Klein, Ingo

**Working Paper**

# Grundlagenstreit in der Statistik

Diskussionspapiere // Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Statistik und Ökonometrie, No. 60/2004

# Friedrich-Alexander-Universität Erlangen-Nürnberg

## Wirtschafts-und Sozialwissenschaftliche Fakultät



**Diskussionspapier**

**60 / 2004**


# Grundlagenstreit in der Statistik


**Ingo Klein**

Lehrstuhl für Statistik und Ökonometrie

Lange Gasse 20 · D-90403 Nürnberg

# Grundlagenstreit in der Statistik

**Ingo Klein**

Lehrstuhl für Statistik und Ökonometrie
Universität Erlangen-Nürnberg
Lange Gasse 20
90403 Nürnberg
ingo.klein@wiso.uni-erlangen.de

## Gliederung

# 1 Einleitung

Mein heutiger Vortrag beschäftigt sich nicht mit unserem täglichen Brot – nämlich der Anwendung der Statistik auf sozio-ökonomische Probleme – sondern hat zum Inhalt, Sie mit älteren und jüngeren Kontroversen über die Grundlagen der Statistik bekannt zu machen. Ein deutliches Indiz dafür, dass sich die Statistik nach ca. vierzig Jahren des – lassen Sie mich es provokativ formulieren – anglo-amerikanisch geprägten Primats der pragmatisch-orientierten Methodenentwicklung wieder auf Ihre Grundlagen besinnt, sind eine Reihe von Vorträgen und Zeitschriftenbeiträge der jüngsten Vergangenheit. So betitelt z.B. der neue Präsident der Royal Statistical Society seine Presidential Address mit der Frage "What is Statistics?". Seine mehrseitige Antwort möchte ich Ihnen ersparen, um den namhaften Vertreter der sog. robusten Statistik Frank Hampel (1994) zu zitieren, der kurz und treffend

"die Statistik als die Wissenschaft **und** Kunst, aus Daten zu lernen"

charakterisiert. Er ordnet sie wegen ihrer Aufgabe, unter Berücksichtigung **unvermeidlicher** Unschärfen und Unsicherheiten, Strukturen in Daten zu erfassen, den grundlegenden Strukturwissenschaften zu – wie die Mathematik und die Informatik. Als eine solche tritt sie – wenn auch oft unbemerkt helfend im Hintergrund – in nahezu allen empirischen Wissenschaften, aber auch in den Geisteswissenschaften in Erscheinung. So vermutet Hampel (1994) zu recht, dass es wenig bekannt sein dürfte, dass unser gesamtes astronomisches Weltbild bis zur Urknalltheorie – wie er sagt – auf dem "kühnen Aufeinandertürmen verschiedener statistischer Extrapolationen" beruht.

Auch in unserer ganz alltäglichen Welt spielt die Statistik eine wichtige Rolle als Mittel der Kommunikation und der Argumentation. Dies ist um so erstaunlicher, als sie gemeinhin nicht vorurteilslos betrachtet wird, wie die weite Verbreitung des bekannten Zitates von Disraeli belegt, in dem die Statistik zur höchsten Form der Lüge befördert wird. Zur Erhärtung dieses Vorurteils möchte ich Ihnen die folgende Äußerung Stephen Leacocks (1869-1944) zur Statistik nicht vorenthalten. Er meint:

*"In der alten Zeit gab es keine Statistik, und daher mussten die Leute lügen. So ist denn die Literatur voll von gewaltigen Übertreibungen – es wimmelt nur so von Riesen und Wundern! Damals log man also, aber heute hat man die Statistik dafür."*

Trotz aller – absichtlichen und unabsichtlichen – missbräuchlichen Verwendungen statistischer Methoden muss es offenbar so viele gelungene Anwendungen geben, dass die Statistik als eine insgesamt recht gefestigte Wissenschaft gelten kann. Das Ergebnis einer vor einigen Jahren unter den betriebswirtschaftlichen Lehrstühlen unserer Fakultät durchgeführten Umfrage, mit dem Inhalt, festzustellen, welche statistischen Lehrinhalte bereits die statistische Grundausbildung bereitstellen sollte, zeigt drastisch, welche herausragende Rolle die Statistik in der betriebswirtschaftlichen Lehre und Forschung spielt.

Umso mehr mag es den Außenstehenden verwundern, dass die philosophischen Grundlagen der Statistik keineswegs geklärt sind. Lassen Sie uns zunächst einen – allerdings sehr kursorischen – Blick auf die historischen Wurzeln der Statistik werfen.

## 2 Universitätsstatistik[1]

Selbst wenn für die Entstehung der Statistik ein ca. zweieinhalbtausend Jahre alter ägyptischer Gedenkstein bemüht wird, der Bevölkerungszahlen dokumentiert die aus einer Erhebung zur Organisation des Pyramidenbaus stammen, kann als Wurzel der wissenschaftlichen Statistik die Universitäts- oder Kathederstatistik des 17. und 18. Jahrhunderts angesehen werden. Geprägt wurde diese von dem 1656 erschienenen Buch "Teutscher Fürstenstaat" des damals berühmten Historikers und Staatsmanns Veit Ludwig von Seckendorff (1626-1692). Daraus entwickelte sich die Lehre von den **Staatsmerkwürdigkeiten**, wobei "merkwürdig" im Sinne von "bemerkenswert" zu interpretieren ist. Eine von dem Jenaer und Hallenser Professor Martin Schmeitzel (1679-1747) gehaltene Vorlesung mit dem Titel " Collegium politico-**statisticum**" prägte den Namen Statistik. Sein Schüler Gottfried Achenwall (1719-1772) bot dann auch im Jahre 1747 an der Universität Göttingen die erste Veranstaltung zur **Staatenkunde** unter dem Titel Statistik an. Dessen Nachfolger August Ludwig von Schlözer (1735-1809) gilt im übrigen als Vater der deutschen Publizistik. Aufgabengebiet der Lehre von den Staatsmerkwürdigkeiten war die Sammlung und Zusammenfassung von Informationen zur Beschreibung geographischer Gegebenheiten, der Bevölkerung, der Wirtschaft, der Verwaltung und des Militärs. Dabei bediente man sich einer mehr **literarischen** Form der Bearbeitung.

## 3 Politische Arithmetik

In England entwickelte sich nahezu zeitgleich die Politische Arithmetik. Diese verwendete nicht nur umfangreiche tabellarische Darstellungen, sondern legte auch bereits erste Analysen vor, in denen induktive Schlüsse verwandt wurden. So publizierte John Graunt im Jahre 1662 eine Schrift, deren Aussagen auf Geburts- und Sterbelisten Londons beruhten. Der bekannte Astronom Edmund Halley (1656-1742) veröffentlichte 1693 die erste vollständige aus Kirchenbüchern der Stadt Breslau erstellte Sterbetafel – mit den daraus resultieren Prämienberechnungen für Lebensversicherungen.

Obwohl die politische Arithmetik sich rasch über Europa verbreitete, hielt sie wegen der Vorherrschaft der Universitätsstatistik nur sehr zögerlich in Deutschland Einzug. Wegen des mehr technischen und formal- analytischen Aspekts wurden die Vertreter der politischen Arithmetik von den Anhängern Achenwalls als "Tabellenknechte" beschimpft, die eine "gemeine" statt einer "höheren" Statistik betrieben. Im Jahre 1806 ist den Göttingschen gelehrten Anzeigen zu entnehmen:

*„Die ganze Wissenschaft der Statistik, eine der edelsten, ist durch die politischen Arithmetiker um alles Leben, um allen Geist gebracht und zu einem Skelett, zu einem wahren Kadaver herabgewürdigt, auf das man nicht ohne Widerwillen blicken kann."*

Stärkste Ausprägung, aber auch größte Übertreibung widerfuhr der politischen Arithmetik durch den Belgier Lambert Adolphe Jacob Quetelet (1796-1874), der 1846 in Belgien die erste Volkszählung

[1] Zur Geschichte der Statistik siehe insbesondere die einschlägigen Kapitel in Zizek (1923), Flaskämper (1956), Kellerer (1979), Heiler & Michels (1994).

nach heutigem Muster durchführte. Auf der Suche nach Gesetzmäßigkeiten in sozialstatistischen Daten glaubte er, der Physik ähnliche Naturgesetze gefunden zu haben. Er entwickelte die Lehre von "mittleren Menschen" als Ideal, das sämtliche Merkmale in mittlerer Ausprägung aufweist.

# 4 Wahrscheinlichkeitsrechnung

Zeitgleich zur Universitätsstatistik und zur politischen Arithmetik entwickelte sich rasch das mathematische Gebiet der Wahrscheinlichkeitsrechnung, zumeist motiviert durch aktuelle Probleme des Glücksspiels. Entscheidend war die Möglichkeit, Zufallsergebnisse durch Wahrscheinlichkeitsmodelle, wie der Binomial- oder der Poissonverteilung zu beschreiben. Die Normalverteilung erhielt ihre Begründung als "**datenerzeugendes Modell**" – wie wir heute sagen würden – aus dem von Gauss postulierten **Fehlergesetz**. Ich möchte hier die These vertreten, dass das Konstrukt des von Quetelet zum Ideal erhobenen "mittleren Menschen" weniger aus einem überzogenen naturwissenschaftlichen Menschenbild als aus der Notwendigkeit entstammte, die Normalverteilung auch für sozialwissenschaftliche Belange über ein Fehlergesetz zu rechtfertigen.

# 5 Kontinentale Schule der Statistik

In Deutschland wurden die mathematischen Methoden der Wahrscheinlichkeitsrechnung und ihre Bedeutung für die Statistik lange Zeit nicht erkannt oder sogar abgelehnt. Ausnahmen waren sicherlich der von Frau Esenwein-Rothe zurecht in Erinnerung gebrachte Nationalökonom und Statistiker Wilhelm Lexis, sein unmittelbarer Schüler Ladislaus von Bortkiewicz und die mittelbar als Schüler einzustufenden russischen Mathematiker Chuprow und Markov.

# 6 Methodenskeptizismus

Vor allem aber Gustav Rümelin (1815-1889) und der als "Altmeister der deutschen Statistik" bezeichnete bayerische Landesstatistiker Georg von Mayr (1841-1925) kritisierten die methodisch orientierte Entwicklung der Statistik auf das heftigste, so dass später der Frankfurter Statistik-Ordinarius Paul Flaskämper ihren Methodenskeptizismus als primär verantwortlich dafür bezeichnete, dass in Deutschland so lange Zeit komplizierte mathematische Methoden in der Statistik vernachlässigt wurden.

# 7 Englische biometrische Schule

## 7.1 Überblick

So entstanden die Grundlagen der modernen Statistik vor allem in England um den Geographen, Meterologen und Genetiker Francis Galton, den Ökonomen Edgeworth und vor allem Karl Pearson

(1857-1936), der zu seiner Zeit als Professor des University College London auf äußerst streitbare Weise die Statistik dominierte.

Neben Galton und Pearson gehört als wohl prominentestes Mitglied Ronald A. Fisher (1890-1962) zur sog. englischen biometrischen Schule. Ihm verdanken wir viele der heute zum Standardrepertoire der Statistik gehörenden Techniken und Konzepte wie Likelihoodfunktion, Maximum-Likelihood-Methode, Suffizienz, Varianzanalyse und statistische Versuchsplanung. Zweifelsohne sind einige dieser Konzepte bereits sehr viel früher – auch in Kontinentaleuropa – entwickelt worden. Fisher hat sie allerdings systematisch erarbeitet und in seinem Standardwerk "Statistical Methods for Research Workers" verständlich und publik gemacht. Die genannte Schrift ist ein schönes Beispiel dafür, dass auch in der Wissenschaft – ich möchte fast sagen eine marketingmäßige Aufbereitung der Ideen benötigt wird – um sie erfolgreich verbreiten zu können. Dies ist den kontinentaleuropäischen, mathematisch orientierten Statistikern nicht gelungen.

Der große Erfolg der Fisherschen Ideen ist u.a. vor allem darauf zurückzuführen, dass sie der statistischen Praxis **konkrete**, **nahezu universell anwendbare** Handlungsweisen vorgeben. So ist die Maximum-Likelihood-Methode mittlerweile **das** Standardschätzverfahren, mit dem die Parameter selbst kompliziertester Modelle der Ökonometrie und der Zeitreihenanalyse geschätzt werden können. Zudem garantiert dieses Verfahren für allerdings **große** Stichproben unter milden Voraussetzungen sehr guten Eigenschaften.

Ein weiterer Erfolgsmotor der Fisherschen Konzepte ist ihre vermeintliche Voraussetzungslosigkeit. Im Gegensatz zur Bayesianischen Schlußweise, die im 19. Jahrhundert vorherrschte, benötigt das Maximum-Likelihood-Prinzip keine a priori Annahme über die Verteilung der unbekannten und zu schätzenden Paramter. Bösartig verkürzt könnte man folgendes Bild benutzen: Der Anwender werfe lediglich die Normalverteilungsannahme in die Schätzmaschine hinein und bekommt vollautomatisch einen Schätzwert.

Diese Vollautomatik stellt aber auch eine große Gefahr dar. Sie hängt nämlich äußerst sensitiv von der benötigten Verteilungsannahme ab. Zudem gibt es eine Vielzahl alternativer Schätzverfahren, die zwar für große Stichproben dieselben guten Eigenschaften besitzen wie die Maximum-Likelihood-Methode, aber für realistische Datensätze zu deutlich anderen Ergebnissen gelangen. Wichtig ist die Suche nach auch für kleine Stichproben optimalen statistischen Methoden. Der aus Polen stammende Mathematiker Jerzy Neyman (1894-1981) arbeitete denn auch in den zwanziger und zu Beginn der dreißiger Jahre zusammen mit Karl Pearsons Sohn Egon (1895-1980) an der entscheidungstheoretischen Grundlegung statistischer Tests.

Zwischen Karl Pearson und Fisher einerseits und Neyman und Fisher andererseits bestand zeitlebens eine zum Teil öffentlich ausgetragene wissenschaftliche Feindschaft. Heute scheint die auf diese Forscher zurückgehende statistische Methodenlehre ein einheitliches Ganzes zu sein. Um so interessanter ist es, die Streitpunkte etwas genauer zu beleuchten, lässt sich doch dann besser verstehen, mit welchen immer noch aktuellen Grundlagenfragen sie sich auseinandergesetzt haben.

## 7.2 Karl Pearson versus R.A. Fisher[2]

1935 forderte der u.a. in der Fischereistatistik tätige Hugo John Buchanan-Wollaston die namhaften Vertreter der englischen biometrischen Schule in der Zeitschrift Nature auf, eine Reihe von kritischen Anmerkungen zu statistischen Tests zu kommentieren. Sowohl Karl Pearson als auch R.A. Fisher antworteten z.T. mehrfach, wobei sie ihre impliziten Grundüberzeugungen zum statistischen Testen offen legten.

Im Detail beklagt Buchanan-Wollaston, dass er im Gespräch mit kontinentaleuropäischen Statistikern – die er allerdings nicht namentlich benannte – ein deutliches Misstrauen gegen die von Pearson, Fisher und anderen entwickelten "modernen" statistischen Tests gespürt habe. Der Hauptgrund für dieses Misstrauen bestehe darin, dass ein Test **simultan** zur Beurteilung der **Falschheit** einer Hypothese und der **Korrektheit** der Gegenhypothese benutzt wird. Er betont wörtlich:

*„There is in fact a large region in the distribution of the criterion for which neither a hypothesis nor its reverse can be assumed to be true."*

Ausgehend von der Hypothese, dass ein Datensatz keiner Normalverteilung folgt, kann z.B. mittels des von Karl Pearson entwickelten $\chi^2$-Anpassungstestes bei einem Signifikanzniveau von z.B. 5% festgestellt werden, dass diese Hypothese nicht abgelehnt werden kann, d.h. die Stichprobenwerte offensichtlich nicht signifikant abnormal sind. Daraus zu schließen, dass die Gegenhypothese gilt, d.h. eine Normalverteilung vorliegt, ist nach Buchanan-Wollastan unzulässig. Der besagte $\chi^2$-Test kann somit über das Vorliegen einer Normalverteilung keine Aussage treffen, obwohl eine Vielzahl der von der englischen biometrischen Schule vorgeschlagenen Testverfahren – wie z.B. $t$- und $F$-Test – notwendig von einer Normalverteilung ausgehen. Er führt weiter aus:

*„The fact the British methods 'work' is due to the prevalence in Nature of distributions similar to the Gaussian rather than to any peculiar virtue in the methods themselves."*

Karl Pearson betont in seiner Replik, dass die $\chi^2$-Statistik lediglich als Diskrepanzmaß – wörtlich "measure for goodness of fit" – fungiert, das misst, wie gut sich die empirische Verteilung eines Datensatzes durch eine "theoretische" Verteilung, die nicht notwendigerweise die Normalverteilung sein muss, **approximieren** lässt. Bitte beachten Sie das Wort "approximieren". Für kleine Stichproben ist eine Diskriminierung unterschiedlicher Verteilungsformen nicht möglich. Er betont weiterhin, dass die theoretischen Verteilungsmodelle die Rolle von "graduation curves", also rein mathematischen, zur Beschreibung verwendeten Konstrukten spielen und keineswegs Hypothesen oder Gegenhypothesen von Naturgesetzen darstellen.

Offensichtlich wendet sich Karl Pearson gegen die im 19. Jahrhundert vorherrschende Einstellung, dass Verteilungsmodelle als **datenerzeugende Modelle** durch die Sachfrage, wie z.B. dem Fehlergesetz von Gauss, zu rechtfertigen sind. So hat er sich bei der Entwicklung seiner 12 Typen von Verteilungsmodellen bewusst von dem Normalverteilungsparadigma abgewendet und damit

[2] Dieser Abschnitt folgt Inman (1994), S2. ff. und den dort angegebenen Quellen.

auch bewusst den Anspruch aufgegeben, die wahrscheinlichkeitstheoretische Herkunft der Daten modellieren zu können. Lassen Sie mich im Zusammenhang mit dem von ihm betonten Approximationscharakter von Wahrscheinlichkeitsmodellen, diese als "**datenbeschreibende Modelle**" bezeichnen.

Die Kritik von Buchanan-Wollastan wendet sich denn auch eher gegen R.A. Fisher, da viele von ihm propagierte, vollautomatische Verfahren, wie z.B. die Varianzanalyse, notwendig von der Normalverteilungshypothese ausgehen. Fisher geht aber auf das Problem der Begründung der Normalverteilungshypothese nicht ein und wendet sich seinerseits gegen die pragmatische Einstellung Karl Pearsons, nicht über die **Wahrheit** von Hypothesen befinden zu wollen. In der Tat: Fishers Anspruch an die Statistik ist ein ungleich höherer als dass er sich mit "datenbeschreibenden Modellen" zufrieden geben kann. Ihm geht es mit dem Konzept der Likelihood um den Aufbau einer Logik induktiver Schlüsse. Dies wird insbesondere in der Kontroverse zwischen Fisher auf der einen und Egon Pearson bzw. Jerzy Neyman auf der anderen Seite deutlich, die auf die auf das Begriffspaar "induktive Schließen versus induktives Verhalten" reduziert werden kann.

### 7.3 R.A. Fisher versus Jerzy Neyman[3]

Die moderne Theorie der Hypothesentests begann mit Students Entdeckung des $t$-Testes im Jahre 1908. Allerdings benötigte man spezielle Tabellen, um eine exakte Testentscheidung durchführen zu können. Die vorher zumeist ausschließlich verwendeten Normalverteilungstabellen reichten nicht mehr aus. Die Berechnung und Tabellierung dieser und anderer Verteilungen ist äußerst aufwendig, so dass sich Fisher entschloss, in seinem Buch "Statistical Methods for Research Workers" nur einige ausgewählte Prozentpunkte anzugeben. Er wählte u.a. das 1% bzw. 5%-Niveau als Standardsignifikanzniveau. Für die praktische Anwendung war dies äußerst bequem. Man schaute – wie auch heute noch – lediglich, ob der empirische Wert des Prüfmaßes den durch das Standardniveau definierten kritischen Wert überstieg, um zu entscheiden, ob die betreffende Hypothese abzulehnen ist.

Offen blieb allerdings die Frage, woher das Testkriterium, im Beispiel die $t$-Statistik, zu kommen hat. Es fehlte ein Entscheidungskriterium. Nach eigenen Angaben beeinflusst durch die Tabellen Fishers mit den vorgegebenen Signifikanzniveaus betrachteten Egon Pearson und Jerzy Neyman Hypothesenpaare mit einer sog. Nullhypothese und einer sog. Alternativhypothese statt wie im Verfahren des Signifikanztestes davon auszugehen, dass die Alternative automatisch das Komplement der Nullhypothese ist. Sie bezeichneten einen Test (und damit ein Testkriterium) als überlegen gegenüber einem anderen Test, wenn dieser bei **vorgegebener** Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie richtig ist (dem sog. Fehler 1. Art) eine kleinere Wahrscheinlichkeit besitzt, die Nullhypothese nicht abzulehnen, obwohl sie falsch ist (dem sog. Fehler 2.Art). Das Lemma von Neyman und Pearson gibt an, wie ein entsprechend dieses Gütekriteriums optimaler Test zu konstruieren ist. Nebenbei sei bemerkt, dass die von Neyman entwickelten Konfidenzintervalle auf einer ähnlichen Fixierung des Konfidenzniveaus basieren und beide Konzepte – Alternativhypothesentest und Konfidenzintervalle – von A. Wald (1939) in eine allgemeine statistische Entscheidungstheorie eingebettet wurden.

[3] Dieser Abschnitt orientiert sich an Lehmann (1993).

Eine unser schwierigsten Aufgaben in der Lehre besteht darin, den Studenten die asymmetrische Behandlung der beiden Fehler statistischer Testentscheidungen klar zu machen. Es werden Fernsehröhren bemüht, deren Zerstörung zwecks Qualitätskontrolle kostspielig ist, zumal wenn diese Röhren vollkommen in Ordnung sind und es deshalb wichtig ist die Wahrscheinlichkeit des Fehlers 1. Art zu kontrollieren. In Wirklichkeit sind lediglich Fishers Tabellen verantwortlich dafür, dass wir beim Testen die Fehler asymmetrisch behandeln: der eine wird fixiert, der andere minimiert.

Fisher akzeptiert zwar die Betrachtung des Fehlers 2. Art als Gütekriterium lehnt aber die Fixierung konkreter Alternativhypothesen ab, bedeutet deren Fixierung doch eine deutliche Einschränkung der Aussagefähigkeit statistischer Tests. Weitaus größere Probleme hat er allerdings mit der folgenden Äußerung Neyman und Pearsons:

*"Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our* **behaviour** *with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong."*

D.h.: Die Wahrscheinlichkeiten der beiden Fehler werden als "langfristige" Häufigkeiten von Fehlentscheidungen, interpretiert. Die Annahme oder Ablehnung einer Hypothese bezeichnen Neyman und Pearson als **"induktives Verhalten"**. Im übrigen wird keine Entscheidung über die Herkunft der Daten postuliert. Der optimale Test gibt lediglich an, welches Wahrscheinlichkeitsmodell besser zu den Daten passt, ohne es generiert haben zu müssen.

Im Gegensatz dazu ist Fisher – wie bereits erwähnt – an der Entwicklung einer Logik des **induktiven Schließens** interessiert. Für ihn spielte das Konzept der Likelihood eine von der Wahrscheinlichkeit losgelöste Rolle eines so wörtlich

*„measure of rational belief when we are reasoning from the sample to the population."*

Die Unterscheidung zwischen induktivem Verhalten und induktivem Schließen hat Fisher Zeit seines Lebens große Probleme bereitet, wie eine späte Äußerung von ihm beweist:

*„there is something horrifying in the ideological movement represented by the doctrine that reasoning, properly speaking, cannot be applied to empirical data to lead to inferences valid in the real world."*

Obwohl wir heute zumeist das Neyman-Pearson-Paradigma und damit einen entscheidungstheoretisch motivierten pragmatischen Zugang zum Problem des Testens verinnerlicht haben, macht die fortwährende wissenschaftstheoretische Diskussion um "Deduktion versus Induktion" die Aktualität der Auseinandersetzung deutlich, zumal noch kein Wort über zumeist bayesianisch orientierte Denkschulen der Statistik verloren wurde.

Trost mag allerdings das Fazit von Hampel spenden, dass sich zwar die von den diversen statistischen Denkschulen benutzten Formalismen deutlich unterscheiden, aber der faktische Kern der

Ergebnisse derselbe sei. Ursächlich für diesen Tatbestand ist weniger eine Vermischung der Ideologien, als vielmehr die Tatsache, dass sich alle wie gute Datenanalysierer verhalten. Dabei hat er sozialwissenschaftliche Fragestellungen nicht ausgenommen.

Zwar wird die Ablehnung mathematisch-statistischer Methoden für die Sozialwissenschaften häufig damit begründet, dass dort weder die Experimentsituation noch die rigorosen Voraussetzungen, wie insbesondere die Normalverteilung vorliegen. Bedenkt man aber die pragmatische Sichtweise Karl Pearsons, aber auch Jerzy Neymans mit der Konzentration auf lediglich datenbeschreibende Modelle mit Approximationscharakter und dem reduzierten Anspruch, keine induktiven Schlüsse durchführen zu wollen, so scheint mir eine Anwendung auch komplizierterer mathematischer Methoden in der sozialwissenschaftlichen Statistik gerechtfertigt zu sein.

# 8 Frankfurter Schule der Statistik

## 8.1 Überblick

Diese soeben von mir vertretene Position war bis zur Mitte dieses Jahrhunderts in Deutschland kein akzeptiertes Paradigma.

Im Gegenteil: Der besagte Methodenskeptizismus hielt auch in der sog. Frankfurter Schule der Statistik an, allerdings eingeschränkt auf sozialwissenschaftliche Problemstellungen und durch den Neokantianismus der südwestdeutschen Philosophenschule wohl fundiert. Es ist sicherlich nicht zulässig, einfach einen Bogen zu spannen von den Universitätsstatistikern um Achenwall, über Georg von Mayr zur Frankfurter Schule der Statistik. Warf doch gerade der zu dieser Schule gehörende Paul Flaskämper Georg von Mayr vor, in Deutschland für einen Methodenskeptizismus verantwortlich zu sein.

Als Begründer der Frankfurter Schule der Statistik kann Franz Zizek (1876–1938) angesehen werden, der mit der Betonung der Begriffslogik die Wichtigkeit sematischer Aspekte für die Statistik herausstellte. Im engeren Sinne gehörten zur der besagten Schule der erwähnte Paul Flaskämper, Adolf Blind und Heinrich Hartwig (1907–1981). In der heutigen Zeit vertritt vor allem der emerierte Frankfurter Kollege Grohmann, der hier in der Fakultät von einem Festvortrag her bekannt sein dürfte, den er vor einigen Jahren zu Ehren von Frau Esenwein-Rothe gehalten hat, die Ideen der Frankfurter Schule.

Als roter Faden durch das Programm der Frankfurter Schule zieht sich nach Grohmann die Forderung nach einem **sachgerechten** Einsatz von statistischen Denk- und Verfahrensweisen zur Untersuchung von Bevölkerung, Wirtschaft und Gesellschaft, wobei als Ausgangspunkt Problemstellungen und Erkenntnisziele der **Substanzwissenschaften** (Bevölkerungs- und Wirtschaftswissenschaft) und der **wirtschaftlichen und sozialen Praxis** dienen. Das **"Vorgehen von der Sache her"** führt zu Problemen, die nicht in der Natur der Statistik liegen und von "Nur-Statistikern" gar nicht gesehen werden können.

Die konsequente Realisierung der Gedanken der Frankfurter Schule erfordert eine strikte Trennung in sozialwissenschaftliche und naturwissenschaftliche Statistik bzw. wie Menges es formuliert, in einen – wohl primär deskriptiven Bereich, in dem das Verstehen im Vordergrund steht und einen –

wohl eher induktiven Bereich, der sich mit dem Aspekt des Erklärens beschäftigt. In letzter Konsequenz wäre dann für sozialwissenschaftliche Fragestellungen die Anlage von Stichprobenplänen und das Ziehen von Stichproben die einzige Möglichkeit, wahrscheinlichkeitsbegründeter Statistik anzuwenden.

Aus den Grundgedanken der Frankfurter Schule möchte ich zwei exemplarisch herausgreifen, nämlich die Forderung nach einem Parallelismus von Sach- und Zahlenlogik und das Adäquationsproblem.

## 8.2 Adäquation

Das Adäquationproblem liegt in einer **grundsätzlichen** Diskrepanz zwischen der Eigenart sozialer Tatsachen (wie z.B. dem holistischen Charakter des sozialwissenschaftlichen Massenbegriffs oder der idealtypischen sozialwissenschaftlichen Begriffsbildung) und dem quantitativen Charakter der statistischen Methode. Adäquation ist der Versuch, diese Diskrepanz zu minimieren. Das Adäquationsproblem wird in der sozialwissenschaftlichen Literatur im Rahmen der Operationalisierung von Begriffen durch Indikatoren diskutiert und ist mit dem Begriff der Validität verbunden. Obwohl auch von dem Mitgliedern der Frankfurter Schule stets gefordert wird, den Adäquationsfehler zu erforschen und wo immer auffindbar zu bekämpfen, kann ich mich des Eindrucks nicht erwehren, als hätten wir es bei dem Adäquationsfehler mit einem Produkt einer – erlauben Sie mir bitte diese Anleihe in der Quantenphysik – **Unschärferelation** zu tun, die sich gar nicht auflösen lässt und statt dessen nach statistischen Methoden verlangt, die dieser unvermeidlichen Unschärfe Rechnung tragen.

## 8.3 Parallelismus von Sach- und Zahlenlogik

Die Forderung nach einem Parallelismus von Sach- und Zahlenlogik wurde von Paul Flaskämper 1933 in seiner Schrift "Die Bedeutung der Zahl für die Sozialwissenschaften" erhoben. Die Zahlenlogik umfasst die mathematischen und stochastischen Voraussetzungen und Eigenschaften von Maßzahlen und die Sachlogik die Verwandlung des "ursprünglich qualitativ-sozialen Tatbestandes in einen quantitativen und die Angabe der sachlichen Bedeutung quantitativer Ergebnisse.

Flaskämper lehnt keineswegs die axiomatische Methode zur Messung statistischer Begriffe ab, verlangt aber eine sachlogisch begründete Wahl der Axiome. Die Anforderungen an den von Irving Fisher konstruierten Idealindex sind für ihn überwiegend zahlenlogisch motiviert. Mit der Vorgabe von Ersatzfunktion und Vergleichsmaßstabsfunktion nimmt er im Gegenzug eine sachlogische, aus Erkenntniszwecken heraus motivierte Bestimmung von Mittelwerten vor. Die Anwendung des arithmetische Mittels beschränkt er explizit auf sog. extensive Merkmale, für die ein sinnvolle Summenbildung möglich ist. Dies schließt z.B. die sachlogisch sinnvolle Berechnung des arithmetischen Mittels von Temperaturen aus.

Sowohl das Adäquationsproblem als auch der Parallelismus von Sach- und Zahlenlogik widmen sich Aspekten der Messung und Operationalisierung von Begriffen, seien es substanzwissenschaftliche, seien es statistische.

## 9 Theorie des Messens

Derartige messtheoretische Fragestellungen werden aber außerhalb der Statistik seit langem diskutiert.

Messtheoretische Betrachtungen im eigentlichen Sinne wurden durch erkenntnistheoretische Fragen der Physik veranlasst. So wendet sich z.B. Heinrich von Helmholtz in seiner berühmten Schrift "Zählen und Messen erkenntnistheoretisch betrachtet" (1887) gegen auf Kant zurückgehende aprioristische Begründungsversuche von Geometrie und Arithmetik aus der transzendentalen Anschauung von Raum und Zeit und axiomatisierte den Begriff der extensiven physikalischen Größe ausgehend von einem physikalischen Gleichheits- und einem physikalischen Additionsbegriff. Helmholtz nahm eine, wenn auch noch informelle, sachlogische Axiomatisierung vor. Diese wurde 1901 von Hölder mathematisch präzisiert, wobei mit dem sog. Archimedischen Axiom eine Forderung aufgenommen werden musste, deren sachlogische Interpretation zumindest fragwürdig ist. Die Formulierung einer Messtheorie aus zumeist sachlogischen begründeten Prämissen, nennt man fundamentale Messung. Der Name Campbells steht für die zu Beginn dieses Jahrhunderts allgemein verbreitete Ansicht, dass es eine fundamentale Messung nur für extensive Größen geben kann, für die eine "physikalischen Addition" verfügbar ist, was die fundamentale Messung sozialwissenschaftlicher Begriffe ausschließt. Genau zu diesem Ergebnis kommt ein von 1932 bis 1940 tagendes Kommitee der British Association for the Advancement of Science mit dem Namen

*„Commitee Appointed to Consider and Report upon the Possibility of Quantitative Estimates of Sensory Events."*

Man ging sogar soweit, die Bedeutung der thermodynamischen Temperaturdefinition zu leugnen.

Gleichzeitig lieferte aber Stevens in Fortsetzung der Arbeiten von Fechner mit Reiz-Stimuli-Versuchen den praktischen Nachweis, dass über sog. psychophysikalische Gesetze eine indirekte Messung psychologischer Größen möglich ist. Er postulierte z.B. eine lineare Beziehung zwischen der subjektiven Tonhöhe und der Lage der von dem betreffenden Ton erregten Stelle der Basilarmembrane.

Es kann deshalb nicht verwundern, dass Stevens sich mit dem Ergebnis des genannten Komitees nicht zufrieden gab und 1946 die folgende allgemeine Definition des Messens vorschlug:

*„measurement is the assignment of numerals to objects or events according to rules – any rule".*

Diese Definition ist derart allgemein, dass sie universell, also auch auf sozialwissenschaftliche Fragestellungen, angewendet werden kann.

Als noch weitreichender erwies sich die von ihm eingeführte und sich an die von Cohen & Nagel (1934) vorgeschlagene Typologie der Begriffe – mit klassifikatorischen, topologischen und quantitativen Begriffen – anlehnende Typologie der Skalen. Grob gesprochen sind dies Nominal-, Ordinal- und Kardinalskalen; die Letzteren mit der Unterteilung in Intervall- (siehe Temperatur) und Verhältnisskalen (siehe physikalische Größen wie Gewicht, Länge etc.). Stevens Typologie basiert auf dem

Konzept der zulässigen Skalentransformationen, die auf der Zahlenebene angeben, wie alternative, aber gleichberechtigte Meßvorschriften, z.B. repräsentiert durch unterschiedliche Dimensionseinheiten, ineinander überführt werden können. Je größer die Menge der zulässigen Skalentransformationen, desto geringer ist der Skalentyp.

Gegen die Stevenssche Definition des Skalentyps lässt sich aber sofort einwenden, dass offenbar der Skalentyp a priori erkennbar sein muss. Er wird den Daten als sog. Metadatum beigefügt und beeinflusst die Analysemöglichkeiten des Statistikers. Stevens leitet nämlich aus dem Skalentyp ein Konzept **erlaubter** statistischer Methoden ab. Diese sollen, um erlaubt zu sein, über Invarianzeigenschaften bezüglich der zulässigen Skalentransformationen verfügen. Dies ist zunächst einleuchtend, wenn man bedenkt, dass der Übergang zu einer anderen Dimensionseinheit zu keinen anderen statistischen Schlüssen führen sollte. Wenn aber die Alternative zu Intervallskalen lediglich Ordinalskalen darstellen, deren zulässige Transformationen sämtlich streng monoton zunehmend und stetig sind, so führt Stevens Kriterium erlaubter Methoden insbesondere für sozialwissenschaftliche Daten zu einer Methodenarmut. Regressionen, Varianzanalyse und ähnliche Instrumente verlangen zumindest intervallskalierte Daten. Es ist deshalb kein Wunder, dass es in den letzten 50 Jahren einen intensiven Streit darüber gegeben hat, ob sich der Statistiker durch ein solch willkürliches Kriterium in seiner Kreativität der Methodenentwicklung und Methodenanwendung beschränken lassen muss. Der vorläufige Höhepunkt dieser Kontroverse ist ein jüngst erschienener Beitrag der namhaften Statistiker Velleman & Wilkinson, die sich gegen eine Anwendung des Stevensschen Kriteriums für die automatisierte Methodenwahl in statistischen Expertensystemen wenden.

Ich möchte im Folgenden zwei Thesen begründen, die mir eine Rechtfertigung der statistischen Methodenwahl aufgrund a priori festgelegter Skaleneigenschaften unserer Daten zu liefern scheinen:

Zum einen ist die dem Kriterium von Stevens zugrundeliegende Invarianzforderung Ausdruck eines allerdings mathematisch zu fassenden Adäquationsproblems, das sich generell ergibt, wenn mathematische Beschreibungen nicht nur empirischer Sachverhalte angestrebt werden. Zum zweiten erlaubt die über den informellen Meßbegriff von Stevens hinausgehende axiomatische Theorie des Messens die Interpretation invarianter statistischer Maßzahlen im Sinne eines Parallelismus von Sach- und Zahlenlogik.

Um die Bedeutung von Invarianzforderungen herausarbeiten zu können, gestatten Sie mir bitte einen kleinen Exkurs zum Erlanger Programm meines berühmten Namensvetters Felix Klein aus dem Jahre 1872 und zur Bedeutung von Galilei- und Lorentztransformationen in der klassischen Newtonschen bzw. Einsteinschen Partikelmechanik.

# 10 Invarianz und Adäquation

## 10.1 Erlanger Programm

Die (synthetischen) Axiome der euklidischen Geometrie verfügen im Sinne von Flaskämper über eine Sachlogik. So legte Euklid die Geometrie als Beschreibung drei-dimensionaler Figuren geradezu an. Im Verlaufe des vorigen Jahrhunderts wurden zum einen nicht-euklidische Geometrien und, was weitaus graviender ist, analytisch (d.h. zahlenlogisch) begründete Axiomatisierungen der

Euklidischen Geometrie gegeben. Es entstand eine Vielzahl von nebeneinander stehenden Geometrien. Im Rahmen eines Berufungsverfahrens auf einen neu gegründeten Mathematik-Lehrstuhl an der Universität Erlangen hat besagter Felix Klein 1872 ein Ordnungskonzept für die gesamte Geometrie vorgelegt, das die Mathematik nachhaltig beeinflusst hat und seit geraumer Zeit als Erlanger Programm bezeichnet wird. Er geht nicht von den üblichen Konzepten, wie Punkten, Geraden oder Parallelen aus, sondern differenziert die Geometrien durch Gruppen von Abbildungen, sog. Transformationen. "Sinnvolle" Bestandteile einer Geometrie sind dann gerade diejenigen algebraischen Konzepte, die von der jeweiligen Abbildungsgruppe nicht bewegt, d.h. invariant gelassen, werden. Je kleiner die jeweilige Gruppe ist, desto mehr Invarianten gibt es und desto reichhaltiger ist die betreffende Geometrie. Mathematisch unkorrekt, aber anschaulich, kann man den Umfang von Transformationsgruppen als Maßstab für die Reichhaltigkeit geometrischer Theorien interpretieren.

## 10.2 Galilei- und Lorentztransformationen

Newton ging bei der Formulierung seiner Mechanik von einem absoluten Raum und einer absoluten Zeit aus, in der die von ihm formulierten Gesetze der Mechanik gelten sollten. Kant erhebt denn auch Raum und Zeit zu synthetischen Kategorien. Vom Gesichtspunkt der mathematischen Beschreibung aus gelten jedoch die Newtonschen Gesetze in jedem mit konstanter Geschwindigkeit bewegten rechtwinkligen Koordinatensystem, wie Fallversuche auf einem Ausflugsdampfer beweisen. Dies wird als das Galileische Prinzip der Relativität bezeichnet. Statt eines absoluten Raumes sind die Newtonschen Gesetze also in einer Vielzahl von Räumen (Koordinatensystemen) gültig, die durch sog. Galilei-Transformationen verbunden sind. Die Newtonschen Gesetze der Mechanik sind dann gerade die Galilei-Invarianten. Das Einsteinsche Prinzip der Relativität gibt zusätzlich das Konzept der absoluten Zeit auf, diskutiert gemeinsame schiefwinklige Raum-Zeit-Koordinatensysteme, die durch sog. Lorentztransformationen verbunden sind und identifziert die Gesetze der Einsteinschen Mechanik als Lorentz-Invarianten.

Die erkenntnistheoretische Bedeutung der Betrachtung von Relativität, Transformationsgruppen und Invarianten wird durch die folgende Äußerung des Physik-Nobelpreisträgers Max Born deutlich, die im folgenden als **Äquivalenzprinzip** bezeichnet werden soll:

*„Es gibt nicht nur das eine, in dem absoluten Raum Newtons ruhende System, wo das der Fall ist, sondern unendlich viele Bezugssysteme, die sämtlich gleichberechtigt sind ... . Nicht der Raum ist da und prägt den Dingen seine Form auf, sondern die Dinge und ihre physikalischen Eigenschaften bestimmen den Raum."*

Es geht also um gleichberechtigte, d.h. äquivalente Bezugssysteme, die im Sinne der Gesetze der jeweiligen Mechanik nicht unterscheidbar sind. Wollte man sie unterscheidbar machen, so benötigte man zusätzliche physikalische Gesetze, die den absoluten Raum festlegen, wenn man sich nicht in eine Kantianische Behandlung des Raumbegriffs flüchten möchte. Soweit sind also die Galilei- und Lorentztransformationsgruppen wiederum semantische Gradmesser der Reichhaltigkeit, oder in der heutigen Sprache des Informationsgehalts konkurrierender physikalischer Theorien.

Lassen Sie mich eine vielleicht gewagte Analogie zum Adäquationsbegriff der Frankfurter Schule und zum Begriff der Unschärferelation ziehen. Eine sachädaquate Operationalisierung des Begriffs des Raumes ist in der Newtonschen Mechanik nicht auf eindeutige Weise möglich. Es gibt viele nicht unterscheidbare Operationalisierungen. Damit erfordert die Untersuchung der Validität von mechanischen Konstrukten und Gesetzen u.a. die Untersuchung der Unabhängigkeit von der willkürlich gewählten Operationalisierung. Diese Invarianzuntersuchung ist Spiegelbild einer theorieimmanenten Unschärfe des Raumkonzeptes in der Newtonschen Mechanik.

## 10.3 Suppes Paradigma empirischer Theorien

Der Philosoph, Physiker und mathematische Psychologe Patrick Suppes erhebt das soeben formulierte Äquivalenzprinzip geradezu zu einem Paradigma der Rekonstruktion empirischer Theorien. Er schrieb 1965:

*„The data of experiments, or the empirical predictions of physical theories, are best expressed not by a particular model of the theory but by an appropriate equivalence class of models. Sentences or propositions of the theory which do not have invariant truth ... over these equivalence classes simply have not any clear meaning."*

Suppes wendete das Äquivalenzprinzip 1959 auf messtheoretische Strukturen an und rechtfertigt damit Stevens informell formuliertes Invarianzkriterium. Ohne auf die Details der axiomatischen Theorie des Messens eingehen zu können, gilt es zu akzeptieren, dass es nach Festlegung der für die Messung empirisch relevanten Axiome, d.h. der Sachlogik im Sinne Flaskämpers, viele äquivalente numerische Beschreibungen, sog. Skalen, dieses Axiomensystems gibt, die genau wie im Erlanger Programm die verschiedenen Geometrien oder in der Newtonschen Mechanik die konstant bewegten rechtwinkligen Koordinatensysteme durch Transformationsgruppen verbunden sind. Je mehr dieser Axiome postuliert werden können, desto kleiner ist die Transformationsgruppe und desto reichhaltiger die Menge der invarianten, d.h. sinnvollen Aussagen. Genau dies versuchen die Skalentypen standardmäßig einzufangen. Sie ordnen den Informationsgehalt von Begriffen ebenso wie die Geometrien im Erlanger Programm durch Transformationsgruppen geordnet werden.

Pfanzagl stellte 1967 in bester Tradition Zizeks, der den Vergleich als die Seele der Statistik bezeichnete, zur Definition messtheoretisch sinnvoller statistischer Maßzahlen die Forderung in der Mittelpunkt, dass diese zumindest für messtheoretisch sinnvolle Vergleiche zu gebrauchen sein müssen. Wenn ich also Zensuren mitteln möchte, so sollte ich im Interesse eines messtheoretisch sinnvollen Vergleichs einen Mittelwert benutzen, der unabhängig von der betrachteten Ordinalskala die Gleichheit der mittleren Leistung zweier Grundausbildungskurse festzustellen erlaubt. Und dies leistet das arithmetische Mittel nicht.

## 10.4 Invarianz- und Definierbarkeit

Innerhalb der axiomatischen Theorie des Messens wird dieses in meinen Augen so plausible Äquivalenzprinzip heftig und kontrovers diskutiert, da auf der einen Seite eine logische Fundierung zu fehlen scheint und auf der anderen Seite jedes im Gefolge des logischen Empirismus konstruierte Signifikanzkriterium gescheitert ist und auch scheitern musste, wie Stegmüller eindrucksvoll gezeigt hat.

Eine partielle logische Fundierung lässt sich allerdings durch das Prinzip von Padoa erreichen, das in der zugegebenermaßen restriktiven Prädikatenlogik 1. Stufe logisch invariante Prädikate definitorisch auf Grundprädikate zurückzuführen erlaubt. Von Narens & Mausfeld wurde jüngst eine Übertragung des Äquivalenzprinzips (1991) unter Zuhilfenahme von Definierbarkeitskonzepten auf ganze psychophysikalische Theorien angewandt. Ich selbst konnte zeigen, dass invariante Relationen sich durch, wenn auch komplizierte, algebraische Berechnungen auf die den Skalentyp definierenden Grundrelationen zurückführen lassen. In diesem Sinne erhalten also invariante statistische Methoden eine Sachlogik, sofern die den Skalentyp definierenden Grundrelationen eine Sachlogik besitzen.

## 10.5 Invarianz als Minimalforderung

Wie erwähnt hat Stegmüller das Scheitern empiristischer Sinnkriterien eindrucksvoll nachgewiesen. Wollte man das Äquivalenzprinzip und die daraus resultierenden Invarianzforderungen zu einem globalen Kriterium empirisch signifikanter Aussagen erheben, so wäre ein Scheitern ebenso unvermeidlich. Schließlich ist eine Axiomatisierung häufig informell vorliegender Sachlogik nur partiell möglich. Ebenso ist die Erfassung des Adäquationsproblems durch die Behandlung äquivalenter mathematischer Beschreibungen und die Beurteilung der Validität durch Invarianzuntersuchungen notwendig unvollständig. Die Adäquationsproblematik vieler holistischer sozialwissenschaftlicher Begriffe ist viel zu komplex als das sie sich durch ein einziges Konzept erfassen ließe. Trotzdem: Angestrebt wird eine Negativauslese. Um in der Sprache der Statistik zu bleiben: Wenn ein arithmetisches Mittel noch nicht einmal in der Lage ist, Zensurendurchschnitte zweier Kurse unabhängig von der verwendeten Zensurenskala zu beurteilen, dann kann erst recht nicht die komplexere und formal nicht zu erfassende Sachlogik des Begriffs der durchschnittlichen Leistung durch arithmetische Mittel erfasst werden. In diesem Sinne sind Invarianzforderungen unverzichtbare und mathematisch handhabbare Minimalforderungen zur Behandlung des komplexen Adäquationsproblems und der komplexen Sachlogik statistischer Konzepte.

## 10.6 Invarianz und Identifizierbarkeit

Während offenbar Geometrie, theoretische Physik und Messtheorie problemlos mit dem Äquivalenzprinzip und den daraus folgenden Invarianzforderungen leben können, geht die Ökonometrie bei der Lösung des verwandten Problems der Identifizierbarkeit anders vor. Ein ökonometrisches System von Gleichungen ist, grob gesprochen, nicht identifizierbar, wenn zu viele Parameter geschätzt werden sollen. Die Konsequenz der Nicht-Identifizierbarkeit ist, dass man für jeden Parameter nicht nur einen Schätzwert, sondern unendlich viele Schätzwerte hat, wobei allerdings die Beziehungen zwischen den Schätzwerten aufrechterhalten bleiben. In einfachen Fällen sind diese Beziehungen wiederum durch Transformationsgruppen gegeben. Wie behilft man sich in der Ökonometrie? Es werden – hoffentlich durch die volkswirtschaftliche Theorie begründete – Restriktionen für die Parameter gesucht, damit unter Beachtung dieser Restriktionen eine eindeutige Schätzung möglich ist. In die Sprache der Newtonschen Mechanik übersetzt, wird also nach "Gesetzen" gesucht, die die Festlegung des absoluten Raumes ermöglichen. Eine Untersuchung der Transformationsgruppen nicht-identifizierter ökonometrischer Systeme ist sicherlich schwierig, aber im Interesse des Äquivalenzprinzips unbedingt erforderlich.

### 10.7 Invarianz und Robustheit

Die dominierende Rolle der Normalverteilung in der Statistik des 19. Jahrhunderts aber auch in der Fisherschen Theorie wurde bereits mehrfach betont. Selbst für astronomische Daten wurde nachgewiesen, dass sie über einen namhaften Anteil von Ausreißern verfügen, der mit einem Normalverteilungsmodell nicht vereinbar ist. Erst recht ist trotz der Anstrengungen Quetelets die Normalverteilung als Modell der Beschreibung sozialwissenschaftlicher Daten ungeeignet. Tukey diskutierte in einer Arbeit aus dem Jahre 1961 statt eines einzigen Verteilungsmodells ein sog. Kontaminationsmodell, das eine gewisse Klasse von gleichberechtigten Verteilungen – oder besser datenbeschreibenden Modellen – umfasst. Konsequent verallgemeinerten der mittlerweile in Bayreuth als Statistiker tätige Peter Huber (1964) und der eingangs erwähnte Hampel (1968) das Konzept der Verteilungsumgebungen. Statt eines dogmatisch einzusetzenden Normalverteilungsmodelles wird eine Umgebung von Verteilungen um die Normalverteilung betrachtet und nach statistischen Verfahren gesucht, die für alle Verteilungen dieser Umgebung "gute" Eigenschaften besitzen, ohne für die Normalverteilung "besonders schlecht" geeignet zu sein.

Auch diese Vorgehensweise ist im Sinne des Äquivalenzprinzips zu interpretieren. Mangels weitergehender Information werden als datenbeschreibende Modelle viele äquivalente Verteilungen zugelassen. Geeignete statistische Verfahren verfügen im Idealfall über eine Invarianz des Güteverhaltens.

## 11 Fazit

Lassen Sie mich abschließend die wichtigsten Punkte festhalten:

1. Statistische Modelle sind sobald sie für empirische Phänomene angesetzt werden, mit wenigen Ausnahmen als datenbeschreibende Modelle zu interpretieren. Sie approximieren im Sinne von Karl Pearson die empirische Verteilung des Datenbefundes. Mit dem Approximationscharakter sind aber unweigerlich Fehler verbunden, die sich meines Erachtens für sozialwissenschaftliche Fragestellung kaum auflösen lassen. Dies schränkt nicht notwendig die Menge der anwendbaren statistischen Methoden ein, macht im Gegenteil spezifische robuste Methoden sinnvoll und wirkt sich allerdings auf die Möglichkeit der Ergebnisinterpretation vehement aus.

2. Ähnlich verhält es sich mit dem Adäquationsfehler als Resultat eines unvermeidbaren Unschärfeproblems bei der Quantifizierung sozialwissenschaftlicher Begriffe .

3. Das Äquivalenzprinzip und nachfolgende Invarianzuntersuchungen können Teilaspekte des Approximationsgedankens und des Adäquationsfehlers erfassen, aber wohl gemerkt, ohne

die Fehler aufzulösen. Es handelt sich um lediglich kleine formalisierbare Teilaspekte, die aber die Invarianzforderungen als Minimalforderungen eines komplexen Adäquationsbegriffs rechtfertigen. Für das Problem der Negativauslese von Methoden sind aber die mathematisch formulierten und damit überprüfbaren Minimalforderungen unerlässlich.

4. Sollten Invarianzuntersuchungen nicht möglich sein, da Transformationsgruppen, die äquivalente Beschreibungen verbinden, nicht angebbar sind, so sind Sensitivitätsanalysen und Alternativrechnungen ein probates Mittel.

5. Skalentypen sind ein unerlässliches und wichtiges Entscheidungskriterium für die Methodenauswahl in statistischen Expertensystemen. Das Metadatum "Skalentyp" ist ein standardisierter Gradmesser der substanzwissenschaftlichen Information, die im Sinne eines Parallelismus von Sach- und Zahlenlogik nur im wirklich vorhandenen Maße in die statistische Analyse einfließen sollten.

## Literatur


1. Blind, A. (1953). Probleme und Eigentümlichkeiten sozialstatistischer Erkenntnis. *Allgemeines Statistisches Archiv* **37**, 301-313.

2. Born, M. (1969). *Die Relativitätstheorie Einsteins.* Berlin.

3. Campbell, P.R. *Physics: the elements.* New York.

4. M.R. Cohen & E. Nagel (1934). *An Introduction to Logic and Scientific Method. New York, Harcourt Brace.*

5. Fechner, G.T. (1860). *Elemente der Psychophysik.* Leipzig.

6. Fisher, R.A. (1925). *Statistical methods for research workers.* Edinburgh.

7. Flaskämper, P. (1929). Das Problem der ,,Gleichartigkeit" in der Statistik. *Allgemeines Statistisches Archiv.* **19**, 393-420.

8. Flaskämper, P. (1933). Die Bedeutung der Zahl in den Sozialwissenschaften. *Allgemeines Statistisches Archiv.* **23**, 68-71.

9. Flaskämper, P. (1956). *Grundriss der Statistik.* Verlag Richard von Meiner, Hamburg.

10. Hampel, F. (1968). *Contributions to the theory of robust estimation.* Ph.D.-thesis. University of California, Berkeley.

11. Hampel, F. (1993). Some thoughts about the foundations of statistics. In: Morgenthaler, S., Ronchetti, E., Stahel, W.A. *New directions in statistical data analysis and robustness.* Birkhäuser, Basel.

12. Hampel, F. (1994). Zur Grundlagendiskussion in der Statistik. Festvortrag. Universität Dortmund.

13. Hartwig, H. (1954). Naturwissenschaftliche und sozialwissenschaftliche Statistik. *Allgemeines Statistisches Archiv.* **37**, 252-266.

14. Heiler, S. & Michels, P. (1994). *Deskriptive und explorative Datenanalyse.* München.

15. Helmholtz, H.v. (1887). *Zählen und Messen erkenntnistheoretisch betrachtet.* Leipzig.

16. Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Maß. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-physikalische Klasse.* **53**, 1-64.

17. Huber, P. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics.* **35**, 73-101.

18. Imam, H.F. (1994). Karl Pearson and R.A. Fisher on statistical tests. A 1935 exchange from *Nature. American Statistician* **48**, 2-11.

19. Kellerer, H. (1979). *Statistik im modernen Wirtschafts- und Sozialleben.* Rowohlt, Reinbeck.

20. Klein, F. (1872). *Vergleichende Betrachtungen über neuere geometrische Forschungen. Antrittsvorlesung. Universität Erlangen.*

*21. Klein, F. (1921). Gesammelte Abhandlungen I. Berlin.*

*22. Klein, I. (1984). Das Problem der Auswahl geeigneter Maßzahlen in der deskriptiven Statistik. Würzburg.*

*23. Klein, I. (1994). Mögliche Skalentypen, invariante Relationen und wissenschaftliche Gesetze. Göttingen.*

*24. Lehmann, E.L. (1993). The Fisher, Neyman-Pearson Theories of testing hypotheses: one theory or two. Journal of the American Statistical Association.* **88**, *1242-1249.*

*25. Manski, C.F. (1995). Identification problems in the social sciences. Harvard University Press. London.*

*26. Menges, G. (1976). Deskription und Inferenz (Moderne Aspekte der Frankfurter Schule). Metrika* **??**, *290-319.*

*27. Narens, L. & Mausfeld, R. (1993). ON the relationship of the psychological and the physical in psychophysics. Psychological Review* **99**, *467-479.*

*28. Pearson, E.S. (1962). Some thoughts on statistical inference. Journal of Mathematical Statistics.* **33**, *394-403.*

*29. Padoa, A. (1901). Essai d'une Théorie Algebraique des Nombres Entier. Précédé d'une Interoduction Logique à une Théorie Déductive Quelqueonque. Bibliothéque do Congrès International de Philosophie.* **3**, *309-365.*

30. *Pfanzagl, J. (1973). Theory of measurement. Würzburg.*

31. *Stegmüller, W. (1957). Das Wahrheitsproblem und die Idee der Semantik. Wien.*

32. *Stegmüller, W. (1970). Probleme und Resultate der Wissenschaftstheorie und Philosophie, Band II.: Theorie und Erfahrung, 1. Halbband: Begriffsformen, Wissenschaftssprache, empirische Signifikanz und theoretische Begriffe. Berlin.*

33. *Stevens, S.S. (1946). On a theory of scales and measurement. Science.* **103**, *677-680.*

34. *Suppes, P. (1959). Measurement, empirical meaningfulness and three-valued logic. In: Churchman, C.W., Ratoosh, P. (Hrsg.). Measurement, definitions and theories. New York, 129-142.*

35. *Suppes, P. (1965). Logics appropriate to empirical theories. In: Addison, J.W., Henkin, L., Tarski, A. (Hrsg.). The theory of models. Proceedings of the 1963 international symposium of Berkeley. Amsterdam, 364-375.*

36. *Suppes, P., Rubin, H. (1954). Transformations of systems or relativistic particle mechanics. Pacific Journal of Mathematics.* **4**, *563-601.*

37. *Tukey, J.W. (1961). The future of data analysis. Annals of Mathematical Statistics.* **33**, *1-67.*

38. *Velleman, P.F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. American Statistician.* **47**, *65-72.*

39. *Zizek, F. (1922). Fünf Hauptprobleme der statistischen Methodenlehre. München.*

40. *Zizek, F. (1923). Grundriss der Statistik. Duncker & Humblot, Berlin.*

41. *Zizek, F. (1929). Gleichartigkeit, Homogenität und Gleichwertigkeit in der Statistik. Allgemeines Statistisches Archiv.* **19**, *393-420.*