Baumdiagramm beim Signifikanztest

RENATE MOTZER, AUGSBURG

Zusammenfassung: *Ergänzend zum Beitrag „Unklare Begriffe und Wunschdenken bei Signifikanztests“ von Matthias Moßburger (SIS 2014(1)) wird vorgeschlagen statt der Vierfeldertafel ein unvollständiges Baumdiagramm zu verwenden. Außerdem wird das Baumdiagramm beispielhaft vervollständigt, um daran zu sehen, warum es häufig wichtig ist, einen zweiten Durchgang eines Tests durchzuführen, um ein aussagekräftiges Ergebnis zu erhalten.*

1 Einleitung

In Schulbüchern wird zur Übersicht über die möglichen (Fehl-)Entscheidungen beim Signifikanztest meist ein Tabellenformat genutzt, welches der den Schülern aus dem bisherigen Unterrichtsgeschehen bereits bekannten Vierfeldertafel sehr ähnelt. Matthias Moßburger stellt heraus, warum diese Darstellung für die Schülerinnen und Schüler irreführend sein kann. In der Vierfeldertafel stehen üblicherweise Wahrscheinlichkeiten der Art $P(A \cap B)$ und keine bedingten Wahrscheinlichkeiten, wie sie beim Entscheidungsverfahren des Hypothesentests gefragt sind.

2 Baumdiagramme im Kontext von Signifikanztests

Bedingte Wahrscheinlichkeiten kann man besser in einem Baumdiagramm darstellen. Ich verwende daher in den letzten Jahren im Unterricht folgendes Baumdiagramm:

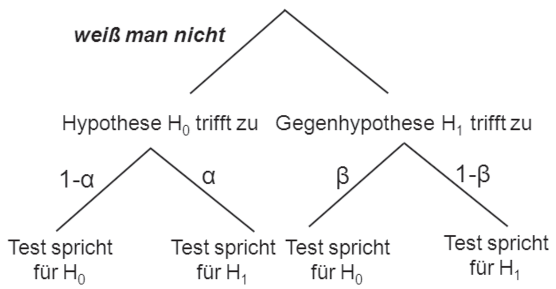

Abb. 1: Baumdiagramm 1

oder unter Verwendung des Ablehnungsbereich (A) bzw. Annahmebereich ($\overline{A}$): (Ich habe mich bei der Benennung von A und $\overline{A}$ an den Gebrauch bei Moßburger gehalten. In meinem Umfeld wird mit A meistens der Annahmebereich der Hypothese H_0 bezeichnet.)

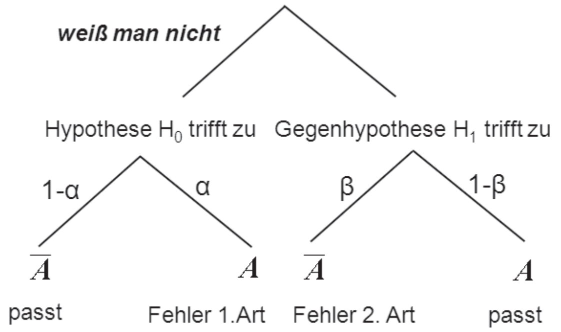

Abb. 2: Baumdiagramm 2

Bei dieser Darstellung sieht man deutlich, dass man über eine Wahrscheinlichkeit von H_0 und H_1 nichts aussagen kann. Ob man Hypothesen wie Ereignisse, die eine Wahrscheinlichkeiten haben, anschauen darf, ist durchaus nicht unumstritten (siehe Absatz 4). Im Kontext von schulischem Unterricht scheint es mir durchaus sinnvoll. Die in Schulbüchern übliche Darstellung in Tabellenform erinnert an die Vierfeldertafel und suggeriert daher auch die Verbindung von zwei Ereignissen (die hier als Hypothesen und mögliche Testausgänge gegeben zu sein scheinen).

Die Deutung der auftretenden Wahrscheinlichkeiten bei einem Signifikanztests fällt oft schwer. Grundlegende Schwierigkeiten zeigen sich schon bei der einfachen Behandlung der Vierfeldertafel.

Ein einfaches Beispiel: 100 Kinder werden befragt.

	Mädchen	Jungen	Insgesamt
mag Fußball	30	35	65
mag Fußball nicht	30	5	35
Insgesamt	60	40	100

Tab. 1: Vierfeldertafel

Es ist hier mit natürlichen Häufigkeiten dargestellt, aber man könnte hinter jede Zahl ein Prozentzeichen setzen.

Die Zahl 30 im oberen linken Feld heißt also: 30 % der befragten Kinder sind Mädchen und mögen Fußball. Sie heißt nicht: „30 % von den Mädchen mögen Fußball“ oder „30 % von den Fußballfans sind Mädchen“.

Die zu diesen Aussagen gehörenden relativen Häufigkeiten könnte man in Baumdiagrammen finden. Man kann sie sich freilich aus den gegebenen Zahlen errechnen: 30 von 60 Mädchen mögen Fußball, also 30/60 = 50 %. 30 von 65 Fußballfans sind Mädchen, also 30/65 = 46 %.

Baumdiagramme erlauben zumindest die Betonung und Visualisierung einer auch sprachlich angemessenen Formulierung bedingter Wahrscheinlichkeiten, welche hier im Unterricht bereits frühzeitig fokussiert werden sollte. So kann diese später im Unterricht im Kontext von Signifikanztests unter der wiederholten Verwendung von Baumdiagrammen wieder aufgegriffen und so deutlicher der Unterschied zwischen „Bei einem Ergebnis im Ablehnungsbereich ist H_0 unwahrscheinlich“ und „Bei H_0 ist ein Ergebnis im Ablehnungsbereich unwahrscheinlich“ hervorgehoben und den hier vielfach dokumentierten bekannten Verständnisschwierigkeiten bei der Interpretation von Testergebnissen vorgebeugt werden.

3 Bayesianisches Hypothesentesten oder die Möglichkeit, etwas über die Wahrscheinlichkeit des Zutreffens der Nullhypothese auszusagen

Die Diskussion mit einem Kollegen hat mich nun auf den Gedanken gebracht, das „Weiß-man-nicht“ im Baumdiagramm probehalber mit konkreten Werten zu besetzen. Der Kollege hat im Unterricht das Beispiel eines Würfeltests verwendet. Bei 120 Würfen wird gezählt, wie oft die „3“ erscheint. Dies sei stolze 32-mal der Fall. Liegt ein gezinkter Würfel vor? Bei einer Trefferwahrscheinlichkeit von $p = 1/6$ beträgt die Wahrscheinlichkeit, so weit oder weiter über dem Erwartungswert von 20 zu liegen (A), nur 0,38 %. Also geht man anschließend davon aus, dass $p > 1/6$ gilt und der Würfel gezinkt ist. Ein Schüler sagte nun aber: „Wir können nicht so vorschnell schließen, dass der Würfel gezinkt ist. Dazu müssen wir doch wissen, wie viele gezinkte und nicht-gezinkte Würfel es überhaupt gibt. Wenn wir unseren Würfel aus einer Kiste mit 9 999 intakten und einem gezinkten Würfel herausgreifen, dann ist die Wahrscheinlichkeit dafür, dass unser Würfel der gezinkte ist, sehr klein, auch wenn er 32-mal die „3“ zeigt.“ Der Schüler hatte bei bedingten Wahrscheinlichkeiten wohl sehr gut aufgepasst.

In dieser Unterrichtssituation war vom Lehrer nicht angedacht, den Hypothesen Wahrscheinlichkeiten zu geben. Der Schüler hat es jedoch getan. Er hat der Gegenhypothese „Der Würfel ist gezinkt“ eine konkrete A-Priori-Wahrscheinlichkeit gegeben: 1/10000. Darüber, wie wahrscheinlich es ist, dass ein gezinkter Würfel als solcher identifiziert würde, hat der Schüler nichts gesagt. Über die Wahrscheinlichkeit β wissen wir also nichts. Sollte es ein Würfel sein, mit dem man fast nur die „3“ würfelt, wären 32 von 120 sogar wenig und dieser Testausgang würde gar nicht für den gezinkten Würfel sprechen. Um einfacher rechnen zu können, nehme ich zunächst dennoch an, dass p deutlich größer als 1/6 wäre, z. B. $p_1 = 1/2$ damit die Wahrscheinlichkeit β für ein Testergebnis im Annahmebereich $\overline{A}$ ziemlich klein bleibt. Für $\overline{A}$ werden hier die Testausgänge mit „weniger als 32-mal die 3“ gewählt.

Berechnen wir mit der Eingangswahrscheinlichkeit 1/10000, mit $\alpha = 0{,}38\ \%$ (für $p_0 = 1/6$) und mit β (fast) 0 % ($p_1 = 1/2$) die Wahrscheinlichkeit dafür, dass es sich bei dem getesteten Würfel tatsächlich um einen gezinkten Würfel handelt, dann beträgt die Wahrscheinlichkeit nur ca. 1/39. 38 der 9 999 nicht-gezinkten Würfel dürften im Schnitt zu einem solchen Würfelergebnis (oder einem noch deutlicheren) führen, der eine gezinkte ebenfalls. Im Baumdiagramm gilt:

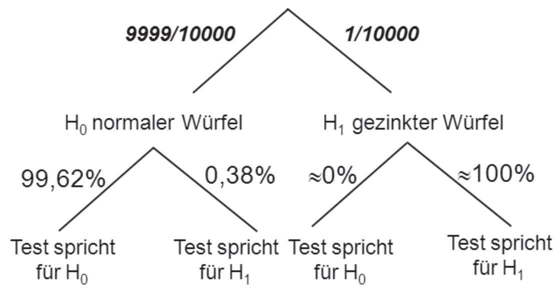

Abb. 3: Baumdiagramm 3

An dieser Stelle dürften auch die Schülerinnen und Schüler erkennen, dass es sinnvoll ist, den Würfel ein weiteres Mal zu testen. Nun liegt die Eingangswahrscheinlichkeit aber nicht mehr bei 1/10000, sondern nur noch bei 1/39.

Bei einem erneuten Ergebnis von „mindestens 32 Dreien“ liegt die Wahrscheinlichkeit dafür, dass es sich um den gezinkten Würfel handelt, schon bei 87,4 % (= 1/39/(38/39 · 0,0038 + 1/39)). Daran ändert sich auch kaum etwas, wenn man β z. B. auf 5 % erhöhen würde (für $p_1 \approx 1/3$): Wir erhielten im ersten Test die möglichen 39 gezinkten Würfel und in einem weiteren Testdurchgang (1/39 · 0,95/(38/39 · 0,0038 + 1/39 · 0,95) = 86,8 %).

Die Höhe von β ist in diesem Beispiel abhängig davon, wie „gezinkt“ der Würfel ist, d. h. wie groß bei ihm die Wahrscheinlichkeit für eine „3“ ist. Liegt die Wahrscheinlichkeit deutlich über 32/120, ist β wesentlich kleiner, als wenn die Wahrscheinlichkeit näher an der sich im Test ergebenden relativen Häufigkeit liegt. Ist p mindestens 1/3, so ist β bei diesem Annahmebereich (mit der Grenze von 31) unter 5 %. Dieser Zusammenhang lässt sich durch die Operationscharakteristik, die die funktionale Abhängigkeit des β-Fehlers von der Wahrscheinlichkeit p beschreibt, darstellen.

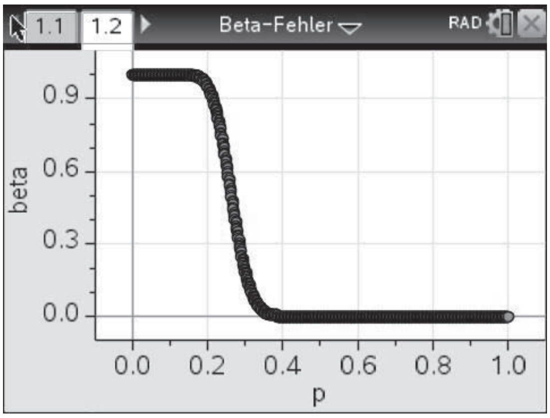

Abb. 4: Operationscharakteristik

Die Wahrscheinlichkeit, dass es sich um den gezinkten Würfel handelt, wenn beim zweiten Durchgang des Test wieder mindestens 32-mal die „3" gewürfelt wird, beträgt zwar immer noch keine 95 % (ein dritter Durchgang würde dies bewirken können). Aber 87 % sind vielleicht doch so viel, dass man geneigt sein darf zu glauben, man habe den gezinkten Würfel gefunden. Die Tatsache, dass β kein fester Wert ist, sondern für die Gegenhypothese „$p > 1/6$" vom tatsächlichen p des gezinkten Würfels abhängt, wurde hier vernachlässigt, um den Bayes-Ansatz (für ein konkretes p) durchrechnen zu können. Es wurden nur der Wert $p = 1/3$ (mit $\beta \approx 5$ %) und der Wert $p = 1/2$ (mit $\beta \approx 0$ %) bedacht. Für p-Werte zwischen 1/6 und 1/3 wäre β größer und damit die errechnete Wahrscheinlichkeit noch kleiner als 1/39 im ersten und 87 % im zweiten Durchgang. Die erste Wahrscheinlichkeit liegt dann bei maximal 1/39 (für $\beta = 0$) ansonsten bei: $(1 - \beta)/(38 + (1 - \beta))$. Läge p bei ungefähr 32/120, würde also zum Grenzwert zwischen A und $\overline{A}$ passen, wäre β bei ungefähr 0,5 und es ergäbe sich nur eine Wahrscheinlichkeit von 1/77, dass es sich um den gezinkten Würfel handelt. Die zweite Wahrscheinlichkeit ergibt sich, wenn man die erste noch mit 1/39 abschätzt auf: $1/39 \cdot (1 - \beta)/(38/39 \cdot 0{,}0038 + 1/39 \cdot (1 - \beta))$. Auch dieser Term wird mit wachsendem β kleiner. Wählt man beide Mal $\beta = 0{,}5$, ergibt sich: $1/77 \cdot 0{,}5/(76/77 \cdot 0{,}0038 + 1/77 \cdot 0{,}5) = 63{,}4$ %.

Auch hier zeigt sich zumindest, dass ein zweiter Durchgang die Wahrscheinlichkeit, dass man den gezinkten Würfel erwischt hat (wenn man voraussetzt, dass wieder so viele Dreien gewürfelt wurden), doch deutlich erhöht hat. Je näher die Wahrscheinlichkeit p_1 der Gegenhypothese an p_0 der Hypothese H_0 liegt, desto leichter führt der Test zu einem falschen Schluss. Dies sollte den Schülerinnen und Schüler auch an anderen Beispielen klar gemacht werden. Dass die Wahrscheinlichkeit für den β-Fehler als Funktion von p angegeben werden müsste, wird in den aktuellen Lehrplänen aber kaum mehr thematisiert. Wenn überhaupt, wird manchmal für einen ausgesuchten Wert von p das zugehörige β berechnet. In manchen Lehrplänen ist bedauerlicherweise aber nicht einmal das der Fall.

4 Fazit

Was diese Rechnung und Betrachtung meines Erachtens zeigen kann, ist, wie wichtig es ist, dass man einen guten Grund hat, warum man eine Gegenhypothese aufstellt. Es reicht nicht, aufgrund eines auffallenden Testergebnisses im Nachhinein eine (Gegen-)Hypothese aufzustellen und zu behaupten, man hätte sie durch dieses Testergebnis damit nachgewiesen. Ein zweiter Durchgang, bei dem schon eine deutliche Apriori-Wahrscheinlichkeit gegen die Nullhypothese sichtbar wird (auch wenn diese nicht immer, wie im konstruierten Beispiel, quantifiziert werden kann), ist oft nötig.

Um dafür Verständnis zu wecken, plädiere ich also für die Verwendung eines Baumdiagramms auch beim Hypothesentest. Und dafür beispielhaft einmal eine Apriori-Wahrscheinlichkeit anzunehmen und aufzuzeigen, warum ein zweiter Test nötig sein könnte.

Literatur

Moßburger, M. (2014). Unklare Begriffe und Wunschdenken bei Signifikanztests. In: *Stochastik in der Schule 34*(1), 2–8.

Anschrift der Verfasserin:

Renate Motzer
Didaktik der Mathematik
Universität Augsburg
Universitätsstr. 10
86135 Augsburg
Renate.Motzer@math.uni-augsburg.de