

## Methoden der Sportwissenschaft 2

### Arbeitsblatt 2.7: $\alpha$ - und $\beta$ -Fehler (aktualisiert 4. Januar 2007)

#### $\alpha$ - und $\beta$ -Fehler

Bei jeder (statistischen) Prüfung einer Hypothese können zwei verschiedene Fehler gemacht werden. Drei Beispiele aus dem Alltag mögen dieses Dilemma verdeutlichen:

**Beispiel 1:** Wenn ein Richter den Grundsatz des römischen Rechts „in dubio pro reo“ (=Nullhypothese) nicht Ernst genug nimmt, und einen Angeklagten nur auf der Basis von Indizien oder zweifelhaften Zeugenaussagen oder gar wegen seiner Hautfarbe verurteilt, kommt es zu einem Justizirrtum (Fehler 1. Art oder  $\alpha$ -Fehler). Solche Justizirrtümer müssen auf jeden Fall vermieden werden. Deshalb darf ein Richter einen Angeklagten nur verurteilen, wenn ausreichend viele Belege für die Schuld des Angeklagten vorliegen, d. h. die Wahrscheinlichkeit sich zu irren (Irrtumswahrscheinlichkeit = Signifikanzniveau) wird sehr tief angesetzt. Dadurch läuft man aber Gefahr, dass der eine oder andere Täter nicht verurteilt wird (Fehler 2. Art oder  $\beta$ -Fehler).

		Urteil	
		Ablehnung	Annahme
		der Unschuldsvermutung	
Angeklagter	Unschuldig	Justizirrtum 1. Art	✓
	Schuldig	✓	Justizirrtum 2. Art

Abbildung 1: In dubio pro reo.

**Beispiel 2:** Ein Feuermelder muss auf jeden Fall im Brandfall Alarm schlagen. Tut er dies nicht, kommt es eventuell zu Todesfällen. Das muss vermieden werden. Ein Fehlalarm ist nicht so schlimm. Die Feuerwehr sagt selbst, dass sie lieber einmal unnötig ausrückt, als dass sie einmal zu spät kommt.

		Alarm	
		Ablehnung	Annahme
		Kein Alarm= Fehler 1. Art	✓
Es brennt	stimmt		
	stimmt nicht	✓	Fehlalarm = Fehler 2. Art

Abbildung 2: Ausgangshypothese: Wenn der Feuermelder Alarm schlägt, brennt es.

**Beispiel 3:** Es ist nicht einfach, den Mann oder die Frau fürs Leben zu finden. Deshalb sagt der Volksbund, man solle prüfen, bevor man sich ewig bindet. Die Ausgangshypothese lautet: Kandidat(in) ist nicht Mann oder Frau fürs Leben. Mann oder Frau sollte versuchen, nicht an die Falsche oder den Falschen zu geraten. Das wäre ein Fehler 1. Art oder  $\alpha$ -Fehler. Durch die Erhöhung des Anspruchsniveaus (Prüfgröße) kann diese Gefahr reduziert werden. Dadurch steigt aber auch die Gefahr, dass der richtige Kandidat oder die richtige Kandidatin nicht erwählt wird ( $\beta$ -Fehler).

		Partnerwahl Entscheidung	
		Ablehnung	Annahme
		Lebensfehler 1. Art	✓
Kandidat ist nicht Mann oder Frau fürs Leben	stimmt		
	stimmt nicht	✓	Lebensfehler 2. Art

Abbildung 3: Wer sich ewig bindet.

**Beispiel 4:** Medikamente sollten Beschwerden lindern. Viele Medikamente haben Schwierigkeiten ihre Evidenz nachzuweisen. Aber oftmals versetzt ja allein schon der Glaube Berge (Placebo-Effekte). Sie dürfen aber auf keinen

Fall zu schwerwiegenden, eventuell sogar tödlichen Komplikationen führen. Nebenwirkungen haben alle Medikamente. Viele Mediziner sagen auch, dass Medikamente, die wirken sollen, auch Nebenwirkungen haben müssen. Mit anderen Worten: Medikamente ohne Nebenwirkungen haben in der Regel auch keine Wirkung. Konsequenter wäre es deshalb, wenn man das Auftreten von Komplikationen als Fehler 1. Art formulieren würde.

Aber hier wird schon deutlich, dass man eine Kosten-Nutzen-Abwägung vornehmen muss. Wenn es keine andere Möglichkeit zur Rettung eines Menschenleben gibt als ein bestimmtes Medikament einzusetzen, das aber in 10 % der Fälle tödliche Nebenwirkungen hat, wird man es einsetzen müssen.

Genauso ist es in der Wissenschaft. Im Allgemeinen werden hochsignifikante Ergebnisse als wissenschaftlich bedeutsamer eingestuft als signifikante Ergebnisse. Hierbei ist aber zu berücksichtigen, dass sich durch die Verringerung des Signifikanzniveaus gleichzeitig die Gefahr erhöht, dass man zu Unrecht die (falsche) Nullhypothese beibehält. Zur Charakterisierung dieser Wechselwirkung werden die Begriffe  $\alpha$ - und  $\beta$ -Fehler bzw. Fehler 1. und 2. Art gebraucht (siehe Abbildung 1).

Der  $\alpha$ -Fehler besagt, dass fälschlicherweise die Nullhypothese abgelehnt wird. Er wird durch das Signifikanzniveau bestimmt. Bei einem 5%igen Signifikanzniveau nimmt man demnach in 5 Prozent der Fälle einen Fehler in Kauf.

Der  $\beta$ -Fehler besagt, dass zu Unrecht die Nullhypothese beibehalten wird, obwohl in Wirklichkeit die Alternativhypothese zutrifft.

		<b>Entscheidung aufgrund der Stichprobe</b>	
		Ablehnung	Annahme
		der Nullhypothese	
<b>Nullhypothese</b>	Richtig	$\alpha$ -Fehler	Richtig
	Falsch	Richtig	$\beta$ -Fehler

**Abbildung 4: Allgemeines Entscheidungsdiagramm, ob ein  $\alpha$ - oder  $\beta$ -Fehler vorliegt.**

Die beiden Fehlerarten bedingen sich gegenseitig. Die Gefahr des Auftretens eines  $\alpha$ - oder  $\beta$ -Fehlers hängt hauptsächlich ab vom Signifikanzniveau  $\alpha$ . Durch eine Verringerung des Signifikanzniveaus  $\alpha$  (= Gefahr eines  $\alpha$ -Fehlers) von beispielsweise 5 % auf 1 % reduziert sich die Gefahr, dass die richtige Nullhypothese abgelehnt wird ( $\alpha$ -Fehler), d. h. ein nicht existierender Unterschied oder Zusammenhang werden nicht so schnell als signifikant eingestuft. Gleichzeitig aber erhöht sich die Gefahr eines  $\beta$ -Fehlers.

Umgekehrt steigt durch eine Erhöhung des Signifikanzniveaus  $\alpha$  von 5 auf 10 % die Gefahr, dass eine falsche Nullhypothese akzeptiert wird ( $\alpha$ -Fehler). Dadurch reduziert sich aber wiederum zwangsläufig die Gefahr des  $\beta$ -Fehlers.

Maßnahme		Fehler 1. Art		Fehler 2. Art
Verringerung von $\alpha$ von 5 % auf 1 %	⇒	$\alpha$ -Fehler ↓	⇒	$\beta$ -Fehler ↑
Erhöhung von $\alpha$ von 5 % auf 10 %	⇒	$\alpha$ -Fehler ↑	⇒	$\beta$ -Fehler ↓

**Abbildung 5: Auswirkungen der Veränderung des Signifikanzniveaus auf das Auftreten von  $\alpha$ - oder  $\beta$ -Fehlern. ↓ = Reduktion; ↑ = Erhöhung.**

Wenn man die Gefahr eines  $\alpha$ -Fehlers reduziert, vergrößert man automatisch die Gefahr eines  $\beta$ -Fehlers und umgekehrt. Man geht im Allgemeinen davon aus, dass die Folgen eines Fehler 1. Art viermal so gravierend sind wie die Folgen eines Fehler 2. Art (Bortz & Döring, 1995, S. 567). Während die Scientific Community für den  $\alpha$ -Fehler einen Wert von 0,05 festgelegt hat, erachtet man für den  $\beta$ -Fehler einen Wert von 0,2 als akzeptabel. Deshalb muss letztendlich neben der statistischen Bewertung immer auch eine Güterabwägung vorgenommen werden. Die anfangs gegebenen Beispiele sollten diese grundsätzliche Aufgabe illustrieren.

**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.

Gigerenzer, G. (2002). *Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken*. Berlin: Berlin Verlag.

Jacobs, B. (1998). *Einführung in die Versuchsplanung* (Version 1.0). Internetauszug vom 28. November 2001, Homepage von B. Jacobs, Medienzentrum der Philosophischen Fakultät der Universität des Saarlandes: <http://www.phil.uni-sb.de/~jakobs/seminar/vpl/index.htm>