

Einer vergleichenden Diskussion modernisierungstheoretischer und marxistischer Ansätze steht Frau SCHISLER zufolge ein prinzipielleres Argument entgegen, und dieses Argument bildet den Angelpunkt der Kritik an meinem Beitrag. Marxistische und modernisierungstheoretische Ansätze unterschieden sich demnach in folgenreicher Weise im Hinblick auf ihr Erkenntnisinteresse und seien insofern „Inkommensurable“ (S. 282). Haben wir es bei den Modernisierungstheorien in der von DEUTSCH und ZAPF repräsentierten Variante mit einem „empirischen Interesse“ zu tun, so bei den Marxisten mit einem „kritisch-analytischen“, das Frau SCHISLER zufolge zwar auch darauf angelegt ist, „den historischen Entwicklungsgang mit der Dialektik von Produktionsverhältnissen und Produktivkräften zu erklären“ (S. 282), jedoch nach dieser Auffassung anscheinend nicht die Formulierung empirisch-analytischer Aussagen zuläßt. Mit dieser überaus handlich anmutenden Konstruktion gerät Frau SCHISLER indessen schnell in Schwierigkeiten. Denn kurz darauf kommt sie im Anschluß an BÜHL darauf zu sprechen, daß die der traditionellen Soziologie entstammenden Autoren und zumal PARSONS dazu neigen, den Entwicklungsprozess „von seinem Endzustand her“ zu definieren (S. 286). Damit bleiben auch sie hinter dem Anspruch einer historisch gerichteten empirisch-analytischen Theorie zurück, und argumentieren ließe sich in der Tat, daß gerade die Ansätze, auf die Frau SCHISLER zuvor sich beruft, dem von ihr über die Marxisten verhängten Verdikt verfallen.

Ohne auf den Positivismus-Streit zurückfallen zu wollen, würde ich diese Schwierigkeiten als Beleg dafür ansehen, daß wissenschaftliche Erkenntnis im Bereich der Sozialwissenschaften sich dem Einfluß vorgängiger erkenntnisleitender Interessen nicht entziehen kann. Mit Frau SCHISLER würde ich allerdings meinen, daß wissenschaftliche Forschung zwar diese Interessen sich zu vergegenwärtigen hat, dabei aber doch auf die Formulierung empirisch überprüfbarer Sätze angewiesen ist. Damit aber ist man durchaus in Übereinstimmung auch mit der Kritik der politischen Ökonomie, die sich wohl als Theorie in praktischer Absicht versteht, aber doch darauf angelegt ist, das „Bewegungsgesetz der modernen Gesellschaft“ aufzudecken, und das heißt so viel wie überprüfbare Aussagen über empirische Entwicklungstrends zu formulieren.

Was die Kategorien instrumentellen Handelns und symbolisch vermittelter Interaktion angeht,

so habe ich sie herangezogen, weil sie mir geeignet erschienen, die hier strittigen Dimensionen von ökonomischen und nicht-ökonomischen Bedingungen gesellschaftlicher Entwicklung genauer zu bestimmen als mit den geläufigen Begriffspaaren. Wie Frau SCHISLER sehe auch ich, daß diese Kategorien zunächst auf der Ebene einer allgemeinen Theorie sozialen Handelns liegen, darin nicht ganz unähnlich den Kategorien „Produktivkräfte“ und „Produktionsverhältnisse“. Wie diese und in angemessenerer Weise als diese erlauben sie es indessen, die spezifisch historischen Formen gesellschaftlicher Arbeit auf jenen doppelten Zweck hin zu bestimmen, von dem ich im Schlußabsatz meines Beitrags spreche. Das ist, wenn ich mich nicht täusche, auch die Intention, die sich bei HABERMAS mit diesen Kategorien verbindet und die sich bei ihm mit der programmatischen Formel von einem „historischen Funktionalismus“ umschrieben findet.

Mir ist völlig bewußt, daß hier die eigentlichen Probleme historisch gerichteter gesamtgesellschaftlicher Analysen erst beginnen. Geklärt werden können sie nach meiner Auffassung nur im Zusammenhang materialer Analysen, denen in der Tat mein eigentliches Interesse gilt. Voraussetzung dafür aber ist, daß man sich anhand der vorliegenden theoretischen Ansätze der kategorialen Bestimmungen einer zulänglichen Entwicklungsforschung versichert. Darum habe ich mich, wie unzulänglich auch immer, bemüht, und gerade darauf will sich Frau SCHISLER unter Rekurs auf Formeln wie „verzerrte Fragestellung“ und „Inkommensurable“ nicht einlassen. Und weil sie sich darauf nicht einlassen will, fällt es so schwer, ihrer Kritik ein konstruktives Moment abzugewinnen.

G. BRANDT, Prof. Dr., Universität Frankfurt
Institut für Sozialforschung/Seminar
6 Frankfurt/M., Myliusstraße 30

Über die Anwendung statistischer Tests

Ein Kommentar zu „Statistische Signifikanz und sozialwissenschaftliche Relevanz“ von Jürgen Kriz

KRIZ (1972) versucht in seinem Beitrag aufzuzeigen, daß die bisherigen statistischen Entscheidungsmodelle nur Entscheidungen ermöglichen, die „sozialwissenschaftlich irrelevant“ sind. Nun

ist durchaus zuzugeben, daß die Entwicklung statistischer Modelle sich im wesentlichen nicht an den Erfordernissen der Soziologie orientiert hat, daß also eine ganze Reihe von Modellen in der Soziologie nicht sinnvoll angewandt werden können. Auf das von KRIZ gewählte Beispiel trifft diese Kritik jedoch nicht zu, zumal er die Frage der Intervallskalierung bewußt akklammert.

Als erstes ist anzumerken, daß eine Nullhypothese in der statistischen Testtheorie nicht notwendig besagt, daß ein Parameter, hier: $\mu_1 - \mu_2$, gleich Null ist. Als Nullhypothese wird eine Hypothese bezeichnet, wenn sie aufgrund früherer Erfahrungen oder theoretischer Einsichten als die wahrscheinlichste erscheint. Was also die Nullhypothese ist, wird in keinem Fall von dem statistischen Modell bestimmt, sondern muß stets aus dem konkreten Forschungsprozeß heraus entschieden werden. Dabei braucht die Nullhypothese keinesfalls einfach zu sein, kann also auch die Form $H_0 : |\mu_1 - \mu_2| \geq d$ haben. Aber selbst wenn eine einfache Nullhypothese vorliegt, etwa $H_0 : \mu_1 = \mu_2$, so ist das Argument, daß eine minimale Abweichung – „möglicherweise erst in der 10. Stelle hinter dem Komma“ – stets vorhanden sein wird und daß deshalb die Hypothese unbrauchbar sei, absurd. Sicher gilt für die von KRIZ genannten Beispiele keine „streng mathematische Gleichheit“, aber ganz sicher würde eine minimale Abweichung bei keinem praktisch relevanten Stichprobenumfang mit einer Wahrscheinlichkeit, die wesentlich über dem Signifikanzniveau liegt, zu Tage treten. Ein Blick auf die von FERRIS et al. (1946) berechneten und graphisch dargestellten Operationscharakteristiken der Standardtests macht schnell die praktische Irrelevanz dieser Argumentation sichtbar. Ist allerdings aufgrund theoretischer Einsichten ein nicht nur minimaler Unterschied zu erwarten, so erscheint es sinnvoll, einem solchen Unterschied mit Hilfe des Prinzips der „proportionalen Irrtumverminderung“ eine inhaltliche Interpretation zu geben und somit eine Möglichkeit zu schaffen, inhaltlich zu beschreiben, was ein „relevanter Unterschied“ ist. Bleiben wir beim KRIZschen Beispiel der beiden Variablen „Intelligenz bei Männern“ und „Intelligenz bei Frauen“, so beträgt die Wahrscheinlichkeit P_2 , aufgrund der Kenntnis der wahren Differenz d der beiden Mittelwerte und der Intelligenz einer untersuchten Person das Geschlecht richtig zu bestimmen

$$P_2 = 2 P \left(z \leq \frac{d}{2\sigma} \right) - 1,$$

wie von KRIZ in etwa angegeben. Ohne diese Kenntnis könnte man nur Gleichverteilung annehmen, also $P_1 = 0.5$. Mithin beträgt die proportionale Irrtumverminderung

$$\lambda = \frac{(1 - P_1) - (1 - P_2)}{(1 - P_1)} = \frac{P_2 - 0.5}{0.5}$$

Daraus ergibt sich beispielsweise für $\lambda = 20\%$ $d = 1.7\sigma$ und für $\lambda = 60\%$ $d = 2.56\sigma$. Diese Ausführungen und bei näherer Betrachtung auch die KRIZschen machen deutlich, daß mit der Wahl von P_2 der minimale relevante Unterschied D bestimmt ist. Wieso KRIZ meint, daß D und P_2 unabhängig voneinander frei wählbar seien, bleibt unerfindlich. Nur als Druckfehler muß man in diesem Zusammenhang wohl anmerken, daß einerseits β die dritte frei wählbare Größe (KRIZ' Empfehlung $\beta = 0.01$ oder 0.001), andererseits aber β aus den „ersten Daten“ geschätzt werden soll.

Abschließend soll kurz skizziert werden, wie „bisherige statistische Entscheidungsmodelle“ auch im KRIZschen Sinne relevante Ergebnisse liefern können.

Wir gehen davon aus, daß es gerechtfertigt ist anzunehmen, daß ein Unterschied besteht.

1. Schritt: Festlegung der gewünschten Irrtumverminderung, d.h. Präzisierung des Begriffs des relevanten Unterschiedes.

Daraus ergibt sich nach der obigen Formel D .

2. Schritt: Formulierung der Nullhypothese, dabei sei bekannt, etwa durch vorherigen Test auf $H_0 : \mu_1 - \mu_2 \geq 0$, daß $\mu_1 \geq \mu_2$:

$$H_0 : \mu_1 - \mu_2 \geq D$$

$$H_1 : \mu_1 - \mu_2 < D$$

3. Schritt: Festlegung des Signifikanzniveaus α , d.h. hier der Grenze für die Wahrscheinlichkeit, einen relevanten Unterschied zu übersehen. KRIZ empfiehlt hierfür 0.01 oder 0.001 . Eine solche Entscheidung sollte jedoch nicht schematisch erfolgen, da die Größe von α den Fehler zweiter Art beeinflusst.

4. Schritt: Festlegung des Stichprobenumfangs, sofern dieser nicht ohnehin schon festliegt. Bei vorgegebenen α bestimmt der Stichprobenum-

fang die Güte des Tests, je höher der Umfang, desto kleiner ein Fehler zweiter Art (β -Fehler), d.h. desto geringer wird die Wahrscheinlichkeit, die Nullhypothese anzunehmen, obwohl sie nicht wahr ist. Es ist einleuchtend, daß man diesen Fehler minimieren möchte, andererseits sind dem Stichprobenumfang aus erhebungstechnischen Gründen Grenzen gesetzt. Absolut unsinnig erscheint in Anbetracht dieser Tatsache die KRIZsche Forderung, unter Umständen „die Anzahl durch Auswahl mit einem Zufallsgenerator“ zu dezimieren. Näheres zu dieser Frage findet sich in den besseren Lehrbüchern, wie beispielsweise HAYS (1970).

Offen bleibt die Frage, ob es nicht unter Umständen sinnvoller ist, Konfidenzgrenzen zu berechnen statt einen Signifikanztest durchzuführen. Jedenfalls stehen beide Modelle seit mehreren Jahrzehnten zur Verfügung.

Literatur:

- FERRIS, C.D., F.E. GRUBBS and C.L. WEAVER, 1946: Operating Characteristics for the Common Statistical Tests of Significance. *Annals of Mathematical Statistics* 17, 178 ff.
- HAYS, W.L., 1970: *Statistics*. London: Holt, Rinehart and Winston.
- KRIZ, J., 1972: Statistische Signifikanz und sozialwissenschaftliche Relevanz. *Zeitschrift für Soziologie* 1, 47–51.

Dr. MANFRED KÜCHLER,
Universität Bielefeld, Fakultät Soziologie
48 Bielefeld, Kurt-Schumacher-Straße 6

Was ist ein Entscheidungsmodell?

Eine Antwort auf Manfred Küchler

Zunächst soll die Grundposition, aus der mein von KÜCHLER kritizierter *Diskussionsbeitrag* (und als solchen möchte ich meinen Artikel verstanden wissen) entspringt, kurz umrissen werden:

Wenn in den Sozialwissenschaften die Frage aufgeworfen wird, ob bestimmte Unterschiede (bzw. Zusammenhänge) existieren oder nicht, sind damit ausschließlich im Hinblick auf irgendeinen Erkenntnisgewinn *relevante* Unterschiede gemeint – ganz sicher jedenfalls nicht mathematische,

denn die Frage bezieht sich zunächst auf das sozialwissenschaftliche Relativ, das aus empirischen Objekten und Relationen zwischen diesen besteht; Begriffe wie „Zahlen“ oder „Mathematik“ sind auf dieser Betrachtungsebene inhaltsleer. Versucht man nun, zur Beantwortung der gestellten Frage ein mathematisch-statistisches Modell – z.B. einen Signifikanztest – zu Hilfe zu ziehen, so ist eine *notwendige* Voraussetzung die isomorphe (homomorphe) Transformation des empirischen in ein numerisches Relativ. Mein Beitrag sollte nun deutlich herausstellen, daß diese Voraussetzung keineswegs *hinreichend* ist, da z.B. „statistische Signifikanz“ keine adäquate Übersetzung der „Relevanz“ darstellt; es müssen daher *zusätzlich* zum Signifikanztest Überlegungen zur Relevanz angestellt werden. Wenn man Signifikanztests also *allein* als Relevanzkriterien in Entscheidungsmodellen verwendet – und dies ist in der Praxis leider eher die Regel – stellt dies einen „Mißbrauch der Signifikanztests“ dar – wie ich in der Zusammenfassung schrieb – bzw. zeugt dies von „einem unzureichenden Verständnis des Anwenders“, wie KÜCHLER formulierte. Mein Artikel fordert daher, den Relevanzbegriff adäquater zu übersetzen, indem in das Entscheidungsmodell *zusätzlich* zum Signifikanztest ein mathematisch formuliertes Relevanzkriterium eingebaut wird. Ein Teil der „Interpretation“ des Tests wird somit durch das Modell vorweggenommen und der Anwender *gezwungen*, sich mit der Relevanz seiner Ergebnisse auseinanderzusetzen.

KÜCHLERS Argumentation geht nun m.E. von einem Mißverständnis aus, der auf einem terminologischen Unterschied zwischen dem, was er und was ich unter einem „Entscheidungsmodell“ verstehen, beruht. Während bei KÜCHLER „Entscheidungsmodell“ und „Signifikanztest“ synonym gebraucht werden, wird von mir der Signifikanztest als *ein Teil* des Entscheidungsmodells verstanden, wobei weitere Teile des Modells die *Anwendung* des Tests bei der Hypothesenprüfung und deren *Interpretation* z.B. im Hinblick auf die Relevanz darstellen. Diese unterschiedliche Terminologie kommt zunächst in KÜCHLERS Inhaltsangabe zum Tragen: „mangelhafte Nähe des (Entscheidungs) Modells zur sozialwissenschaftlichen Fragestellung“ und „unzureichendes Verständnis (des Signifikanztests) des Anwenders“ sind dann keine gegensätzlichen Interpretationen mehr für die Ursache des mangelhaften Gebrauchs von Signifikanztest in der Praxis.