bei größer werdendem Stichprobenumfang sinkt die Chance auf Bestätigung eines Modells
- Vorschlag MacCallum: Test, bei dem ein sign. Ergebnis einen guten Modell-Fit zeigt
→ basiert auf RMSEA-Index u. zeigt, wie schlecht ein Modell durch die Daten angepasst wird

Kausale Mikromediatoren

- wenn eine Maßnahme wirkt, weiß man noch nicht, was die tats. Wirkmechanismen waren
- Dekomposition globaler Wirkprozesse in kausale Mikromediatoren kann dabei helfen
= präzisiert die Überprüfung kausaler Hypothesen erheblich
- man versucht z.B. durch qualitative Interviews die erlebten Ursachen des vermeintlich durch das Treatment ausgelösten Verhaltens zu ergründen

Metaanalyse

- um eine Kausalhypothese zu stützen u. zu generalisieren, sind wiederholte Prüfungen der gleichen Kausalhypothese wichtig
- aus den Ergebnissen vergleichbarer Untersuchungen erfährt man wie stark der Kausaleffekt ist, bei welchen Gruppen/Bed. er auftritt u. wie seine Generalisierung eingeschränkt ist

Fazit:

- querschnittl. Interdependenzanalysen können vermutete Gemeinsamkeiten zw. Merkmalen absichern (solange keine Kausalinterpretationen intendiert sind)
→ interne Validität ist jedoch meistens gering
- interne Validität korrelativer Studien kann durch folgende Maßnahmen erhöht werden:
 - zeitversetzte Messungen von Prädiktor- u. Kriteriumsvariablen (Cross-lagged Panel)
 - Neutralisierung der Wirkung von Kontroll-/Störvariablen (Partialkorrelation)
 - Detailanalysen v. Wirkungspfaden in komplexen Kausalmodellen (Pfadanalyse, SEM)

→ dadurch ist es möglich, kausale Modelle zu interpretieren (wenn auch nicht zu beweisen)

3) Unterschiedshypothesen

- Vergleich mehrerer Stichproben, die sich hinsichtlich einer/mehrerer UV unterscheiden
→ Bsp.: UV ist behandelt/unbehandelt (Treatmentgr.: behandelt, Kontrollgr.: unbehandelt)
- Hypothesen, die sich auf die Wirksamkeit einer Maßnahme beziehen

Kontrolltechniken

= Stichproben müssen hinsichtlich aller relevanten Merkmale vergleichbar/äquivalent sein

1. Kontrolle personengebundener Störvariablen

- Randomisierung:
 - zufällige Zuordnung von Personen zu Untersuchungsbedingungen
 - beste Möglichkeit, personengebundene Störvariablen zu kontrollieren
 - es werden auch Störvariablen neutralisiert, die noch nicht bekannt waren
- Grenzen der Randomisierung:
 - viele Treatments entziehen sich einer künstlichen Manipulation u. beziehen sich auf bereits bestehenden Gruppen/Populationen
= sie sollen in ihrer natürlichen Zusammensetzung untersucht werden
 - in diesen quasiexp. Untersuchungen müssen besondere Maßnahmen ergriffen werden, um personengebundene Störvariablen zu kontrollieren
→ UV kann mit weiteren Variablen konfundiert sein, die d. AV ebenfalls beeinflussen
- Konstanthalten:
 - werden Störvariablen konstant gehalten, beeinflussen sie nicht die Unterschiedlichk. von Vergleichsgruppen
 - es verringert sich jedoch die externe Validität
- Parallelisierung:
 - parallele Vergleichsgruppe weist hinsichtlich der Störvariablen gleiche Mittelwerte u. Streuungen auf (z.B. Alter, Geschlecht, Schulbildung)
 - Parallelisierung geht häufig zu Lasten externer Validität
= werden z.B. nur Personen mit der gleichen Berufsdauer verglichen, repräsentiert diese Stichprobe die Grundpopulation nicht mehr
- Matched Samples:
 - Vpn der einen Stichprobe wird einem Vpn einer anderen Stichprobe in Bezug auf die zu kontrollierenden Störvariablen zugeordnet („passender Partner“)
 - bei mehreren Störvariablen schwierig (Einschränkung der externe Validität)
- Mehrfaktorielle Pläne:
 - Störvariable wird als gesonderter Faktor in einem mehrfakt. Untersuchungsplan mit berücksichtigt
- Kovarianzanalytische Kontrolle:
 - kontrolliert rechnerisch die Beeinflussung einer AV durch personengeb. Störvariablen

2. Kontrolle untersuchungsbedingter Störvariablen

- Ausschalten:

- Untersuchungen aller Vergleichsgruppen sollte störungsfrei verlaufen
 - typ. für Laboruntersuchungen
 - geht auf Lasten der externen Validität (keine natürlichen Rahmenbed. mehr)
 - Konstanthalten
 - Störungen sollten zumindest in allen Versuchsdurchführungen gleich sein
 - Registrieren
 - Art u. Intensität von Störungen sollten möglichst genau protokolliert werden, um ggf. nachträglich die Ergebnisse bezüglich des Einflusses der Störvariablen zu korrigieren
 - v.a. bei Feldexp., bei denen aufs Ausschalten/Konstanthalten verzichtet wird
- Voraussetzung: untersuchungsbedingten Störvariablen müssen bekannt sein

Zweiggruppenpläne

- einfachste einfaktorielle Plan (zweifach gestufte UV, eine AV)

Exp. Untersuchung

- Ziehung einer repräsentativen Stichprobe, daraus werden Vpn zufällig in zwei gleich große Gruppen aufgeteilt (Vergleichbarkeit kann durch Vortests geprüft werden)
 - Treatmentgruppe: erhält das Treatment
 - Kontrollgruppe: bleibt unbehandelt
- Hypothesen, die sich auf die untersch. Wirkung von zwei Treatments beziehen, werden ebenfalls mit einem Zweigruppenplan geprüft
 - statt Kontrollgruppe wird eine zweite Treatmentgruppe untersucht

Quasiexp. Untersuchungen

- Zugehörigkeit der Vpn zu den zwei Stufen einer UV ist zugegeben
- aufgrund fehlender Randomisierung: Gefahr der Konfundierung der UV mit Störvariablen
 - Überprüfung von Unterschiedshypothesen: mit t-Test für unabhängige Stichproben
 - wählt man Matching als Kontrolltechnik: t-Test für abhängige Stichproben

Extremgruppenvergleich

- es werden nur Vpn berücksichtigt, die bezüglich einer kontinuierlichen UV besonders hohe/niedrige Ausprägungen aufweisen
- eigentlich eine explorative Studie u. keine hypothesenprüfende Untersuchung

Mehrgruppenpläne

- untersucht Hypothesen, die sich auf mehr als zwei Treatments beziehen
- exp. Vorgehensweise entspricht dem Zweigruppenplan (nur: mehrfach gestufte UV, eine AV)
- Überprüfung erfolgt mittels einfaktorieller Varianzanalyse
- Einzelvergleiche überprüfen, ob sich best. Treatments sign. voneinander unterscheiden
 - a-priori-Einzelvergleich: setzen die Formulierung gezielter Einzelvergleichshypothesen vor der Untersuchung voraus
 - a-posteriori-Einzelvergleich: man stellt im Nachhinein fest, welche Treatments sich sign. voneinander unterscheiden

→ Einzelvergleiche verwendet man auch, um Kombis einzelner Treatments mit anderen zu vergleichen (v.a. wenn mehreren Treatment- u. mehrere Kontrollgr. untersucht werden)

- weitere Zusatzauswertungen sind möglich, wenn UV u. AV intervallskaliert sind (Trendtests)
- Mehrgruppenpläne sind auch bei quasiexp. Untersuchungen einsetzbar
 - statt versch. Treatmentgruppen werden Stichproben aus den relevanten Populationen miteinander verglichen
 - Kontrolle durch Matching ist nur bei höchstens 3-4 kleineren Stichproben möglich (Auswertung mittels Varianzanalyse für abhängige Stichproben)

Faktorielle Pläne

- man untersucht mehrere UV (nicht nur Stufen von einer) u. eine AV

Zweifaktorielle Pläne

- einfachste Variante faktorieller Pläne (zwei UV, eine AV)
= Bsp. der UV: Faktor A (p-fach), Faktor B (q-fach) -> man benötigt $p \cdot q$ Stichproben
- informiert auch über die Interaktion/Wechselwirkung der beiden UV
- Auswertung mittels einer zweifaktoriellen Varianzanalyse
→ Prüfung von Haupteffekt A, Haupteffekt B u. die Interaktion AxB

→ Haupteffekte u. Interaktion

- einfacher Fall einer zweifaktoriellen Varianzanalyse: je zwei Stufen für A u. B
 - Bsp.: Placebo gegen herkömmliches Beruhigungsmittel bei Männern u. Frauen
 - Haupteffekt A: Placebo u. Medikament wirken unabhängig vom Geschlecht untersch.
 - Haupteffekt B: Männer u. Frauen wirken in Bezug auf beide Medikamente untersch.
 - Interaktion AxB: z.B. reagieren Frauen stärker auf das Placebo als Männer
 - Interaktion:
 - Wirkung eines Faktors auf AV hängt von der Ausprägung des anderen Faktors ab
 - sign. Interaktionseffekt bedeutet, dass beide Faktoren nicht einfach additiv zusammenwirken, sondern in anderer Weise (verstärken sich, schwächen sich ab)
 - Interaktionsdiagramme:
 - um die Art des Zusammenwirkens sichtbar zu machen
 - Werte der AV werden auf der y-Achse, Stufen der Faktoren auf x-Achse eingetragen
 - es werden dieselben Zellenmittelwerte dargestellt, nur jew. anders gruppiert
 - ohne Interaktion sind die Grafen parallel
 - drei Arten der Interaktion:
 - ordinal: Grafen verlaufen nicht parallel, aber doch gleichsinnig (beide fallen ab etc.)
 - hybrid: Grafen verlaufen nur in einem der beiden Interaktionsdiagramm gleichsinnig
 - disordinal: Grafen verlaufen nicht gleichsinnig
- ist für die Interpretation der sign. Haupteffekte relevant
- bei ordinal: Haupteffekte dürfen global interpretiert u. generalisiert werden
 - bei hybrid: nur ein Faktor darf global interpretiert werden
 - bei disordinal: keiner der beiden Faktoren darf global interpretiert werden
- während sich Haupteffekte auf die gesamte Zielpopulation beziehen, richten sich Interaktionshypothesen auf die differenzielle Wirkung der Treatments

= Wirkung hängt von der Art der untersuchten Subpopulationen ab

- Prüfung von Interaktionen soll eine möglichst gezielte Hypothese vorangestellt werden

→ Experimentelle u. quasiexperimentelle Pläne

- für eine rein exp. zweifaktorielle Untersuchung müssen beide Faktoren die Zuordnung der Vpn zu allen Faktorstufenkombis zufällig ermöglichen
- rein quasiexp. Untersuchung überprüft hingegen die Unterschiede zw. natürlichen Teilpopul., die sich hinsichtl. zwei Faktoren unterscheiden (Randomisierung nicht möglich)

→ Kontrollfaktoren

- zweifaktorielle Untersuchungen überprüfen drei versch. Unterschiedshypothesen, es müssen jedoch nicht immer alle drei explizit formuliert sein
= meist steht nur eine Hypothese im Vordergrund u. der zweite Faktor dient Kontrollzwecken
- UV kann durch andere Merkmale überlagert sein (Störvariablen)
= lässt sich hierbei ein Merkmal benennen, das mit hoher Wahrsch. konfundiert, kann diese als Kontrollfaktor aufgenommen werden (obwohl sich Hyp. auf den anderen Faktor bezieht)

Drei- u. mehrfaktorielle Pläne

- wenn drei/mehrere Faktoren sowie deren Interaktionen simultan kontrolliert werden
- Stufen jedes Faktors müssen mit den Stufen aller anderen Faktoren kombiniert werden u. unter jeder Faktorstufenkombination wird eine Zufallsstichprobe untersucht
- ebenfalls varianzanalytische Auswertung
→ Teststärke hängt in hohem Maße davon ab, welche/wieviele Hypothesen geprüft werden

→ Interaktionen zweiter Ordnung

- wenn die Art der Interaktion zw. zwei Faktoren von den Stufen eines 3. Faktors abhängt
= besteht nicht, wenn das Muster der AxB-Interaktion auf allen Stufen des Faktors C gleich ist

Solomon-Viergruppenplan

- wurde zur Kontrolle von Pretesteffekte entwickelt
→ Erweiterung des klass. Pretest-Posttest-Designs
- erfordert vier randomisierte Gruppen
= Experimentalgruppe, Kontrollgruppe, One-Shot-Case-Design mit Posttest, nur Posttest
- Vergleich der Gruppe 2 u. 4 dient der Abschätzung von Pretesteffekten
= beides mal ohne Treatment, nur einmal mit Pretest u. einmal ohne
- er lässt sich auch in mehrfaktorielle Pläne einbauen
→ entscheidend ist, dass immer ein weiterer Faktor einbezogen wird, der vorgetestete u. nicht vorgetestete Vpn unterscheidet

Hierarchische Pläne

- unvollständige Pläne, die nur einige der möglichen Faktorstufenkombis berücksichtigen
- es können nur Haupteffekte geprüft werden
= diese sollten nur dann interpretiert werden, wenn Interaktionen unwahrscheinlich sind

Zweifaktorielle hierarchische Pläne

- nicht alle Stufen des Faktors B werden mit allen Stufen von A kombiniert
= Stufen von B sind unter die Stufen von A geschachtelt
- Vorteil:
 - viel weniger Stichproben u. Vpn als bei vollständigen Plänen
 - z.B. bei multizentrischen Studien sinnvoll
- Nachteil:
 - Unterschiede zw. den Stufen des einen Faktors sind nur in Verbindung mit den jew. Stufen des anderen Faktors interpretierbar
 - Interaktionen sind in hierarchischen Plänen nicht überprüfbar
- ähnlich bei dreifaktoriellen hierarch. Plänen: hier sind nur die drei Haupteffekte überprüfbar

Teilhierarchische Pläne

- z.B. wenn neben den Faktoren noch das Geschlecht überprüft wird
= ist mit allen drei Faktoren kombinierbar
- es werden nicht nur die drei Haupteffekte überprüft, sondern auch die Interaktionen zw. den Faktoren, die vollständig miteinander kombiniert sind

Quadratische Pläne

- Sonderform der unvollständigen Pläne, bei denen alle Faktoren die gleiche Stufenzahl haben
= zweifakt. Plan mit zwei Faktoren mit gl. Stufenzahl kann als Quadrat dargestellt werden
- es können nur Haupteffekte überprüft werden

Lateinische Quadrate

- jede Stufe jeden Faktors ist mit allen Stufen der übrigen Faktoren vollständig kombiniert
= Plan ist bezügl. der Haupteffekte ausbalanciert
- als Konstruktionsprinzip wählt man sog. zyklische Permutationen
- es können keine Interaktionseffekte geprüft werden u. Haupteffekte sind nur dann eindeutig interpretierbar, wenn die Interaktionen zw. den Faktoren zu vernachlässigen sind

Griechisch-lateinische Quadrate

- wenn man zwei lateinische Quadrate kombiniert u. diese zueinander orthogonal sind
→ orthogonal: wenn deren Kombi zu einer neuen Anordnung führt, in der jede Zweierkombi der Faktorstufen genau einmal vorkommt
- es können vier Faktoren kontrolliert werden (weniger aufwändig als vierfakt. Plan)
- ebenfalls nur Haupteffekte prüfbar

Pläne mit Kontrollvariablen

- Kontrollvariable:
= Störvariablen, deren Einfluss mittels Kovarianzanalyse aus der AV heraus-partialisiert wird
- Möglichkeit mehrere Kontrollvariablen zu berücksichtigen: Partialkorrelation höherer Ordn.
- Kovarianzanalyse wird eher bei quasiexp. Untersuchungen verwendet
→ ist sie unangemessen, sollte das Blockbildungsverfahren vorgezogen werden

Multivariate Pläne

- es werden simultan mehrere AV überprüft (unabhängig davon, wieviele UVs es gibt)
- Vorteil: wechselseitigen Beziehungen der AVs werden berücksichtigt u. aufgedeckt
- lieber eine multivariate Analyse anstelle mehrerer univariater Analysen (mehr Erkenntnisse)
- Auswertung durch multivariate Varianzanalyse (MANOVA)

Zusammenfassende Bewertung

- exp. Zwei-/Mehrgruppenpläne haben eine hohe interne Validität
 - bei großen Stichproben sorgt die Randomisierung dafür, dass sich die verglichenen Stichproben vor der Behandlung nicht unterscheiden
 - bei kleinen Stichproben sind jedoch Pretestunterschiede nicht auszuschließen
- Sicherung der internen V. durch:
 - Parallelisierung
 - Matched Samples
 - Kontrolle von Störvariablen

→ Kontrolltechniken können jedoch zu Lasten der externen Validität gehen
- bei quasiexp. Untersuchungen sind ergänzende Kontrolltechniken nötig:
 - personengebundene Störvariablen sollen vorsorglich miterhoben u. kovarianz-analytisch berücksichtigt werden
 - Blockbildung homogener Gruppen bei nominalska. Störvariablen mit wenig Stufen
- faktorielle Pläne
 - tragen zur Erhöhung der internen Validität bei
 - erklären AV durch die Berücksichtigung der Interaktionen besser als einfakt. Pläne
- multivariate Pläne
 - haben eine höhere externe Validität als univariate Pläne
 - erfassen das Konstrukt über viele Indikatoren (generalisierende Ergebnisse)
- bei exp. Querschnittsuntersuchungen sind externe zeitl. Einflüsse u. Testübung irrelevant = bei quasiexp. Untersuchungen können sie die Ursache von Pretestunterschieden sein u. müssen daher kontrolliert werden