

α - und β -Fehler, Macht eines Tests

MÖGLICHE FEHLER BEI SIGNIFIKANZPRÜFUNGEN

Bei den Tests, die in diesem Kapitel behandelt wurden, werden stets zwei Hypothesen (H_0 und H_1) aufgestellt, von denen am Ende des Tests eine verworfen und die zweite angenommen wird. Die Testergebnisse sind dabei immer einer gewissen Irrtumswahrscheinlichkeit unterworfen, d.h. es handelt sich um statistisch begründete Vermutungen, nicht um Tatsachen.

Es gibt zwei Möglichkeiten, welche Hypothese *stimmt*, und ebenso zwei, welche aufgrund des Testergebnisses *für richtig gehalten* wird. Also ergeben sich 4 denkbare Kombinationen:

		Testergebnis	
		H_0	H_1
Wirklichkeit	H_0	✓	F.1.Art
	H_1	F.2.Art	✓

FALLS DIE H_0 STIMMT,

gibt es zwei Möglichkeiten:

Fehler 1. Art, „ α -Fehler“

Die fälschliche Verwerfung der Nullhypothese (zugunsten der H_1) nennt man *Fehler 1. Art*. Er passiert, falls die H_0 gilt, mit Wahrscheinlichkeit α , weswegen man ihn auch als *α -Fehler* bezeichnet.

H_0 richtig beibehalten

Die Wahrscheinlichkeit, dass der Test erkennt, dass die Nullhypothese stimmt und die Alternativhypothese zu verwerfen ist, ist demnach $1 - \alpha$.

FALLS DIE H_0 ZUGUNSTEN DER H_1 ZU VERWERFEN IST,

gibt es ebenfalls zwei denkbare Ergebnisse:

Fehler 2. Art, „ β -Fehler“

Unter einem *Fehler 2. Art* versteht man die Entscheidung, H_0 beizubehalten, obwohl eigentlich die H_1 stimmt. Dieser Fehler, der sog. *β -Fehler*, tritt - falls H_1 stimmt - mit der Wahrscheinlichkeit β auf.

H_0 richtig verworfen

Der Test erkennt mit einer Wahrscheinlichkeit von $1 - \beta$ eine falsche Nullhypothese, d.h. die Wahrscheinlichkeit, richtigerweise signifikant zu werden, ist $1 - \beta$.

Zur Veranschaulichung wurden hier die Wahrscheinlichkeiten in obige Tabelle eingetragen:

		Testergebnis	
		H_0	H_1
Wirklichkeit	H_0	$1 - \alpha$	α
	H_1	β	$1 - \beta$

MACHT EINES TESTS

Als Testmacht wird die Wahrscheinlichkeit bezeichnet, mit der ein Test eine falsche Nullhypothese entlarvt, d.h. wie wahrscheinlich das Ergebnis signifikant ausfällt, falls H_1 gilt. Dies geschieht (siehe obige Tabelle) mit Wahrscheinlichkeit $1 - \beta$.

Die Macht eines Tests hängt unmittelbar mit dem Signifikanzniveau zusammen: Je größer α , umso größer auch $1 - \beta$ und umgekehrt. Man muss also einen geeigneten Kompromiss zwischen möglichst großer Testmacht einerseits, damit möglichst viele falsche Nullhypothesen verworfen werden können, und möglichst kleiner Irrtumswahrscheinlichkeit andererseits finden. In der Psychologie hat sich $\alpha = 5\%$ als übliches Signifikanzniveau durchgesetzt.

EIN BEISPIEL ZUM BESSEREN VERSTÄNDNIS

Ein Schüler muss für eine schriftliche Prüfung die Antworten auf 1000 (gleich schwierige) Fragen lernen. Beim Test werden aus Zeitgründen aber nur 10 Fragen vorgegeben. Der Lehrer wünscht sich, dass der Schüler, falls er „genug“ (mindestens 500 Fragen) gelernt hat, durchkommt und dass er, falls er zu wenig gelernt hat (also höchstens 499 Fragen kann) den Test nicht besteht.

Hypothesen

H_0 ... Der Schüler hat genug gelernt (≥ 500 Fragen).

H_1 ... Der Schüler hat nicht genug gelernt (< 500 Fragen).

Design des Tests

Der Lehrer überlegt vor der Prüfung, wie streng er bewerten soll. Er hat vor, den Schüler positiv zu benoten, falls dieser mindestens 5 der 10 Fragen richtig beantwortet.

Welche Fehler können ihm bei dieser Bewertung unterlaufen?

α -Fehler

Auch wenn der Schüler eigentlich genug gelernt hat, besteht - sofern er nicht mindestens 995 der 1000 Fragen beantworten kann - die Gefahr, dass zum Test 6 oder mehr Fragen kommen, die er nicht gelernt hat. In diesem Fall würde die H_0 zu unrecht verworfen.

β -Fehler

Der Schüler hat, sofern er wenigstens 5 Fragen gelernt hat, eine Chance, den Test zu bestehen, auch wenn er nicht die vom Lehrer erwarteten 500 Fragen kann. Tritt dies ein, so nimmt der Lehrer zu unrecht an, dass die H_0 gilt.

Änderung des Designs?

Eine strenge Kollegin weist auf die Gefahr hin, dass der Schüler im Falle eines β -Fehlers für seine Faulheit auch noch belohnt werden würde. Das ginge doch wirklich nicht, meint sie, und will den Lehrer überzeugen, die Grenze für eine positive Note auf 8 richtige Fragen anzuheben (d.h. das Signifikanzniveau α zu erhöhen).

Die Konsequenz wäre eine größere Testmacht, d.h. eine ungültige H_0 hat eine geringere Chance, unentdeckt zu bleiben. Allerdings steigt das Risiko für einen α -Fehler ebenfalls stark an.

Ein junger Kollege meint daraufhin, dass die negativen Folgen eines α -Fehlers für den Schüler wesentlich drastischer wären. Eine durch Glück erlangte gute Note könnte auch Positives bewirken (z.B. gesteigertes Selbstvertrauen), hat der Schüler trotz hohen Lernaufwands ein Misserfolgserlebnis, so würde sich das auf die weitere Lernmotivation hingegen nachhaltig negativ auswirken. Die Grenze sollte nach der Meinung des Junglehrers auf 3 richtig zu beantwortende Fragen gesenkt werden (d.h. man sollte α verringern).

Die Konsequenz wäre ein geringeres Risiko für einen α -Fehler auf Kosten der Testmacht.

Schlussfolgerung

Der Lehrer muss sich entscheiden, welchen Fehler er für schwerwiegender hält und einen dementsprechenden Kompromiss treffen.