

# Statistik für Sozialwissenschaftler

Bortz, J. (1999): Statistik für Sozialwissenschaftler. Berlin/Heidelberg/...: Springer

## Einleitung: Empirische Forschung und Statistik

Die **Inferenzstatistik** ermöglicht im Unterschied zur **deskriptiven Statistik** die Überprüfung von Hypothesen.

Für den sinnvollen Einsatz der Inferenzstatistik ist es erforderlich, daß vor Untersuchungsbeginn eine theoretisch gut begründete Hypothese oder Fragestellung formuliert wurde.

### Phasen der empirischen Forschung

- ◆ **Erkundungsphase.** Einordnung der eigenen Fragestellung in einen theoretischen Rahmen. Anknüpfen an bewährte Theorien oder wissenschaftliches Neuland?
- ◆ **Theoretische Phase.** Vor der ›Prüfung‹ einer Theorie, bzw. daraus abgeleiteter Hypothesen, ist die Theorie auf ihre Entwickeltheit und den Grad der Formalisierung zu überprüfen (was auf die wenigsten sozialwissenschaftlichen und psychologischen Theorien zutrefte). Fragen an die Theorie: Präzise formuliert? Innere logische Konsistenz? Informationsgehalt? Logische Vereinbarkeit mit anderen Theorien? Empirische Überprüfbarkeit der Theorie?
- ◆ **Präzision.** Die verwendeten Begriffe müssen mindestens so präzise und verständlich sein, dass diejenigen mit Kenntnis der (Fach-)Sprache sie verstehen können.
- ◆ **Informationsgehalt.** Erkundung der logischen Struktur der Aussagen der Theorie: ›Wenn-dann‹-, bzw. ›Je-desto‹-Sätze (*Konditionalsätze*). »Der Informationsgehalt eines Wenn-dann-Satzes (entsprechendes gilt für Je-desto-Sätze) nimmt zu, je mehr Ereignisse denkbar sind, die mit der Aussage des Dann-Teils in Widerspruch stehen. Ereignisse, die mit dem Dann-Teil der Aussage in Widerspruch stehen, werden als *potentielle Falsifikatoren* bezeichnet.« (5) Die Verwendung unpräziser Begriffe kann den Informationsgehalt eines Satzes einschränken, da die Verwendung unscharfer Begriffe die Eindeutigkeit bei der Einordnung eines Ereignisses herabsetzt.
- ◆ **Logische Konsistenz.** »Führt die logische Überprüfung einer theoretischen Aussage zu dem Ergebnis, daß diese entweder immer wahr, oder immer falsch ist, so ist die entsprechende Aussage logisch inkonsistent. Wir bezeichnen derartige Aussagen als *tautologische* bzw. *kontradiktorische Sätze*. Ein tautologischer Satz besitzt keine potentiellen Falsifikatoren, ein kontradiktorischer Satz besitzt keine potentiellen Konfirmatoren.« (5) »Neben tautologischen und kontradiktorischen Sätzen gibt es Sätze, die deshalb unwissenschaftlich sind, weil sie aus anderen Sätzen *logisch falsch abgeleitet* worden sind.« (6) (Diese Fragestellung fällt in den Bereich der formalen Logik.)
- ◆ **Logische Vereinbarkeit.** Bei gleicher logischer Konsistenz, aber widersprüchlichen Geltungsansprüchen ist die Theorie vorzuziehen, die empirisch besser abgesichert ist. »*Der Wahrheitsgehalt einer Theorie kann nur durch empirische Überprüfungen ermittelt werden.*« (6)
- ◆ **Empirische Überprüfbarkeit.** »Die Forderung nach empirischer Überprüfbarkeit einer Theorie ist eng an die Forderung nach ihrer Falsifizierbarkeit gekoppelt. Es sind dennoch Aussagen denkbar, die zwar im Prinzip falsifizierbar, aber (noch) nicht empirisch überprüfbar sind.« (6) Beispiel: ›Verdrängte‹ Aggressivität sei (noch) nicht messbar, weil nicht nachweisbar.
- ◆ **Planungsphase.** Vor dem Beginn der Untersuchung soll der Aufbau und der Ablauf der Untersuchung geplant werden.
- ◆ **Auswahl der Variablen.** *Variablen* können, im Unterschied zu *Konstanten*, in mindestens zwei Ausprägungen vorkommen. »Unter den *unabhängigen Variablen* werden diejenigen Merkmale verstanden, deren Auswirkungen auf andere Merkmale – die *abhängigen Variablen* – überprüft werden. Im allgemeinen ist bereits aufgrund der Fragestellung festgelegt, welche der relevanten Variablen als abhängige und welche als unabhängige Variablen in die Untersuchung eingehen sollen.« (7f). Darüber hinaus werden noch *Kon-*

*trollvariablen* erhoben, diese konstant gehalten oder überhaupt nicht berücksichtigt werden sollen. Potentiell bedeutsame, aber nicht erhobene Variablen werden *Störvariablen* genannt.

## Labor- oder Felduntersuchung

»Diese Untersuchungsvarianten markieren die Extreme eines Kontinuums, das durch eine unterschiedlich starke *Kontrolle untersuchungsbedingter Störvariablen* gekennzeichnet ist.« (8, Herv. J.G.) Laboruntersuchungen sind i.d.R. eher *intern valide*, während Felduntersuchungen eher *extern valide* sind.

Eine Untersuchung ist **intern valide**, wenn ihr Ergebnis eindeutig interpretierbar ist. Die interne Validität sinkt mit wachsender Anzahl plausibler Alternativerklärungen für das Ergebnis auf Grund nicht kontrollierter Störvariablen.

Eine Untersuchung ist **extern valide**, wenn ihr Ergebnis über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen hinausgehend generalisierbar ist. Die externe Validität sinkt mit wachsender Unnatürlichkeit der Untersuchungsbedingungen bzw. mit abnehmender Repräsentativität der untersuchten Stichproben.

## Experimentelle oder quasiexperimentelle Untersuchung

»Während das Kontinuum Labor vs. Feld das Ausmaß der Kontrolle untersuchungsbedingter Störvariablen beschreibt, kennzeichnet die Unterscheidung von experimenteller und quasiexperimenteller Untersuchung das Ausmaß der *Kontrolle von personenbedingten Störvariablen*.« (9, Herv. J.G.) In experimentellen Untersuchungen ist die ›Gleichausprägung‹ der Störvariablen durch *Randomisierung* zu erreichen.

Unter **Randomisierung** versteht man die zufällige Zuordnung der Untersuchungsteilnehmer zu den Untersuchungsbedingungen.

»Bei einer quasiexperimentellen Untersuchung muß auf eine Randomisierung verzichtet werden, da hier ›natürliche‹ bzw. bereits existierende Gruppierungen untersucht werden.« (9)

Experimentelle Untersuchungen haben eine höhere interne Validität als quasiexperimentelle Untersuchungen.

»Die interne Validität einer quasiexperimentellen Untersuchung läßt sich jedoch erhöhen, wenn es gelingt, die zu vergleichenden Gruppen nach relevanten Störvariablen zu parallelisieren.« (10)

## Operationalisierung

Nach der Festlegung, welche Variablen erfaßt werden sollen, geht es nun darum zu bestimmen, *wie* die Variablen erfaßt werden sollen. »Durch die Operationalisierung wird festgelegt, welche Operationen (Handlungen, Reaktionen, Zustände usw.) wir als indikativ für die zu messenden Variable ansehen wollen und wie diese Operationen quantitativ erfaßt werden. (10) Das *Wie* und *Ob überhaupt* sind ganz, ganz schwierige Fragen...

## Stichprobengröße

Die Frage nach der Anzahl benötigter ›Versuchspersonen‹ ist am direktesten mit: ›So viele wie möglich.‹ zu beantworten. Für genauere Angaben muss der Kontext der Untersuchung berücksichtigt werden (Wie groß ist der Effekt, der vom Untersuchenden für praktisch bedeutsam gehalten wird? Was für Folgen hätte eine Fehleinschätzung?)

## Planung der statistischen Auswertung

»Die Planungsphase endet mit *Überlegungen zur statistischen Auswertung des Untersuchungsmaterials*. Es müssen diejenigen statistischen Auswertungstechniken bestimmt werden, die es in optimaler Weise gestatten, auf Grund der zu erhebenden Daten über die Brauchbarkeit der Hypothese zu entscheiden.« (11) »*Oberstes Ziel sollte es sein, eine Untersuchung möglichst nach inhaltlichen Kriterien zu gestalten, ohne jedoch auf die Präzision in der statistischen Auswertung zu verzichten,*« (11)

## Untersuchungsphase

Bei genauer Planung seien keine Problem zu erwarten – außer evtl. Versuchsleiter-(VI)-Effekte.

## Auswertungsphase

Zunächst wird sich ein Überblick über die *testtheoretische Brauchbarkeit der Daten* verschafft (u.a. Quantifizierbarkeit). Zudem muss im Sinne der *Objektivität* überprüft werden, »ob verschiedene Auswerter Vpn auf Grund der Untersuchungsergebnisse die gleichen Zahlenwerte zuordnen.« (11) Problematisch wird diese Frage i.d.R. bei wenig standardisierten Verfahren mit ungenauen Operationalisierungen. Stellt sich das Problem als gravierend da, ist die Untersuchung abzubrechen. In größer angelegten Untersuchungen ist zudem noch die *Reliabilität* – Meßgenauigkeit als technisches Gütekriterium – zu überprüfen.

Werden die *Daten* als brauchbar angesehen, so werden sie *tabellarisch zusammengestellt* und je nach Ziel der Untersuchung weiterverarbeitet. War das Ziel die Überprüfung einer Hypothese, so wird *mittels inferenzstatistischer Verfahren* bzw. eines »Signifikanztests« eine *Entscheidung* über die zu prüfende Hypothese festgelegt. »Hierzu errechnet man eine sog. *Irrtumswahrscheinlichkeit p*, die angibt, mit welcher Wahrscheinlichkeit man sich irren würde, wenn man die fragliche Hypothese akzeptiert. Um die Hypothese annehmen zu können, sollte diese Irrtumswahrscheinlichkeit natürlich möglichst klein sein.

Die Größe der maximal tolerierbaren Irrtumswahrscheinlichkeit liegt allerdings nicht im Ermessen des Untersuchenden, sondern ist durch eine allgemeingültige Konvention festgelegt. Man bezeichnet diese Grenze, die von der Irrtumswahrscheinlichkeit  $p$  nicht überschritten werden darf, als »*Signifikanzniveau*« und verwendet hierfür das *Symbol*  $\alpha$ . Die üblichen Werte für das Signifikanzniveau sind  $\alpha = 5\%$  oder gar  $\alpha = 1\%$ . Der Untersuchende muß vor Durchführung der Untersuchung des Signifikanztests festlegen, welches  $\alpha$ -Niveau er für die Untersuchung für angemessen hält.

Die Auswertungsphase endet mit der Bestimmung der Irrtumswahrscheinlichkeit für eine Entscheidung zugunsten der getesteten Hypothese unter Zuhilfenahme des in der Planungsphase festgelegten Signifikanztests.« (12, Herv. J.G.)

## Entscheidungsphase

»Ein Vergleich der ermittelten Irrtumswahrscheinlichkeit  $p$  mit dem zuvor festgelegten Signifikanzniveau  $\alpha$  zeigt, ob das Ergebnis der Untersuchung signifikant ( $p < \alpha$ ) oder nicht signifikant ( $p > \alpha$ ) ist.« (12)

Bei einem **nicht signifikanten** Ergebnis gilt die Hypothese nicht als bestätigt, der/die ForscherIn ist jedoch dazu berechtigt zu überprüfen, inwieweit eventuell im *Verlaufe der Untersuchung Fehler* aufgetreten sind, die für das Ergebnis verantwortlich sind. Dann darf der Versuch noch einmal wiederholt werden.

»Problematischer ist ein nicht signifikantes Ergebnis, wenn Untersuchungsfehler praktisch ausgeschlossen werden können. Ist der deduktive Schluß von der Theorie auf die überprüfte Hypothese korrekt, muß an der allgemeinen Gültigkeit der Theorie gezweifelt werden.« (13) Konsequenterweise muß die *Theorie dann verändert werden*. Die Erweiterung der Wenn-Komponente um einen weiteren Teil, der das Ergebnis der Prüfung plausibel macht, wird *Exhaustion* genannt. Dabei stellt sich natürlich die Frage, wieweit eine Theorie mit derartigen Erweiterungen belastbar ist, bevor sie zu verwerfen ist (bzw. sie in ihrem Geltungsbereich so weit eingeschränkt ist, dass sie praktisch nutzlos ist).

Bei einem **signifikanten** Ergebnis kann sich der bzw. die UntersucherIn ziemlich (5% bzw. 1%) sicher sein, dass die Hypothese zutrifft. »Dennoch ist es eine Konvention, die geprüfte Hypothese in diesem Falle als bestätigt anzusehen.« (13) »Hinsichtlich der Theorie besagt eine durch ein signifikantes Ergebnis bestätigte Hypothese, daß wir keinen Grund haben, an der Richtigkeit der Theorie zu zweifeln, sondern daß wir viel mehr der Theorie nach der Untersuchung her trauen sollten als vor der Untersuchung.« (14) Die Theorie kann jedoch auch durch noch so viele Untersuchungen nicht bestätigt, sondern nur falsifiziert werden.

»So ist die empirische Basis der objektiven Wissenschaft nichts »Absolutes«; die Wissenschaft baut nicht auf Felsengrund. Es ist eher ein Sumpfland, über dem sich die kühne Konstruktion ihrer Theorien erhebt; sie ist ein Pfeilerbau, dessen Pfeiler sich von oben her in den Sumpf senken – aber nicht bis zu einem natürlichen »gegebenen« Grund. Denn nicht deshalb hört man auf, die Pfeiler tiefer hineinzutreiben, weil man auf eine feste Schicht gestoßen ist: Wenn man hofft, daß sie das Gebäude tragen werden, beschließt man, sich vorläufig mit der Festigkeit der Pfeiler zu begnügen« (POPPER, 1966; S. 75f.)« (14).