Stochastik in der Schule, Heft 2, 7. Jg. (1987)

SEQUENTIELLES TESTEN - AUCH DIDAKTISCH VIELLEICHT EINE GUTE ALTERNATIVE

von Raphael Diepgen, Bochum

1. Vorbemerkung

Statistisches Hypothesentesten hat sich zwar in vielen Wissenschaften als forschungsmethodisches Instrumentarium etabliert, aber seine Praxis zeigt allzuoft eine Fülle von Irrationalismen; insbesondere in den eher mathematikfernen Humanwissenschaften hat das statistische Hypothesentesten den Charakter eines in seiner Logik unverstandenen, durch bloße Konvention definierten Rituals angenommen, dem zu unterwerfen sich gleichwohl der eigenen wissenschaftlichen Karriere wegen empfiehlt (vgl. etwa BREDENKAMP, 1972, WITTE, 1980). Vor allem der Umgang mit dem Standardverfahren Signifikanztest leidet an zwei Mängeln:

1. Der Signifikanztest wird in der Praxis zumeist nicht, wie von seiner Logik gefordert, als ein <u>vor</u> der Datenerhebung oder zumindest -inspektion festgelegtes Entscheidungsverfahren angewandt, sondern vielmehr bei der nachträglichen Analyse schon bekannter Daten. Dabei hat dann die Kenntnis der Daten Einfluß auf die konkrete Auswahl eines bestimmten Signifikanztestes, so daß dessen Signifikanzniveau alles andere angibt als die Fehlerwahrscheinlichkeit des Entscheidungsverhaltens des Forschers. Die so ex post erzielten Signifikanzniveaus sind lediglich die hypothetischen Wahrscheinlichkeiten dafür, daß - die Geltung der Nullhypothese vorausgesetzt - Ereignisse auftreten, die mindestens so extrem sind wie das beobachtete Ereignis. Diese nachträglich ermittelten Überschreitungswahrscheinlichkeiten haben aber nicht die Aussagekraft wie Fehlerwahrscheinlichkeiten von Tests, zumindest wenn man den konzeptionellen Rahmen nur eines einzigen Tests verläßt.

2. Der Fehler zweiter Art oder β-Fehler, also ein fälschliches Beibehalten der Nullhypothese, bleibt unbeachtet. Es ist in der Forschungspraxis nicht üblich, die Wahrscheinlichkeit

eines solchen Fehlers, eine Funktion natürlich des "wahren" Parameters (Operationscharakteristik OC), zu analysieren. Vielfache Appelle von Statistikern an die Forschungspraktiker, bei jedem Signifikanztest zunächst eine solche Analyse voranzuschicken, haben nichts gefruchtet, vor allem wohl, weil eine solche Analyse numerisch oft recht aufwendig, in den entsprechenden Statistikprogrammen nicht implementiert und in der Publikation - als Funktionsgraph - nur umständlich darstellbar ist; sicherlich aber auch, weil den meisten Empirikern schon auf der gedanklichen Ebene mangels Verständnis' der Testlogik die Bedeutung einer solchen Analyse nicht gegenwärtig ist. Da aber in der Regel die inhaltlichen Forschungshypothesen den zusammengesetzten Alternativhypothesen der Signifikanztests entsprechen, führt dies zu einem seltsamen Zustand: Der Forscher kontrolliert zwar mit seinem Signifikanzniveau, sofern er nicht nur - siehe oben - nachträglich analysiert, die Wahrscheinlichkeit, eine Nullhypothese zu verwerfen, wenn diese richtig ist. Nur: Daß diese punktförmige Nullhypothese (exakt) richtig ist, ist ohnehin so gut wie ausgeschlossen. Aber die eigentlich interessierende Wahrscheinlichkeit, nämlich die Forschungshypothese zu verwerfen, obwohl sie richtig ist, und damit die Wahrscheinlichkeit, eine wahre Forschungshypothese auch als solche zu erkennen, diese Wahrscheinlichkeit bleibt im Dunkeln. Und dies hat Folgen: Ein signifikanter Untersuchungsbefund läßt keine Rückschlüsse darauf zu, wie groß denn wohl die Abweichungen der Realität von der Nullhypothese sind. Da die Forschung in der Regel nicht an irgendwelchen minimalen Abweichungen von der Nullhypothese interessiert ist, sondern nur an nennenswerten Unterschieden, die auch praktische oder wissenschaftliche Relevanz beanspruchen können, bedeutet die Signifikanz eines Befundes allein noch gar nichts; erst in Verbindung mit "praktischer Signifikanz" wird sie wichtig. Diese "praktische Signifikanz" läßt sich aber nur beurteilen, wenn man weiß, welche Abweichungen von der Nullhypothese mit welchen Wahrscheinlichkeiten zu einem signifikanten Ergebnis führen, also nur in Kenntnis der β-Fehler-Funktion. Diese OC-Funktion hängt bekanntlich vom Stichprobenumfang ab: Je größer die Stichpro-

be, desto steiler fällt die OC-Funktion ab, wenn man sich vom nullhypothetischen Parameter entfernt. Das bedeutet: Je größer die Stichprobe, mit desto größerer Wahrscheinlichkeit werden auch minimale und praktisch wie wissenschaftlich unbedeutende Abweichungen von der Nullhypothese signifikant. (Und deswegen degeneriert die mit hoher Reputation versehene Signifikanz in der Praxis zu einem Maß für den empirischen Aufwand, den ein Forscher zu leisten bereit und in der Lage ist.) In Unkenntnis des genauen Verlaufs der OC-Funktion läßt sich daraus aber lediglich eine vage Skepsis gegenüber signifikanten Befunden aus großen Stichproben folgern; quantitative Abschätzungen sind kaum möglich. Kurzum: Die Mißachtung des β-Fehlers ist Kernpunkt der weitverbreiteten Irrationalität im Umgang mit dem Signifikanztest.

Für einen Statistikunterricht auf der Schule, der sich als Propädeutik für eine aufgeklärtere inferenzstatistische Praxis versteht (vgl. DIEPGEN, 1984), hat das natürlich Konsequenzen: Es verbietet sich ein ausschließlich signifikanzorientierter Unterricht (DIEPGEN, 1985a), der Untersuchung der β-Fehler-Funktion sollte im Unterricht breiter Raum gegeben werden (DIEPGEN, 1985b), und optimal scheint ein entscheidungtheoretisch orientierter Einstieg in das Hypothesentesten (DIEPGEN, 1985c), der von Anfang an sowohl den Fehler erster Art, als auch den Fehler zweiter Art zum Gegenstand macht. Indessen: Die didaktisch wichtige entscheidungstheoretische Konstruktion von Tests nach dem Bayes-Prinzip gelingt im Unterricht nur für den Test zweier einfacher, punktförmiger Hypothesen gegeneinander, nicht aber für den in der Praxis viel wichtigeren Test einer einfachen Nullhypothese gegen eine zusammengesetzte Alternativhypothese. Schön wäre daher der Einstieg in das Hypothesentesten über eine Testkonstruktion, die sowohl von Anfang an den Charakter des Tests als ein zwei Fehler riskierendes Entscheidungverfahren betont, als auch zu einem für die Forschungspraxis wichtigen Test mit zusammengesetzter Alternativhypothese führt. Solche Tests finden sich im Rahmen der von Abraham WALD entwickelten Sequentialstatistik. Es sei daher im folgenden diese Sequentialstatistik kurz charakterisiert, der dort zentrale sequentielle Quotiententest einschließlich eines Beipiels dargestellt und schließlich in didaktischer Hinsicht andiskutiert.

2. Sequentialstatistik: Allgemeines

Die Sequentialstatistik wurde bereits gegen Ende des zweiten Weltkrieges von Abraham WALD entwickelt, und zwar vorrangig mit dem Ziel, Stichprobengrößen bei der Qualitätskontrolle in der Rüstungsprodunktion zu minimieren. Die sequentialstatistischen Grundideen galten daher zunächst als Militärgeheimnis und wurden erst mit Verzögerung von WALD (1945) veröffentlicht. Eine ausdrücklich für den mathematischen Laien geschriebene Darstellung dieser Grundideen stammt von WALD (1966) selbst. Für den Mathematiker interessant ist vor allem WALDs (1950) entscheidungstheoretische Interpretation der Statistik, auch der Sequentialstatistik. Moderne Darstellungen der inzwischen weit expandierten Sequentialstatistik finden sich etwa bei GHOSH (1970), WETHERILL (1975) und GOVINDARAJULU (1975). Deutschsprachige Darstellungen einiger sequentialstatistischer Ideen und Verfahren finden sich bei WEBER (1967, p. 395-482), SACHS (1968, p. 217-223), BÜNING & TRENKLER (1976, p. 301-310), FISZ (1976, p. 676-708) und HARTUNG (1984), Darstellungen indessen, die meines Erachtens häufig nicht die didaktische Klarheit von WALDs (1966) eigenen - freilich englischsprachigen - Ausführungen erreichen. Einige Veröffentlichungen zu den mathematischen Grundlagen der sequentiellen Statistik kommen aus der DDR, etwa von HECKENDORF (1982) und EGER (1985). Für den Schulbereich findet sich eine kurze Darstellung sequentiellen Testens bei FREUDENTHAL (1975, p. 97-100). Verbreitet ist die Sequentialstatistik vor allem dort, wo die Datenerhebung von einem kostenbewußten Finanzier - und nicht etwa nur wie in der Universität von der anonymen öffentlichen Hand - bezahlt wird, etwa in der industriellen Qualitätskontrolle (vgl. UHLMANN, 1966). Denn die Sequentialstatistik kommt im Schnitt mit deutlich kleineren Stichproben aus als die klassische Teststatistik. Verblüffend ist - und für den Unterricht wichtig -,

daß die Grundidee der Sequentialstatistik einerseits mathematisch wirklich so einfach ist, daß sie ohne besondere mathematische Vorkenntnisse verstanden werden kann, und dabei andererseits in ihrer argumentativen Brillianz sogar einen gewissen ästhetischen Reiz hat. (Dieser sei den Schülern nicht vorenthalten.) Zum Namen "Sequentialstatistik" : Er bezieht sich auf den sequentiellen Charakter des Entscheidungsverfahrens. Es werden nämlich - im Unterschied zum klassischen Testen auf der Basis der Neyman-Pearson-Testtheorie - sukzessive Daten erhoben, und nach jedem erhobenen Datum wird erneut entschieden zwischen der Nullhypothese, der Alternativhypothese und der Fortsetzung des Versuchs durch Erhebung eines weiteren Datums.

Grundlegend für die gesamte Sequentialstatistik ist der von WALD entwickelte Quotiententest (Sequential Probability Ratio Test SPRT).

3. Der sequentielle Quotiententest

Wir wollen zwischen zwei einfachen Hypothesen über die Wahrscheinlichkeit p eines Merkmals entscheiden, nämlich zwischen H_0: $p = p_0$ und H_1: $p = p_1$, wobei wir ohne Beschränkung der Allgemeinheit $p_0 \leq p_1$ voraussetzen wollen. (Die Erweiterung der Argumentation für zusammengesetzte Hypothesen der Form H_1: $p \geq p_1$ folgt im anschließenden Beispiel.) Gesucht ist ein Entscheidungsverfahren, bei dem wir im Unterschied zur klassischen Statistik sukzessive zufällige Daten erheben und nach jeder Datenerhebung drei Möglichkeiten haben, nämlich:

1. Annahme von H_0 (symbolisch "H_0"),
2. Annahme von H_1 (symbolisch "H_1"),
3. Erhebung eines weiteren Datums und dann erneute Entscheidung.

Dieses sequentielle Entscheidungsverfahren soll nun folgende Eigenschaften haben:

(1) Es soll mit Sicherheit irgendwann einmal zu einem Ende

kommen, d.h. es soll mit Wahrscheinlichkeit 1 zu einer Entscheidung für H_0 oder H_1 führen. (Den exakten Beweis für diese plausible Eigenschaft unseres unten entwickelten Entscheidungsverfahrens übergehe ich, da er für den Unterricht etwas zu kompliziert sein dürfte.)

(2) Die Wahrscheinlichkeit für einen Fehler erster Art, also einer fälschlichen Entscheidung für H_1, beträgt α. Formal $P("H_1"/p=p_0) = \alpha$. Dabei kann α beliebig gewählt werden, in der Regel aufgrund von Erwägungen über die Folgen einer solchen Fehlentscheidung natürlich klein.

(3) Auch die Wahrscheinlichkeit **eines Fehlers zweiter Art**, also einer fälschlichen Entscheidung für H_0, ist begrenzt durch ein - wieder frei wählbares - β. Formal $P("H_0"/p=p_1) = \beta$.

Aus (2) und (3) ergibt sich in Verbindung mit (1)
$P("H_0"/p=p_0) = 1-\alpha$ und $P("H_1"/p=p_1) = 1-\beta$.

Es erscheint nun sinnvoll, das Entscheidungsverfahren auf die Häufigkeit des Auftretens des Merkmals zu beziehen. Wenn in n unabhängigen Versuchsdurchgängen (in einer bestimmten Reihenfolge) m mal das Merkmal auftritt - und damit genau n-m mal nicht -, so geschieht dies, wenn H_0 stimmt, mit der Wahrscheinlichkeit ("Likelihood") $p_0^m(1-p_0)^{n-m}$, dagegen wenn H_1 stimmt, mit der Wahrscheinlichkeit $p_1^m(1-p_1)^{n-m}$. Nennen wir diese offensichtlich leicht aus den Daten zu berechnenden Likelihoods l_0 bzw. l_1. Ist nun der Quotient l_1/l_0 klein, so spricht dies für die Annahme von H_0: Denn die Wahrscheinlichkeit l_0 unseres beobachteten **Ereignisses** unter H_0 ist dann weitaus größer als die Wahrscheinlichkeit l_1 unseres **Ergebnisses** unter H_1. (Und natürlich ist es vernünftig, sich für die Hypothese zu entscheiden, die dem beobachteten Geschehen höhere Wahrscheinlichkeit zuspricht.) Umgekehrt: Ist der Likelihood-Quotient l_1/l_0 groß, so spricht dies für die Annahme von H_1. Bewegt sich der Likelihood-Quotient schließlich im mittleren Bereich, so favorisiert dies keine der beiden fraglichen Hypothesen; es erscheint somit sinnvoll, weitere Daten

zu erheben. "Groß" und "klein" bedürfen noch der Konkretisierung, aber wir können mit noch näher zu bestimmenden Grenzen A und B (selbstverständlich mit $B < 1 < A$) die Form des Entscheidungsverfahrens festlegen:

1. Falls $l_1/l_0 \leq B$, nimm H_0 an.
2. Falls $l_1/l_0 \geq A$, nimm H_1 an.
3. Falls $B < l_1/l_0 < A$, setze das Experiment fort, d.h. erhebe eine weitere Beobachtung und entscheide danach erneut.

Jetzt gilt es nur noch, A und B numerisch so zu bestimmen, daß die geforderten Eigenschaften gewährleistet sind. Dazu zwei Überlegungen:

a) Sei E_0 die Menge aller möglichen Experimente, die nach unserer Entscheidungsregel zur Annahme von H_0 führen. Für jedes dieser Experimente muß dann offensichtlich gelten $l_1/l_0 \leq B$, d.h. $l_1 \leq l_0 B$. Summieren wir nun die Likelihoods aller Experimente aus E_0, so erhalten wir insgesamt natürlich die Wahrscheinlichkeit, H_0 anzunehmen. Diese Wahrscheinlichkeit ist aber mit $1-\alpha$ bzw. β vorgewählt. Also:

$P("H_0"/p=p_0) = 1-\alpha = \sum_{E_0} l_0$ und
$P("H_0"/p=p_1) = \beta = \sum_{E_0} l_1$.
Insgesamt erhalten wir
$\beta = \sum_{E_0} l_1 \leq \sum_{E_0} (l_0 B) = (\sum_{E_0} l_0) B = (1-\alpha) B$
und damit
(i) $\beta/(1-\alpha) \leq B$.

b) Analog: Sei E_1 die Menge aller Experimente, die zur Annahme von H_1 führen. Für jedes dieser Experimente gilt demnach $l_1/l_0 \geq A$, d.h. $l_1 \geq l_0 A$. Summieren wir hier die Likelihoods aller Experimente aus E_1, so erhalten wir insgesamt die Wahrscheinlichkeit, H_1 anzunehmen. Diese Wahrscheinlichkeit ist mit α bzw. $1-\beta$ vorgewählt. Also:

$P("H_1"/p=p_1) = 1-\beta = \sum_{E_1} l_1$ sowie
$P("H_1"/p=p_0) = \alpha = \sum_{E_1} l_0$.
Und damit:
$1-\beta = \sum_{E_1} l_1 \geq \sum_{E_1} l_0 A = \alpha A$.

Schließlich:

(ii) $(1-\beta)/\alpha \geq A$.

Leider ergeben die aus den geforderten Eigenschaften abgeleiteten Ungleichungen (i) und (ii) noch keine numerischen Werte für A und B, sondern Höchst- bzw. Tiefstwerte für ihre Wahl. Und nun kommt ein schöner Trick von WALD: Wir wählen einfach einmal willkürlich $A = (1-\beta)/\alpha$ und $B = \beta/(1-\alpha)$. Wir erhalten so ein konkretes Entscheidungsverfahren, denn A und B sind jetzt feste Zahlen, aber wir wissen der willkürlichen Setzung wegen leider nicht, ob dieses Entscheidungsverfahren die geforderten Fehlerwahrscheinlichkeiten α bzw. β hat. Nennen wir daher die tatsächlichen und uns unbekannten Fehlerwahrscheinlichkeiten dieses willkürlich festgelegten Entscheidungsverfahrens α' und β'. Die obige Argumentation gilt jetzt also für diese neuen, tatsächlichen Fehlerwahrscheinlichkeiten, und damit ergibt sich

gemäß (i) $\beta'/(1-\alpha') \leq B$, also $\beta'/(1-\alpha') \leq \beta/(1-\alpha)$,

gemäß (ii) $(1-\beta')/\alpha' \geq A$, also $(1-\beta')/\alpha' = (1-\beta)/\alpha$.

Insgesamt ergibt sich daraus

$\beta' \leq (\beta/(1-\alpha))(1-\alpha') \leq \beta/(1-\alpha) \approx \beta$, sofern α klein,

und

$\alpha' \leq (1-\beta')/((1-\beta)/\alpha) = ((1-\beta')\alpha)/(1-\beta) \leq /(1-\beta) \approx \alpha$,

sofern β klein.

Fazit: In der Praxis, in der α und β natürlich klein gewählt werden, führt die zunächst willkürlich anmutende Setzung von A und B zu einem Entscheidungsverfahren, dessen tatsächlichen Fehlerwahrscheinlichkeiten α' und β' die vorgewählten Grenzen α und β allenfalls unwesentlich überschreiten. Wir haben also das gesuchte Entscheidungsverfahren - bis auf praktisch unbedeutende Ungenauigkeiten - gefunden: Die Fehlerwahrscheinlichkeiten sind vermutlich kleiner, allenfalls unwesentlich größer als gefordert.

4. Beispiel

Wir entscheiden zwischen H_0: $p=0.5$ und H_1: $p=0.6$ zu $\alpha=0.05$

und $\beta=0.1$. Wir erhalten also die Grenzen $A=0.9/0.05=18$ und $B=0.1/0.95=0.105$. Tritt in n unabhängigen Versuchen das Merkmal m mal auf, so gilt für die Likelihoods $l_0=0.5^m\ 0.5^{m-n}$ und $l_1=0.6^m\ 0.4^{m-n}$. Zur einfachen Formulierung der Entscheidungsregel helfen äquivalente Umformungen der entsprechenden Ungleichungen (durch beiderseitiges Logarithmieren):
$l_1/l_0 \leq 0.105$ ist äquivalent zu $m \leq 0.54n-5.49$, und $l_1/l_0 \geq 18$ entspricht $m \geq 0.54n+6.95$. Wir erhalten also das äußerst simple Entscheidungsverfahren: Falls wir in den ersten n Versuchsdurchgängen höchsten $0.54n-5.49$ mal das Merkmal beobachten, nehmen wir H_0 an. Falls wir dagegen in den ersten n Versuchsdurchgängen mindestens $0.54n+6.49$ mal das Merkmal beobachten, nehmen wir H_1 an. Andernfalls erheben wir im n+1-ten Versuchsdurchgang ein weiteres Datum. Dieses simple Entscheidungsverfahren läßt sich schließlich graphisch in Form eines "Entscheidungsspaziergangs" organisieren (Abb. 1) , der mit dem Bleistift auf dem Papier ohne irgendwelche Berechnung zur Entscheidung führt.

Nun macht man sich schnell klar, daß dieses Entscheidungsverfahren natürlich auch einen Test von H_0 gegen die zusammengesetzte Alternativhypothese H_1: $p \geq 0.6$ darstellt. Denn je größer das "wahre" p, desto weniger wahrscheinlich ist es, daß das Merkmal so selten auftritt, daß wir für H_0 entscheiden. Das vorgewählte β ist hier demnach die "maximale" Wahrscheinlichkeit einer fälschlichen Entscheidung für H_0 (vgl. auch unten Abb. 2).

Anders ausgedrückt: Wir haben einen einseitigen Test der Nullhypothese H_0: $p=0.5$ mit einer "Mindesteffektgröße" oder "Indifferenzzone" oder "praktischen Signifikanzgröße" $\delta=0.1$. Wir halten nämlich eine fälschliche Entscheidung für die Nullhypothese nur dann für problematisch, wenn der wahre Parameter um mindestens δ über dem nullhypothetischen Wert liegt. Und genau die Wahrscheinlichkeit einer solchen "relevanten" Fehlentscheidung ist durch unser vorgewähltes β begrenzt.

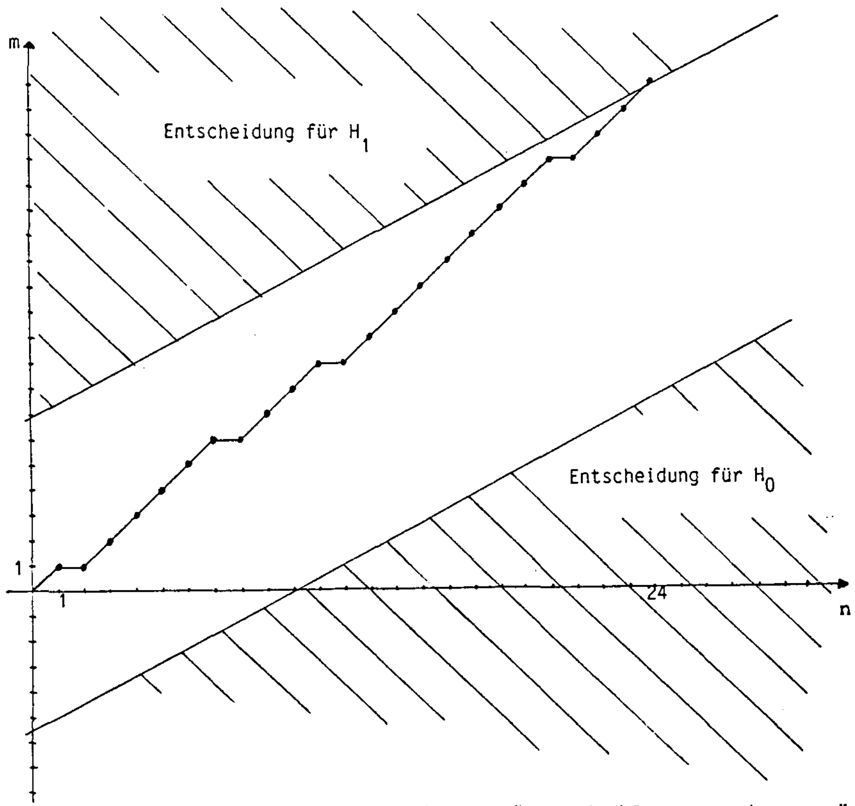

Abb. 1: Plan für einen sequentiellen "Entscheidungsspaziergang" entsprechend dem Beispiel im Text. Jedes erhaltene Datum wird eingetragen. Gerät man dabei in das Gebiet oberhalb der durch $m=0.54n+6.95$ definierten Geraden, so beende man die Datenerhebung und entscheide für H_1. Gerät man in das Gebiet unterhalb der durch $m=0.54n-5.49$ definierten Geraden, so beende man die Datenerhebung und entscheide für H_0. Eingetragen ist als Beispiel ein Experimentverlauf, der nach 24 Durchgängen, in denen das Merkmal 20 mal auftrat, mit der Entscheidung für H_1 endet.

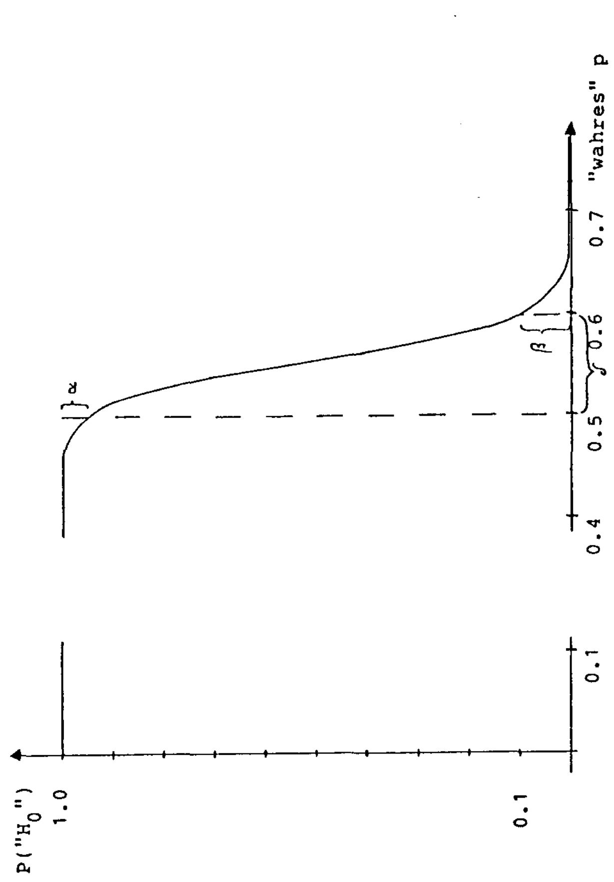

Abb. 2: Verlauf der Wahrscheinlichkeit einer Entscheidung für die Nullhypothese in Abhängigkeit vom "wahren" Parameter (OC-Funktion) beim sequentiellen Quotiententest des Beispiels im Text.

5. Diskussion

Dieser WALDsche Wahrscheinlichkeitsverhältnistest ist natürlich ein sequentielles Analogon zum bekannten Vorzeichen- oder Binomialtest aus der Neyman-Pearson-Statistik. Inzwischen gibt es eine unübersehbare Fülle sequentialstatistischer Testverfahren für die unterschiedlichsten Fragestellungen; insbesondere gibt es etwa auch ein sequentialstatistisches Analogon zum F-Test - und damit die Möglichkeit zum sequentiellen Testen im Rahmen des allgemeinen linearen Modells, also parametrischer Statistik. Häufig lassen sich Sequentialtests in Form von graphischen "Entscheidungsspaziergängen" organisieren und sind dann kinderleicht zu handhaben. In der Forschungspraxis bieten sequentielle Verfahren zwei Vorteile. Erstens den ökonomischen Vorteil der geringeren Stichprobengrößen: Der Erwartungswert für die Stichprobengrößen eines Sequentialtests - diese ist ja hier eine Zufallsgröße - hängt natürlich von der Fragestellung, von α, β, δ und schließlich dem "wahren" Parameter ab; aber er ist deutlich, z.T. bis zu 50 % geringer als die notwendige Stichprobenlänge eines entsprechenden konventionellen Signifikanztestes. Zweitens einen forschungsmethodischen Vorteil: Sequentaialtests funktionieren eben auch in der Praxis nur als im Vorhinein festgelegte Entscheidungsverfahren, und die dafür notwendige explizite Wahl der Fehlerwahrscheinlichkeiten α und β sowie der Mindesteffektgröße δ zwingt zur Reflexion und Transparenz. Insbesondere steigt bei einem Sequentialtest konstruktionsbedingt die OC-Funktion, also die Wahrscheinlichkeit einer Entscheidung zugunsten der Nullhypothese in Abhängigkeit vom "wahren" Parameter, schnell an, wenn man sich von der durch δ definierten Grenze des Indifferenzbereiches dem nullhypothetischen Parameterwert nähert (vgl. Abb. 2). Die Gefahr signifikanter, aber völlig irrelevanter Befunde ist daher bei der Sequentialstatistik deutlich geringer als in der herkömmlichen Neyman-Pearson-Statistik. Kurzum: Sequentialstatistik impliziert eine aufgeklärtere inferenzstatistische Praxis auch im Forschungsalltag; dies macht sie natürlich auch didaktisch interessant, jedenfalls unter der Zielvorgabe einer Propädeutik für eben diese Praxis.

Es erscheint daher lohnenswert, einmal die didaktische Aufbereitung der oben dargestellten Logik des sequentiellen Quotiententests für den Unterricht zu versuchen, wie es ja bereits FREUDENTHAL (1975, p. 97-100) angedeutet hat. Dieser Versuch erscheint erfolgversprechend, werden doch nur relativ einfache Konzepte der Wahrscheinlichkeitsrechnung benutzt, nämlich lediglich die Additivität von Wahrscheinlichkeiten und die Multiplikationsregeln für unabhängige Ereignisse, daneben Regeln für das Rechnen mit Logarithmen. Daß in der Argumentation die Wahrscheinlichkeiten über abzählbar unendlich viele Ereignisse summiert werden, also im Grunde konvergente Reihen vorliegen und strenggenommen nicht nur auf die Additivität, sondern auf die σ-Additivität der Wahrscheinlichkeit rekurriert wird, dürfte angesichts der Plausibilität der Argumentation zu keinem Problem führen. Vielmehr könnte dies eher ein reizvoller Anlaß für eine nachträgliche, wenngleich durchaus verzichtbare begriffliche Präzisierung durch Rückgriff auf Konzepte aus der Infinitesimalrechnung sein. Auch das Übergehen des exakten Nachweises dafür, daß der Test mit Wahrscheinlichkeit 1 irgendwann zu einer Entscheidung zwischen H_0 und H_1 führt, erscheint tragbar; bei Vorliegen des Plans für den graphischen Entscheidungsspaziergang läßt sich im Nachhinein plausibel machen, daß die Wahrscheinlichkeit, daß die Schwankung des Graphen niemals den schmalen Streifen der Indifferenzzone sprengen, 0 beträgt. Ein Nachteil ist sicherlich, daß die Bestimmung der OC-Funktion mathematisch zu schwierig ist, um im Unterricht erarbeitet werden zu können; sie müßte bei Bedarf vom Lehrer einfach mitgeteilt werden. (Hierbei übrigens die Formeln: Beim einseitigen sequentiellen Quotiententest bestimme man für ausgewählte Werte h mit $-\infty < h < +\infty$ - interessant sind eigentlich nur Werte im Bereich zwischen -2 und +2 - zugehörige Werte p nach der Formel

$$p = \frac{1 - \left(\frac{1-p_1}{1-p_0}\right)^h}{\left(\frac{p_1}{p_0}\right)^h - \left(\frac{1-p_1}{1-p_0}\right)^h} .$$

Dann berechnet sich der Wert der OC-Funktion an der Stelle p nach

$$\frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} .$$

Daß schließlich die WALDsche Lösung nur approximativ ist, dies erscheint nicht als Nachteil, sondern eher als Vorteil: Hier gewinnen die Schüler endlich einmal einen Eindruck, wie leistungsfähig und wichtig solche Annäherungen in der Praxis sind, und sie erfahren, daß es neben logischer Stringenz und Exaktheit auch noch andere wichtige Normen in der Mathematik gibt, jedenfalls in der angewandten Mathematik.

Alles in allem erschiene es als lohnenswerte didaktische Aufgabe, im Detail einen unterrichtlichen Einstieg in das Hypothesentesten direkt mit der Sequentialstatistik zu konzipieren. Denn hier wäre von Anfang an der Charakter eines Tests als eines im Vorhinein definierten Entscheidungsverfahrens, das zwei Fehlentscheidungen riskiert und unter Maßgabe eines Relevanzkriteriums deren Risiko begrenzt, überdeutlich; den vielfältigen Irrationalismen im Umgang mit dem Hypothesentesten würde also schon auf der Ebene der einführenden Begrifflichkeiten und Konzepte - also der "Ankerbegriffe" oder "advance organizers" im Sinne AUSUBELS - vorgebeugt. Überdies erschiene der Einstieg in das Hypothesentesten über den Sequentialtest das "natürlichere" Verfahren - jedenfalls in der ent-

scheidungstheoretischen Auffassung des Hypothesentestens -, auch wenn dadurch die statistikhistorische Abfolge auf den Kopf gestellt würde: Denn es ist für die Schüler mit Recht wenig plausibel, weshalb man beim Testen im Vorhinein schon Stichprobengrößen festlegen soll, anstatt sich dabei flexibel der Empirie anzupassen. An diese "Unnatürlichkeiten" des traditionellen Signifikanztestes könnte übrigens ein Unterricht über sequentielles Testen organisch anknüpfen, wenn er nicht als Einstieg, sondern als Erweiterung eines Unterrichtes über Hypothesentesten durchgeführt würde.

Es erscheint daher zumindest erwägenswert, ob nicht der statistikdidaktischen Diskussion - bislang fixiert auf den klassischen Signifikanztest - durch die stärkere Beachtung der Sequentialstatistik neue Ideen und Impulse gegeben werden könnten. Denn, dies als letzte didaktische Spekulation, ein Unterricht über sequentielles Testen könnte überdies in nahezu einmaliger Weise die Paradigmen gleich mehrerer statistischer Schulen repräsentieren, nämlich

erstens den ursprünglichen Signifikanzgedanken Fishers, eine (Null-)Hypothese dann zu verwerfen, wenn Ereignisse auftreten, die - ihre Geltung unterstellt - sehr unwahrscheinlich sind,

zweitens das Testgütekonzept von Neyman und Pearson, denn in der Sequentialstatistik spielt die β -Fehler-Kontrolle eine hervorragende Rolle,

drittens die moderne entscheidungtheoretische Sichtweise, denn in der Sequentialstatistik spielen zumindest indirekt die Kosten der Stichprobennahme eine Rolle, nämlich in ihrer Zielsetzung, die Stichprobe klein zu halten,

und nicht zuletzt viertens den Grundgedanken der Bayes-Statistik, nämlich sich Schritt für Schritt von der Erfahrung belehren zu lassen.

Literatur

BREDENKAMP, J.: Der Signifikanztest in der psychologischen Forschung. Frankfurt: Akademische Verlagsgesellschaft, 1972.

BÜNING, H. & TRENKLER, G.: Nichtparametrische statistische Methoden. Berlin, New York: De Gruyter, 1978.

DIEPGEN, R.: Anwendungsorientierung im Unterricht über statistisches Hypothesentesten - Orientierung an Anwendung oder Orientierung der Anwendung? In: Beiträge zum Mathematikunterricht 1984. Bad Salzdetfurth: Franzbecker, 1984

DIEPGEN, R. 1985a: Probleme eines Statistikunterrichtes nach STRICK-Muster. In: Stochastik in der Schule (1985) Nr. 2, S. 32 - 36.

DIEPGEN, R. 1985b: Signifikanz - na und? Politik und Computer im Statistikunterricht. In: Mathematik lehren Nr. 13, S. 54 - 57.

DIEPGEN, R.: 1985c: Eine Aufgabensequenz zum Hypothesentesten. In: Stochastik in der Schule (1985) Nr. 2, S. 22 - 27 (Teil 1) sowie (1985) Nr. 3, S. 17 - 38 (Teil 2).

EGER, K. H.: Sequential tests. Leipzig: Teubner, 1985.

FISZ, M.:Wahrscheinlichkeitsrechnung und mathematische Statistik. Berlin: VEB Deutscher Verlag der Wissenschaften, 1976.

FREUDENTHAL, H.:Wahrscheinlichkeitsrechnung und Statistik. München, Wien: Oldenbourg, 1975.

GHOSH, B.: Sequential tests of statistical hypotheses. Reading, Mass.: Addison-Wesley, 1970

GOVINDARAJULU, Z.: Sequential statistical procedures. New York: Academic Press, 1975

HARTUNG, J.: Statistik: Lehr- und Handbuch der angewandten Statistik. München, Wien: Oldenbourg, 1984.

HECKENDORF, H.: Grundlagen der sequentiellen Statistik. Leipzig: Teubner, 1982.

SACHS, L.: Statistische Auswertungsmethoden. Berlin, Heidelberg: Springer, 1968.

UHLMANN, W.: Statistische Qualitätskontrolle. Stuttgart: Teubner, 1966

WALD, A.: Sequential tests of statistical hypotheses. In: Annales of Mathematical Statistics 16, S. 117 -186, 1945

WALD; A.: Statistical Decision Functions. New York: John Wiley, 1950

WALD, A.: Sequential analysis. New York: John Wiley, 1966

WEBER, E.: Grundriß der Biologischen Statistik. Anwendungen der mathematischen Statistik in Naturwissenschaft und Technik. Stuttgart: Fischer, 1967.

WETHERILL, G. B:: Sequential Methods in Statistics. London, 1975

WITTE, E. H.: Signifikanztest und statistische Inferenz. Analysen, Probleme, Alternativen. Stuttgart: Enke, 1980.