

Die Bedeutung des β -Risikos

MANFRED BÖRGENS, WÖLFERSHEIM

Zusammenfassung: Bei der didaktischen Behandlung und der Anwendung von Parametertests wird in der Regel das Hauptaugenmerk auf das α -Risiko (Risiko 1. Art) gelegt. Dieses Risiko ist die Wahrscheinlichkeit, die Nullhypothese H_0 irrtümlich zu verwerfen. Das β -Risiko (Risiko 2. Art) gilt dagegen der irrtümlichen Beibehaltung von H_0 und wird oft vernachlässigt. Dieses Risiko lässt sich jedoch geeignet abschätzen, wenn man die „Effektgröße“ in die Konzeption des Tests mit einbezieht, d. h. den Abstand der Hypothesen H_0 und H_1 zugrunde liegenden Parameter. Damit kann dann im Schulunterricht das Konzept der „Testgüte“ ohne großen Aufwand eingeführt werden.

Für die Umsetzung im Mathematikunterricht der Sekundarstufe 2 werden nur elementare Voraussetzungen benötigt. Wegen der Bedeutung der Normalverteilung für zahlreiche Anwendungen wird hier der Gauß-Test als Einführung für das β -Risiko gewählt. Die Berechnung von β erfolgt beispielhaft anhand des Zwei-Stichproben-Tests, um das Thema auf eine breitere Basis zu stellen – der im Schulunterricht häufiger behandelte Ein-Stichproben-Test ist dann durch eine leichte Modifikation mit abgedeckt. – Als Anwendungsbeispiel dient die Messung der Intelligenz.

1 Einführung: Gründe für die Dominanz des α -Risikos

Fragt eine Schülerin oder ein Schüler nach den Gründen der asymmetrischen Bedeutung, die α - und β -Risiko (für sie oder ihn offenkundig) beigemessen wird, so lauten häufig gegebene Antworten sinngemäß so:

- Das β -Risiko sei *prinzipiell nicht berechenbar*.
- Das α -Risiko stelle das „wichtigere“ der beiden Risiken dar, d. h. ein „falscher Alarm“ habe

schwerwiegendere Konsequenzen als ein „Verfehlen“ (siehe Tafel 1 für einen Parametertest als Beispiel).

Test auf Unterschied zweier Erwartungswerte	Entscheidung	
	H_0	H_1
Hypothese H_0 $\mu_1 = \mu_2$	korrekt	„Falscher Alarm“ Risiko α
Hypothese H_1 $\mu_1 \neq \mu_2$	„Verfehlen“ Risiko β	korrekt

Tafel 1: Richtige Entscheidungen vs. Fehlentscheidungen

Diesen beiden Gründen ist entgegenzuhalten:

- Es ist wahr, dass die Berechnung von β nur unter Angabe zusätzlicher Bedingungen möglich ist; es wird noch gezeigt werden, dass die Vorgabe einer „Effektgröße“ eine solche Bedingung sein kann. Die Angabe des Zahlenpaars (α, β) macht den Test transparenter. In der Testtheorie dient β zur Bestimmung der „Testgüte“ und bietet somit die Möglichkeit eines Gütevergleichs verschiedener Tests. In vielen Anwendungsfällen wird man auch β vorgeben wollen (und nicht berechnen). Es wird gezeigt werden, dass dies mittels der Wahl des Stichprobenumfangs n möglich ist. – Bei der Vermittlung von Parametertests in einem Lehrmodul Statistik/Stochastik ist dies eine einfache und interessante Ergänzung des Lehrstoffes.
- Es hängt ausschließlich vom jeweiligen Anwendungsfall ab, ob das Risiko, Unterschiede festzustellen, wenn in Wirklichkeit keine vorliegen, das kritischere gegenüber dem Risiko ist, vorhandene

ne Unterschiede nicht aufzudecken. Aber selbst wenn die „Kosten“ eines falschen Alarms diejenigen eines Verfehlens übersteigen, sollte nicht auf die wichtige Information verzichtet werden, die durch die Angabe des β -Risikos geliefert wird.

Diese Argumente sind im Kontext der historischen Entwicklung der Stochastik zu sehen. Der ältere Ansatz von R. A. Fisher verzichtet ganz auf die Alternativhypothese. Auch heute noch wird dieser Ansatz verwendet, wenn es lediglich darum geht, ob die erhobenen Daten das Vorliegen eines „Signals“ nahelegen oder nicht. Dabei wird nur das Risiko α vorgegeben, um die Wahrscheinlichkeit von „falschen Alarmen“ (siehe Tafel 1) zu begrenzen. Die Wahrscheinlichkeit β , das Signal nicht zu erkennen, wird nicht in Betracht gezogen. – Dagegen steht der Ansatz von J. Neyman und E. S. Pearson, der vom Vorliegen je einer konkreten Wahrscheinlichkeitsverteilung für die beiden Hypothesen ausgeht. In diesem Fall werden beide Risiken betrachtet, sowohl α für die fälschliche Annahme von H_1 als auch β für die fälschliche Annahme von H_0 . Allerdings sehen wir bei dem Beispiel, das in Tafel 1 gewählt wurde, dass sich dort H_1 nicht mit einer einzigen Verteilung beschreiben lässt, da der Abstand von μ_1 und μ_2 nicht bekannt ist. Mit diesem Problem werden wir uns also noch beschäftigen müssen. Ausführlich befasst sich Lenhard (2006) mit den Ansätzen von Fisher und Neyman/Pearson.

Ein verbreitetes und einfaches Testverfahren – der *Gauß-Test* – und ein zugehöriges Beispiel sollen im folgenden die Einführung des β -Risikos motivieren. Alle Schritte lassen sich mit der elementaren Stochastik nachvollziehen. Der Gauß-Test ist gut geeignet, die Berechnung von β und damit der Testgüte paradigmatisch darzustellen; er weist darüber hinaus zahlreiche Möglichkeiten zur Anwendung und Generalisierung auf.

2 Gauß-Test für zwei unabhängige Stichproben

Dieser Test macht folgende Grundannahmen: Aus zwei Grundgesamtheiten wird je eine zufällige Stichprobe gezogen, mit den Umfängen n_1 bzw. n_2 . In beiden Stichproben wird dasselbe Merkmal gemessen; die zugehörige Zufallsvariable in den Grundgesamtheiten ist jeweils normalverteilt mit gleicher, bekannter Standardabweichung σ . Der Test ist ein Parametertest; er dient dem Vergleich der Erwartungswerte μ_1 und μ_2 der beiden Normalverteilungen, in Kurzform: Ist $N(\mu_1, \sigma) = N(\mu_2, \sigma)$? Es soll eine Entscheidung zwischen zwei Hypothesen H_0 und $H_1 = -H_0$ getroffen werden.

„Nullhypothese“ $H_0 : \mu_1 = \mu_2$

„Alternativhypothese“ $H_1 : \mu_1 \neq \mu_2$

Die Stichprobenmittelwerte \bar{x}_1 und \bar{x}_2 sind

$N\left(\mu_i, \frac{\sigma}{\sqrt{n_i}}\right)$ -verteilt ($i = 1, 2$);

ihre Differenz ist somit $N(\mu_1 - \mu_2, \sigma h)$ -verteilt mit

$h = \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$, dafür ist die Summe der Varianzen zu bilden: $\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \frac{n_1 + n_2}{n_1 n_2}$. Also ist

$$(1) \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma h} N(0, 1)\text{-verteilt}$$

(standardnormalverteilt).

Für das Risiko 1. Art gibt sich der Anwender eine (gerade noch tolerierbare) Wahrscheinlichkeit α vor (Irrtumswahrscheinlichkeit). Mit $\pm z_{\alpha/2}$ werden die Koordinaten bezeichnet, für die gilt:

$$F(-z_{\alpha/2}) = \alpha/2, F(z_{\alpha/2}) = 1 - \alpha/2$$

Dabei ist F die Verteilungsfunktion der Standardnormalverteilung (siehe Tafel 2).

(2) Auf H_1 wird genau dann entschieden, wenn

$$\frac{|x_1 - x_2|}{\sigma h} \geq z_{\alpha/2}.$$

α	$z_{\alpha/2}$
0,001	3,2905
0,005	2,8070
0,01	2,5758
0,05	1,9600
0,1	1,6449

Tafel 2: Quantile der Standardnormalverteilung

Risiko 1. Art (α):

H_0 liegt vor, d. h. $\mu_1 = \mu_2$, aber H_1 wird entschieden (*falscher Alarm*). Wegen (1), (2) gilt dafür die bedingte Wahrscheinlichkeit $\alpha = p(H_1 | \mu_1 = \mu_2)$.

Risiko 2. Art (β):

H_1 liegt vor, d. h. $\mu_1 - \mu_2 = d \neq 0$, aber H_0 wird entschieden (*Verfehlen*). Dann ist

$$\frac{\bar{x}_1 - \bar{x}_2}{\sigma h} \in (-z_{\alpha/2}, z_{\alpha/2})$$

$$\Leftrightarrow \frac{\bar{x}_1 - \bar{x}_2 - d}{\sigma h} \in \left(-z_{\alpha/2} - \frac{d}{\sigma h}, z_{\alpha/2} - \frac{d}{\sigma h}\right)$$

Nach (1) ist $\frac{\bar{x}_1 - \bar{x}_2 - d}{\sigma h} N(0, 1)$ -verteilt.

Wegen $\beta = p(H_0 | \mu_1 - \mu_2 = d)$ folgt:

$$(3) \quad \beta = F\left(z_{\alpha/2} - \frac{d}{\sigma h}\right) - F\left(-z_{\alpha/2} - \frac{d}{\sigma h}\right)$$

3 Die Rolle des β -Risikos

(3) ist die zentrale Formel zur Berechnung des Risikos 2. Art. Mit Hilfe einer Normalverteilungstabelle oder einer Statistik-Software (z. B. EXCEL) ist β leicht zu berechnen. Das entscheidende Hindernis ist aber, dass d in (3) dem Testanwender unbekannt ist. Dennoch lassen sich aus (3) zunächst einige Folgerungen ziehen, die die Lernenden leicht anhand selbstgewählter Beispiele verifizieren können (in EXCEL verwendet man NORMVERT für die Funktion F und NORMINV für $z_{\alpha/2}$):

- β sinkt, wenn α wächst (und umgekehrt)
- β sinkt, wenn $|d|$ wächst (und umgekehrt)
- $\beta < 1 - \alpha$
- $\lim_{d \rightarrow 0} \beta = 1 - \alpha$

Bild 1 veranschaulicht diese Zusammenhänge.

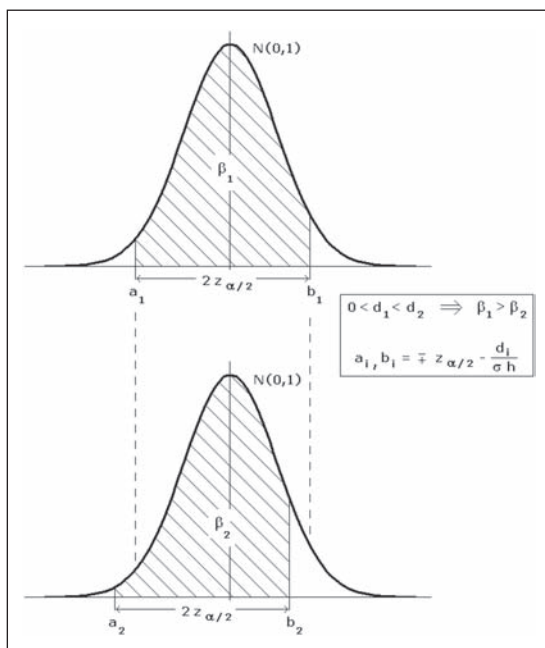


Bild 1: Annahmebereiche (schraffiert) für zwei verschiedene Werte von d

Die Grafik im Bild 1 und die Schlüsse, die daraus zu ziehen sind, können im Unterricht schrittweise behandelt werden (siehe auch Schwarze (2013, Kap. 12.5)):

- In der oberen Standardnormalverteilung wird (3) für einen konkreten Abstand $\mu_1 - \mu_2 = d_1$ dargestellt. b_1 steht dann für die obere Grenze $z_{\alpha/2} - d_1/\sigma h$ und a_1 für die untere Grenze $-z_{\alpha/2} - d_1/\sigma h$. Somit entspricht β_1 dem Inhalt der schraffierten Fläche.

- Die untere Standardnormalverteilung zeigt, welchen Effekt eine Vergrößerung von d hat. Hier soll $\mu_1 - \mu_2 = d_2 > d_1$ sein. Dadurch rücken die beiden zugehörigen Grenzen a_2 und b_2 nach links.
- Die Grafik macht plausibel, dass durch die Ersetzung von d_1 durch d_2 der Inhalt der schraffierten Fläche, also β_2 gegenüber β_1 , kleiner geworden ist. Die Abnahme des schraffierten Bereichs auf der rechten Seite wird nicht kompensiert durch die Zunahme auf der linken Seite. Dies lässt sich natürlich auch nachrechnen, z. B. mit EXCEL. Für negative d wären die schraffierten Flächen symmetrisch an 0 gespiegelt, so dass insgesamt erkannt wird, dass β sinkt, wenn $|d|$ wächst.
- Das Verständnis des Zusammenhangs von α und β kann ebenfalls durch die Grafik in Bild 1 unterstützt werden. Wenn α wächst, wird $z_{\alpha/2}$ kleiner (siehe Tafel 2), und damit auch der schraffierte Bereich, der für β steht. Dessen Grenzwert $1 - \alpha$ für $d \rightarrow 0$ entspricht in der oberen Skizze dem symmetrischen Fall $a_1 = -b_1$. Dann verbleiben links und rechts dieser Grenzen jeweils die Flächeninhalte $\alpha/2$. Im vorigen Punkt wurde schon dargelegt, dass jegliche Verschiebung des Intervalls $[a_1, b_1]$ nach außen zu einer Verringerung von β führt, so dass auch $\beta < 1 - \alpha$ folgt.

Da β nicht unabhängig von d berechnet werden kann, ist zu überlegen, welche Rolle d bei der Anwendung des Testverfahrens spielt. Es sollen *Unterschiede aufgedeckt* werden, wobei die Risiken für Fehlentscheidungen akzeptabel bleiben sollen. Diese Unterschiede werden durch d repräsentiert. Man kann aber davon ausgehen, dass die Anwender eine Vorstellung davon haben, welche Unterschiede relevant sind, d. h. für welche Werte von d sich eine Aufdeckung überhaupt lohnt. Bei minimalen Unterschieden zwischen μ_1 und μ_2 ist in der Regel die (Fehl-)Entscheidung H_0 tragbar. In der Literatur findet man u. a. für $|d|/\sigma$ den Begriff der *Effektstärke* (Cohen 1988). Relevante Unterschiede lassen sich so durch die Effektstärke ausdrücken.

Für die Berechnung des β -Risikos bei festgelegtem Stichprobenumfang n legt man also außer dem Wert α noch ein $d_{\min} > 0$ fest, für das aus Anwendersicht gilt:

Relevante Unterschiede: $|d| \geq d_{\min}$

β sinkt mit wachsendem $|d|$, also erhält man mit (3) für das Risiko 2. Art unter der Bedingung $|d| \geq d_{\min}$ das folgende *bedingte* β -Risiko:

$$(4) \quad \beta = p(H_0 | |\mu_1 - \mu_2| \geq d_{\min}) \leq F\left(z_{\alpha/2} - \frac{d_{\min}}{\sigma h}\right) - F\left(-z_{\alpha/2} - \frac{d_{\min}}{\sigma h}\right)$$

Auf diese Weise berechnet man also die Wahrscheinlichkeit, dass man einen *nennenswerten* Unterschied übersieht. Das entspricht dem Ansatz von Neyman und Pearson aus dem Abschnitt 1, wenn man dort die alternativen Testverteilungen für $d = 0$ und $d = d_{\min}$ zugrunde legt.

Wegen der Symmetrie der Standardnormalverteilung kann man wieder $+d_{\min}/(\sigma h)$ statt $-d_{\min}/(\sigma h)$ einsetzen.

Bei praktischen Anwendungen von Parametertests wird in der Regel der Stichprobenumfang n nicht vorgegeben sein, sondern den jeweiligen Anforderungen angepasst werden. Diese Anforderungen könnten dann die Vorgabe von (α, β) und d sein. Für gleiche Stichprobenumfänge $n = n_1 = n_2$ ist $h = \sqrt{(2/n)}$. Setzt man dies in (3) ein, so erkennt man aus Bild 1, dass mit wachsendem n das Intervall $[a_i, b_i]$ weiter vom Zentrum wegrückt, also β kleiner wird.

Bei der Vermittlung von Parametertests im Unterricht stößt man oft auf ungläubiges Erstaunen, wie groß das Risiko β ausfällt. Große β bedeuten, dass man mit hoher Wahrscheinlichkeit relevante Unterschiede nicht aufdeckt. Bei den Lernenden sollte eine Diskussion angeregt werden, wie man mit diesem Sachverhalt umgeht. Die Kernfrage dabei ist, wie man α, β, n und d_{\min} richtig „austariert“. Dabei sind die „Kosten“ der beiden möglichen Fehlentscheidungen „Falscher Alarm“ und „Verfehlen“ einzubeziehen. Ist etwa ein falscher Alarm doppelt so teuer wie ein Verfehlen, so darf das Risiko 1. Art nur halb so hoch sein wie das Risiko 2. Art.

Um zu einem ausgewogenen Testkonzept zu gelangen, ist zu berücksichtigen:

- α und d_{\min} werden in der Regel von den Anwendern vorgegeben. Diese sollten darauf hingewiesen werden, dass diese Werte nicht unnötig klein gewählt werden sollten, weil sonst n oder β inakzeptabel hoch ausfallen können.
- Über die Stichprobengrößen lässt sich zwar die Größe von β steuern, aber größere n_1, n_2 bedeuten auch höhere Kosten.

4 Fallbeispiel: IQ-Messung

Die standardisierten Testverfahren zur Messung des Intelligenzquotienten sind so konzipiert, dass die Messwerte in guter Näherung normalverteilt sind mit $\mu = 100$ und $\sigma = 15$. (Dies gilt zumindest für die Bevölkerung der westlichen Welt.) Die IQ-Messung soll die Anwendung des Gauß-Tests und die Berechnung des β -Risikos demonstrieren. Zwei Gruppen von Pro-

banden sollen auf Intelligenzunterschiede untersucht werden. Es soll davon ausgegangen werden, dass sich die beiden Gruppen möglicherweise im Mittel von $\mu = 100$ unterscheiden, dass aber die Standardabweichung von $\sigma = 15$ unverändert bleibt (das ist eine Voraussetzung des Gauß-Tests, die ggfs. in einer separaten Untersuchung bestätigt werden muss).

Schon ein einfaches Zahlenbeispiel zeigt die Problematik des β -Risikos in drastischer Form. Es sollen aus beiden Gruppen gleich große Stichproben genommen werden; wir setzen $n = n_1 = n_2 = 100$, also $h = \sqrt{0,02}$, und $\alpha = 0,05$, also nach Tabelle 2 $z_{\alpha/2} \approx 1,96$. IQ-Unterschiede von weniger als 3 sollen als vernachlässigbar gelten, also wird $d_{\min} = 3$ gesetzt. Das entspricht nach den Konventionen von Cohen (1988) einem Minimaleffekt von $|d|/\sigma = 0,2$. Dann gilt nach (4):

$$\beta \leq F\left(1,96 - \frac{3\sqrt{50}}{15}\right) - F\left(-1,96 - \frac{3\sqrt{50}}{15}\right) \approx 0,71$$

Dass das Risiko, einen relevanten Unterschied zwischen den beiden Gruppen nicht zu erkennen, weniger als 71 % beträgt, ist wohl wenig hilfreich. Dabei sind die Werte für die Stichprobengröße, das Risiko 1. Art und d_{\min} keineswegs ungewöhnlich gewählt worden.

Im Zusammenhang mit dieser Berechnung kann man sich ein Bild von der „Testgüte“ $1 - \beta$ machen, indem man mit (3) das Risiko 2. Art gegen d abträgt. Man erhält so eine fallende Kurve (Bild 2), die „Operationscharakteristik“ (Schwarze 2013, Kap. 12.5) für den Gauß-Test:

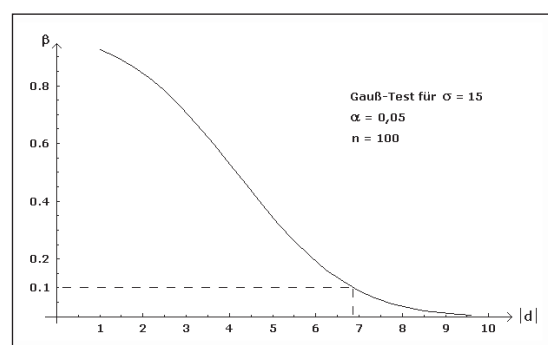


Bild 2: Operationscharakteristik: β in Abhängigkeit von $|d|$

Erst ab IQ-Unterschieden von etwa 7 wäre nach Bild 2 $\beta < 0,1$ und damit die Testgüte $> 0,9$. Hier setzt nun das „Design“ der verschiedenen Parameter für den Test ein. Innerhalb einer Lerngruppe wird sich schnell eine Diskussion über Vor- und Nachteile der einzelnen Parameterwahlen ergeben. In einer solchen Diskussion muss auch zur Sprache kommen, wie bedeutsam für den Untersucher von IQ-Unterschieden

eine Fehlentscheidung 1. Art (es wird fälschlich ein Unterschied behauptet) im Vergleich mit einer Fehlentscheidung 2. Art (es wird ein vorhandener und relevanter Unterschied nicht erkannt) ist. So kann man durchaus der Auffassung sein, dass eines der beiden Risiken deutlich größer als das andere sein darf. Es sollte im Unterricht herausgearbeitet werden, dass dies keine mathematische Frage ist, sondern vom Anwendungsfall abhängt. Die Ergebnisse in Tafel 3 wurden nach (4) mit EXCEL berechnet. $d_{\min} = 7,5$ entspricht nach Cohen (1988) einer mittleren Effektstärke von $|d|/\sigma = 0,5$.

$d_{\min} = 3 \quad n = 100$		$d_{\min} = 7,5 \quad n = 10$	
$\alpha =$	$\beta \leq$	$\alpha =$	$\beta \leq$
0,010	0,877	0,010	0,927
0,025	0,796	0,025	0,869
0,050	0,707	0,050	0,799
0,075	0,642	0,075	0,744
0,100	0,590	0,100	0,698

$d_{\min} = 3 \quad n = 500$		$d_{\min} = 7,5 \quad n = 50$	
$\alpha =$	$\beta \leq$	$\alpha =$	$\beta \leq$
0,010	0,279	0,010	0,530
0,025	0,179	0,025	0,398
0,050	0,115	0,050	0,295
0,075	0,084	0,075	0,236
0,100	0,065	0,100	0,196

$d_{\min} = 3 \quad n = 1000$		$d_{\min} = 7,5 \quad n = 100$	
$\alpha =$	$\beta \leq$	$\alpha =$	$\beta \leq$
0,010	0,029	0,010	0,169
0,025	0,013	0,025	0,098
0,050	0,006	0,050	0,058
0,075	0,004	0,075	0,040
0,100	0,002	0,100	0,030

Tafel 3: Beispielrechnungen

Man erkennt, dass $d_{\min} = 3$ eine restriktive Forderung ist und große Stichproben verlangt, um zu annehmbaren Risiken zu gelangen.

5 Andere Testverfahren

Das grundsätzliche Vorgehen zur Bestimmung des Risikos 2. Art wurde hier für den *2-Stichproben-Gauß-Test bei Varianzhomogenität* ausführlich dargestellt. Verallgemeinerungen liegen auf der Hand, auch wenn die Testverteilungen bei anderen statistischen Testverfahren nicht immer so leicht zu handhaben sind wie bei der Normalverteilung. Für die Didaktik der angewandten Statistik ist es aber zunächst wesentlich, die Grundidee des Antagonismus α -Risiko/ β -Risiko zu vermitteln und damit eine typische „Falle“ der Statistik konkret darzustellen und rechne-

risch zugänglich zu machen – dafür ist der Gauß-Test bestens geeignet.

Die naheliegendste Generalisierung ist die Änderung der Rahmenbedingungen für den Gauß-Test; der Rechenweg, der zu (3) führte, muss dann nur unwesentlich abgeändert werden. Wird bei der Vermittlung statistischer Tests der *1-Stichproben-Gauß-Test* behandelt (der sich für den Einstieg sehr gut eignet), so wird unterstellt, dass die Stichprobe mit Umfang n aus einer $N(\mu, \sigma)$ -verteilten Grundgesamtheit stammt; H_0 lautet dann $\mu = \mu_0$. In (3) ist dann lediglich $d = \mu - \mu_0$ und $h = 1/\sqrt{n}$ zu setzen. – Die universellste Variante ist der *2-Stichproben-Gauß-Test bei Varianzheterogenität*. Dabei werden wieder die Erwartungswerte μ_1 und μ_2 verglichen, aber in Normalverteilungen $N(\mu_1, \sigma_1)$ und $N(\mu_2, \sigma_2)$ mit verschiedenen Standardabweichungen σ_1, σ_2 . Dann ist in (3) $\sigma \cdot h$ durch $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ zu ersetzen.

Sind die Standardabweichungen σ_1, σ_2 unbekannt, arbeitet man mit den empirischen Standardabweichungen s_1, s_2 der Stichproben. Dies führt auf den *t-Test*, der ebenfalls gut mit EXCEL durchgeführt werden kann. Da die *t*-Verteilung für wachsende Stichprobengrößen asymptotisch in die Standardnormalverteilung übergeht (Schwarze 2013, Kap. 9.11), lässt sich das β -Risiko zumindest näherungsweise wie beim Gauß-Test berechnen, indem (für große n_1, n_2) s_1, s_2 als Schätzwerte für σ_1, σ_2 eingesetzt werden.

Software

Berechnungen zur Testgüte können mit dem Open-Source-Programm *G*Power3* der Universität Düsseldorf durchgeführt werden (<http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>).

Literatur

- Cohen, J. (1988): *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Lenhard, J. (2006): *Models and Statistical Inference: The Controversy between Fisher and Neyman-Pearson*. In: *Brit. J. Phil. Sci.* (57), S. 69–91.
- Schwarze, J. (2013): *Grundlagen der Statistik, Band 2*. Herne/Berlin: NWB-Verlag.

Anschrift des Verfassers

Manfred Börgens
 Technische Hochschule Mittelhessen
 Wilhelm-Leuschner-Straße 13
 61169 Friedberg
 manfred.boergens@mnd.thm.de