

Deutsche Zusammenfassung

Die vorliegende Dissertation schlägt Ansätze zur Verbesserung des statistischen Denkens vor. Warum ist Forschung zum Thema "statistisches Denken" notwendig? In seinem Buch *So lügt man mit Statistik* vermutet Walter Krämer (1998), dass in einer durchschnittlichen Zeitung mehr Statistiken zu finden sind, "als Goethe oder Schiller sahen, so lange sie lebten, da heutzutage ein Trommelfeuer von Daten, Zahlen, Fakten, Tabellen, Kurven, Trends und Tests den modernen Medienkonsumenten eindeckt". Er stellt sogar die These auf, dass das Wort "Prozent" mittlerweile das häufigste Substantiv der deutschen Sprache ist.

Die Fähigkeit statistisch zu denken ist natürlich nicht nur im "Alltagsdschungel der Prozente" erforderlich, sondern auch beim Umgang mit empirischen Befunden in der Wissenschaft. Die vorliegende Dissertation geht auf beide Aspekte des statistischen Denkens ein. Der gemeinsame rote Faden aller Kapitel ist das als schwierig geltende Konzept der bedingten Wahrscheinlichkeit. Die gemeinsame Struktur der drei Kapitel ist: In jedem Kapitel wird eine relevante Aufgabe zum Thema bedingte Wahrscheinlichkeit vorgestellt und die Schwierigkeiten, die Menschen beim Lösen dieser Aufgabe haben, werden empirisch untersucht. Die kognitiven Merkmale dieser Schwierigkeiten werden daraufhin ausführlich analysiert. Diese Analyse mündet dann am Ende jedes Kapitels in einen didaktischen Vorschlag zur Überwindung der jeweiligen Verständnisschwierigkeiten. In Kapitel 1 und 2 werden diese Vorschläge empirisch überprüft und in Kapitel 3 wird eine komplette Unterrichtseinheit bereit gestellt.

Die Dissertation behandelt folgende Themen: In der Einleitung (*Introduction*) wird ein historischer Überblick über die wissenschaftliche Erforschung des statistischen Denkens gegeben – beginnend bei der Erfindung der Wahrscheinlichkeitsrechnung bis zum heutigen Stand der Forschung. Das Konzept der bedingten Wahrscheinlichkeiten wurde im Wesentlichen aus 2 Gründen als Untersuchungsthema gewählt: Zum Einen gilt es als ein "besonders verwirrendes Konzept" (Moore, 1997); es herrscht großer Bedarf an didaktischen Vorschlägen auf diesem Gebiet. Zum Anderen ist dieses Konzept ein elementarer Grundbaustein für die Fähigkeit des statistischen Schließens in der Wissenschaft. Es gibt in der Inferenzstatistik zwei zentrale Paradigmen des

Hypothesentestens: Die Bayes-Statistik und das klassische Signifikanztesten. Beide Paradigmen fußen auf der Betrachtung von bedingten Wahrscheinlichkeiten. Während die "Bayesianer" die Wahrscheinlichkeit des Zutreffens einer Hypothese unter der Bedingung der vorliegenden Daten berechnen, betrachtet man z.B. bei Nullhypothesentests die Wahrscheinlichkeit der gefundenen (oder noch extremeren) Daten unter der Bedingung "die Nullhypothese trifft zu".

Im ersten Kapitel (*Chapter 1*) wird ein Ansatz zur Verbesserung Bayesianischen Denkens vorgestellt, der von Gigerenzer und Hoffrage (1995) vorgeschlagen wurde. Gigerenzer und Hoffrage konnten empirisch zeigen, dass sich die Urteile von Versuchspersonen deutlich verbessern, wenn man in sogenannten "Bayesianischen Aufgaben" die gegebenen (bedingten) Wahrscheinlichkeiten durch "natürliche Häufigkeiten" ersetzt. In Kapitel 1 wird dieser Ansatz vorgestellt und dann verallgemeinert: Bislang wurden in der kognitionspsychologischen Forschung zu diesem Thema nur solche Bayesianische Aufgaben betrachtet, bei denen von einem *binären Cue* (z.B. ein medizinischer Test mit den möglichen Ausprägungen "positiv" und "negativ") auf ein ebenfalls *binäres Kriterium* (z.B.: liegt eine Krankheit vor oder nicht?) geschlossen werden sollte. Die Realität ist in der Regel aber vielschichtiger: Oftmals ist ein medizinischer Test nicht binär (es könnten z.B. auch unklare Testbefunde vorkommen) oder ein Test könnte sensitiv für mehrere Krankheiten sein (z.B.: liegt Krankheit A, B oder C vor?). Auch die Annahme, dass eine Krankheit nur mit einem Test diagnostiziert wird, ist oftmals nicht erfüllt.

In der Literatur wurde bezweifelt, ob sich das Konzept der "natürlichen Häufigkeiten" von Gigerenzer und Hoffrage (1995) auch in komplexen Situationen zur Verbesserung der statistischen Einsicht anwenden lässt (Massaro, 1998). In Kapitel 1 wird empirisch gezeigt, dass sich "natürliche Häufigkeiten" in diesen komplexen Situationen nicht nur anwenden lassen, sondern dass ihr Effekt sogar noch größer ist als in einfach strukturierten Bayesianischen Situationen. Kapitel 1 endet mit einer Diskussion des Konzepts der "natürlichen Häufigkeiten" und ihrer Definition für den allgemeinen Fall.

Das zweite Kapitel (*Chapter 2*) beschäftigt sich mit der wohl berühmtesten "Bayesianischen Kopfnuss", dem "Monty Hall Dilemma" (bzw. "3-Türen-Problem" bzw.

"Ziegenproblem"). Dieses stochastische Problem erhitze 1991 die Gemüter in den USA, nachdem Marilyn vos Savant die richtige, aber scheinbar widersinnige Lösung dieses Problems in ihrer Kolumne "Ask Marylin" im *Parade Magazine* veröffentlichte. Marylin vos Savant war zu dieser Zeit im Guinness Buch der Weltrekorde als die intelligenteste Frau der Welt aufgeführt (IQ = 228) und Leser konnten sie fragen, was immer sie wollten. Ein Leser stellte ihr folgende Frage: *Sie nehmen an einer Spielshow teil, bei der Sie eine von drei verschlossenen Türen auswählen sollen. Hinter einer Tür wartet der Preis, ein Auto, hinter den anderen beiden stehen Ziegen. Sie zeigen auf eine Tür, sagen wir Nummer eins. Sie bleibt vorerst geschlossen. Der Moderator (Monty Hall) weiß, hinter welcher Tür sich das Auto befindet; mit den Worten "Ich zeige Ihnen mal was" öffnet er eine andere Tür, zum Beispiel Nummer drei, und eine Ziege schaut ins Publikum. Er fragt: "Bleiben Sie bei Nummer eins, oder wählen Sie Nummer zwei?"* Als Marylin vos Savant in ihrer Kolumne antwortete, es sei von Vorteil für den Kandidaten, die Türe zu wechseln, erhielt sie ca. 10.000 (!) Protestbriefe, darunter auch viele Beschimpfungen von Mathematikern. In der Folgezeit entflammte in den USA eine öffentliche Debatte und das Problem wurde sogar auf der Titelseite der New York Times diskutiert.

In den letzten 10 Jahren gab es sowohl in der Didaktik der Stochastik als auch in der Kognitionspsychologie zahlreiche Versuche, die richtige Lösung ("es ist besser zu wechseln") zu verdeutlichen. In Kapitel 2 der Dissertation wird empirisch gezeigt, dass durch den Einsatz verschiedener kognitionspsychologischer Konzepte ("natürliche Häufigkeiten", "mentale Modelle", "Perspektivenwechsel" und "Weniger-ist-mehr"-Effekt) das Problem so formuliert werden kann, dass 60% der Versuchspersonen sich für "Wechseln" entscheiden. Dieser Wert stellt innerhalb des Paradigmas "Problemumformulierung" eine neue Bestmarke dar. Theoretisch relevante Ergebnisse liefert auch die Untersuchung der wechselseitigen Abhängigkeiten der 4 untersuchten Konzepte. Bislang wurden diese Konzepte in der Kognitionspsychologie nur getrennt betrachtet.

Im dritten Kapitel (*Chapter 3*) werden Verständnisschwierigkeiten bei der Interpretation eines signifikanten Testergebnisses beleuchtet. Mit einem Signifikanztest berechnet man die Wahrscheinlichkeit der vorliegenden (oder noch unwahrscheinlicherer) Daten unter der Voraussetzung, dass die Nullhypothese stimmt. Aussagen über die

Wahrscheinlichkeit von Hypothesen sind mit Signifikanztests nicht möglich. Diese und andere Fehlvorstellungen über ein signifikantes Testergebnis sind unter Studierenden aber weitverbreitet (z.B. Falk & Greenbaum, 1995). In diesem Kapitel wird erstmals gezeigt, dass auch Dozenten aus dem Bereich "Statistik für Psychologen" diesen Fehlvorstellungen unterliegen. In einer empirischen Untersuchung ließen sich 80% der Statistikdozenten von den vorgegebenen Misinterpretationen verleiten (lediglich die an der Untersuchung teilnehmenden Professoren zeigten keine Fehlinterpretationen). Diese Ergebnisse sind alarmierend, wenn man bedenkt, welche Bedeutung Signifikanztests in der methodischen Ausbildung von Psychologiestudenten haben.

Ausgehend von den häufigsten Misinterpretationen wird in Kapitel 3 eine didaktisches Konzept vorgestellt, das diesen Fehlvorstellungen entgegenarbeitet. Eine Implementation dieses Konzepts in eine Statistikvorlesung sollte gewährleisten, dass das Wort "Signifikanz" für Studenten nicht nur ein diffuser, mechanisch verwendeter Begriff bleibt, sondern dass klar wird, welche Aussagen sich mit Signifikanztests machen lassen und welche nicht.

In einer abschließenden Diskussion (*General Discussion*) werden die wichtigsten Ergebnisse der drei Kapitel noch einmal zusammengefasst und es wird ein Ausblick auf weiterführende Fragestellungen gegeben.