



Methoden der Sportwissenschaft 2

Arbeitsblatt 4: Interne und externe Validität - Gültigkeitsanspruch der Untersuchungsbefunde (aktualisiert 07.12.2006)

Interne Validität

Überlegungen zur internen Validität befassen sich mit Fragen, ob die Untersuchungsbefunde eindeutig interpretiert werden können. Es stellen sich folgende Fragen:

- In wie fern ist die Variation in der abhängigen Variable (AV) auf die Variation in der unabhängigen Variable (UV) zurückzuführen?
- Wie gut ist es gelungen, in der Untersuchung nur die Auswirkungen der UV in der AV zu finden?
- Können mögliche Alternativhypothesen zur Erklärung der Variation der AV ausgeschlossen werden?

Ein Versuchsplan ist dann **intern valide**, wenn die Variation der AV (bis auf unsystematische Fehlervarianz) nur auf die Variation der UV zurückgeht.

Das MAX-KON-MIN-Prinzip als Maßnahme zur Verbesserung der internen Validität (vgl. Bös, Hänsel & Schott, 2000, S. 50 - 60)

Bei allen Messreihen erhält man ein Maßzahlen für die zentrale Tendenz (Modus, Median, Mittelwert) und Streuungsmaße (Prozentränge, Range, Varianz, Standardabweichung). Diese Streuung eines Wertwertes um den Mittelwert herum hängt von verschiedenen Faktoren ab.

Beispiel: Zusammenhang zwischen Körpergewicht und Körpergröße.
 Das Körpergewicht als AV ist in hohem Maße von der Körpergröße (UV) abhängig. Dies kommt beispielsweise in der Verwendung des Body Mass Index (BMI) zur Einschätzung des Normalgewichts zum Ausdruck. Den Anteil an der Streuung der Messwerte des Körpergewichts, der durch die Körpergröße zustande kommt, bezeichnet man als Primärvarianz.
 Das Körpergewicht ist nicht nur von der Körpergröße, sondern auch von der Muskelmasse und dem Körperfettanteil abhängig. Die Muskelmasse bzw. der Körperfettanteil führen zu einer systematischen Veränderung des Zusammenhangs zwischen Körpergröße und Körpergewicht und werden deshalb als Sekundärvarianz bezeichnet.
 Bei jeder Messung können Fehler gemacht werden. Deshalb legt man so großen Wert auf die Objektivität und Reliabilität von Messverfahren. Das ist die sog. Fehlervarianz. Im Gegensatz zur Primär- und Sekundärvarianz ist ihr Einfluss unsystematisch, d. h. man kann nicht sagen in welche Richtung die Messwerte beeinflusst werden.
 Bei einer Untersuchung an Rehatteilnehmern wurden Körpergröße, Körpergewicht und der Körperfettanteil bestimmt. Eine mit diesen Variablen durchgeführte Regressionsanalyse erbrachte eine Gesamtvarianzaufklärung von 77 %, d. h. 77 % der zu beobachtenden Unterschiede im Körpergewicht können erklärt werden durch die Körpergröße (46 %), den Körperfettanteil (17 %) und das Geschlecht (14 %). Die restlichen 23 % sind bedingt durch die intraindividuellen Unterschiede bzw. die Fehlervarianz.

Daraus folgt das Max-Kon-Min-Prinzip:

- **Maximiere** die Primärvarianz!
- **Kontrolliere** die Sekundärvarianz!
- **Minimiere** die Fehlervarianz!



MAXimierung der Primärvarianz der UV

Primärvarianz maximieren bedeutet, die Stufen der UV so zu wählen, dass sich in der AV möglichst klare Unterschiede zwischen den Bedingungen zeigen.

- **Wahl von extremen Stufen** der UV - bei Altersabhängigkeit einer Variable Junge und Alte untersuchen.
- **Wahl von optimalen Stufen** der UV - setzt Kenntnis der optimalen Stufen voraus; z. B. bei Krafttests hängt das Ergebnis von der Wiederholungszahl ab. Es werden wiederholte Messungen durchgeführt bis die Kraftleistung nachlässt. Das beste Ergebnis wird gewertet.
- **Wahl vieler Stufen** der UV - besonders wichtig bei Variablen, wie z. B. Wohlbefinden, die sich verändern.

KONtrolle der Sekundärvarianz aus den Störvariablen

Störvariablen (SV) sind Variablen, die neben der interessierenden unabhängigen Variablen (UV) einen unerwünschten Einfluss auf die abhängige Variable (AV) ausüben.

- **Eliminierung von Störvariablen** - Ausschaltung von z. B. Lärm; Arbeit im Labor.
- **Konstanthaltung** - Tageszeit, Wochentag, Ort der Untersuchung etc.
- **Umwandlung von Störvariablen in UV** - Auswertung unter Berücksichtigung der Störvariable (z. B. des Geschlechts, Alter, Nationalität etc.; der Umgebungsbedingungen, der Tageszeit etc.).
- **Vermeidung von Konfundierungen** (Vermengung zweier oder mehrerer unabhängiger Variablen, so dass nicht eindeutig gesagt werden kann, wodurch die beobachteten Effekte verursacht wurden)

Beispiel Konfundierung: Bei einem Stretchingexperiment (Wallin et al., 1985) wurde die Verbesserung der Dehnfähigkeit der ischiokruralen Muskelgruppe durch dynamisches bzw. statisches Dehnen überprüft. Die dynamisch trainierende Trainingsgruppe dehnte die ischiokrurale Muskulatur durch Rumpfbeugen. Die statisch trainierende Gruppe dehnte die ischiokrurale Muskelgruppe, indem das zu dehnende Bein im Stehen auf einen Tisch aufgelegt wurde. Das Ergebnis der Untersuchung lautete, dass das statische Dehnen effektiver sei als das statische. Es handelt sich um eine typische Konfundierung zweier unabhängiger Variablen: die Stretchingtechniken (statisch bzw. dynamisch) waren mit verschiedenen Ausführungsformen (Rumpfbeuge, Fuß auf dem Tisch) konfundiert. Wenn man die Frage der Effektivität der Stretchingtechnik beantworten will, muss man die gleiche Übungsausführung wählen.

MINimierung der Fehlervarianz

Keine definierten Kontrolltechniken; Objektivität und Reliabilität der Messverfahren und der Auswertung. Im Einzelnen können folgende Faktoren zu einer unsystematischen Erhöhung der Fehlervarianz beitragen (vgl. Sarris, 1992, S. 147):

- **Zeiteinflüsse (history):** Was geschah zwischen Vor- und Nachtest? (z. B. Ereignisse des 11. September auf die allgemeine Ängstlichkeit einer Population)
- **Entwicklung (maturation):** Vor allem bei längeren Experimenten von Bedeutung (z. B. Einstellung und Verhalten von Jugendlichen in der Pubertät; siehe hierzu die Ausführungen zu entwicklungspsychologischen Untersuchungen weiter unten)
- **Testeffekte (test sophistication):** Einflüsse des Vortestes auf den zweiten Test (Lern- und Trainingseffekte). Alleine schon durch die Durchführung des Vortests werden unter Umständen Einstellungs- und Verhaltensänderungen hervorgerufen, die wiederum einen Einfluss auf den Nachtest haben können (siehe Solomon-Vier-Gruppenplan, Arbeitsblatt 2.3).
- **Veränderung der Messinstrumente (instrumentation):** Versuchsleiterwechsel und mangelnde Reliabilität des Messverfahrens.
- **Statistische Regression (regression):** Boden- und Deckeneffekte: Pbn. mit sehr guten Leistungen, z. B. Leistungssportler, können sich kaum noch verbessern, während sich Pbn. mit einer schlechten Leistungsfähigkeit, z. B. Rehabilitanden, fast nur noch verbessern können.
- **Auswahlverzerrung (selection):** Vpn. unterscheiden sich hinsichtlich wichtiger Kriterien vor allem bei nicht-randomisierter Auswahl. Bei Untersuchungen zur Motorik muss das Geschlechterverhältnis ausgewogen sein und die sportliche Aktivität muss berücksichtigt werden, da diese beiden Variablen einen direkten Einfluss auf die motorische Leistungsfähigkeit haben (siehe Parallelisierung, Arbeitsblatt 2.3).
- **Experimentelle Einbußen (mortality):** Testmortalität (Ausscheiden von Vpn. durch Verletzungen). Je länger eine Untersuchung dauert und mit mehr Aufwand eine Untersuchung für die Pbn. verbunden ist, umso größer ist die Gefahr, dass Pbn. aus einer Untersuchung aussteigen. Die Testmortalität sollte deshalb bei der Untersuchungsplanung abgeschätzt und die Stichprobengröße entsprechend angepasst werden.
- **Versuchsleitereffekte (experimenter-bias effects):** Verhalten und Eigenschaften der Versuchsleiter. Optimal ist es, wenn wie in Doppelblindexperimenten die Versuchsleiter nicht wissen, welcher Pbn. welches Treatment bekommt.
- **Interaktive Effekte (interactive-effects):** Verschiedene experimentelle Bedingungen führen zu Carry-Over-Effekten.

Externe Validität

Eine Untersuchung ist **extern valide**, wenn ihre Ergebnisse über die untersuchten Personen (Populationsvalidität), über die besonderen Bedingungen der Untersuchungssituation (Situationsvalidität) und über den Zeitpunkt der Untersuchung (zeitliche Validität) hinausgehend generalisiert werden können. Störfaktoren der externen Validität sind Wechselwirkungen zwischen der UV und den Versuchspersonen, den Messzeitpunkten und den sonstigen situativen Bedingungen.

Populationsvalidität (Wechselwirkung zwischen Personen-Selektion und UV). Ein Kernproblem bei vielen sportwissenschaftlichen Untersuchungen besteht darin, dass für die Untersuchungen Sportstudenten als Pbn. herangezogen werden. Es stellt sich die Frage, ob die bei jungen, gesunden und hochleistungsfähigen Menschen gefundenen Ergebnisse auf andere Populationen (ältere, kranke und wenig leistungsfähige Menschen) übertragen werden können.

Zeitliche Validität in Abhängigkeit von besonderen Ereignissen in der Zeit (Konfundierung Alter und Generation bei Entwicklungsstudien)

Beispiel - Entwicklungspsychologie: Bei der Analyse von altersabhängigen Entwicklungsverläufen ist zu beachten, dass es methodisch so gut wie unmöglich ist, Menschen entlang ihrer gesamten Lebensspanne zu untersuchen. Die Ergebnisse entsprechender Untersuchungen (siehe Abbildung 1) stammen zumeist aus Querschnittsuntersuchungen. Bei solchen Daten spielt aber nicht nur das biologische Lebensalter eine Rolle, sondern auch die Zugehörigkeit zu einer bestimmten Generation, d. h. wann die entsprechenden Individuen geboren worden sind. So haben die generationstypischen Bedingungen (Einflüsse von Kriegen, Ernährungsbedingungen, Schulsystem etc.) einen Einfluss auf den Verlauf der entsprechenden individuellen Kennlinien. Das biologische Alter ist also mit der Zugehörigkeit zu einer bestimmten Generation (Kohorte) konfundiert

Man muss deshalb versuchen, die Alternseffekte und Kohorteneffekte durch die Kombination von Querschnitts- und Längsschnittuntersuchungen voneinander zu trennen.

Um eine Aussage über die Entwicklung einer altersabhängigen Variablen, wie z. B. der körperlichen Leistungsfähigkeit zwischen dem 5 und 20 Lebensjahr, machen zu können, müssen Querschnitts- und Längsschnittuntersuchungen kombiniert werden (siehe Abb. 2). In den Jahren 1970 und 1975 wurden jeweils Querschnittsuntersuchungen durchgeführt, wobei die jeweils 5-, 10-, 15- und 20-Jährigen erfasst wurden. Des Weiteren werden die Entwicklungsverläufe der Kohorten der 1965 bzw. 1970 geborenen über Längsschnittanalysen bis zum 20. Lebensjahr erfasst. Werden mehrere Querschnittsuntersuchungen zeitlich hintereinander geschaltet, können Menschen eines bestimmten Lebensalters zu verschiedenen Zeitpunkten analysiert werden (Zeitwandelmethode).

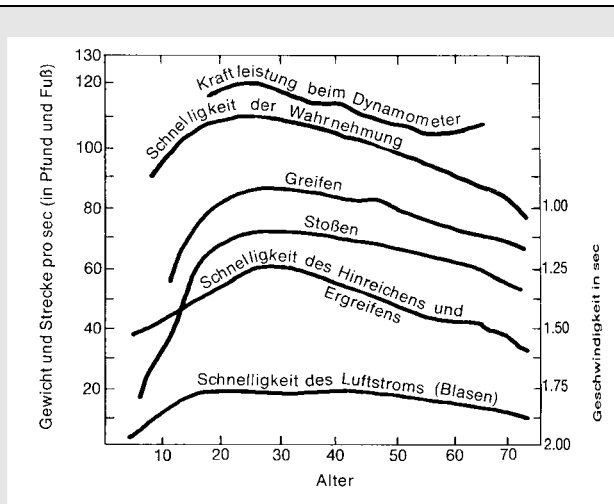


Abbildung 1: Altersabhängige Veränderung einiger motorischer und körperlicher Fähigkeiten.

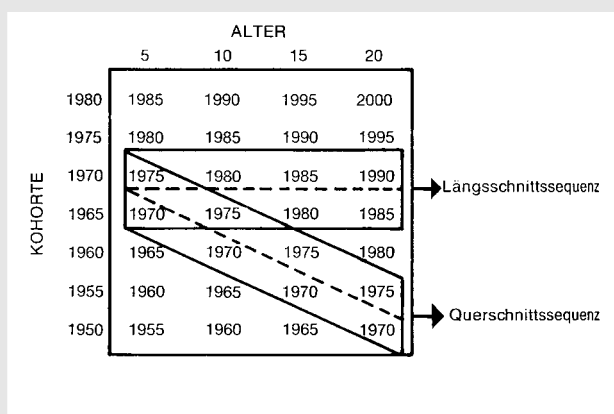


Abbildung 2: Querschnitts- und Längsschnittsequenzen

Situationsvalidität

- Interaktion Situation und UV
 - Placebo-, Neuigkeits- und Hawthorne-Effekt
 - Versuchsleitereffekt
 - Ökologische Validität (Natürlichkeit der Untersuchungsbedingung)

Die externe Validität sinkt mit wachsender Unnatürlichkeit der Untersuchungsbedingungen bzw. mit abnehmender Repräsentativität der untersuchten Stichproben. Bei Experimenten erhöht sich die externe Validität, wenn sie nicht im Labor, sondern im Feld durchgeführt werden (siehe Abbildung 3).

	Quasiexperiment	Experiment
Labor	Geringe externe Validität Geringe interne Validität	Geringe externe Validität Hohe interne Validität
Feld	Hohe externe Validität Geringe interne Validität	Hohe externe Validität Hohe interne Validität

Abbildung 3: Kombination der Untersuchungsvarianten „experimentell vs. quasiexperimentell“ und „Felduntersuchung vs. Laboruntersuchung“ im Hinblick auf die interne bzw. externe Validität.

Die beste Möglichkeit zur Verbesserung der externen Validität besteht darin, Untersuchungen unter anderen Bedingungen mit anderen Versuchspersonen zu replizieren!

Literatur:

- Bös, K., Hänsel, F. & Schott, N. (2000). *Empirische Untersuchungen in der Sportwissenschaft. Planung - Auswertung - Statistik*. Hamburg: Czwalina.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler* (2. erweiterte Aufl.). Berlin: Springer.
- Jacobs, B. (1998). *Einführung in die Versuchsplanung* (Version 1.0). Medienzentrum der Philosophischen Fakultät der Universität des Saarlandes: Zugriff am 28. November 2001: <http://www.phil.uni-sb.de/~jakobs/seminar/vpl/index.htm>
- Ludwig-Mayerhofer, W (2003). ILMES - Internet-Lexikon der Methoden der empirischen Sozialforschung. Zugriff am 2.12.2003 unter <http://www.lrz-muenchen.de/~wlm/ilmes.htm>
- Sarris, V. (1992). *Methodologische Grundlagen der Experimentalpsychologie. 2: Versuchsplanung und Stadien*. München: Reinhardt.
- Wallin, D., Ekblom, B., Grahn, R. & Nordenbrog, T. (1985). Improvement of muscle flexibility: A comparison between two techniques. *The American Journal of Sports Medicine*, 13, 263 - 268.