

STATISTISCHE FEHLER*)

ULRICH NIEMEYER

Mit Statistiken läßt sich nach landläufiger Meinung jede Behauptung beweisen - genauso wie auch ihr Gegenteil. Diese Skepsis gegenüber einem Spezialgebiet der Mathematik scheint auch begründet. Allzuoft haben Werber der Wirtschaft oder der Politik mit Tabellen und Graphiken zweifelhaften Aussagen den Anschein gegeben, sie seien wissenschaftlich abgesichert. Die groben Tricks, die in der beschreibenden Statistik angewandt werden, fallen zumindest dem wissenschaftlich Geschulten ins Auge.

Da gibt es den Maßstabeffekt: Ein Anstieg der Arbeitslosenzahl etwa von 1,1 auf 1,2 Millionen erscheint in einer Graphik viel steiler und damit bedrohlicher, wenn man auf der Ordinate nur die Werte oberhalb einer Million findet. Auch die Wahl der Maßeinheiten kann das Ergebnis optisch beeinflussen. So wurde jüngst in Sachen Umweltschutz darüber gestritten, ob ein bestimmter Verschmutzungsgrad z.B. als 5ppm (parts per million) oder als 5000 ppb (part per billion = 10^0) anzugeben sei. Das eine verharmlose, das andere übertreibe. (Korrekt sind übrigens beide Schreibweisen nicht: Die DIN-Vorschrift verlangt die Angabe von Volumenverhältnissen, also z.B. Milliliter pro Kubikmeter.

Neben solchen Eingriffen in die Darstellung einer Statistik gibt es aber auch Möglichkeiten, das Ergebnis selbst im gewünschten Sinne zu beeinflussen, etwa durch die geeignete Wahl der Basisdaten: Hat man etwa Werte aus mehreren Jahrzehnten zur Verfügung, so mag ein langfristiger Trend zwischen 1965 und 1980 berechnet, vielleicht ganz anders, willkommener aussehen als der zwischen 1960 und 1980. Tut er das nicht, so liefert vielleicht ein Vergleich von 5-Jahresdurchschnitten ein günstigeres Ergebnis. Oder nach einer alten Physiker-Faustregel werden der höchste und der niedrigste Wert aus der Datenreihe als 'Ausreißer' gestrichen und nicht in die Auswertung mit einbezogen.

*) Erschienen in 'UMSCHAU IN WISSENSCHAFT UND FORSCHUNG'
(1981) Heft 12, Brönner Verlag, Frankfurt/Main.

Nachdruck mit Genehmigung des Verlags.

Wie vielfältig sind erst die Fehlermöglichkeiten, wenn auch die Datenerhebung selbst der Willkür des Statistikers unterworfen ist. Auch bei bester Absicht kann da manches schief laufen. Dabei wollen wir gar nicht von den Demographen sprechen, die sich oft auf subjektive Personenbefragungen stützen - denken wir eher etwa an einen Agronomen, der die Ergiebigkeit einer neuen Weizensorte nachprüfen will. Soll er eine Hilfskraft aufs Versuchsfeld schicken mit der Anweisung, 'zufällig' hundert Halme auszusuchen? Für diesen Fall müssen neben einem sorgfältig aufgestellten Stichprobenplan auch noch Hilfsgeräte wie z.B. eine Dreieckskonstruktion bereitgestellt werden. ("Ausgewählt" sind die Halme, die sich innerhalb einer aufs Feld geworfenen Triangel befinden.) Ein solcher Aufwand ist sicher im Interesse eines unverfälschten Untersuchungsergebnisses in anderen Fachbereichen ebenso notwendig, aber ist er auch die Regel?

Weniger offensichtlich sind die Fehler, die sich - versehentlich oder mit Absicht - bei der Anwendung der schließenden Statistik einschleichen. Mit Verfahren der schließenden Statistik kann nach der Erprobung eines Medikaments seine Qualität mit der seiner Konkurrenzpräparate verglichen werden. In solchen statistischen Testverfahren versucht man, zu möglichst aussagekräftigen Vergleichen zu kommen und dabei Irrtümer weitgehend auszuschließen. Je nach Ausgangslage stehen dafür dem Statistiker eine Reihe von Rechenmodellen zur Verfügung. Hier ist die Versuchung groß, zusätzliche, nicht gerechtfertigte Annahmen zu machen. Wenn etwa die beobachteten Werte sich um einen bestimmten Mittelwert häufen, fühlt man sich an die Gaußsche Glockenkurve erinnert und setzt in den weiteren Berechnungen voraus, daß es sich um eine Normalverteilung handelt. Diese zusätzliche unkorrekte Voraussetzung ermöglicht im Ergebnis eine schärfere Aussage, als sie eigentlich gerechtfertigt wäre.

Ganz besonders verlockend wird die Versuchung zur Manipulation, wenn alternative Verfahren als Rechenprogramme schon auf Abruf bereitstehen. War man in der Vor-Computerzeit wegen des hohen Rechenaufwandes gezwungen, sich a priori für ein bestimmtes Testverfahren zu entscheiden, so ist es jetzt möglich, gleich

mehrere durchzurechnen. Das am Ende für die Veröffentlichung ausgewählte Programm ist dann oft dasjenige, welches am ehesten das erwartete Ergebnis liefert. Warum fehlen bei statistischen Aussagen so oft die Informationen, die ihre Herleitung nachvollziehbar machen? Zugegeben, das fordert vom Verfasser mehr Sorgfalt, vom Leser mehr Zeit und mehr Vorwissen. Die Glaubwürdigkeit der Wissenschaft müßte dies Opfer wert sein.