

# Statistik und Manipulation

Vortrag bei einer Klausurtagung der Geschäftsstelle des Wissenschaftsrates am  
31.10.2013 in Bonn (Haus der Geschichte) von

Peter von der Lippe

Der Vortrag versucht zu zeigen, dass das so beliebte Gerede über "Lügen mit Statistik" oder "Lügen mit Zahlen" Unsinn ist, oder zumindest wenig durchdacht ist. Wenn man andere der Lüge bezichtigt, sollte man auch in der Lage sein, zu sagen, was die Wahrheit ist. Aber was ist das "wahre" Maß der Teuerung, der Arbeitslosigkeit usw.? Das Reden über "Manipulation" und "Lügen" ist die Folge unseres noch nicht so langen und daher noch nicht so gewohnten Umgangs mit Zahlen. Über Jahrhunderte beherrschten nicht Zahlen, sondern Worte (Geschichten), Bilder, und Analogien unser Denken. Unser heutiges Streben, alles messen zu wollen (was zu einer Obsession geworden ist, die wohl ihren Höhepunkt noch nicht erreicht hat), haben wir die längste Zeit unserer Geschichte als eher als unnötig und abwegig empfunden. Früher geriet man auch nicht in Panik, weil es eine Statistik (noch) nicht (rechtzeitig) gab.

Vor diesem Hintergrund wird auch die Widersprüchlichkeit vieler Menschen im Umgang mit Statistik verständlich: sowohl immer mehr (und für immer komplexere Sachverhalte) Zahlen zu fordern, auch wenn einem deren Interpretation schwer fällt, als auch vorhandene Zahlen mit emotionaler Rhetorik wegzuwischen (und dann solchen Reden mehr zu glauben als den Zahlen), und sowohl zahlengläubig sein, als auch überall dunkle Machenschaften zu vermuten. Hinter letzterem steht die immer wieder enttäuschte, aber gleichwohl unverwüsthliche Illusion (der die Statistik ewiges Leben verdanken dürfte) von "Fakten" mit ultimativer Objektivität, Exaktheit und Gewissheit in der Politik. Viele empfinden sich als Opfer einer sie angeblich manipulierenden Statistik, sie sind aber in Wahrheit nicht selten Mittäter (als Zahlenfetischisten), und sie vergessen, dass man mit Worten viel leichter und wirkungsvoller "manipulieren" kann als mit Zahlen und dass nicht alles so leicht zu zählen ist wie Erbsen.

Abschnitt	Seite
1. Wirklichkeitserkenntnis mit Zahlen	2
2. Zählen und erzählen	5
3. Manipulation und Zahlengewinnung in der Statistik: grundsätzliche Fragen	5
a) Manipulation ohne Statistik	5
b) Was ist Manipulation mit Statistik und wem ist sie vorzuwerfen?	6
c) Es kommt darauf an, wer für wen Statistik macht	7
4. Voraussetzungen einer Statistik	8
a) Zählen und messen und wie kann man zu Zahlen gelangen	8
b) Biased sample und "Repräsentativität"	8
5. Messung, Dimensionen und Aussagefähigkeit von Zahlen	9
a) Definieren und operationalisieren	9
b) Eine und mehrere Dimensionen	10
c) Zahlen von außen und Zahlen von innen gesehen, Sinnhaftigkeit von Zahlen	11

Abschnitt	Seite
7. Warum Wahrscheinlichkeitsrechnung und Stichproben?	16
a) Wahrscheinlichkeiten, Schätzen und Testen	16
b) Stochastische Modelle	17
c) Daten und datengenerierende Prozesse	18
8. Widersprüchlichkeiten im Umgang mit Zahlen	19
a) Ruf nach immer mehr Zahlen, fundierte und nicht fundierte Schätzung	19
b) Computopia: hätten wir nur die Zahlen, könnten wir auch die richtige Politik machen	20
c) Noch mehr Zahlen von zweifelhaftem Wert: Data Mining und Big Data	22
d) Der impressionistische Gegenschlag und Immunität gegen Zahlen	23
9. Rating, Evaluation und Selektion	24
a) Kleines und großes Beurteilen	24
b) Strategisches Beeinflussen des Rankings	25
c) Replikations- und Metastudien	26

d) Statistik und Astrologie und warum im Zweifel die Statistik falsch ist und nicht die Überzeugung	12
6. Deskription, Vergleiche und Zusammenhänge	12
a) Spielregeln der Deskription	13
b) Vergleiche und Bezugsgrößen	13
c) Scheinkorrelation und Milchmädchenrechnung	15
d) Operational nutzbare Zusammenhänge	15

10. Zukunft der Statistikausbildung und weitere Trends	26
a) Das Geschäft mit der Not mit Statistik	27
b) Ausgieben, auswendig lernen und vergessen	27
c) Widerwillig lehren und Protzen mit Mathematikkenntnissen	28
d) Modisches, Ultimatives und Albernies	28
Fazit und zwei Nachträge	30 - 33

## 1. Wirklichkeitserkenntnis mit Zahlen

Anders als "Statistik und Manipulation" wäre "Mathematik und Manipulation" kein Thema. Warum nicht? Was ist es, was bei der Statistik zur Mathematik hinzukommt und vielleicht Manipulation erst möglich und reizvoll macht?

Es ist der Anspruch, nicht mit rein gedanklichen Gegenständen umzugehen, sondern die Wirklichkeit beschreiben und erklären zu wollen. Zumindest bei der vom Menschen gemachten "Wirklichkeit" ist aber die Frage, wie es wirklich ist, keine emotionsfreie Angelegenheit. Sie kann auch nicht mit unbezweifelbarer Autorität einer "Offenbarung" beantwortet werden, sondern sie verlangt eine Argumentation und Erklärung (als Auseinandersetzung mit [Hypo-] Thesen).

Wir halten heutzutage mit Zahlen untermauerte Argumente<sup>1</sup> für besonders überzeugend. Aber die Realität mit Zahlen zu erfassen, wie es die Statistik tut, ist nicht die einzige Art und schon gar nicht die Art, die von den meisten Menschen als die natürlichste und verständlichste Art empfunden wird. Denn verglichen mit Worten, Bildern und Gleichnissen (Analogien), womit wir eine jahrhundertelange Übung haben, ist die zahlenmäßige Darstellung einer Sache relativ modern und viel weniger vertraut. Er ist "wie ein Baum, der keine Früchte trägt",<sup>2</sup> galt als gute Beschreibung eines Menschen. An einer Zahl für seine vielleicht bescheidenen geisti-

gen Fähigkeiten, etwa in Form eines Intelligenzquotienten bestand kein Interesse.

Es ist unser Ziel, im Folgenden deutlich zu machen, dass das lange Dominieren von Worten in unserem Denken, verglichen mit dem relativ modernen und ungewohnten Umgang mit Zahlen viel erklärt, was unser ambivalentes Verhältnis zur Statistik betrifft, unsere Neigung, bei Zahlen sofort entweder dunkle Machenschaften, oder aber wahre Wunder zu erwarten.

Selbst für einen so universellen und auch mathematisch so genialen Geist, wie Gottfried Wilhelm Leibniz (1646 - 1716),<sup>3</sup> war noch die Analogie zu Transformationen (Metamorphosen) im Tierreich ein "Beweis" für die Unsterblichkeit der Seele ("...wie ja die Seidenraupe und der Schmetterling das gleiche Lebewesen sind", oder "Auch Schlaf und Ohnmacht zeigen, dass der Tod nicht ein Aufhören aller Funktionen ist, sondern nur eine Unterbrechung gewisser besonders deutlich bemerkbarer Funktionen.")<sup>4</sup>.

Es lohnt sich, kurz noch etwas weiter auszuholen, um die Eigenart der relativ modernen Errungenschaft "Statistik" besser verstehen zu können. Für uns heutzutage mag es schwer vorstellbar sein, dass dem Menschen über viele Jahrhunderte die "Wirklichkeitserkenntnis" (und erst recht eine zahlenmäßige, "numerische") gar kein Bedürfnis war.

<sup>3</sup> Bekanntlich gilt Leibniz neben Newton (1642 - 1726) als Begründer der Infinitesimalrechnung.

<sup>4</sup> Betrachtungen über die Lehre von einem einzigen allumfassenden Geiste (Considérations sur la doctrine d'un esprit universel unique) 1702. Für Leibniz war Geburt und Tod in dieser Schrift vergleichbar mit Vergrößerung und Verkleinerung von einer Art Seelen-Substanz.

<sup>1</sup> Entsprechend dem Titel des sehr lesenswerten Buches von Robert P. Abelson (1928 - 2005) "Statistics as Principled Argument", Hillsdale NJ, 1995.

<sup>2</sup> So Dante (1265 - 1321) in seiner Schrift über die Monarchie (um 1316).

Wissenschaftliche Neugier galt bis ins 17te Jahrhundert eher als Laster und Frevel. In Unbekanntes, dem Menschen von Gott (natürlich von diesem aus gutem Grund) nicht zugedachtes Wissen vordringen zu wollen, galt als Hybris oder als sündhafter Mangel an Glauben, und die Pflege von Okkultem und Mystischem durch obskure Geheimgesellschaften stand für viele auf gleicher, wenn nicht sogar höherer Stufe als die seinerzeit beginnende moderner Wissenschaft.<sup>5</sup>

Als die drei Wurzeln (Anfänge) der Statistik werden, in vielen Statistikbüchern genannt

1. die vergleichende Staatenkunde (daher auch das Wort "Statistik"), die noch kaum zahlenmäßig war, die
2. politische Arithmetik (Suche nach Regelmäßigkeiten einer natürlichen, letztlich aber von Gott gestifteten und *verborgenen*<sup>6</sup> "Ordnung" in Massenerscheinungen, wie Geburt und Tod) und
3. die Wahrscheinlichkeitsrechnung, ein übrigens überraschend spät entwickelter Teil der Mathematik.

Die "politische Arithmetik" entspringt einer Geisteshaltung, deren Bedeutung nicht hoch genug eingeschätzt werden kann. Für einige markiert das "Quantifizieren" nicht nur den Beginn der Wissenschaft, wie wir sie heute verstehen,<sup>7</sup> sondern den Beginn einer Absatzbewegung des "Westens" (d.h. Europa, dann später v.a. die USA) gegenüber dem Rest der Welt, insbesondere dem nahen und

fernen Osten.<sup>8</sup> Sie führte zu einer wirtschaftlichen, politischen und auch kulturellen (z.B. auch in der Wissenschaft) Hegemonie des "Westens", von deren Früchten wir im Westen noch heute zehren.

Als man damit begann, alles "quantifizieren" zu wollen war dies nicht nur der Beginn von Wissenschaft und Technik, sondern auch von Buchhaltung, Fernreisen (Navigation!), Bürokratie und modernem Militär und auch von dem modernen Politikverständnis,<sup>9</sup> kurz es war der Auslöser für eine Entwicklung von weltgeschichtlicher Dimension. Es war weit mehr als nur der Beginn der Statistik mit ihrer anfänglichen Fokussierung auf Daten für die Verwaltung der seinerzeit überall entstehenden großen Staaten.<sup>10</sup>

Dass "Quantifizieren", also Zählen und Messen eine so immense Bedeutung gewinnen konnte liegt vor allem darin begründet, dass dadurch erst die

- Mathematisierung vieler Gebiete und die
- Rechenhaftigkeit, Mechanisierung und Automatisierung des Lebens möglich wurde, und damit auch der Siegeszug des Computers.

Misst man das Alter mit der Anzahl vollendeter (Lebens)Jahre, dann wird das Alter zu einer Variable  $x$ , die die Werte 0, 1, 2, ... annehmen kann, und so in einer Funktion  $y = f(x)$  erscheinen kann, mit der dann jedem Wert von  $x$  (etwa  $x = 18$ ) ein Wert von  $y$  (etwa  $y = 490$ ) "zugeordnet" wird, so dass dann  $y$  in einem gewissen Sinne durch  $x$  "erklärt" wird. Damit gewinnt das Feststellen und "Erklären" von empirischen Zusammenhängen eine ganz andere Qualität als in einer Welt, die nur Worte, nicht Zahlen kennt.<sup>11</sup>

<sup>5</sup> Philip Ball, Curiosity, How science became interested in everything, London 2012.

<sup>6</sup> Wichtig für die Entstehung der Statistik: für den einzelnen in seinem begrenzten Erfahrungshorizont nicht, wohl aber durch (bzw. überhaupt erst durch) die große Masse gesammelter Daten erkennbar.

<sup>7</sup> Wissenschaft oder besser Philosophie im Mittelalter, war weniger interessiert an einer Beschreibung der Natur, als an hinter und unabhängig von den Fakten existierenden universellen Wahrheiten, worüber Aussagen auf typisch scholastische Manier (durch Besinnung über Begriffe und logische Deduktionen) entwickelt wurden. Über etwas durch Gegenüberstellung des Deduzierten mit (durch Beobachtung oder auf andere Art gewonnenen) Fakten zu entscheiden, hätte man damals wohl eher als ein abwegiges Unterfangen empfunden. Das wäre auch bei seinerzeit beliebten Themen wie Gottesbeweise kaum möglich gewesen

<sup>8</sup> Das ist die zentrale These des sehr lesenswerten Buchs von Alfred W. Crosby, The Measure of Reality, Quantification and Western Society 1250 – 1600, Cambridge (USA) University Press, 1997.

<sup>9</sup> Dazu später mehr (Stichwort "Computopia").

<sup>10</sup> Es gibt auch die Auffassung, dass die Statistik ihr schlechtes Image ihrer über Jahrhunderte währenden Identifikation mit dem Obrigkeitsstaat verdankt.

<sup>11</sup> Mit Worten kann man nicht addieren also "zählen" und "rechnen" und auch römische Zahlen wie LX, MDCII, VII usw. taugen nicht viel als Variablen  $x$  und  $y$  in irgendwelchen Gleichungen.

## 2. Zählen und erzählen

Statistik, also Zählen und Messen ist – wie gesagt – im Zusammenhang mit der "Mathematisierung" der Wissenschaft zu sehen. Das bringt auch Voraussetzungen und Grenzen des Erkennens mit sich. Daher ein paar kurze Bemerkungen zur "Zahl".

Eine interessante Auffassung von der Zahl findet man im folgenden Zitat aus dem Jahre 1440, einer Zeit also, in der die zahlenmäßige Beschreibung der Realität noch ganz unüblich war:

"Ohne Zahl kann die Vielheit der Dinge nicht bestehen, denn ohne Zahl gibt es keine Unterscheidung, Ordnung, Proportion, Harmonie."<sup>12</sup>

Diese Worte von Nikolaus von Kues könnte man dahingehend missverstehen, als habe die Zahl erst die Vielheit (also Unterschiedlichkeit) der Dinge geschaffen. Das ist natürlich nicht der Fall. Was aber gilt, ist dass man mit der (auch ohne Zahlen bestehenden) Vielheit besser umgehen kann, wenn man Zahlen hat, etwa so, wie blau und grün auch ohne die Worte "blau" und "grün" existieren (es gibt auch Sprachen, die diese Unterscheidung nicht kennen), aber die Welt mit diesen Worten besser beschreibbar wird.

Interessant ist nicht, ob die Vielheit der Dinge *ohne, oder erst mit* Zahlen existiert, sondern *wie* man von der Vielheit der A-Dinge und der Vielheit der B-Dinge zu Zahlen  $Z_A$  und  $Z_B$  für die A- und für die B-Dinge gelangen kann. Das setzt die Wahrnehmung dessen (und Verständigung darüber) voraus,

- was unter den einzelnen A-Dingen gleich ist,<sup>13</sup> und
- was die A- Dinge von den B-Dingen unterscheidet.

Der Hinweis auf die "Unterscheidung" in dem Zitat, ist also wichtig weil es das Konzept gleich/ unterschiedlich ist, was das Zäh-

len überhaupt erst möglich macht.<sup>14</sup> Eine Zahl ist nicht eine Eigenschaft, die ein beobachteter Gegenstand hat. Dieser kann blau oder rund sein, aber er kann nicht 4 sein; 4 kann sich nur beziehen auf eine durch Abstraktion gebildete *Menge* von Gegenständen.

Es mag in vielen Fällen einfach sein, sicherzustellen, dass die Anzahl  $Z_A$  der A-Dinge nur aus "gleichen" Zählobjekten besteht (nur dann macht die Zahl  $Z_A$  "Sinn"). Das gilt z.B. wenn man "nur" die Anzahl der Männer oder Frauen in einer "Gesamtheit" feststellen will. Ist  $A_1$  und  $A_2$  jeweils ein Mann, so haben wir es mit zwei Männern zu tun. Wenn man aber die Anzahl von Promovierten feststellen möchte, könnten schon Diskussionen auftreten, ob man Dr.  $A_1$  und Dr.  $A_2$  wirklich *in einen Topf A werfen*<sup>15</sup> darf, hier also von zwei Doktoren sprechen darf, wo doch bei genauerem Hinsehen die Qualität der Dissertationen sehr unterschiedlich ist. Man kann hier zwar den Einwand der Unterschiedlichkeit ignorieren, dies aber nur mit der Konsequenz, dass man dann später die "Anzahl der Doktoren" als wenig "aussagefähig" kritisieren wird. Wenn man sich aber auf evtl. lange Diskussionen über die Vergleichbarkeit der Dissertationen von Dr.  $A_1$  und Dr.  $A_2$  einlässt, wird deutlich, dass man irgendwo auf dem Weg von der "Anzahl der Männer" zur "Anzahl der Doktoren" die Grenze zwischen "zählen" und "erzählen" überschritten hat.

Die etymologische Verwandtschaft der beiden Wörter ist kein Zufall. Auch hinter "tale" und "to tell" steht das Wort "Zahl" (bzw. ein gemeinsamer Vorfahr dieses Worts in den beiden Sprachen), so wie auch hinter "town" und Zaun.<sup>16</sup> Früher verstand man unter "erzählen" eine geordnete (der Reihe nach, chronologisch) verbale Wiedergabe von Geschehenem. Das wurde von "zählen" (wo es

<sup>14</sup> Genauer: es setzt die Existenz der Äquivalenzrelation ( $=$ ,  $\neq$ ) voraus. Der Hinweis mit den A-Dingen und den B-Dingen zeigt bereits, dass Zählen Zuordnen von Elementen zu einer Menge bedeutet und in diesem Sinne auch eine "Ordnung" und "Harmonie" erzeugt, von der Nikolaus von Kues sprach.

<sup>15</sup> Die Formulierung macht erneut deutlich, dass es um das Zuordnen von Elementen zu einer Menge geht.

<sup>16</sup> Man bedenke auch: "to tell x from y" heißt x und y *unterscheiden* können.

<sup>12</sup> Nikolaus von Kues (1401 – 1464), Über das belehrte Nichtwissen (De docta ignorantia) 1440, Kap. 5.

<sup>13</sup> Erkennen der Gleichheit der A-Dinge ( $A_1, A_2, \dots$ ), obgleich sie in bestimmter Hinsicht untereinander auch wieder verschieden sind, verlangt eine Abstraktion (von der Individualität von  $A_1$  und von  $A_2$  usw.).

ja auch um eine Reihenfolge geht) und sogar auch von "rechnen" nicht unterschieden.<sup>17</sup>

In den seltensten Fällen ist Statistik nur "zählen" und fast immer ist "zählen" auch mit "erzählen" verbunden, d.h. es sind vor der Gewinnung einer empirischen Zahl und danach, bei der Interpretation einer Zahl Worte nötig. Das zeigt auch das Beispiel mit Dr. A<sub>1</sub> und Dr. A<sub>2</sub>: man muss unwesentliche Unterschiede ignorieren, wenn man jetzt von zwei Doktoren spricht; aber wenn man wesentliche Unterschiede ignoriert, kommen die Worte (das "Erzählen") später wieder in Gestalt von Fragen zur "Aussagefähigkeit".

Bei der Statistikproduktion haben wir Worte-Zahlen-Worte, und nicht, wie vielleicht die meisten denken mögen, nur Zahlen-Zahlen-Zahlen. Die Worte vor den Zahlen betreffen alle (verbalen) Überlegungen, die angestellt werden müssen, um überhaupt etwas zählen und messen zu können, also die "Operationalisierung" von Merkmalen. Die Worte nach den Zahlen betreffen die Interpretation, die Frage "was sagt uns das Ergebnis?"

Es ist die Frage nach dem "Sinn" der Zahlen, die es in der Statistik, nicht aber in der Mathematik gibt: es ist nicht üblich, zu fragen, was die 5 bedeutet bei der Gleichung  $2 + 3 = 5$ , und was man mit der 5 "anfangen" kann, und man vermutet hinter der 5 auch keine "Manipulation" oder gar "Lüge". Dass es in der Statistik nicht nur "zählen", sondern auch "erzählen" gibt, nicht nur Zahlen sondern auch Worte, wird gerne vergessen, von

- Statistiknutzern, die schnell von einem Extrem (Zahlengläubigkeit) ins andere (jede Statistik ablehnen) verfallen, und
- von Studenten, deren Energie und Gekuld voll mit dem Lernen von Formeln und Rechenbeispielen aufgebraucht ist, so dass nichts mehr übrig bleibt für ein "tieferes" (und auch haften bleibendes) Verständnis für das, was das alles soll.<sup>18</sup>

Mehr Eigentümlichkeiten der Statistik werden deutlich, wenn wir uns im Folgenden mit "Manipulation" ganz allgemein beschäftigen.

### 3. Manipulation und Zahlengewinnung in der Statistik: grundsätzliche Fragen

Wir wollen zeigen, dass Manipulation (vor allem die ohne Statistik) viel verbreiteter und alltäglicher ist, als man zunächst glauben mag, und dass es andererseits aber sehr schwer ist, exakt zu bestimmen, wann man von Manipulation im Zusammenhang mit Statistik sprechen kann (und was dagegen z.B. nur eine Inexaktheit ist, die in der Natur der Sache liegt), geschweige denn, wann man hier jemandem etwas vorwerfen kann.

#### a) Manipulation ohne Statistik

Oft denkt man bei "Manipulation" nur an eine ideologische Dauerberieselung durch ein autoritäres Regime oder vielleicht an eine subtile, die Tiefenpsychologie ausnutzende Werbung. Tatsächlich sind wir aber im Alltagsleben oft ganz anderen Manipulationen ausgesetzt, und das weitgehend unbemerkt und nicht etwa nur initiiert von perfiden Menschen, die Macht über uns haben. Hierzu gehören z.B. Versuche, jemand durch Geschenke, Erregen von Mitleid oder durch Gruppenzwang "moralisch" in die Enge zu treiben. Wie wir noch sehen werden, sind solche Manipulationen meist wirksamer als eine unvermeidlich etwas dröge wirkende Statistik, die das Denken und Fühlen der Menschen ohnehin in den seltensten Fällen nachhaltig beeinflusst. Vor allem wird dauernd mit der passenden Wortwahl manipuliert. Wenn man weniger als geplant Schulden macht, nennen Politiker das "sparen". Oder man sagt nicht gerne, dass der deutsche Steuerzahler "zahlen" soll; er soll stattdessen mehr "Verantwortung" übernehmen, oder "Solidarität" (mit Südeuropa) zeigen.<sup>19</sup>

<sup>17</sup> Nach Davis und Hersh, Descartes Traum, (engl. Titel: Descartes' Dream. The World According to Mathematics) 1986/88 S. 257 gab es die auch Ziffern 1, 2, ... und die Zeichen + und = erst 1481 bzw. 1557.

<sup>18</sup> Gerade in der Statistik kann man auch besonders wenig mit dem anfangen, was man nur auswendig

gelernt hat, oder von dem man nur mal etwas "gehört" hat, ohne es verstanden zu haben.

<sup>19</sup> Was natürlich heißt, dass er trotzdem zahlt. Nächstes Jahr jährt sich zum einhundertsten Mal der Ausbruch des Ersten Weltkriegs. Man wird dann schon hierzulande entsprechende Worte zu hören bekommen.

### *b) Was ist Manipulation mit Statistik und wem ist sie vorzuwerfen?*

Das soll nicht heißen, dass man nicht auch mit geschickt "getarnter" einseitiger Statistik manipulieren kann. Man kann es. Die Frage ist nur, ob es genauso wirksam ist, wie andere Arten zu manipulieren (z.B. mit emotionaler Rhetorik) – was wohl zu bezweifeln ist –, und ob hier überhaupt jemandem ein Vorwurf zu machen ist, wenn wir einmal annehmen der Manipulationsversuch sei als solcher überhaupt erkannt worden. Selbst dann könnte so etwas nämlich leicht mit guten Gründen bestritten werden. So bleibt am Ende von dem Vorwurf einer Manipulation oft nur übrig, dass man vielleicht eine "schlechte Statistik" gemacht hat. Aber wie kann man das jemandem vorwerfen, der es vielleicht nur nicht besser gekonnt hatte, oder der sich darauf berufen kann, dass auch alle anderen genau die gleichen umstrittenen Methoden anwenden?

Die Kosten-Nutzen-Analyse der "Manipulation" mit Statistik, oder vielleicht besser, der Produktion "schlechter Statistik" zeigt also, dass es wenig wahrscheinlich ist, dass dem Produzenten Sanktionen drohen, und wenn, wären sie kaum gravierend. Worin sollten sie denn schon bestehen? Schlimmstenfalls vielleicht darin, dass er eine neue Statistik vorlegen muss. Es spricht also alles dafür, dass "schlechte Statistiken" eine Dauererscheinung sind, weil per saldo der Nutzen größer ist als die Kosten.

Dann bleibt noch die Frage, wie das "moralisch" zu bewerten ist. In vielen populären Büchern oder Sprüchen über Statistik ist hier von "Lügen" die Rede. Manche Autoren unterscheiden sogar direkte Lügen (es werden bewusst falsche Zahlen als "Daten" ausgegeben, was selten genug passiert und wohl noch seltener nachgewiesen werden kann<sup>20</sup>)

<sup>20</sup> Ein Beispiel sind Fälschungen von Zahlen zum Außenhandel mit den "kapitalistischen Ländern" in der amtlichen Statistik der ehemaligen DDR, mit deren Nachweis aus Archivunterlagen ich mich einige Zeit beschäftigt hatte. Aber selbst dort haben unermüdlige Kämpfer für die Sache des Sozialismus versucht, mit den abenteuerlichsten Konstruktionen den Vorwurf der Fälschung zu entkräften. Danach habe – so

und indirekte Lügen. Für letztere ist kennzeichnend, dass das "Opfer" zum Mittäter wird, weil es die generelle Problematik bestimmter Methoden nicht durchschaut.

Von "Lügen" oder gar von "indirekten Lügen" zu sprechen halten wir für Unsinn und wir wollen im Folgenden noch etwas weitergehen und zu zeigen versuchen, dass die meisten sog. "Lügen" auf das Konto der Statistikenutzer (also der vermeintlichen "Opfer") gehen und dass es meist nicht die Statistiker sind, die hier die eigentlichen Täter sind. Sie wurden oft eher zu fragwürdigen Zahlen gedrängt, dann missverstanden und kritisiert

Was ist das z.B. für eine Geisteshaltung, wenn man mit einem Schlagwort wie "Beyond the GDP" von der Statistik verlangt, nicht nur die Wirtschaftsleistung, das GDP<sup>21</sup> (dessen Aggregation schon komplex genug ist), sondern auch soziale und ökologische Nachhaltigkeit, Gerechtigkeit und sogar das "Glück" der Menschen mit einer einzigen, alles zusammenfassenden Zahl (in Mrd. Euro) zu messen und wenn man sich dann darüber wundert (oder gar darüber beklagt), dass diese Zahl nur von zweifelhaftem Wert wäre, weil sie sehr verschiedene Dinge zum Ausdruck bringt und sich in ihr z.B. mehr Wohnungen und mehr Lärm gegenseitig aufheben, zusammen also zu Null addieren.<sup>22</sup> An der einen, alles umfassenden Zahl sind inzwischen zwar Zweifel entstanden, es bleibt aber dabei, dass Politiker ungerne für eine Sache mehr als eine Zahl haben und bei ihnen

wurde gesagt – Günter Mittag, dem die Wirtschaft (und damit auch die Statistik) der DDR unterstand, die richtigen Außenhandelszahlen gekannt, weil er auch andere, geheime Quellen hatte, aber er habe seine Statistiker, die diese Quellen nicht hatten, erst einmal herum rechnen lassen und dann deren Ergebnisse berichtet. Was nach einer von oben diktierten politisch motivierten Fälschung einer Statistik aussah, war danach also in Wahrheit nur deren "Berichtigung. Es mag Leute geben, die so etwas glauben.

<sup>21</sup> Gross domestic product, also Bruttoinlandsprodukt.

<sup>22</sup> Ganz abgesehen davon hatten wir entsprechende Diskussionen schon in den 70er Jahren des letzten Jahrhunderts unter dem Stichwort "soziale Indikatoren" gehabt. Und wir werden sie wohl auch künftig immer wieder bekommen, weil wohl jede Generation von Statistikenutzern die Möglichkeiten und Grenzen der Statistik für sich wieder neu entdecken muss.

schon zwei Zahlen, die sich evtl. sogar noch widersprechen Unmut erzeugen.

Dass man dieses Spiel ("alles" mit einer Zahl messen zu wollen), das man nie wirklich gewinnen kann, trotzdem immer wieder spielt, liegt wohl daran, dass es in der Politik zu einer Obsession geworden ist, zu glauben, "lieber irgendeine Zahl, als gar keine Zahl" haben zu müssen.<sup>23</sup>

Was ist an einer Zahl der Arbeitslosen von den Statistikern (direkt oder indirekt) "gelogen", wo es doch allgemein bekannt ist, oder sein sollte, dass der zugrundezuliegende Begriff der "Arbeitslosigkeit" notwendig stets problematisch ist (schon allein, weil es sich ja auch um ein *unfreiwilliges* Los des *arbeitswilligen* und *arbeitsfähigen* Betroffenen handeln muss), oder weil das zugrundeliegende Messkonzept (Befragung, oder Auswertung einer Kartei) seine Grenzen hat?<sup>24</sup> Wenn das "gelogen" sein soll, hat man nur die Wahl zwischen gar keiner Arbeitslosenzahl oder einer immer irgendwie "gelogenen" Zahl. Das allein zeigt schon, dass der Vorwurf der "Lüge" hier nicht ganz sauber durchdacht sein kann.

Oder was ist "gelogen" an einer Statistik, wenn es bei der Auswertung erst auf dem zweiten Blick erkennbar wird, dass es sich hier vielleicht um eine "Scheinkorrelation" handelt, oder wenn die Aussagefähigkeit eines empirischen Befunds überschätzt wurde weil, wie so oft, wieder einmal der Begriff "Signifikanz" missverstanden wurde?

Bei Statistik geht es nicht um geoffenbarte unzweifelhafte Wahrheiten,<sup>25</sup> sondern "nur"

<sup>23</sup> Vgl. unten die Ausführungen zum Stichwort "Computopia". Die m. E. richtige, bloß in der Politik nicht durchzuhaltende Gegenposition ist "Abelson's 5th Law: If you have nothing to say, don't say anything."

<sup>24</sup> Wer taucht in einer Kartei auf und wer nicht? Kann man den Absichtserklärungen ("ja, ich bemühe mich um einen Job") der Befragten Glauben schenken?

<sup>25</sup> Gedacht ist hier nicht nur an Religion und "Glauben", sondern auch an den (gescheiterten) Versuch, der sog. "Frankfurter Schule der Statistik", Statistik als eine Art "Logik" zu konstruieren. Wenn so etwas möglich wäre, könnte man seinem Gegner mangelndes Denkvermögen vorwerfen, weil er etwas bestreitet, was "logisch" nicht zu bestreiten ist (etwa so, wie  $2 + 2 = 4$  ist). Aber so liegen die Dinge in Sachen "Statistik"

um zahlenmäßige Feststellungen zur Realität, die mehr oder weniger fundiert und damit glaubwürdig sind. "Gelogen" oder auch nur "falsch" sind Kategorien, die hier nicht anwendbar sind.

### c) *Es kommt darauf an, wer für wen Statistik macht*

Von meinem Statistikerkollegen Werner Neubauer habe ich den folgenden Gedanken übernommen, der mir sehr einleuchtete, wofür ich aber auch von anderen kritisiert wurde:<sup>26</sup> Es ist ein entscheidender Unterschied, ob es Manipulationen (was immer das heißen mag) gibt bei der *amtlichen* Statistik (wie in dem erwähnten Beispiel der DDR) oder bei der *nichtamtlichen, also stets privaten* Statistik. Neubauer verglich die Situation der Statistik hier mit der der Rechtswissenschaft. Von einem Jurist als *Richter* erwartet man Unparteilichkeit und Verpflichtung auf Wahrheit und nichts als die Wahrheit. Anders liegt es beim Jurist als *Anwalt*. Er verhält sich vollkommen korrekt, wenn nicht das sein Ziel ist, sondern die Vertretung der Interessen seines Mandanten (selbst dann wenn diese moralisch fragwürdig sind). Er darf nicht nur, er muss sogar parteiisch sein.

Die amtliche Statistik ist mit dem Gericht vergleichbar. Sie soll allen dienen und von allen als seriös anerkannt werden, und sie muss dazu ein herausgehobenes Ansehen genießen,<sup>27</sup> wofür es zweierlei bedarf:

- der fachlichen Unabhängigkeit vom übrigen Staatsapparat, und der
- "professionellen Integrität" (Verwendung anerkannter Methoden nach dem neustem Stand der Wissenschaft).

tik" nicht. Wer hier anderer Meinung ist, muss deswegen nicht dumm, böswillig oder uneinsichtig sein.

<sup>26</sup> So z.B. von Constatin Birnstiel, Amtliche Statistik im Spannungsfeld zwischen wissenschaftlichem Anspruch und politischen Vorgaben, Diss. Universität der Bundeswehr München 2001, S. 12.

<sup>27</sup> Es sei dahin gestellt, ob sie es auch verdient hat. Aber wie auch immer: das zunehmende crowding out der amtlichen Statistik am lukrativen Datenmarkt durch immer mehr als gleichwertig angesehene nicht-amtliche Anbieter ist so gesehen ein ernstes Problem. Wir kommen darauf in Abschn. 8c zurück.

So wie ein Anwalt im Auftrag seiner Mandanten arbeiten kann, so kann auch ein Statistiker sein Fachwissen in den Dienst von Unternehmen, Verbänden, Ministerien usw. und deren Interessen stellen. Es gibt hierfür auch prominente Beispiele, etwa Sir Ronald A. Fisher (1890 – 1962), der statistische Erhebungen und Analysen im Auftrag der Tabakindustrie durchführte.

Genau darin aber, dass Statistik zur Unterstützung von Standpunkten und Interessen einer Partei genutzt werden kann, wird von nicht wenigen sogar der eigentliche Nutzen der Statistik gesehen. Statistik ist eben kein *l'art pour l'art*, und das ist auch nicht verwerflich.

#### 4. Voraussetzungen einer Statistik

Das mit Statistik stets verbundene Bemühen, Zahlen für die Beschreibung und Analyse der Wirklichkeit zu gewinnen verlangt bestimmte Voraussetzungen. So wie es "Kategorien" (wie Qualität, Quantität, Modalität<sup>28</sup> etc.) für das Denken gibt, so gibt es auch solche für die Gewinnung von empirischen Zahlen.

##### a) Zählen und messen und wie kann man zu Zahlen gelangen?

Statistische Aussagen betreffen stets Eigenschaften (Merkmale) von statistisch erfassten Einheiten (Personen, Betriebe, Städte etc.) einer "Gesamtheit" von solchen Einheiten. Übertragen auf "Erhebungen" als immer noch häufige Art der Zahlengewinnung, heißt das: *Wen* (Einheiten) soll man über *was* (Merkmale) *befragen*<sup>29</sup> und *wie viele* (z.B. eine Stichprobe) müssen befragt werden?

Bei den Merkmalen interessiert die Art der Ausprägungen, die ein Merkmal hat. Sind es nur endlich viele und gibt es zwischen ihnen nicht beliebig viele Abstufungen (wie z.B. bei Geschlecht oder Familienstand), dann ist das Merkmal "diskret"<sup>30</sup>, gibt es dagegen

(zumindest theoretisch) unendlich viele feine Abstufungen (z.B. bei der Körpergröße, die man ja auch in Millimeter oder noch feiner messen kann), spricht man von "stetig" (oder "kontinuierlich"). Diese Unterscheidung betrifft auch den Unterschied zwischen *zählen* und *messen* (vgl. hierzu auch den Nachtrag auf S. 31). Man kann zwar die *Anzahl* der Ledigen, Geschiedenen etc. feststellen, hat damit aber nicht eine "Variable" (etwa "Familienaffinität") "gemessen". Messen setzt voraus, dass eine *Metrik* (Nullpunkt und Maßeinheit, wie kg, m etc.) definiert ist, was bei stetigen Merkmalen meist der Fall ist.<sup>31</sup>

Das Zählen führt zu einer Häufigkeit (absolut als Anzahl der Fälle, bzw. relativ als Prozentzahl). Zur *Anzahl* fügt man meist keine Maßeinheiten (etwa "Stück") hinzu. Das Ergebnis einer Messung ist eine Anzahl von Maßeinheiten für eine Dimension, etwa 80 kg für die Dimension "Gewicht". Die heute häufige Verwendung merkwürdiger Phantasieeinheiten, wonach z.B. etwas 80 "Zähler" beträgt, zeigt allein schon, dass hier wohl nicht wirklich eine "Messung" vorliegt. Wir haben soweit nur formale Aspekte von zählen und messen besprochen. Auf inhaltliche Aspekte gehen wir im nächsten Abschnitt ein und wir kommen nun zu den Fragen: wer wird befragt und wie viele werden befragt?

##### b Biased sample und "Repräsentativität"

Als ein eher scherzhaftes Beispiel für eine verzerrte Stichprobe (biased sample) gilt das Ergebnis von nahezu 100% "ja" bei der Frage "Antworten Sie gerne auf eine statistische Befragung?" Klar, denn die, die nein geantwortet hätten, haben den Fragebogen einfach

---

sichtsvoll, zurückhaltend, verschwiegen etc. Im Englischen gibt es hier zwei Wörter – mit gleicher Aussprache – *discrete* (im obigen statistischen Sinn) und *discreet* (als taktvolles Verhalten).

<sup>31</sup> Genau genommen geht es an dieser Stelle um die Unterscheidung verschiedener Skalenarten, was wiederum damit zusammenhängt, welche Methoden in der Statistik sinnvoll angewendet werden können, welche Eigenschaften der Zahlen (z.B. bei den natürlichen Zahlen die Unterscheidbarkeit  $1 \neq 2$  und die Reihenfolge  $1 < 2 < 3$  usw.) man sich dabei zunutze macht und welche Transformation zulässig sind, die den Informationsgehalt der Zahlen nicht verändern.

<sup>28</sup> Es geschieht z.B. zufällig oder "mit Notwendigkeit".

<sup>29</sup> Der Begriff ist hier allgemein im Sinne von Daten *erheben* gemeint.

<sup>30</sup> Das Wort bereitet regelmäßig Verständnisschwierigkeiten, weil diskret auch so viel heißt wie rück-

in den Papierkorb geworfen (ein Fall von Selbstselektion durch Nichtbeantwortung).<sup>32</sup>

Den wenigsten fällt wahrscheinlich auf, dass das Ergebnis, dass ca. 90% der Deutschen für höhere Steuern der Reichen sind, von gar nicht so grundsätzlich anderer Qualität ist. Denn wenn ca. 90% die Anderen für die Reichen halten, hätte man auch fragen können: "Sind Sie dafür, dass die *Anderen* höhere Steuern zahlen?" Natürlich sind wir dafür; denn wir haben dann ja auch mehr Kindergärten, mehr Schulen usw., und das haben alles ja "die Anderen" bezahlt.

Der Grund für eine *Zufalls*auswahl in Gestalt einer Stichprobe ist u.a. gerade der Versuch, eine Verzerrung (engl. "bias") zu vermeiden, wie sie z.B. entsteht, wenn man nur ganz bestimmte Leute befragt. Wenn eine bewusst einseitige Bevorzugung oder Benachteiligung, kurz ein Bias vermieden werden soll, lassen wir gern das Los, also den Zufall entscheiden. In diesem Sinne (Objektivität, Rechenbarkeit) sind Zufall und Wahrscheinlichkeit zentrale Konzepte bei Stichproben und Experimenten, aber nicht nur dort, sondern auch in anderen Gebieten der Statistik.

Im Zusammenhang mit Stichproben tritt immer die Frage nach der "Repräsentativität" auf.<sup>33</sup> Das Wort wird gerne so benutzt, als gäbe es ein mehr oder weniger "repräsentativ". Fragt man dann aber, wie man feststellt, um wie viel größer die Repräsentativität in einem Fall ist als in einem anderen, entsteht oft eine Verlegenheit. Das ist auch kein Wunder, weil Repräsentativität kein sinnvoller Begriff und deshalb auch kein Fachbegriff ist.<sup>34</sup> Fragen, wie "wie viele Befragte"

und "wie auswählen" führen uns zur Wahrscheinlichkeitsrechnung (Abschn. 7), die für die Statistik von immenser Bedeutung ist.

## 5. Messung, Dimensionen und Aussagefähigkeit von Zahlen

Betrachten wir nun die bisher noch ausgeblendeten mehr *inhaltlichen* Aspekte von "zählen" und "messen", so zeigt sich, dass noch einiges mehr zu den "Zahlen" (und deren Interpretation) gesagt werden kann als das, was in Abschnitt 2 gesagt wurde. Dazu muss man zunächst den Begriff "Merkmal" (oder "Variable") etwas genauer betrachten.

### a) Definieren und operationalisieren

Wie könnte man z.B. die "Bildung" von Personen messen: Durch einen Test, dem sich die Befragten unterziehen müssen? Durch Selbsteinschätzung der Befragten (wenn diese gering/mittel/groß oder ähnliche Antwortvorgaben im Fragebogen ankreuzen)? Durch den höchsten Schulabschluss (wenn die Abschlüsse überhaupt so von "niedrig" bis "hoch" auf einer Dimension abzubilden sind)? Oder durch die Anzahl der Jahre des Schulbesuchs (je mehr, desto "gebildeter")?

Fragen dieser Art sind bei jedem Merkmal zu klären; es ist nur so, dass die Definition eines Merkmals mehr oder weniger kompliziert sein kann. Sich jetzt lange über das "Wesen" einer Sache den Kopf zu zerbrechen (womit man gerade bei "Bildung" sicher viel Zeit verbringen könnte) führt meist nicht sehr weit. Deshalb sind Merkmale in der Statistik i.d.R. "operational" (durch Angabe von Operationen) definiert. Eine häufig (aber nicht immer) gegebene Möglichkeit etwas operational zu definieren<sup>35</sup> ist z. B. aufzulisten, was einbezogen werden soll.

Die Financial Times meldete im April 2013, dass eine Revision des Bruttoinlandsprodukts (BIP) durch das US Bureau of Economic Analysis (BEA) ein um ca. 3% höheres

<sup>32</sup> Das Beispiel stammt von D. Huff, *How to Lie with Statistics*, London (Penguin Books) 1973, S. 17.

<sup>33</sup> Gerade weil "zufällig" bei einer Stichprobe bedeutet, dass es bei der Einbeziehung einer Einheit in die Stichprobe keine Bevorzugung oder Behinderung gibt, kann angenommen werden, dass eine einbezogene Einheit auch eine nichteinbezogene "repräsentieren" kann. "Repräsentativität" kann also eigentlich nichts anderes als "Zufälligkeit" der Auswahl bedeuten.

<sup>34</sup> Ich habe mich wiederholt an anderen Stellen dazu geäußert. Was der Statistiknutzer mit Repräsentativität meint ist eigentlich eine Frage des Stichprobenfehlers (der anders als die Repräsentativität sehr wohl zu quantifizieren ist).

<sup>35</sup> Spiele, wie z.B. das Schachspiel sind nur so, d.h. "operational" definiert, indem man festlegt, welche Bewegungen (Operationen) mit den Figuren zulässig sind und nicht indem man endlos über das "Wesen" der Figuren Turm, Pferd usw. philosophiert.

BIP für die USA ergibt. Dahinter steht vor allem eine andere Verbuchung der F&E Ausgaben (als Investition) aber auch eine geänderte Behandlung von künstlerischen Originalen (Manuskripte etc.), Nebenkosten des Immobilienerwerbs und Pensionszusagen.<sup>36</sup> Nicht nur die Auswirkung (auf das BIP), auch die Begründung für revidierte Verfahren war auch hier wieder Anlass für die üblichen Diskussionen zur Operationalisierung. Wir gehen nicht auf Details ein und ziehen nur einen sehr allgemeinen Schluss hieraus.

Wenn die Operationalisierung problematisch ist (und das ist sie bei allen etwas komplizierten Merkmalen), dann sind es natürlich auch die so gewonnenen Zahlen. Aber wenn man einmal eine Zahl hat, vergisst man irgendwie immer gerne, dass sie nicht vom Statistiker als naturgegeben vorgefunden wurde, sondern ganz entscheidend durch eine Operationalisierung gewonnen wurde. Natürlich kann man mit guten Gründen die Revision einer Abgrenzung und Behandlung (Verbuchung, Bewertung etc.) einzelner Positionen, wie im Falle des BIP (oder GDP) kritisieren, aber was könnte man dann anderes tun, als sich vielleicht für eine neue Abgrenzung, Bewertung etc. stark machen, die dann auch wieder kritisiert werden wird?

Es ist nun wirklich nicht sehr tiefgründig, wenn man hier von "Manipulation" oder gar "Lügen" der Statistik spricht, nur weil es Dinge gibt, die grundsätzlich (und wohl auch für alle Zeiten) immer schwieriger zu zählen sind als z.B. Erbsen.

### **b) Eine und mehrere Dimensionen**

Die meisten Merkmale sind genau genommen "komplex", indem sie mehrere Dimensionen umfassen. Betrachtet man die Variable  $x$  als "Qualität", "Leistung" o.ä. eines Professors, so ist jedem schon "gefühlsmäßig" völlig klar, dass es hier gewaltige Unterschiede zwischen den Professoren gibt. Aber vom Gefühl zur einer als korrekt akzeptierten Zahl (an der ohne Zweifel ein berechtigtes Interesse besteht) ist noch ein weiter Weg.

<sup>36</sup> Im Weltmaßstab wurde der Effekt der Revision als "equivalent to adding a country as big as Belgium to the estimated size of the world economy" angegeben.

Die Variable  $x$  hat mindestens zwei Dimensionen, die Leistungen in der Forschung (F) und in der Lehre (L). Angenommen, es wäre eindeutig klar, ob ein Professor hier gut (+) oder schlecht (-) ist, was aber "in der Realität" bekanntlich nicht der Fall ist, dann gibt es vier Professorentypen:

	F -	F +
L -	A --	B - +
L +	C +-	D ++

Klar ist  $x_A < x_D$  (also -- ist eindeutig schlechter als ++), aber wie sind die "hybriden" Fälle von Prof. B und Prof. C einzuordnen? Das hängt natürlich davon ab, ob man die Lehre oder die Forschung höher "gewichten" soll. Wir haben hier nur zwei Dimensionen betrachtet. Mit mehr, etwa  $\lambda$  Dimensionen (was ja eine bessere Messung liefern soll)<sup>37</sup> wird das Verhältnis der eindeutigen zu den hybriden Fällen immer ungünstiger. Wir haben stets nur zwei eindeutige, aber eine wachsende Zahl von  $2^\lambda - 2$  hybriden Fälle (bei  $\lambda = 3$  zwei eindeutige, --- und +++, aber schon  $2^3 - 2 = 8 - 2 = 6$  hybride Fälle).

Mit einem Trick kann man sich aus allen diesen Schwierigkeiten heraus retten und sogar bei jeder Dimension (wie F und L) nicht nur zwei, sondern sehr viel mehr Abstufungen einführen. Man kann sogar sagen, wie groß der Unterschied zwischen den eindeutigen Fällen ist, also die Differenz  $x_D - x_A$  ist, nicht nur dass  $x_D > x_A$  ist. Mehr noch: wir haben kein Problem mehr mit hybriden Fällen und alle Professoren lassen sich schön in eine Reihe anordnen.

Der Trick ist allgemein bekannt: man vergibt Punktzahlen für die einzelnen Dimensionen und addiert (oder mittelt) diese Punktzahlen (mit oder ohne Gewichte, mit denen die unterschiedliche Wichtigkeit der Dimensionen zum Ausdruck gebracht werden soll).

<sup>37</sup> Nach einer allgemeinen Vorstellung wird es beim Ranking umso besser, je mehr Dimensionen man einbezieht. Das ist aber deshalb nicht ganz so selbstverständlich, weil es auch darauf ankommt, ob und wie die Dimensionen miteinander korreliert sind.

Die Rechtfertigung dieser sehr beliebten Methode ergibt sich (wie auch sonst ziemlich oft) praktisch nur aus dem "Prinzip des mangelnden Grundes" (principle of insufficient reason): ich weiß nicht, wie ich es sonst machen sollte (oder: ich kenne nichts Besseres).<sup>38</sup> Denn es machen hier nur rein inhaltliche, nicht statistische Überlegungen aus, ob Sinn oder Unsinn produziert wird. "Von außen gesehen" ist nämlich eine Summe immer eine Zahl und man kann Zahlen (wenn es nicht gerade komplexe Zahlen sind) an einem Strahl (also in *einer* Dimension) anordnen.

Man kann die Professoren nach ihrer "Leistung" aber auch dann wieder auf einem Strahl anordnen, wenn man bei der Punktschritte als weitere "Leistungen" noch Alter (A) und Schuhgröße (S) addiert hätte. Es ist nur eine Frage der inhaltlichen Interpretation, ob man eine so um A und S erweiterte Punktschritte als sinnvolle "Messung" akzeptieren möchte. Es wird gerne vergessen, dass dies aber auch schon für die Punktschritte vor ihrer Erweiterung durch A und S galt.

### c) Zahlen von außen und Zahlen von innen gesehen, Sinnhaftigkeit von Zahlen

Halten wir einige Einsichten über Zahlen fest:

1. Beim Rechnen erhält man immer Zahlen (also ein "Ergebnis" in Gestalt von Zahlen [von außen gesehen]). Es gibt nichts, was "nicht messbar" wäre, man erhält immer einen Messwert (auch wenn er faktisch nichts wert ist).
2. aber die Zahlen sind (quasi "von innen gesehen") nicht immer sinnvoll<sup>39</sup> und man kann es den Zahlen von außen nicht ansehen, ob sie sinnvoll sind oder nicht,
3. aber man sucht trotzdem immer nach einem "Sinn" in den Zahlen und hat

<sup>38</sup> Man könnte z.B. mit der Faktorenanalyse prüfen ob den Messungen (mit Punktzahlen) für die einzelnen Dimensionen überhaupt nur eine (wie bei der Punktschritte als selbstverständlich vorausgesetzt) oder nicht vielmehr mehrere "latente" Dimensionen (Faktoren) zugrundeliegen.

<sup>39</sup> Ein Beispiel ist das von mir wiederholt zitierte "Gesamtgewicht von Brüssel" mit allen Häusern, Bäumen Menschen usw. von dem Statistiker A. Quetelet.

4. lieber *irgendeine* (also auch eine sinnlose) Zahl (Messung) als keine Zahl.

Mit der Innensicht und Sinnsuche bei Zahlen berühren wir ein anderes, im Zitat von Nikolaus von Kues anklingendes Verständnis der Zahl: Es ist der Zusammenhang mit "Ordnung" und "Harmonie".

Hat man für eine wichtige Sache noch keine Zahl, dann hat man das Gefühl, in die Sache noch keine "Ordnung" gebracht zu haben. Und auch wenn man sehr genau weiß, dass die Zahl äußerst problematisch ist, so hat man doch irgendwie ein besseres Gefühl, als wenn man so ganz ohne eine Zahl dastünde.

Die Zahlen selbst können auch "Gefühle" von Stabilität (es gab "immer schon" 4 Jahreszeiten, 4 Himmelsrichtungen)<sup>40</sup> und Harmonie tangieren. Es ist üblich, dass man in der Politik ein "Zehnpunkte Programm" hat. Selten hört man von einer 9 oder 11 Punkte Erklärung eines Politikers. So haben auch wir hier 10 Abschnitte; 9 oder 11 sähe irgendwie nicht ganz so "harmonisch" aus.

Was "sinnvoll" ist dürfte also auch viel eine Sache von Gefühl und Gewohnheit sein. Es ist außerdem eine Frage der "Erklärung", der "Geschichte" hinter den Zahlen. Früher war es ja – wie gesagt – nicht nur üblich, in Bildern und Gleichnissen zu reden, sondern auch Inhalte in Dialoge und Geschichten zu verpacken. Es gibt ein Bedürfnis, das Zustandekommen einer Zahl durch eine "story" zu verstehen, sich aus Zahlen "einen Reim zu machen".

Das zeigt sich am Phänomen der "Gestern-Erklärer" des Börsengeschehens. Diese Leute "erklären" warum sich die Kurse so entwickeln "mussten" (nachdem sie gerade gesehen haben, wie sie sich entwickelt hatten), etwa wegen einer "Blase" oder einer "Seitwärtsbewegung". Voraussehen ist wegen der Ungewissheit der Zukunft schwierig, aber

<sup>40</sup> In einem sehr viel tieferen Sinn fasziniert die Menschen "Stabilität" im Zusammenhang mit Zahlen: der berühmte Physiker Richard Feynman (1918 – 1988) bekannte z.B. dass ihn das Staunen darüber, dass die Konstante  $\pi$  praktisch überall vorkommt (auch dort, wo es gar nicht um Kreise geht; z.B. bei der Elektrizität) maßgeblich bewogen hatte, Physiker zu werden.

"erklären" der Vergangenheit geht immer. Das ist allgemein bekannt. Trotzdem lacht man nicht über "Gestern-Erklärer" (after-the-fact story teller), sondern man hört ihnen zu.

Das Bedürfnis, einer Zahl einen "Sinn" zu geben ist meist (erst) dann befriedigt, wenn sie "erklärt" werden kann (d.h. dass etwas so sein muss wie es ist, bzw. so kommen musste wie es gekommen ist) und wenn sie zu "Erklärungen" in Gestalt von Erwartungen, Erfahrungen oder Vorurteilen passt.

Man hört deshalb das, was man *gerne* hört und man neigt dazu, die Zahl herauszugreifen, die zu den Vorstellungen passt, die man von einer Sache hat.

#### **d) Statistik und Astrologie und warum im Zweifel die Statistik falsch ist und nicht die Überzeugung**

Dass die Zahlen der Statistik nur "von außen gesehen", oder "formal" Produkt der Mathematik sind, bei ihrer "Sinnggebung" aber vieles andere (nicht selten auch Emotionales) hinzukommt, und die Möglichkeit, sich Passendes auswählen zu können, hat im Falle der Statistik zur oft zu hörenden Aussage geführt, man könne mit Statistik "alles beweisen".<sup>41</sup> Das legt einen vielleicht etwas ketzerischen Vergleich nahe. Danach ähneln sich Statistik und Astrologie in zwei Punkten:

1. auch wenn eine Sache viel mit Mathematik zu tun hat, kann sie trotzdem Humbug sein, und
2. weil alles zwei oder noch viel mehr Seiten hat, kann man auch für alles (z.B. für jeden Charakter) eine passende Beschreibung (oder – bei Statistiken – einen passenden Gutachter) finden und aus den Zahlen das herauslesen, womit man sich bestätigt fühlen kann.

Früher (etwa zur Zeit von J. Kepler 1571 – 1630) galt Astrologie als eine sehr mathematische und auch ernstzunehmende Disziplin. Heutzutage gilt das, anders als bei der Statis-

tik wohl nicht mehr. Trotzdem gibt es "Humbug" auch mit Statistik, aber das kommt erst nach der Mathematik, mit falsch *angewandten* (weil nicht verstandenen) Methoden und falsch *interpretierten* (Interpretieren ist nicht Sache der Mathematik) Ergebnissen ins Spiel.

Dass man sich bei der Beschreibung von Tierkreiszeichen gerne das über seinen Charakter aussucht, was man schön findet, erklärt, warum sich so viele Menschen in der Beschreibung "wiederfinden". Auch in der (angewandten) Statistik kann man sich oft das Passende aussuchen. Wenn einer lobt, dass A, B, C ... gestiegen ist, kann man oft damit dagegenhalten, dass D, E, F, ... gesunken ist, ganz abgesehen davon, dass man bei jeder Statistik methodische Mängel reklamieren kann.

Auch wenn jeder vor den gleichen Zahlen steht und keiner mit anderen Zahlen aufwarten kann wird jeder seinen eigenen Reim daraus machen können. Bei den Wahlabenden erlebt man regelmäßig das Spiel, dass die Prozentzahlen so gedeutet werden können, dass eigentlich jede Partei gewonnen hat.

Es gibt grundsätzlich keine zwingende Notwendigkeit, aufgrund von Statistiken die Dinge anders zu sehen als man sie sehen möchte.

Das erklärt auch, warum auch noch so viele empirische Studien, z.B. über positive Wirkungen von Studiengebühren oder negative Wirkungen von einem einheitlichen gesetzlichen Mindestlohn die Leute nicht davon abbringen werden gegen das eine und für das andere zu sein.

## **6. Deskription, Vergleiche und Zusammenhänge**

Wir kommen nun zu einigen einfachen Überlegungen aus der sog. "Deskriptiven Statistik" (im Unterschied zur "Induktiven Statistik", auf die wir im folgenden Abschnitt 7 eingehen) und darauf, wie man gerade in diesem Bereich schnell mit dem Vorwurf der "Manipulation" bei der Hand ist (wohl weil die meisten Statistiknutzer auch nur mit diesem Teil der Statistik zu tun haben).

<sup>41</sup> Man kann darüber streiten, ob dies nicht sogar bei dieser Materie bis zu einem gewissen Grade wahr ist. Es scheint zumindest so zu sein, dass man im Falle der Statistik hin und wieder mit Effekthascherei und Hochstapelei durchaus Erfolg haben kann.

### a) Spielregeln der Deskription

Das Ziel dieses Teils der Statistik besteht darin, aus der Fülle von erhobenen Daten für einzelne Einheiten zusammenfassende Kennzahlen zu bestimmen, die charakteristisch sind für die Gesamtheit der erhobenen Einheiten (z.B. befragten Personen). Jeder kennt das im Falle des Mittelwerts: die fünf Einzelwerte 2, 4, 5, 7, und 12 werden zum arithmetischen Mittel (Durchschnitt) 6 "verdichtet". Es ist nun wirklich keine große Leistung, zu erkennen, dass bei einem Mittel von 6 nicht alle Einzelwerte 6 sein müssen, sondern z.T. sehr viel kleiner (wie etwa 2) oder auch sehr viel größer (wie 12) als 6 sein können. Natürlich geht die Unterschiedlichkeit der Einzelwerte verloren, aber dafür gewinnt man ja auch etwas durch eine zusammenfassende Kennzahl, wie den Mittelwert. Das war sozusagen, die Spielregel, auf die man sich mit Berechnung eines Mittelwerts einließ. Für viele ist es ja auch – wie gesagt – von unschätzbarem Gewinn, nur noch eine Zahl, statt fünf Zahlen betrachten zu müssen. Trotzdem halten es dann immer wieder einige für geistreich, darauf hinzuweisen, dass trotz des Mittels von 6 die Einzelwerte sehr viel kleiner oder auch größer sein können als 6, und dass 6 damit nur wenig aussagt. Das läuft natürlich darauf hinaus, die Spielregeln im Spiel zu ändern: zuerst will man *eine* Zahl, und dann klagt man darüber, dass man *nur eine* Zahl (und nicht mehr alle ihr zugrundeliegenden Zahlen) hat.<sup>42</sup>

### b) Vergleiche und Bezugsgrößen

Eine "stand alone" Zahl aus der Statistik ist i.d.R. nicht viel wert. Was man stets braucht ist ein "standard of comparison" (Abelson), um einschätzen zu können, ob etwas groß oder klein ist. Ein solcher Maßstab wird durch

- Erwartungen und/oder

<sup>42</sup> Bekanntlich hängt das mit der Streuung zusammen. Dass es i.d.R. nützlich ist, neben einem Mittelwert auch die Streuung zu betrachten, dürfte trivial sein. Aber die Wichtigkeit der Streuung wird dann auch gerne gleich wieder vergessen, wenn es heißt ein Mittel von 6, errechnet aus fünfzig entsprechenden Einzelwerten sei notwendig immer besser als eines aus nur fünf Werten.

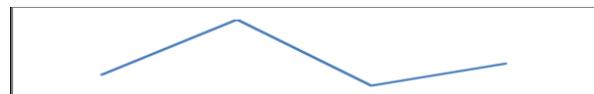
- andere Beobachtungen

geliefert. Die anderen Beobachtungen können auch in allgemeinen Erfahrungen liegen.

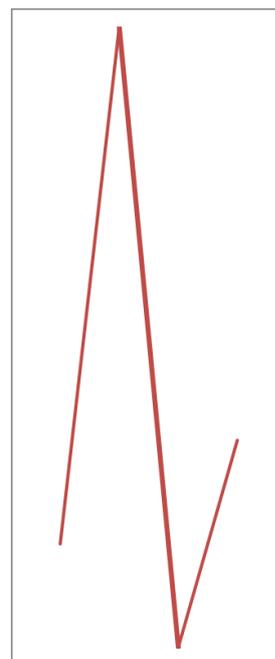


Das eröffnet natürlich auch der Manipulation Möglichkeiten, indem man dem Nutzer der Statistik einen falschen Vergleichsstandard nahelegt, oder gar keinen anbietet. Bei diesem Bild (E. Laspeyres 1834 - 1913) sieht man sofort, dass die senkrechte Achse in die Länge gezogen wurde; denn so sieht ja kein Mensch aus. Bei einem Gesicht haben wir ein natürliches Empfinden für Proportionen und wir erkennen sofort eine Manipulation am Maßstab einer Achse. Bei einer Graphik in der Statistik ist das nicht so.

Selbst wenn es einen Maßstab mit Zahlen auf der senkrechten Achse gäbe, ist es schwer zu entscheiden, ob diese blaue Kurve



oder aber die rote Kurve die "richtige" ist (weshalb auch beide gleich gut sind).



Denn anders als beim Gesicht, für dessen Proportionen man ja ein Gefühl hat (so dass sich auch schnell das "Gefühl" meldet, wenn etwas nicht zu stimmen scheint oder nicht plausibel ist) haben wir bei einer statistischen Graphik oft keinen standard of comparison aus der Erfahrung oder aus Erwartungen.

Ein Beispiel für einen offenbar falschen Vergleichsstandard, in Ge

stalt einer falschen Erwartung ist das Stauen darüber, dass der Vatikanstaat die weltweit höchste Kriminalitätsrate (pro Kopf) hat (bzw. haben soll, wie es heißt). Die hohe Kriminalitätsrate ist unerwartet, weil man doch beim Vatikanstaat an Priester denkt; sie ist aber andererseits auch dann wieder nicht so sensationell, wenn man bedenkt, dass der Vatikan ein besonders kleiner Staat ist mit stets sehr vielen Touristen und deshalb auch vielen Taschendieben.

Vergleiche haben eine formale und eine inhaltliche Komponente. Betrachten wir zunächst den formalen Aspekt und erst später den inhaltlichen. "Formal" heißt hier: wie bringt man rechnerisch zum Ausdruck, wie groß  $x$  im Verhältnis zu  $y$  ist?

Solche Vergleiche werden oft vorgenommen mit Hilfe der Prozentrechnung, bei der die "Basis" dann "gleich 100 gesetzt" wird. Man sollte nicht meinen, wie oft hier schon die einfachsten Rechnungen Schwierigkeiten bereiten. Man kann fast täglich hierfür Demonstrationsbeispiele mit den entsprechenden Nonsens-Zitaten finden. In einem Artikel über die Hochschulfinanzierung steht

"Der Anteil der Länder an der Finanzierung der Hochschulhaushalte ist seit 1995 von etwa 89 Prozent auf knapp 84 Prozent im Jahr 2012 kontinuierlich gesunken. Das klingt nicht dramatisch, aber wenn man die die konkreten Summen betrachtet, hat der Bund seine Zuwendungen in dem Zeitraum von 1,8 Milliarden Euro auf zuletzt 4 Milliarden Euro angehoben – eine Steigerung um über 60 Prozent."<sup>43</sup>

Man würde meinen, dass von 1,8 auf 4 eine Steigerung um 122,2% ist, denn  $4/1,8 = 2,22$ . Auf jeden Fall ist sofort mühelos erkennbar, dass es eine Zunahme von über 100% (also mehr als eine Verdoppelung) sein muss, denn eine Verdoppelung von 1,8 ist 3,6 und das liegt immer noch unter 4.<sup>44</sup>

<sup>43</sup> Marion Schmidt und Martin Spiewak, Die fetten Jahre sind vorbei. Der Bund soll die Universitäten retten, DIE ZEIT, 13.6.2013, Nr. 25/2013

<sup>44</sup> Wie man hier tatsächlich gerechnet hat und auf über 60% gekommen ist, bleibt ein Rätsel. Vielleicht wurde  $4 - 1,8 = 2,2$  durch 4 geteilt, was dann aber wegen  $2,2/4 = 0,55$  unter, nicht über 60% wäre. Es fragt sich auch, ob man so ohne weiteres vom Finanzierungsanteil (89 und 84%) *der Länder* (an der "Finanzierung der Hochschulhaushalte") zu den Ausga-

ben (in Mrd. €) *des Bundes* überleiten kann, zumal es in dem Artikel auch um Ausgaben (des Bundes) für das Bafög und außeruniversitären Einrichtungen geht.

Was die inhaltliche Seite betrifft, so ist zu fragen, welche anderen Zahlen als Vergleichsstandard wirklich "vergleichbar" sind und welche Kausalvermutungen evtl. implizit hinter einem Vergleich stecken.

Bei Verhältniszahlen  $x/y$  geht es um die Wahl der richtigen Bezugsgröße  $y$  im Nenner für die Größe  $x$  im Zähler. Ein Beispiel: Um das Todesrisiko des Reisens im Flugzeug mit dem der Deutschen Bahn zu vergleichen, könnte man als richtige Bezugsgröße  $y$  (für die Unfalltoten  $x$ ) denken an:

1. die Zahl der transportierten Personen (P),
2. die geflogenen (gefahrenen) Kilometer (= Strecke S),
3. die Personenkilometer (Produkt P·S)
4. die Zeit, die man im Zug/Flugzeug bringt,
5. die Anzahl der Flüge (Bahnfahrten),
6. die Anzahl der entsprechenden Fahrzeuge (Züge bzw. Flugzeuge)?

Jede dieser sechs Bezugsgrößen mag einen Sinn haben. Aber was ist der "Sinn" hier wirklich? Hinter Nr. 2 steckt offenbar die Vermutung, dass (tödliche) Unfälle umso wahrscheinlicher sind, je weiter die Reise ist. Man könnte aber auch sagen, dass dahinter die Vorstellung steht, dass der Tod auf einem Flug von München nach Tokyo weniger tragisch ist als auf einer Bahnfahrt nach Hannover, nur weil Tokyo weiter weg ist als Hannover. Oder: ist Tokyo  $n$ -mal so weit weg von München wie Hannover, dann sind  $n$  Tote auf dem Flug "so gut wie" ein Toter auf der Bahnfahrt, was wohl eine etwas makabere Logik ist.

Es ist auch zweifelhaft ob ein Unglück weniger tragisch wird,<sup>45</sup> wenn sich mehr Menschen dem entsprechenden Risiko aussetzen (wie dies Nr. 1 impliziert), oder dadurch,

<sup>45</sup> Der Bruch (und damit die Gefährlichkeit) wird bei gegebenem Zähler kleiner wenn der Nenner größer wird (mehr Menschen und mehr Zeit in der Bahn als im Flugzeug, also ist die Bahn weniger gefährlich als das Flugzeug).

dass die Menschen mehr Zeit in der Bahn als im Flugzeug verbringen (wie bei Nr. 4). So gesehen wäre auch der Tod im Bett, obgleich er dort – wie allgemein bekannt ist – viel häufiger auftritt, als andernorts, relativ betrachtet, gar nicht so dramatisch, weil die meisten Menschen (v.a. Alte und Kranke) dort weit mehr Zeit verbringen als in der Bahn oder im Flugzeug.<sup>46</sup>

Noch ein kleines Problem: die Todesfälle haben um 20% zugenommen  $x_1 = fx_0 = 1,2x_0$ , was an sich Anlass zur Besorgnis sein sollte, aber weil die Bezugsgröße  $y$  um mehr als 20% zugenommen hat, z.B. um 50%, kann die Relation  $x_1/y_1$  im Verhältnis zu  $x_0/y_0$  abgenommen haben (im Beispiel um 20% weil  $1,2/1,5 = 0,8$ ); die gleiche Sache "relativ gesehen" also gar nicht so bedrohlich aussehen.

### c) Scheinkorrelation und Milchmädchenrechnungen

Kennzeichen einer Scheinkorrelation zwischen zwei Variablen  $X$  und  $Y$  ist, dass  $X$  und  $Y$  korreliert sind, obgleich keine (direkte) Kausalbeziehung besteht, sondern nur eine indirekte über eine dritte Variable  $Z$ . Was demgegenüber eine Milchmädchenrechnung ausmacht ist, dass man etwas als unkorreliert vorausgesetzt, obgleich es kausal verknüpft ist. Das ist somit in gewisser Weise das Gegenteil der Scheinkorrelation.

Die dritte Variable  $Z$  kann auch in Gestalt eines sehr vage formulierten Hintergrundfaktors  $H$  bestehen oder in einer Strukturveränderung. So mag z.B. ein Zusammenhang zwischen Rauchen und Krebs auch dadurch zustande kommen, dass die typische "Raucherpersönlichkeit" risikofreudiger ist und einen entsprechenden Lebenswandel pflegt.<sup>47</sup>

<sup>46</sup> Dass die meisten Menschen im Bett sterben und nicht bei Verkehrsunfällen ist auch ein "klassisches" Beispiel für einen nicht entscheidungsrelevanten (operational nutzbaren) Zusammenhang: soll man denn im Stehen schlafen? Außerdem wäre auch "Krankenhausbett aussagefähiger als nur "Bett".

<sup>47</sup> Vgl. Walter Krämer, So lügt man mit Statistik, überarbeitete Neuauflage, München, Zürich (Piper), 2011 S. S. 169 – 172 für derartige Beispiele. Auch hinter einem Zusammenhang zwischen einem hohem Ausländeranteil und hoher Kriminalität kann ein  $Z$  stehen, etwa in Gestalt der Bevorzugung (sowohl von

Eine Strukturveränderung kann hinter einer steigenden Rentabilität von Unternehmen stehen, weil die Grenzanbieter vom Markt verdrängt worden sind.

Ich habe für beides, Scheinkorrelationen und Milchmädchenrechnungen an anderer Stelle<sup>48</sup> und auch im mündlichen Vortrag Beispiele und weitere Erklärungen gebracht, so dass wir diesen Punkt hier nicht weiter ausführen müssen.

### d) Operational nutzbare Zusammenhänge

In einem Bericht über den sog. Gehaltsreport 2013 steht in der Schlagzeile auch (was kein Witz ist) "... Er erklärt auch, was Sie tun können, um mehr zu verdienen"<sup>49</sup> Es wurden folgende Durchschnittsgehälter (Jahresgehälter in €) zitiert

Berlin	47.500*
München	62.500
Sozialwissenschaftler	47.500*
Mediziner	92.500

\* offenbar zufällig genau der gleiche Wert

Der Rat, den man einem Berliner Soziologen geben könnte (im Sinne der Schlagzeile) könnte dann lauten: nach München umziehen und Mediziner sein. Da er letzteres aber nicht so ohne weiteres kann, könnte der Rat auch lauten: 20 Jahre jünger werden und dann neu anfangen zu studieren (nämlich Medizin). Es ist damit klar, dass der Zusammenhang zwischen Fachrichtung und Einkommen nicht so ohne weiteres operational nutzbar ist.

Anders ist es beim Zusammenhang zwischen Essen und einem friedlichen (oder durch das Essen friedlicher werdenden) Verhalten der Menschen. Er ist operational nutzbar und wurde genutzt in Gestalt von Bordverpflegung der Passagiere in Flugzeugen (etwas,

Ausländern als auch Kriminellen, wobei dies nicht die gleichen Personen sein müssen) der Großstadt gegenüber anderen Gemeindegrößen.

<sup>48</sup> Vgl. meine Satire "Was tun, wenn einem eine Statistik nicht passt? Zwölf Grundsätze, die nützlich sein könnten, wenn man gegen statistische Daten und Analysen eines Redners argumentieren möchte".

<sup>49</sup> SPON (Spiegel online) vom 24.5.2013.

was es ja so bei der die Menschen viel weniger ängstigen Bahnfahrt nicht gibt).

Nach dem kurzen Ausflug in die Deskriptive Statistik machen wir nun einen ebenfalls sehr kurzen Ausflug in die Induktive Statistik, bei der es um "Induktion", d.h. den Schluss vom Teil (Stichprobe) aufs Ganze (die Grundgesamtheit, aus der die gezogen wurde) geht.

## 7. Warum Wahrscheinlichkeitsrechnung und Stichproben?

Im Statistikerunterricht umfasst die Wahrscheinlichkeitsrechnung meist zwei Teile, das Rechnen mit Ereignissen und Zufallsvariablen (also Wahrscheinlichkeitsverteilungen) einerseits und Grenzwertsätze andererseits. Beides ist mehr nur als eine Vorbereitung für die Schätz- und Testtheorie, also für die Analyse von Stichproben (durch Schließen auf die Grundgesamtheit) zu verstehen.

Die Nützlichkeit und praktische Relevanz von Stichproben wird vielen erst so richtig deutlich am Beispiel der Qualitätskontrolle: ein Textilfabrikant kann nicht alle seine Textilien zerreißen, sondern nur einen kleinen Teil auf Reißfestigkeit überprüfen, weil er ja den weitaus größeren Teil seiner Produkte noch verkaufen will.

### a) Wahrscheinlichkeiten, Schätzen und Testen

Aufgaben, in denen es um die Berechnung von Wahrscheinlichkeiten geht, werden von den Studenten – nicht ganz zu Unrecht – meist nur als reine Denksportaufgaben aufgefasst, die schwer und von nur geringem praktischen Nutzen sind. Es ist in der Tat nicht ganz einfach zu sagen, was damit gewonnen ist, wenn man weiß, dass etwas eine Wahrscheinlichkeit von 0,376 hat.

Betrachten wir einmal als Beispiel die Aussage: "college educated women who are still single at the age of thirty five have only a 5 percent chance of ever getting married".<sup>50</sup> Die Frage ist jetzt, womit man diese 5% vergleichen soll, und welche Schlüsse eine un-

verheiratete Akademikerin daraus *für sich* ziehen kann?<sup>51</sup>

Sehr viel besser (was den erkennbaren Nutzen betrifft) ist es sicher beim Schätzen und Testen von "Parametern" (etwa einem Mittelwert  $\mu$  in der Grundgesamtheit aufgrund des Mittelwerts  $\bar{x}$  der Stichprobe). Es leuchtet ein, dass  $\bar{x}$  und  $\mu$  zwar zufällig gleich groß sein können, dass aber auch bei  $\bar{x} = 490$  der entsprechende wahre Wert  $\mu$  in der Grundgesamtheit sehr viel kleiner oder auch sehr viel größer sein kann als 490.<sup>52</sup> Er könnte mit einer bestimmten Wahrscheinlichkeit (z.B. von 90%) irgendwo zwischen 470 und 510 liegen.<sup>53</sup> Statt ein Intervall für  $\mu$  als Ergebnis einer Schätzung von  $\mu$  anzugeben kann man auch eine Hypothese über  $\mu$  "testen",<sup>54</sup> etwa die Hypothese, dass  $\mu = 500$  ist ( $\mu$  ist nicht bekannt, hier kann man also nur eine Annahme treffen; aber bei  $\bar{x}$  gibt es nichts anzunehmen,  $\bar{x}$  ist ja bekannt als 490).

Man fragt sich nun: wenn die Hypothese  $\mu = 500$  richtig ist, wie wahrscheinlich ist es dann, mit einer Stichprobe ein Mittel von  $\bar{x} = 490$  oder weniger zu erhalten?<sup>55</sup> Wenn eine solche Abweichung von nur 10 (bei 500 und 490) noch "im Rahmen des Zufalls" ist, was

<sup>51</sup> Man könnte solche Betrachtungen zum Anlass nehmen, nach Ursachen für das Ereignis  $M = \text{get married}$  zu suchen. Es ist klar, dass die 5% eine bedingte Wahrscheinlichkeit darstellen, bedingt durch ein Alter  $A$  von 35 und College Abschluss  $C$ , also  $P(M|AC) = 0,05$ . Hinter einer bedingten Wahrscheinlichkeit steht eine Kausalvermutung, dass nämlich die Wahrscheinlichkeit für  $M$  vom Alter  $A$  und Bildungsabschluss  $C$  abhängt.

<sup>52</sup> Wir haben ja 490 nur mit eine Stichprobe erhalten. Mit einer andern Stichprobe aus der gleichen Grundgesamtheit könnte man durchaus auch einen größeren Wert bekommen. "Zufall" bei der Auswahl heißt ja auch, dass wir zufällig besonders viele kleine (große)  $x$ -Werte und dadurch auch einen relativ kleinen (großen) Mittelwert bekommen haben können.

<sup>53</sup> Diese Formulierung ist genau genommen nicht richtig, weil  $\mu$  ja keine Zufallsvariable ist, über die man eine Wahrscheinlichkeitsaussage machen kann.

<sup>54</sup> Es wird oft vergessen, dass (zumindest bei der obigen Fragestellung) ein Hypothesentest und ein "Konfidenzintervall" nur zwei Darstellungsarten des gleichen Sachverhalts sind.

<sup>55</sup> Das Prinzip ist also: man tut so als ob die Hypothese richtig ist und fragt dann, wie wahrscheinlich das ist, was man beobachtet hat.

<sup>50</sup> Wieder nach Abelson a.a.O., S. 3.

vor allem bei einer kleinen Stichprobe gut möglich ist, wird die Hypothese "angenommen". Ist die Wahrscheinlichkeit aber sehr gering (quasi nicht mehr mit Zufall zu "erklären"), wird sie "verworfen", d.h. der empirische Befund  $\bar{x} = 490$  ist dann "signifikant" verschieden von der Annahme  $\mu = 500$ .

Das wird notorisch dahingehend missverstanden, dass damit gezeigt sei, dass  $\mu$  nicht 500 ist, die Hypothese also falsch war. Davon kann aber keine Rede sein, denn wir kennen ja nach wie vor die Grundgesamtheit nicht und haben nur eine Stichprobe gezogen. Was damit gezeigt wurde ist nur: *wenn*  $\mu = 500$  *wäre*, dann *wäre*  $\bar{x} = 490$  sehr unwahrscheinlich, also "signifikant" (ein Wort, das leider dazu verführt, mehr dahinter zu vermuten als tatsächlich der Fall ist).

Mit einem Test wird nicht *festgestellt*, ob eine Hypothese richtig oder falsch *ist*, sondern es wird *entschieden*, ob man sie für richtig oder falsch *halten soll*.

### b) Stochastische Modelle

Es ist aber nicht die eben beschriebene Aufgabe, bei der das Ergebnis "signifikant" regelmäßig die Gemüter über Gebühr bewegt, sondern die folgende Fragestellung, bei deren Bearbeitung aber genau die gleiche Verfahrensweise angewendet wird. Die jetzt primär interessierende Frage ist, ob die "erklärende" Variable (oder "Regressor")  $x$  in einer (hier: linearen) Regressionsfunktion einen "signifikanten" Einfluss auf die abhängige Variable (oder Regressand)  $y$  hat, oder ob  $y$  nur zufällig nach Maßgabe der Zufallsvariable  $u$  um  $\alpha$  schwankt (so die Annahme der "Nullhypothese"  $H_0 \beta = 0$ )<sup>56</sup>. Das Modell lautet also

$$y_i = \alpha + \beta x_i + u_i \quad (\text{bei } i = 1, 2, \dots, n \text{ Beobachtungen})$$

und es ist ein stochastisches Modell, weil  $u$  eine Zufallsvariable ist. Die aufgrund einer Stichprobe bestimmten Schätzwerte für  $\alpha$  und  $\beta$  nennen wir  $\hat{\alpha}$  und  $\hat{\beta}$ . Die Hypothese

<sup>56</sup> Die Größe  $\alpha$  ist i.d.R. nicht von Interesse. Das Wort "Nullhypothese" stammt nicht aus diesem Zusammenhang, sondern aus dem eines Tests mit zwei Stichproben, nicht mit einer wie im obigen Beispiel.

$H_0$  besagt, dass  $x$  keinen Einfluss auf  $y$  hat (weil  $\beta = 0$  ist) und die Unterschiedlichkeit der Werte  $y_1, y_2, \dots$  (allgemein  $y_i$ ) allein durch Zufall (durch die Unterschiedlichkeit der Werte  $u_1, u_2, \dots$ ) zu "erklären" ist. Das Modell wäre dann  $y_i = \alpha + u_i$ , und es ist die "all-chance-explanation" (Abelson), also eigentlich etwas, was man im Alltagsleben eher eine Nichterklärung nennen würde.<sup>57</sup>

Auch hier ist es wieder so, dass bei einem  $\beta$  von Null trotzdem  $\hat{\beta} > 0$  oder  $\hat{\beta} < 0$  sein kann (so wie ja auch bei  $\mu = 500$   $\bar{x}$  kleiner als 500, nämlich 490 sein kann). Und auch jetzt geht es wieder nur um die Wahrscheinlichkeit für ein  $\hat{\beta} \neq 0$ , *wenn* (das "wahre")  $\beta = 0$  *wäre* ( $H_0$  ist ja eine *Hypothese*). Was hat man nun faktisch in der Hand, wenn sich herausstellt, dass die Nullhypothese ( $H_0$ )  $\beta = 0$  abgelehnt werden kann und  $\beta$  "signifikant" auf dem 5% oder 1% Niveau ist? Es heißt auch hier wieder nicht, dass man gezeigt hat, dass  $H_0$  falsch ist, sondern nur dass  $\hat{\beta} \neq 0$ , wenig wahrscheinlich *wäre* (weniger als 5% bzw. 1%) *wenn*  $H_0$  richtig *wäre*, wenn also  $x$  wegen  $\beta = 0$  tatsächlich nicht "relevant" wäre. Was man gezeigt hat ist nur: Es könnte mehr als nur Zufall im Spiel sein. Das heißt aber noch nicht, dass nicht auch ganz andere systematische Einflüsse neben  $x$  im Spiel sein, die auch im höheren Maße relevant sein könnten als  $x$ .

Die folgenden Probleme werden hier meist nicht gesehen, womit dann Fehlinterpretationen entstehen und *dadurch* (nicht durch die Statistik als Methode) "Manipulationen" ermöglicht werden (evtl. auch ganz ohne böse Absicht, nur durch Nichtwissen):

- man glaubt *inhaltlich* etwas "*Bedeutsames*" gefunden zu haben und erkennt, dass "signifikant" viel weniger aussagt (nämlich nur, dass es Gründe gibt, anzu-

<sup>57</sup> Sie besagt, dass keine systematischen Einflüsse am Werk sind und alles "nur Zufall" sein kann. Außerhalb der Statistik heißt "Erklären" aber i.d.R., Ursachen dafür angeben, warum etwas so ist wie es ist. Da es meist mehrere Ursachen gibt und es oft nicht klar ist, welche wichtiger, und welche weniger wichtig ist, dürfte es hier im Allgemeinen Diskussionen geben, die es bei statistischen Tests nicht gibt.

nehmen, dass mehr als nur Zufall, noch etwas anderes als "all chance" im Spiel ist),

- das Ergebnis ist sehr abhängig vom Stichprobenumfang; bei großen Stichproben wird vieles signifikant, was es bei kleinen Stichproben nicht ist;
- die Schätzung von  $\hat{\beta}$  (bzw.  $\hat{\beta}_1, \hat{\beta}_2, \dots$  bei mehreren Regressoren  $x_1, x_2, \dots$ ), also die Regressionsrechnung als Methode, verlangt ihrerseits, dass für die Grundgesamtheit bestimmte Voraussetzungen gelten (so die  $H_0$ ), die auch statistisch getestet werden sollten (worüber dann auch nur *mit einer Wahrscheinlichkeit entschieden* wird);<sup>58</sup>
- Testen wird besonders dann unseriös, wenn man mehrere Stichproben gezogen hat und sich dann eine davon aussucht, nämlich die, die am deutlichsten das Ergebnis "signifikant" liefert;
- das Unterschlagen nichtsignifikanter<sup>59</sup> und selektive Publizieren signifikanter Zusammenhänge bedeutet faktisch, mit einem anderen (höheren) als dem angegebenen Signifikanzniveau zu rechnen;
- oft werden auch Signifikanztests in Fällen angewendet, bei denen schwer zu sagen ist, aus welcher Grundgesamtheit hier etwas ausgewählt wurde, geschweige denn *zufällig* ausgewählt wurde.

"Erklären" einer Variablen  $y$  bedeutet in der Statistik nur, dass die Varianz<sup>60</sup> von  $y$  in eine "systematische" und eine "zufällige" Varianzkomponente zerlegt werden kann, bzw. – was auf das Gleiche hinausläuft – dass eine Größe  $y$  als Funktion anderer ("erklärender" oder "systematischer") Einflüsse bzw. Variablen  $x_1, x_2, \dots$  mit Koeffizienten  $\beta_1, \beta_2 \dots$  und einer Zufallsvariable  $u$  (als nicht erklär-

ter Rest) darzustellen ist. Die Nullhypothese zu testen, heißt zu fragen, wie wahrscheinlich die beobachtete Stichprobe ist, wenn es keine "systematische", sondern nur die zufällige Varianzkomponente gibt. "Erklären" hat aber – wie gesagt<sup>61</sup> – außerhalb der Statistik eine viel weiter gehende Bedeutung, nämlich etwas auf seine Ursachen zurückführen und, mehr noch (im Idealfall), den Mechanismus aufzeigen, mit dem die fragliche Ursache die beobachtete Wirkung hervorbringt.

Es gibt also nicht wenig Anlass für Missverständnisse und Missbräuche, des "Testens" in der Statistik. Aber es ist Unsinn, hier von "Manipulation" zu sprechen und dafür zu plädieren, auf das Testen zu verzichten.

### c) Daten und datengenerierende Prozesse

Es ist nicht nur der Begriff "signifikant", bei dem es erfahrungsgemäß Verständnisschwierigkeiten gibt, sondern es ist auch schon der Umstand, dass Wahrscheinlichkeitsaussagen von anderer Art sind, als die uns viel gewohnteren zahlenmäßigen Feststellungen (an konkreten Einheiten) im Rahmen der deskriptiven Statistik. Dass sie von anderer Art sind, ist schon erkennbar an einem Phänomen, das früher einige Philosophen als ein Paradoxon faszinierte: Zufall bedeutet, dass man gerade *nicht* etwas vorhersagen und berechnen kann, und trotzdem gibt es hier aufgrund der Wahrscheinlichkeitsrechnung etwas zu berechnen. Das ist aber deswegen nicht paradox, weil sich die Aussage (bzw. Berechnung) nicht auf das Gleiche bezieht. Einmal ist es *eine konkrete einzelne* Beobachtung (der Würfel fällt jetzt gleich auf die Augenzahl 4), wo es nichts zu rechnen gibt, und zum anderen ist es die *Gesamtheit aller möglicher Beobachtungen* unter den gleichen Bedingungen (was passiert, wenn man den Würfel unendlich oft wirft?), wo man sehr wohl etwas rechnen kann. Zur weiteren Verdeutlichung mögen Gegenüberstellungen nützlich sein:

- Erhebungen vs. Gedankenexperimente;
- Feststellungen an konkreten *einzelnen* Einheiten vs. Aussagen über *alle* Einheiten unter

<sup>58</sup> Das sind Tests, die von den Tests über die Parameter des Modells zu unterscheiden sind.

<sup>59</sup> Auch ein nichtsignifikantes Ergebnis hat einen Wert, insofern nämlich, als es geeignet ist, eine potenzielle "Erklärung" auszuschließen.

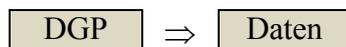
<sup>60</sup> Warum eine "Konstante", etwa  $y = 14$  stets gleich groß ist, bei allen Befragten oder im Zeitablauf, kann man nicht mit den Mitteln der Statistik "erklären".

<sup>61</sup> vgl. Fußnote 57.

gegebenen (konstanten) Bedingungen (Wahrscheinlichkeitsaussagen werden daher auch nicht durch eine einzelne Beobachtung verifiziert oder falsifiziert);

- Ist-Ergebnisse vs. zahlenmäßige Implikationen eines Modells (bzw. einer Hypothese);
- *Eingetretene* Tatsachen vs. Gesamtheit der *möglichen* (Potenzial!) Beobachtungen.

Das Neue an der die Wahrscheinlichkeitsrechnung "vereinnahmten" Statistik ist der Fokus nicht mehr nur auf Daten, sondern auf die stochastischen datengenerierenden Prozesse (DGP) und ihr entstammenden Daten:



Der Schluss  $\Rightarrow$  von der latenten Ebene eines angenommenen DGP auf die manifeste (beobachtbare) Ebene der damit möglichen Zahlen ist eindeutig. Aber der Schluß von der bekannten rechten Seite (Daten) zurück ( $\Leftarrow$ ) auf den DGP ist es nicht. Den gleichen Daten können ganz verschiedene DGPs zugrundeliegen, die den Daten gleich gut "angepasst" sein können.

Man kann die durch Wahrscheinlichkeitsbetrachtungen mit DGP entstandene Bereicherung der Statistik nicht hoch genug einschätzen. Ähnlich wie seinerzeit am Beginn der Statistik die Wahrnehmung der (sozialen) Umwelt durch *Zahlen* neben den bis dato vorherrschenden Worten erheblich bereichert wurde, so wurde auch das "Verstehen" und "Erklären" durch die Wahrscheinlichkeitsrechnung, mit Gedankenexperimenten in Gestalt von Modellen (Gleichungen) bereichert.

Auch hier ist es wieder so, dass das Neue weniger "natürlich" und "plausibel" erscheint als das Altbekannte, wenn es nicht schlicht missverstanden wird. Es ist z.B. von Statistiker\*innen immer wieder darauf hingewiesen worden, dass wir wohl kein Sensorium für die Größe von Wahrscheinlichkeiten haben; wir verschätzen uns hier im Alltagsleben oft, und es fällt vielen auch schwer, sich von Fehlvorstellungen über den Charakter von Wahrscheinlichkeitsaussagen frei zu machen.<sup>62</sup>

<sup>62</sup> Ein oft diskutiertes Phänomen ist bekannt als "gambler's fallacy": wenn die Roulettekugel bereits

Wir kommen damit von den Methoden der Statistik<sup>63</sup> wieder zurück zum Verhalten der Menschen gegenüber Statistiken.

## 8. Widersprüchlichkeiten im Umgang mit Zahlen

Es ist bezeichnend für unserem Umgang mit Zahlen, dass immer wieder sowohl die Neigung besteht, nach immer mehr Zahlen zu rufen, als auch ganz ohne Zahlen, "impressionistisch" zu argumentieren, oder gar Positionen einzunehmen, die immun gegen Zahlen sind, weil man von etwas überzeugt ist, was immer auch empirisch dagegen sprechen mag. Mehr Zahlen und Immunität gegen jede Zahl ist nur auf dem ersten Blick widersprüchlich. Tatsächlich ist es Ausdruck unseres ambivalenten Verhältnisses zu Zahlen.

### a) Ruf nach immer mehr Zahlen, fundierte und nicht fundierte Schätzungen

Die Verarbeitung immer größerer Datenmengen in immer kürzerer Zeit verleitet zu illusionären Vorstellungen über die grenzenlosen Möglichkeiten des Rechnens: Der Computer stiftet z.B. glückliche Ehen durch Kombination von Persönlichkeitsprofilen (je detaillierter die Profile, desto glücklicher und dauerhafter die Ehe), oder man erhält z.B. eine optimale Auswahl aus Bewerbern durch Eignungstests usw.<sup>64</sup>

Vor allem wächst auf diesem Boden die Illusion, alles "messen" zu können und zu sollen.<sup>65</sup> In der Regel ist aber etwas, bei dem es

---

10-mal auf rot gefallen ist, müsste sie jetzt eigentlich auf schwarz fallen.

<sup>63</sup> Wir verlassen damit den methodischen Teil, der in der Praxis für Anwender ohnehin oft nicht dominierend ist. Methodendiskussionen sind oft nur vorgeschoben: Es geziemt sich nicht, offen zuzugeben, dass man etwas ablehnt, weil man es nicht "glaubt", man muss deshalb bei einer Kritik oft an der Methode ansetzen.

<sup>64</sup> Hier führen Illusionen bekanntlich auch jeweils zu lukrativen Geschäftsideen.

<sup>65</sup> Eine bemerkenswert oberflächliche Betrachtung, nach der alles, wirklich auch alles, messbar sein soll, und zwar einfacher als man denkt, ist D. W. Hubbard, *How to measure anything. Finding the Value in "Intangibles" in Business*, 2nd. ed. (Wiley), 2010, ein Buch, das es immerhin zum Bestseller in den USA geschafft hat.

schon schwer ist, in Worten zu *sagen*, was es überhaupt ist, auch schwer zu *messen*. Selbst bei einer so anspruchslosen Methode, wie die Berechnung einer Punktschuld hat man hier schon gleich zu Beginn, bei der Auswahl und Skalierung der "Indikatoren" Probleme.

In den 1970er Jahren gab es im Zusammenhang mit den sog. "sozialen Indikatoren" eine Diskussion darüber, ob eine hohe Scheidungsquote "positiv", weil modern und weltoffen, oder aber negativ als Sittenverfall zu werten ist. Und jetzt, 40 Jahre später, haben wir ähnliche Probleme unter dem Stichwort "Beyond the GDP", wenn es im weitesten Sinne um eine Messung des "Glücks" geht. Welche Indikatorqualität hat z.B. die "Homo-Ehe" bei der Messung des sog. "societal progress", was immer man darunter verstehen mag?

Das Problem solcher permanenter Aufforderungen, mit dem "Messen" immer mehr in Bereiche nebulöser Konzepte vorzudringen, die schon mit Worten kaum klar beschrieben werden können, ist dass Messungen verlangt (bzw. massiv eine gefordert) werden,

1. obgleich man praktisch noch nicht die geringste Idee hat, wie man das messen will, was man fordert, und
2. auch offenbar nicht weiß, dass die geforderten "Messungen" dann auch oft notwendig nur von mangelhafter Qualität sein können.

Zu 1: Wenn man inhaltlich nicht weiter weiß, ist die Vorstellung oft die, dass die Messung von etwas umso besser wird, je mehr Daten zu ihrem Zweck im Computer verarbeitet werden.<sup>66</sup> Oder kurz gesagt: Wo mehr Nachdenken nicht hilft, da helfen mehr Zahlen. Man beginnt also oft mit einem "brainstorming" und fragt sich, was z.B. alles für das "Glück" oder den "societal progress" irgendwie relevant sein könnte. Auch ohne

<sup>66</sup> Bei "Beyond the GDP" versucht es Eurostat z.B. mit 111 Indikatoren, deren "Zusammenschau", wenn man so etwas überhaupt versucht, naturgemäß eine echte "Herausforderung" sein dürfte. Nicht nur wegen seiner Rechenleistungen, auch wegen des Automatismus "ermutigte der Computer die Auffassung, Zweck von Berechnungen sei es, die Menschheit von der Notwendigkeit tieferen Nachdenkens zu entbinden" (Davis u. Hersh, S. 378).

viel Fantasie fällt einem hierzu schnell eine enorme Zahl möglicher Indikatoren ein,<sup>67</sup> wo doch "Glück" bekanntlich so ein "weites Feld" ist. Und weil man für jeden Indikator empirische Daten braucht, hat man schnell einen großen Zahlenfriedhof angelegt.

Eine neue Parole auszugeben, *was* man messen sollte, ist einfach (natürlich ist das "Glück" letzten Endes der Zweck allen Wirtschaftens und aller Politik), aber zu wissen, *wie* man es messen soll ist schwierig und wird i.d.R. erst einmal vertagt.

Zu 2: Aber es kommt noch toller: obwohl man es noch nicht weiß, fordert man schon, dass der noch zu entwickelnde "happiness" Index (HI) nicht das GDP *ergänzen*, sondern auch gleich *ersetzen* soll, nicht wissend, dass die Schätzung des GDPs sehr viel besser abgesichert ist, als es die eines HI jemals sein könnte.

Für das GDP gelten nämlich Zusammenhänge (aufgrund der Kreislauftheorie, einem nichtstochastischen Modell<sup>68</sup>), die es erlauben, ein und die gleiche Größe auf verschiedene Art zu schätzen. Das leuchtet unmittelbar ein, denn die Ausgaben des einen sind die Einnahmen des anderen. Wenn es verschiedene gut begründete Möglichkeiten gibt, etwas zu schätzen, kann man die Ergebnisse (die theoretisch gleich groß sein müssten) gegenüberstellen und "abgleichen". Eine Schätzung kann so "fundiert" werden. Für das Aggregat "Glück" gibt es kein Modell (Gleichungssystem) und keine "fundierte Schätzung", man kann froh sein, wenn man es überhaupt auf *eine* Art schätzen kann. Auch wenn man einen HI hätte, wäre er also von statistisch viel geringerer Qualität als das GDP, das man so vorschnell wegwerfen will.

### ***b) Computopia: hätten wir nur die Zahlen, könnten wir auch die richtige Politik machen***

Wir kommen nun von der Manie alles messen zu wollen zur nächst höheren Stufe des

<sup>67</sup> Oder man fragt einfach die Bevölkerung entsprechend unpräzise Fragen (in der Annahme, dass jeder selbst am besten beurteilen kann, ob er sich glücklich fühlt). Wir kommen darauf im Abschn. 10 zurück.

<sup>68</sup> Ein anderer Fall einer Messung mit einem nichtstochastischen Modell ist die Messung der Lebenserwartung mit dem Modell der "stationären Bevölkerung" (auch "Sterbetafelbevölkerung" genannt).

Zahlenfetischismus, nämlich zum Gebrauch von Zahlen in der Politik. Dort heißt es nicht selten: wir können nicht die richtige Politik machen, weil wir nicht die statistischen Daten haben, die wir dazu brauchen. Der Optimierung der Politik steht danach nur der Mangel an Statistiken entgegen.

Das ist eine Geisteshaltung, die ich schon vor über zwanzig Jahren "Computopia" genannt habe.<sup>69</sup> Politik ist danach die rechnerische (Computer) Verarbeitung eines Zahleninputs zu einem optimalen (Utopia) Politikoutput, letztlich nur ein Informations- und Rechenproblem, und nicht auch ein Motivations-, Organisations- und Durchsetzungsproblem. Der Ablauf ist danach



und es gibt nur einen Pfeil ( $\Rightarrow$ ) hin zur optimalen Politik, der vom Computer zur Politik.

In diesem Sinne forderte z.B. 1980 der damalige Bundesarbeitsminister Herbert Ehrenberg in einem Buch eine "computerunterstützte, tagesaktuelle Übersicht aller offenen Stellen" (vermutlich auch noch regional tief gegliedert).<sup>70</sup> Ohne so etwas, was wir bis heute noch nicht haben, glaubte er, sei eine ordentliche Arbeitsmarktpolitik nicht zu machen. Mit welchen statistischen oder sonstigen Methoden man im Ministerium die tägliche Übersicht am Computer "verarbeiten" wollte, hatte er allerdings nicht verraten (vermutlich hatte er sich darüber auch noch keine Gedanken gemacht).

Beispiele dieser Denkungsart gibt es in Hülle und Fülle. Gibt es ein Problem, wird ein Gremium gebildet, das regelmäßig "Berichte" abliefern soll und dessen erste Amtshandlung dann i.d.R. darin besteht, über unzureichende Statistiken zu klagen und mehr Zahlen zu fordern, die alle angeblich unabdingbar seien für die Arbeit des Gremiums.

An der dabei üblichen Theatralik und Dramatisierung hat sich zwischenzeitlich nichts

geändert. So heißt es jetzt, dass ein Politiker, der nur mit dem GDP (oder BIP) "ausgerüstet" ist und über nichts "beyond" verfügt, vergleichbar sei mit einem Piloten, der ohne Kompass fliegt, womit man beiläufig und wohl ganz unbeabsichtigt zu erkennen gibt, dass man nicht nur von Statistik, sondern auch von der Fliegerei keine Ahnung hat.

Es gibt, wie ich andernorts schrieb,<sup>71</sup> immer

"Leute, die es schaffen, den Eindruck zu erwecken, dass wir einer Katastrophe entgegengehen, weil wir über irgendetwas immer noch keine Statistik haben, oder weil die entsprechende Statistik erst in einer Woche kommt und nicht schon jetzt da ist. Wenn sie dann aber die angeblich so alles entscheidende Zahl 'endlich' sehen, dann fällt ihnen kaum mehr dazu ein, als 'interessant'...".<sup>72</sup>

Eine Variante dieses Glaubens an Wunder durch Statistik ist die Erklärung eines Professors der Volkswirtschaftslehre (VWL)<sup>73</sup>, dass die Unfähigkeit, die weltweite Finanzkrise (ab 2007) und ihre Schwere vorauszu- sehen, mitnichten "Schwächen der VWL-Theorie" oder gar deren Versagen offenbart habe. Schuld sei nur, dass sich die Ökonomen "über die Faktenlage im Unklaren befanden". Hätte der IMF die nötigen Zahlen präsentiert, dann hätten alle Ökonomen die gleichen Politikempfehlungen gegeben.<sup>74</sup>

Auch hier, in der Wissenschaft, nicht nur in der Politik wird die Alibi-Funktion fehlender Statistik und die an noch mehr Statistik geknüpfte Heilserwartung deutlich:

<sup>71</sup> vgl. meinen satirischen Text: Was tun, wenn einem eine Statistik nicht passt?

<sup>72</sup> Ähnlich pervers ist dann das Thema "Genauigkeit" einer Zahl: je mehr Nachkommastellen, desto besser. Man vergisst dabei, dass eine Zahl, etwa für den societal progress auch schon ganz ohne Nachkommastellen hochgradig problematisch ist, und dass sie mit Nachkommastellen nicht besser wird.

<sup>73</sup> Lutz Arnold, Die VWL steckt nicht in der Krise, ifo Schnelldienst 14/2009.

<sup>74</sup> Es fällt schwer, so etwas zu glauben, wo doch jeder weiß, dass Professoren der VWL auch schon dann sehr unterschiedlich denken, wenn es "nur" um ihre Modelle, und gar nicht um konkrete Fakten geht. Außerdem: auch *wenn* es die (!) Zahlen vom IMF gegeben hätte, hätte es immer noch genug Professoren gegeben, die sie nicht ernst genommen hätten.

<sup>69</sup> In der vierten Auflage meines Buchs "Wirtschaftsstatistik" (UTB-Reihe Nr. 209, Stuttgart 1990). Die Wortschöpfung stammt nicht von mir, sondern aus der Literatur über Methoden der Wirtschaftsplanung.

<sup>70</sup> v. d. Lippe, Wirtschaftsstatistik, a.a.O., S. 216.

- Weil man mit jedem Standpunkt immer an mindestens einer Stelle aneckt, braucht man zu seinem Schutz *Unterstützung*, bzw. zu seiner Entlastung einen *Sündenbock*, und
- was für einen Standpunkt" in der Wissenschaft gilt, das gilt ganz entsprechend für Entscheidungen, Pläne und Programme in der Politik

Statistische Daten liefern beides; wenn es sie gibt, kann man bei ihnen Unterstützung suchen, und wenn es sie nicht gibt (und das wird immer der Fall sein, wenn es nur hinreichend spezielle Daten sind, um die es geht), kann man sich damit retten, dass man ihre Nichtexistenz zum Sündenbock macht. Beides funktioniert, weil viele von Zahlen Wunder erwarten und auf das Fehlen von Zahlen hysterisch reagieren.

Ein erster Motor, der "Computopia" antreibt ist die immer mehr alles beherrschende Streitkultur, in der alles, bestritten, zerredet und "skandalisiert" wird und in der sich deshalb auch kaum noch jemand traut (bzw. es sich leisten kann), etwas als seine persönliche, subjektive und "einsame" Entscheidung bzw. Meinung zu verkaufen<sup>75</sup> Also braucht man eine Unterstützung seiner Meinung.

Die Unterstützung ist aber gerade wegen der "Objektivität" und der "Unpersönlichkeit" der Zahlen so gefragt. Am wenigsten angreifbar ist etwas, was automatisch produziert wird, wo kein Mensch mehr involviert ist, und das ist das "Comp" in "Computopia".

Als zweiter Motor von "Computopia" wirken deshalb die rasant gestiegenen technischen Möglichkeiten der (automatischen) Erzeugung und Verarbeitung von zahlenmäßigen Informationen.

### c) *Noch mehr Zahlen von zweifelhaftem Wert: "Data Mining" und "Big Data"*

Unter "Big Data" versteht man zunehmend automatisch, laufend quasi als Nebenproduk-

te, erzeugte besonders große Datenmengen,<sup>76</sup> die die Dimensionen herkömmlicher Datenbanken sprengen. Die "objektive" Art der Erzeugung der Daten (ohne dass Menschen involviert wären), ihre schnelle Verfügbarkeit und ihr schier gigantisches Volumen machen die Sache (geschäftlich) attraktiv. Diese Entwicklung und die damit verwandten Trends zu Data Warehouse Systems (Vorhalten von Daten und Analysen) auf Basis von "Data Mining" (automatisches Suchen nach "etwas Wertvollem", wie Muster, Zusammenhänge etc. in einem Datenberg) sind vor allem eine Herausforderung für die amtliche Statistik und eine Bedrohung ihrer Sonderstellung. Wenn man – was nicht selten geschieht – Unterschiede in der Qualität (inhaltlich korrekte [der Fragestellung adäquate] und vergleichbare Konzepte bei der Datenerfassung usw.) ignoriert oder für irrelevant erklärt, erscheinen solche Daten den amtlichen oft überlegen. Sie werden schon jetzt oft als gleichwertig betrachtet und z.T. schon mehr geschätzt als die "altbackenen" amtlichen Zahlen. So werden z.B. Scannerdaten im Einzelhandel und Internetrecherchen zu privaten Preisindizes verarbeitet, als Konkurrenzprodukte zur amtlichen Inflationsrate. Bei ihnen dürfte zwar vermutlich weniger Sorgfalt darauf verwendet werden, auszuschließen, dass sie neben der Preisbewegung auch laufende Qualitäts- und Sortimentsveränderungen im Angebot und Änderungen der Gewichte (Mengen) seitens der Nachfrager widerspiegeln. Aber wen interessiert das (außerhalb der Statistikämter), wo doch auch nicht unerhebliche private geschäftliche Interessen im Spiel sind?

Das alles mag primär die Statistikämter in Schwierigkeiten bringen, die schon lange Anlass haben, über eine zu geringe Würdigung (und finanzielle Unterstützung) ihrer Arbeit zu klagen, aber es betrifft uns auch alle, wenn wir noch mehr Zahlen bekommen werden, die wir mit noch mehr Fragezeichen versehen müssen. Vielleicht stehen wir am

<sup>75</sup> Zu den Auswüchsen dieser Entwicklung in Politik, Medien, Justiz usw. in den USA siehe Deborah Tannen, *The Argument Culture*, New York, 1998.

<sup>76</sup> Quellen sind Web-Logs, Protokolle von Telekommunikationsverbindungen, radio frequency identification (RFID) mit elektromagnetischen Wellen, von Steuerungssystemen, Kameras und sonstige Sensoren generierte Daten usw.

Anfang einer neuen Arbeitsweise in der Statistik von einem bisher üblichen Ablauf



zum neuen Arbeitsablauf



Wir können uns hier diesen interessanten Entwicklungen nicht weiter widmen. Fest stehen dürfte aber, dass wir in Zukunft noch weit mehr statistische Daten haben werden und es noch dringlicher wird, sich zu fragen, ob wir

- auch fähig sind, die Qualität der Statistiken zu beurteilen und *wenn* ja, ob wir angesichts des Überangebots auch
- von ihnen einen vernünftigen Gebrauch machen und auch bereit sind, bei wichtigen Themen und Kontroversen darauf gestützte Argumente gelten zu lassen.

Beides ist mit guten Gründen zu bezweifeln. Gerade in der Politik, denen die statistischen Zahlen ja dienen sollen, erlebt man es nicht selten, dass Politiker es vorziehen, von ihnen besser gar nicht erst Gebrauch zu machen.

#### **d) Der impressionistische Gegenschlag und Immunität gegen Zahlen**

Interessant ist nämlich, dass wir nicht nur ständig Forderungen nach mehr Daten haben, ganz ohne Maß und Ziel, sondern dass es auch eine Neigung zum genauen Gegenteil, zum Ignorieren, wenn nicht gar Verachten und Verächtlichmachen jedweder bereits vorhandener Zahlen gibt, dann nämlich, wenn es um "Höheres" geht. Es ist immer wieder erstaunlich, wie viele Menschen bereit sind, alle Zahlen zu vergessen, wenn sie durch geschickte emotionale Rhetorik den Eindruck vermittelt bekommen, dass alles doch ganz anders sei und die Statistik naturgemäß nie die volle Wahrheit sein kann.

Das sieht nach Widerspruch aus, aber ein Politiker kann sehr wohl beide Seiten personifizieren, ohne schizophoren zu erscheinen. Er kann das Fehlen von Statistiken beklagen, womit die Statistik dann – wie gesagt – bei Angriffen auf seine Politik als Sündenbock herhalten muss und er kann genauso alle

Zahlen als irrelevant wegwischen, wenn es für ihn um eine Herzensangelegenheit geht, oder um eine Position, die um jeden Preis zu halten ist, weil man sich auf sie nun einmal festgelegt hat (z.B. die Euro-Rettung).<sup>77</sup>

Dass das, was nach Widerspruch aussieht sehr wohl gleichzeitig in einer Person vorgelebt werden kann, ist auch wieder ein Erbe unserer eingangs erwähnten langen Übung im Hören von Geschichten, Gleichnissen, Predigten usw., und unserer im Vergleich dazu noch kurzen Übung im Quantifizieren und wissenschaftlichen Argumentieren.

Ich habe an anderer Stelle die Eigenarten der "impressionistischen" – im Unterschied zur "statistischen" – Betrachtungsweise und ihre Steigerung in Gestalt von Immunität gegen Zahlen jedweder Art durch emotional bewegende Rhetorik dargestellt,<sup>78</sup> so dass ich mich hier kürzer fassen kann.

Der impressionistische Gegenschlag (gegen Zahlen) besteht darin, zahlenmäßige Feststellungen durch selektive, "aus dem Leben gegriffene" persönliche Eindrücke und Beobachtungen zu entwerten. Die impressionistische Betrachtungsweise ist in vieler Hinsicht das Gegenteil der statistischen. Die impressionistische (nicht aber die statistische) Methode funktioniert auch dort, wo

- es gar keine Daten gibt,
- eine Variable gar nicht klar definiert ist (z.B. "soziale Kälte") und wegen überzeugender (emotionaler) Rhetorik einer politischen Autorität bei vielen Menschen auch gar kein Bedürfnis nach deren Messung besteht,

<sup>77</sup> Dass mit einem Schlag, wenn es politisch opportun ist, von Politikern alle Überlegungen der Wissenschaft vom Tisch gefegt werden, haben – gerade auch bei "der Euro-Krise, in der sich niemand mehr um Regeln geschert hat" – nicht nur die Statistiker, sondern auch die Juristen erleben müssen. Plötzlich war z.B. der "Stabilitätspakt" schon lange "Unsinn". W. Mussler, Der dumme Pakt, FAZ v. 29. 5. 2013.

<sup>78</sup> Vgl. wieder meinen Text "Was tun ...". Mit Sprache kann man auch mehr manipulieren als mit Zahlen, was aber niemand zu kritischen Bemerkungen über die Sprache veranlasst.

- wo es entscheidend darauf ankommt, *wer* (aus welcher Position heraus) etwas sagt,<sup>79</sup> und wo
- es um Themen geht, bei denen Vorurteile oder schlicht Ideologien bedient werden können.

Sie kultiviert Subjektivität und macht noch nicht einmal den Versuch, objektiv zu sein, wofür aber gerade die Statistik steht. Der Erfolg des impressionistischen Gegenschlags steht und fällt mit Mechanismen, die alle typisch für Manipulation sind:

- dramatisieren, emotionalisieren und moralisch überhöhen, und
- an die Angst appellieren, Außenseiter zu sein (d.h. sich den Wertvorstellungen der Mehrheit nicht anzuschließen, nicht zu fühlen, was andere fühlen).

Das alles ist nicht nur typisch für Manipulation, sondern auch generell für ein unwissenschaftliches Vorgehen, gegen das man nicht *argumentieren* kann. Bei Rhetorik geht es um *Überzeugen* (und überzeugt sein), in der Wissenschaft um *Beweisen* (und Wissen).<sup>80</sup> Wissen und Zahlen nützen nichts, wenn jemand seine Überzeugungen bedroht sieht. Man kann niemandem mit Vernunft Überzeugungen nehmen, die er ohne Vernunft gewonnen hat.

## 9. Rating, Evaluation und Selektion

Die Statistik als Instrument der Entdeckung und des Verstehens von Tatsachen hat Grenzen, einfach deshalb, weil Zahlen nicht alles ausdrücken können, was die Tatsachen beinhalten.

<sup>79</sup> Statistik dagegen ist in dem Sinne demokratisch, als es bei ihr auf *viele* Auskünfte von (bzw. Beobachtungen an) quasi *gleichen* Durchschnittsmenschen der großen Masse ankommt. Wenn es dagegen gelingt, sie und ihre Zahlen mit guter Rhetorik, Impressionen und Emotionen zu überspielen, dann meist nur weil sich *wenige* Menschen engagieren, die einen *Bonus* durch Prominenz haben (wir haben es quasi mit einer Wiederkehr des Autoritätsbeweises zu tun).

<sup>80</sup> Zu überzeugen ist *die Mehrheit*; es kann und darf durchaus Außenseiter geben, die nicht verstehen können, warum andere so denken. Aber mit einem Beweis muss etwas *für alle* bewiesen sein.

Mit ihrem Programm der *zahlenmäßigen* Darstellung und Analyse kann sie auch den Boden bereiten für Übertreibungen nach Art von "Computopia". Sie aber deshalb der Manipulation oder gar der "Lüge" zu bezichtigen ist unsinnig. Schlechter Umgang mit Statistik mindert den Wert der Statistik als solche nicht; denn wir brauchen Statistik.

Wir brauchen sie immer dann, wenn etwas zu evaluieren und zu entscheiden ist und dabei jeweils der persönliche Eindruck nicht ausreicht und wohl auch unter dem Verdacht steht, zu subjektiv und unreflektiert zu sein. Und was ist denn auch die Alternative? Etwa über viele Menschen und große Projekte mit großen Ausgaben allein aufgrund von Einflüsterungen aller Art, von dezent bis massiv zu entscheiden?

### a) Kleines und großes Beurteilen

Persönliche "Eindrücke" und Verzicht auf Quantifizierung sind dann angebracht und wohl auch leistungsfähiger als die statistische Herangehensweise, wenn die Dimensionen, um die es geht kleiner sind. So haben wir bis jetzt noch nichts Besseres gefunden, als "peer reviews" bei der Beurteilung der Veröffentlichungswürdigkeit einer wissenschaftlichen Arbeit. Die Beurteilung ganzer Fachbereiche oder gar Hochschulen und von Forschungsprojekten, an denen viele Personen mitwirken und die über viele Jahre "laufen" ist dagegen eine ganz andere Dimension. Man könnte hier von wholesale ranking sprechen im Unterschied zum kleiner dimensionierten retail ranking bei peer reviews.

Die rein verbale Begutachtung einer *vorliegenden* einzelnen Arbeit durch einen Gutachter, quasi der Einzelhandel, um in dem Bild zu bleiben, kann gut und hilfreich sein, sie kann eine schlechte Publikation verhindern und eine gute vielleicht noch verbessern, *wenn* es denn der "richtige" Gutachter ist. Es kann aber auch ein Gutachter sein, der offensichtlich auf dem fraglichen Gebiet wenig kompetent ist, der mit aller Gewalt später in der Veröffentlichung seine höchst eigenen Ansichten wiedersehen möchte, oder der sich erst nach zwei Jahren herablässt, zwei Zeilen zu schreiben. Dann steht der Autor schlech-

ter da als jeder Student. Der Student weiß wenigstens, wer ihm "eine Fünf" gegeben hat und es ist i.d.R. sichergestellt, dass es auch jemand ist, der mehr von der Sache versteht als der Student.

Die Beurteilung ohne Statistik kann also durchaus gut und der Sache angemessen sein, aber es ist nicht sicher, ob im Einzelfall die Voraussetzungen dafür auch gegeben sind.

Beim "wholesale ranking" dagegen geht es wohl nicht ohne Statistik und es tritt hier typischerweise eine ganz andere Gefahr auf, als die des schlechten Gutachters: wer Zahlen entscheiden lässt darf sich über "kreative Buchführung" nicht wundern.

### **b) Strategisches Beeinflussen des Rankings**

Es ist klar, dass es bei jedem Ranking, wie z.B. dem Ranking von Hochschulen darauf ankommt, *was* man zählt (Anzahl der Hörsäle, der Bücher in Bibliotheken, der Publikationen der Professoren usw.). Wer sich an solchen Entscheidungen beteiligt wird natürlich ein Interesse daran haben, dass das gezählt ist womit er/sie glänzen kann und es ist auch klar, dass sich dabei die Vorstellungen bedeutender, weithin bekannter Hochschulen mehr durchsetzen, als die von unbekanntem Lehranstalten in der Provinz.

Wegen der erwähnten "kreativen Buchführung" ist es kein Wunder, dass man oft post festum sieht, dass wieder einmal gilt

"Denn wer hat, dem wird gegeben, dass er die Fülle habe; wer aber nicht hat, von dem wird auch genommen, das er hat" (Matth. 13, Vers 12).<sup>81</sup>

In einem Artikel in der DUZ (31.5.2013),<sup>82</sup> in dem auch vom "Matthäuseffekt" (im Sinne dieses Zitats) die Rede war, wurde deutlich gemacht, dass beim allseits so sehr beliebten Ranking von Universitäten große, ältere Hochschulen "im englischsprachigen

Raum" eindeutig im Vorteil sind<sup>83</sup> und es wurde in dem Artikel sogar potentiellen Spielern gesagt, sie "sollen lieber auf ein Mitmachen verzichten, sie haben sowieso keine Chancen".

So wie es Medienberater gibt, die zu einem optimalen Auftritt in den Medien verhelfen, so gibt es inzwischen auch Berater, die dazu da sind, um ihrem Arbeitgeber zu einem vorderen Platz in einem Ranking zu verhelfen. Und um ihnen das aber nicht zu leicht zu machen, halten die Produzenten von Rankings und Ratings wiederum, ganz ähnlich wie die Ratingagenturen bei Geldanlagen, viel von ihrer Methode geheim. Es ist ja praktisch ihr "Kapital", *wie* sie zu ihren Zahlen kommen. Die Dienstleistung, die eigentlich der Transparenz dienen soll, ist also selber keineswegs transparent.

Damit haben wir aber mittlerweile mit unserer Art des Umgangs mit Zahlen etwas erzeugt, was praktisch das Gegenteil dessen ist, was ursprünglich mal mit Statistik beabsichtigt war, nämlich etwas *transparent, nachvollziehbar* und frei von der Einflussnahme der Mächtigen festzustellen, was zudem auch wegen der Verwendung von Zahlen frei ist von den Mehrdeutigkeiten und emotionalen Konnotationen der Worte einer Sprache.

Bedenklich ist nicht nur das strategische Verhalten ganzer Institutionen (z.B. Hochschulen), sondern auch der so ausgelöste Herdentrieb ihrer einzelnen Akteure: es dürfte wohl genug Leute (v.a. junge Leute) in der Wissenschaft geben, die etwas tun, was sie für falsch halten (oder etwas nicht tun, was sie eigentlich lieber täten), nur weil es alle tun und sie fürchten, sonst keine Karriere zu machen, keine Forschungsgelder zu bekommen oder nicht mehr zum Kreis der Experten auf ihrem Gebiet gerechnet zu werden.<sup>84</sup>

<sup>81</sup> Es lohnt sich zu sehen, wie es mit dem obigen Zitat in Vers 13 weitergeht: "Denn mit sehenden Augen sehen sie nicht, und mit hörenden Ohren hören sie nicht; denn sie verstehen es nicht." Der aus Überzeugung oder von anderen hierzu gedrängte Quantifizierer sieht nicht, dass er es ist, der (vielleicht sogar gegen seinen Willen) Zahlenfetischismus sät und Manipulation erntet.

<sup>82</sup> Auch SPON 5.6.2013

<sup>83</sup> So wurde in Kommentaren zu dem Artikel auch darauf hingewiesen, dass in den USA Forschungsinstitute mehr mit den Hochschulen verzahnt sind und deshalb deren Leistungen – viel mehr als in Deutschland – den entsprechenden Hochschulen zugutekommen.

<sup>84</sup> Entsprechend beschäftigen sich wohl auch viele nicht mit etwas, was sie eigentlich interessieren würde, und worin sie vielleicht auch etwas leisten könnten.

Es ist auch beobachtet worden, dass in der Wissenschaft, Fälschungen oder bloß mangelnde Sorgfalt beim Publizieren zugenommen haben, wohl deshalb weil der Druck größer geworden ist, viel zu publizieren und das dann jeweils auch als erster zu tun.<sup>85</sup> Das alles mag bedauerlich sein und es ist letztlich auch Konsequenz unserer Art des Umgangs mit Zahlen und des Vorrangs der Quantität vor allem anderen. Aber – wie gesagt – was ist die Alternative zur Quantifizierung, wenn es um Weichenstellungen durch Ablehnen oder Bewilligen großer Geldbeträge geht?

### c) Replikations- und Metastudien

Eigentlich wäre es in dieser Situation zu begrüßen, wenn mehr Arbeiten unternommen würden, in denen Studien untereinander verglichen und in ihren Methoden und empirischen Feststellungen überprüft werden. Es gibt aber verschiedene Gründe, weshalb daran wenig Interesse besteht, und zwar sowohl seitens des Herausgebers einer wissenschaftlichen Zeitschrift als auch seitens eines potenziellen Autors. Wenn fremde Arbeiten kritisiert werden, müssen sie zitiert werden, und das nützt dem impact factor der entsprechenden anderen Zeitschriften, in denen sie erschienen sind, nicht aber der eigenen Zeitschrift.<sup>86</sup> und für einen Autor ist eine solche "Metastudie" auch schon deshalb nicht sehr reizvoll, weil sie

- als nicht originell gilt und deshalb nicht viel "zählt",
- ein für eine faire Replikation erforderlicher Zugriff auf Originaldaten und Originalsoftware oft schwierig ist,
- Kritik – vor allem von einem jungen noch nicht bekannten Autor – oft nicht ernst genommen wird, und weil

- diesem dann, wie die Erfahrung gezeigt hat, schließlich sogar auch noch mit rechtlichen Konsequenzen gedroht wird.

Die Rahmenbedingungen sind also so, dass die Nachvollziehbarkeit, die generell die Wissenschaft und speziell auch die Statistik auszeichnet, mehr und mehr nur eine theoretische Möglichkeit ist.

Fazit: Ausgerechnet dort, wo es angesichts der Tragweite der Entscheidungen schwer ist, ohne Zahlen und ohne Statistik zu arbeiten, haben wir mit unserer Neigung zur Übertreibung und zum Zahlenfetischismus den Boden bereitet für Missbrauch und Manipulation, die aber nicht der Statistik als solche anzulasten ist. An der Neigung zur Übertreibung wird sich kaum etwas ändern lassen. Was wohl eher helfen könnte, ist mehr Wissen über statistische Methoden, um deren Missbrauch besser zu durchschauen.

## 10. Zukunft der Statistikausbildung und weitere Trends

Einige Beispiele für "innumeracy",<sup>87</sup> wie ein oben zitierter Fall von Stümpfern mit Prozentzahlen zeigen, dass es notwendig sein dürfte, auf diesem Gebiet mehr zu tun, auch wenn es zum Klassenziel, Mängel einer Statistik zu durchschauen, nicht ganz reichen wird.<sup>88</sup> Skepsis dürfte hier angebracht sein, denn es gibt nicht, oder nur schwer zu beeinflussende Vorbehalte gegen Statistik (woran sich zumindest auf kurze Sicht wohl nichts ändern wird), sondern auch aktuell Erscheinungen, die nicht sein müssten, und die die Probleme mit der Statistikausbildung eher noch verstärken.

### a) Das Geschäft mit der Not mit Statistik

Eine nicht zu unterschätzende Hypothek, die das Fach "Statistik" wohl auf Dauer belastet, ist die in Deutschland immer noch verbreitete Einstellung, dass man sich als Schüler seiner schwachen Leistungen in Mathematik

<sup>85</sup> Gefälscht? Warum Forscher immer öfter Fehler und Fälschungen publizieren – und wie sich das verhindern lässt, Wirtschaftswoche Nr. 23/2013, S. 72.

<sup>86</sup> Es ist abenteuerlich zu sehen, wie es oft bei Diskussionen über den "impact" eines "journals" nur um ein oder zwei Zitate mehr oder weniger in einem Jahr geht und wie Programme die Autoren zwingen, jede namentlich erwähnte Person mit einer "reference" zu belegen. So werden von vielen Autoren Schriften aus dem 17. und 18. Jahrhundert zitiert, die garantiert keiner der sie zitierenden Autoren jemals in der Hand gehabt hat.

<sup>87</sup> In Analogie zu "illiteracy".

<sup>88</sup> Das erwähnte Stümpfern wird wohl immer bleiben. Sicher ist wohl auch, dass es immer mehr Statistiken geben wird, über die wohl auch in der besagten Weise geredet, geschrieben wird.

nicht schämen muss, weil das so verbreitet ist und (angeblich) nicht viel aussagt. Viele bekennen sich sogar mit Stolz, in der Schule "in Mathe" eine Fünf gehabt zu haben. In anderen Ländern ist das sehr oft ganz anders.

Mir missfällt aber mindestens genauso das Spiel mit umgekehrten Vorzeichen, wenn

- mehr und mehr Professoren der BWL und VWL den, mit Sicherheit oft nicht gerechtfertigten Eindruck zu erwecken trachten, alles in puncto Statistik und Ökonometrie zu beherrschen so dass ihnen nichts zu schwer sein kann,
- und wenn immer wieder neue Lehrbuchautoren auftauchen, die meinen, mit ihrem Opus sei Statistik kinderleicht.

Die Wahrheit ist, dass Statistik nicht leicht ist und mit ihr auch intelligente Menschen Schwierigkeiten haben. Und offenbar sind es nicht nur Studenten, sondern es ist ein sehr viel größerer Personenkreis, der sich genötigt sieht, sich widerwillig mit Statistik zu beschäftigen. Daher auch die offenbar ungebremste Produktion von Büchern, die versprechen, die Materie nur mit Worten und ganz ohne Formeln abzuhandeln, und die angeben, mehr Hintergründe und Zusammenhänge aufzuzeigen und ein tieferes Verständnis zu vermitteln als die üblichen Lehrbücher dies tun.<sup>89</sup> Die Not mit Statistik scheint also sehr groß zu sein (und damit auch das Geschäft mit ihr).

Betrachtet man solche Bücher genauer, so sieht man, dass in ihnen nichts anderes steht, als in den "üblichen" Büchern auch, es werden nur ganz unverantwortlich die üblichen Vorurteile gepflegt und dabei viele Worte gemacht, meist aufmunternde Worte eines sich als verständnisvollen Kumpel anbietenden Autors. Aber Statistik wird keineswegs leichter, wenn man sich erst einmal durch viele Seiten mit albernem Geschwätz

<sup>89</sup> Die Sache geht sogar so weit, dass Bücher von Autoren gekauft werden, die offenbar überhaupt keine Manieren mehr haben, weil sie ihre Leser mehr oder weniger als Deppen bezeichnen, wie David M. Levine und David F. Stephan, *Even you can learn statistics. A guide for everyone who has ever been afraid of statistics*, 2009 (2<sup>nd</sup> edition 2010).

kämpfen muss, bis man zu den Formeln kommt.<sup>90</sup> Und es ist nicht richtig, dass Statistik "without the use of higher mathematics or mathematical symbols"<sup>91</sup> leichter ist. Das Gegenteil ist der Fall. Die Formeln sind nicht leichter zu verstehen, wenn nur über sie (quasi "in Abwesenheit") lang und breit *geredet* wird.<sup>92</sup> Ohne Formeln wird es nicht leichter, sondern schwieriger, zu verstehen, was bei einer Methode gerechnet wird und damit auch zu verstehen, was als Ergebnis dabei herauskommt.

### *b) Aussieben, auswendig lernen und vergessen*

Kein Wunder, dass es gerade die Statistik ist, eine eher als unwichtig empfundene Teildisziplin innerhalb der VWL oder BWL, bei der dann viele Studenten erkennen müssen, dass sie vielleicht für ein anderes Studium besser geeignet wären. Stolperstein sein macht das Fach natürlich unbeliebt. Diese Rollenverteilung, bei der die Statistik ihnen viel Arbeit und Ärger mit der Selektion von Studenten abnimmt ist den Professoren der VWL und BWL meist ganz recht. Hinzu kommt, dass diese Professoren oft nicht wenig auf ihre eigenen Statistikkenntnisse stolz sind und sie schon deshalb schlecht die Relevanz des Faches herunterspielen können. Außerdem weiß jeder, dass Statistik immer mehr ein Muss ist, wenn man in den Wirtschaftswissenschaften etwas werden will.

Die Sicht der Studenten ist hier natürlich anders. Bei einem Fach, dem ein schlechter Ruf vorausgeht, ist ein ineffizientes Lernverhalten zwangsläufig. Studenten scheinen oft zu glauben, dass die schnellste Art, etwas zu lernen darin besteht, es auswendig zu lernen. Sie denken so vor allem, wenn sie etwas nicht verstanden haben und es mag dann

<sup>90</sup> Es gibt immer wieder Autoren, die das nicht glauben. Ein schlimmes neuerliches Beispiel hierfür ist M. Oestreich u. O. Romberg, *Keine Panik vor Statistik, Erfolg und Spaß im Horrorfach nichttechnischer Studiengänge*, Wiesbaden 2009.

<sup>91</sup> Levine/Stephan in der Einleitung.

<sup>92</sup> Andererseits hat man aber auch wenig von Statistik, wenn man sich nur mit den Formeln beschäftigt oder gar diese auswendig lernt. Studenten scheinen oft diesem Missverständnis zu erliegen.

vielleicht auch in der Tat die einzige Möglichkeit sein, etwas zu lernen, indem man es auswendig lernt. Sie vergessen dabei aber, dass nur einen Rechengang gelernt zu haben (und das vielleicht sogar noch mit den konkreten Zahlen eines speziellen Zahlenbeispiels) gerade in der Statistik gleichbedeutend damit ist, eigentlich gar nichts gelernt zu haben. Und weil das Gelernte dann auch immer wieder schnell vergessen wird, ist auch die Lehre auf diesem Gebiet oft nicht sehr effizient. Es ist schwer, voranzukommen, wenn auf jeder Stufe nichts oder wenig da ist, auf das man aufbauen könnte.

Das sind Missstände, die vermutlich schwer abzustellen sind, solange Statistik so ein Negativimage hat. Anders könnte es wohl mit den Bedingungen auf Seiten der Lehrenden sein auf die wir jetzt zu sprechen kommen.

### *c) Widerwillig lehren und Protzen mit Mathematikkenntnissen*

Studenten die Anfangsgründe in Statistik beizubringen ist keine Arbeit, die viel Ansehen genießt. Für Dozenten ist es daher eher eine lästige Pflicht, die möglichst wenig Energie und Zeit absorbieren sollte. Die Produktion von Powerpoint Folien mit copy & paste aus einem US Lehrbuch erfüllt diese Bedingungen sehr gut und ist daher nicht selten aus Sicht der Lehrenden die Methode der Wahl: sie ist zeitsparend und wirkt nicht so provinziell wie die frühere Art, Statistik zu lehren, auf Deutsch am Overheadprojektor oder gar an der Tafel. Auch aus Sicht der Studenten ist Powerpoint gut; man braucht nicht mitzuschreiben und hat ein Paket von Folien zum Auswendiglernen (je dicker das Paket, umso mehr hat man gelernt). Schlecht ist nur, dass Überlegungen hinter den auswendig gelernten Stichworten oft nicht verstanden worden sind (wenn sie denn überhaupt erklärt worden sind), und dass sie wohl auch selbst dann nicht verstanden worden wären, wenn man sie in deutscher Sprache vorgetragen hätte.

Aus Sicht der Professoren spielt Statistik aber auch noch eine ganz andere Rolle. Was in der Lehre eher lästig ist, das ist hinsichtlich der Karriere als Forscher sehr viel posi-

tiver zu sehen: mit Kenntnissen in Statistik kann man Eindruck machen. Statistik ist also vor allem eine Profilierungsplattform.

Es ist etwas paradox, aber die Anfänge und Grundlagen dieses Faches sind für viele unter ihrer Würde, nur das Schwierige und Fortgeschrittene zählt und zu einer Art obligatorischen Befähigungsnachweis im Rahmen einer wissenschaftlichen Karriere geworden.<sup>93</sup> Ihr, der Statistik, sagt hier keiner "Manipulation" nach. Was hier aber an Manipulation erinnert, ist der Gruppenzwang, (vorgeben) alles zu "verstehen", wenn man sich profilieren muss. Aus solchen Gründen ist von Seiten der Lehrenden keine Änderung zu erwarten, die es den Lernenden leichter machen könnte, mehr Verständnis für (und eine entspanntere Einstellung zur) Statistik zu gewinnen.

### *d) Modisches, Ultimatives und Albernes*

Dabei sind die Gegenstände der Statistik, mit denen man sich profilieren kann in nicht unerheblichem Maße auch typischen Moden unterworfen. In den 1970er Jahren brach unter den Professoren der BWL die Begeisterung für die "Faktorenanalyse" aus, dann kam die "Spektralanalyse" als eine Zeitreihenanalyse im Frequenzbereich, die inzwischen aber kaum noch jemand interessiert, wo sich doch im Rahmen der Ökonometrie eine ganz andere Art Zeitreihenanalyse entwickelte. Das Interesse verlagerte sich deshalb auf stochastische Prozesse, später dann auf VAR Modelle, und inzwischen ist es ein Muss geworden, etwas über Panelökometrie zu wissen, die es vor 20 oder 30 Jahren noch gar nicht gab, zumindest nicht in den Seminarveranstaltungen der Ökonomen. Unabhängig von ihrem Nutzen ist der Statistik somit auch in Zukunft eine große und wohl immer größer werdende Aufmerksamkeit in der Wissenschaft (jedenfalls in der Ökonomie) sicher.

Das gilt auch, mit etwas anderen fachlichen Schwerpunkten, für die angewandte (v.a. amtliche) Statistik, die im Dienst des Mess-,

<sup>93</sup> Mehr zu dem absurden Theater mit der Statistik im akademischen Bereich vgl. mein Papier "Statistik für Schaumschläger".

Planungs- und Kontrollwahn der Politik steht, und womit wir zum Thema "Computopia" zurückkommen. Es ist genau dieser Wahn, der der Statistik zum ewigen Leben verhilft.

Der Witz ist nämlich, dass jedes Messproblem selbst wieder weitere Messprobleme, Positionspapiere, Machbarkeitsstudien und wohl auch Dissertationen generiert. Bei der Einführung der Gesundheitsberichterstattung hielt man auch Zahlen über psychische Risiken (von denen doch eigentlich jeder zu jeder Zeit mehr oder weniger bedroht ist) und die Konstruktion eines Maßes für den "Gesundheitsstatus" für unverzichtbar. Diesen Status definierte man als "zu erwartende Dauer sozialer Funktionsfähigkeit", was natürlich nicht nur ein Prognoseproblem implizierte, sondern auch die Frage, was für Fähigkeiten bei welcher Rolle verlangt werden, um in ihr (noch) "funktionieren" zu können und bei welchen Diagnosen das nicht mehr der Fall ist.<sup>94</sup>

Ein anderes Beispiel ist die ökonomische Bewertung der Tätigkeit einer "Hausfrau", weil ja immer wieder eine ultimative, alles umfassende Gesamtrechnung gefordert wird, in der natürlich diese wohlfahrtsrelevante Produktion<sup>95</sup> mit einbezogen werden sollte. Weil die Messung über die Arbeitszeit der Frauen erfolgen soll, wurden dann allen Ernstes in der deutschen amtlichen Statistik Themen diskutiert wie: Ist die Zeit für "Beziehungsarbeit" mit dem Partner produktiv? Ja, sie ist es, aber die Zeit für eine telefonische Verabredung mit einer Freundin zum Besuch des Schwimmbads ist es nicht. Rückblickend

"wirkt die Dringlichkeit und Größe des Datenbedarfs, die Planungseuphorie und die Liebe zum statistischen Detail nicht selten geradezu grotesk."<sup>96</sup>

Aber nicht nur rückblickend, auch in der Gegenwart ist manches grotesk. Nur ein Bei-

spiel für etwas, was m.E. unter die Rubrik "Albernes" fällt: Ab April 2011 werden – ganz im Sinne der oben schon wiederholt erwähnten Glücksforschung – regelmäßig ca. 200000 Briten gebeten, auf einer Skala von 0 bis 10 anzukreuzen, was sie empfinden bei Fragen wie<sup>97</sup>

- Overall, how satisfied are you with your life nowadays ? oder
- Overall, to what extent do you feel that the things you do in your life are worthwhile?

Im Rahmen dieses happiness research Projekts will man ein measure of general well being (GWB) gewinnen, das dann das GDP ergänzen, wenn nicht gar ersetzen soll.

Angenommen, viele Briten schätzen ihr Tun und Lassen sub specie aeternitatis als nicht sonderlich bedeutungsvoll ein, so dass die zweite Frage nur ein Mittel von 3,64 erzielte. Was sagt uns das über die Qualität der aktuellen Wirtschafts- und Sozialpolitik? Wäre sie wirklich besser, wenn der Wert 4,82 statt 3,64 wäre? Wer kann aus dem negativen Ergebnis von  $3,64 - 4,82 = -1,18$  politische Entscheidungen z.B. über dadurch evtl. notwendige Ausgabenprogramme ableiten oder auch ganz individuell Konsequenzen für seinen Lebenswandel ziehen? Was kann ich (oder gar die Regierung) dagegen tun, wenn ich die Sachen, die ich tue nicht wirklich "worthwhile" finde? Mit anderen Worten: wie kann man die Ergebnisse der Statistik "operativ nutzen"? Ist das der oben genannte Kompass, den man angeblich nicht hat, wenn man kein Maß "beyond the GDP" hat?

<sup>94</sup> Auch hier gilt wieder, dass alles davon abhängt, wie genau man hinschaut. Nach einem Bonmot der Mediziner gibt es keine gesunden Menschen, sondern nur noch nicht ausreichend diagnostizierte Menschen.

<sup>95</sup> Es ist zweifellos eine wohlfahrtsrelevante, nicht aber eine marktmäßig erbrachte Leistung.

<sup>96</sup> P. v. d. Lippe, Wirtschaftsstatistik, 5. Aufl., Stuttgart 1996, S. 262.

<sup>97</sup> zitiert nach Niels aus dem Moore u. Christoph M. Schmidt, On the Shoulders of Giants: Tracing Back the Intellectual Sources of the Current Debate on "GDP and Beyond" to the 19th Century, in Jahrbücher für Nationalökonomie und Statistik Bd. 233/3, S. 266 ff. (271).

### **Fazit: Nicht Manipulation, sondern unsere Obsession alles zu quantifizieren ist das Problem**

Es zeigt sich, dass Manipulation im Zusammenhang mit Statistik (wenn man denn überhaupt von so etwas sprechen kann, und dabei jemand zur Rechenschaft ziehen kann) wohl das geringere Problem in Sachen Statistik sein dürfte, verglichen mit den Problemen, die wir haben mit

1. unserer Neigung (um nicht zu sagen *Obsession*), *überall Zahlen* (Rankings und Messungen) *zu verlangen*, und das
  - auch dann, wenn deren Notwendigkeit und Aussagefähigkeit kaum durchdacht ist, und die Zahlen, schon vom (nebulösen) Messziel her gesehen, auf Treibsand gebaut sind (wo die meisten Menschen Schwierigkeiten haben, zu sagen, was es überhaupt ist, wie z.B. das "Glück", wird auch eine Zahl hierfür kaum von Nutzen sein können),
  - obgleich unsere Fähigkeiten, selbst mathematisch einfache statistische Methoden zu verstehen und deren Ergebnisse inhaltlich zu interpretieren eher gering sind, und wohl kaum in dem Maße zunehmen werden, wie die Anzahl der angebotenen statistischen Daten und Analysen
  - und man bei dem Zwang überall nur noch auf der Basis von Zahlen entscheiden zu können, "kreative Buchführung" und Intransparenz der Methoden geradezu herausfordert;
2. dem *Mess-, Planungs- und Kontrollwahn der Politik ("Computopia")* obgleich Politiker
  - nicht selten eher die Statistik als ihre Überzeugungen für falsch halten und deshalb kaum bereit sind, Überzeugungen aufgrund von statistisch untermauerten Argumenten zu überdenken, sondern vielmehr
  - gerade bei wichtigen Themen und Kontroversen von Statistiken besser erst gar keinen Gebrauch machen und impressionistisch argumentieren, und
3. unserer *Neigung, Statistik für andere Zwecke zu instrumentalisieren*,
  - uns sich ihrer selektiv zur Rechtfertigung vorgefasster Meinungen zu bedienen, und sich mit dem Nichtvorhandensein entsprechender Statistiken zu exkulpieren und<sup>98</sup>
  - ihre Methoden exzessiv und ritualisiert anzuwenden und sich ihrer zur Profilierung als mathematisch versierter Wissenschaftler zu bedienen,
4. unseren enorm gestiegenen *technischen Möglichkeiten* quasi automatisch *noch mehr Daten* und noch mehr (kaum hinterfragte) Analysen (*big data*) zu produzieren.<sup>99</sup>

Hat man eine Zahl oder ist etwas "signifikant", so muss das nicht gleich Wunder bewirken. Umgekehrt ist keine Zahl zu haben nicht immer gleich eine Katastrophe. Quantifizierung kann auch zu einer Obsession werden, die keinen Wert hat, wenn unsere Fähigkeiten, Statistiken zu verstehen nicht im gleichen Maße wachsen, wie das Angebot an Statistiken.

<sup>98</sup> Siehe oben die Exkulpation der VWL-Professoren bei der Finanzkrise ab 2007 mit fehlenden Daten des IMF.

<sup>99</sup> "More data inevitably results in more time spent arguing, validating, reconciling and replicating. There is a real danger that Big Data moves us backward, not forward. It threatens to take science back to the Dark Ages, as bad theories gain ground by gathering bad evidence and drowning out good theories." Kaiser Fung, Numbersense, How to use Big Data to Your Advantage, 2013.

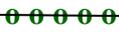
## Einige Nachträge

### zu Abschn. 2 und 4a:

#### Zählen, messen und rechnen (bzw. schätzen)

Weil es beim *Messen* von Merkmalen, die stetig und nicht diskret sind, beliebig viele Abstufungen gibt, kann nicht einfach "gezählt" werden und man muss vom Konzept der *natürlichen* Zahlen, das beim Zählen ausreicht, übergehen zu dem allgemeineren (mehr als nur die natürlichen Zahlen umfassenden) Konzept der *reellen* Zahlen.

Es ist jetzt die Einführung einer Maßeinheit (z.B. ein Meter), die es erlaubt, *das Messen gleichwohl auf ein Zählen zurückzuführen*. Um in noch feineren Abstufungen messen zu können, kann man eine "ein k-tel" Maßeinheit einführen, z.B. 1/100 Meter, also ein Zentimeter (bei  $k = 100$ ). Das verlangt den Übergang zu gebrochenen Zahlen (das ist dann nicht mehr eine natürliche, sondern eine rationale Zahl), etwa  $17/100 = 0,17$  wenn eine Länge von 17 cm gemessen wurde. Da der Zahlenstrahl soweit noch nicht überall "dicht" ist<sup>100</sup>, kann auch ein Hinzukommen von irrationalen Zahlen (wie z.B.  $\sqrt{2}$  oder  $\ln(3)$  usw.)<sup>101</sup> gefordert werden, womit man dann zu den "reellen" Zahlen gelangt.

Das Zählen ist zunächst ein einfacher Vorgang, wenn die zu zählenden Gegenstände in einer Reihe angeordnet sind und den Menschen war über lange Zeit nur diese Art des (Auf-) Zählens möglich. Es ist klar, dass es sich bei  um fünf Erbsen



handelt, aber um festzustellen wie viele Erbsen in einer Tüte sind, muss man dazu über-

gehen (z.B. das Volumen der Tüte) zu *berechnen*, bzw. zu *schätzen*.

<sup>100</sup> Das ist genau genommen nicht ganz richtig (weil er schon mit den rationalen Zahlen) "dicht" ist, sondern hier nur anschaulich gemeint.

<sup>101</sup> Eine irrationale Zahl ist damit definiert, dass sie nicht als Bruch (von natürlichen Zahlen) dargestellt werden kann. Das ist ein sehr schönes Beispiel für eine "operationale" (hier mit dem Hinweis auf eine nicht zielführende "Operation") Definition.

### zu Abschn. 3b und 5a, noch einmal:

#### Das Gerede über "Lügen mit Statistik"; welche Art Wissenschaft ist Statistiki?

Wenn man andere der Lüge bezichtigt, sollte man auch in der Lage sein, zu sagen, was die Wahrheit ist. Aber wo ist die Wahrheit in Sachen Statistik, wenn man unter Statistik (wie hier im Folgenden) nicht gerade den Teil versteht, der reine Mathematik ist (wie z.B. die Wahrscheinlichkeitsrechnung), sondern die Entwicklung geeigneter Methoden zur Beschreibung und Analyse der Wirklichkeit und die Interpretation von Statistiken, also numerischen Ergebnissen? Gibt es einen wahren und einen falschen Korrelationskoeffizienten, eine wahre und eine falsche Tabelle?

Es gibt Wissenschaften, wie die Mathematik, wo man über "wahr" und "falsch" durch einen Beweis entscheiden kann.<sup>102</sup> Die Statistik ist nicht so eine Wissenschaft. Man kann nicht beweisen, dass die Preisindexformel von Irving Fisher die einzig wahre Formel ist und dass alle anderen Preisindexformeln logisch "falsch" sind.<sup>103</sup>

Es gibt Wissenschaften, wie die Physik, wo man durch ein Experiment über "wahr" und "falsch" entscheiden kann. In diesem Sinne ist z.B. die Äthertheorie des Lichts durch das Michelson-Morley Experiment (1881 u. 1887) eindeutig und endgültig wi-

<sup>102</sup> Die Existenz von etwas, wie der Goldbachschen Vermutung (jede gerade Zahl  $> 2$  kann als Summe von zwei Primzahlen geschrieben werden, wie  $4 = 1+3$ ,  $6 = 1+5$ ,  $8 = 3+5$  usw.), die bisher niemand beweisen konnte, für die aber auch niemand ein Gegenbeispiel finden konnte, zeigt, dass es auch in der Mathematik nicht ganz so einfach ist. Wenn man daran festhält, dass es nichts gibt, was nicht entweder wahr oder falsch ist, muss man wohl zwischen "wahr" und "beweisbar" unterscheiden.

<sup>103</sup> Es hat – wie schon in Fußnote 25 angedeutet – in Deutschland den Versuch gegeben, die allein richtige Statistik aus einer "Logik des Vergleichs" zu entwickeln und man kann – gerade auch an den Positionen der Vertreter dieser Richtung zur Indextheorie – sehr schön zeigen, dass das Vorhaben, "Statistik" als eine spezielle "Logik" begreifen zu wollen notwendig scheitern musste. Vgl. v. d. Lippe, Die "Frankfurter Schule" in der Statistik und ihre Folgen, in: AStA Wirtschafts- und Sozialstatistisches Archiv 7 (2013), S. 71 – 89.

derlegt worden. Es gibt auch Wissenschaften, wie die Volkswirtschaftslehre, bei denen Thesen und Theorien durch Konfrontation mit Zahlen der Statistik und durch Berechnungen mit Methoden der Statistik "überprüft" werden, auch wenn das Ergebnis vielleicht für die Meinungen vieler irrelevant sein mag (bzw. immer geltend gemacht werden kann, eine Bestätigung oder Widerlegung sei nicht gelungen).

Die Statistik ist auch nicht so eine Wissenschaft.<sup>104</sup> Es gibt keinen empirischen Befund, aus dem folgt, dass man Berechnungen mit der Varianzanalyse besser bleiben lassen sollte, weil die Methode im Lichte dieses Befunds nicht haltbar ist. Es kann auch nicht die Erfahrung zeigen, dass wir bisher das Inlandsprodukt zu niedrig geschätzt haben, weil die zu messende Sache selbst, das Inlandsprodukt weitgehend eine Frage der Konvention ist.

Man kann nachgerade wohl auch die Auffassung vertreten, dass Statistik eher ein Instrument, oder eine Technologie (oder eine Kunst)<sup>105</sup> sei, was deswegen genauso ehrenvoll ist wie eine Wissenschaft, aber nicht an Wahrheit sondern an Nützlichkeit zu messen ist. "Lügen" ist dann schon von vornherein die falsche Kategorie.

Jemandem, der ein Flugzeug konstruiert kann man auch nicht vorwerfen, dass er "lügt", sondern nur, dass das Flugzeug vielleicht nichts taugt, oder dass man so etwas nicht braucht.

Andererseits kann man auch das sog. "Lügen" abmildern und sagen, man habe eigentlich nur "manipulieren" gemeint und man sei deshalb auch nicht in der Pflicht, sagen zu müssen, was hier die "Wahrheit" sei. Aber auch das ist nicht weniger unsinnig, wenn es denn als Vorwurf zu verste-

hen sein soll. Dass Statistiknutzer nicht vor Fehlinterpretationen sicher sind, kann man der Statistik nicht vorwerfen, solange es jedem möglich ist, durch mehr Wissen und mehr Nachdenken (und vielleicht auch kritisches Nachfragen) das "Falsche" an seiner Interpretation zu erkennen.

Die Statistik bietet sozusagen ein Produkt an, das viele Nachfrager nicht richtig "bedienen" können. Aber das ist nicht unehrenhaft, es sei denn man nutzt Unfähigkeit und Unwissen vorsätzlich aus.

Es bleibt somit vom "Lügen" nur die unterstellte böse Absicht übrig. Aber kann denn eine *Wissenschaft* eine böse Absicht haben, oder haben es vielleicht nur einige der betreffenden *Wissenschaftler*? Vom "Lügen mit *Statistik*" bleibt bestenfalls das "Lügen eines *Statistikers*" übrig.

Aber auch dann haben wir ein Problem. Denn wir haben es beim "Lügen-" oder "Manipulationsvorwurf" mit genau den drei Beweisproblemen zu tun, die wir im Zusammenhang mit der Kosten-Nutzen-Analyse einer schlechten Anwendung der Statistik<sup>106</sup> aufgezählt hatten:

1. *richtig/falsch*: gibt es eine "falsche" oder vielleicht nur eine "schlecht gemachte" statistische Analyse?
2. *vorwerfbar*? kann man jemand eine "schlecht gemachte" Statistik vorwerfen, wenn er es vielleicht nur nicht besser gekonnt hat?
3. *Schaden*? Wem entsteht ein Schaden durch eine "schlechte" Statistik, die er auch ohne Angabe von Gründen einfach ignorieren kann (so wie auch gute Statistiken ignoriert werden).

Weil die "Anklage" auf allen diesen Ebenen schlechte Karten hat, haben auch schlecht gemachte Statistiken einen positiven Nettonutzen. Wir werden also auch in Zukunft mit ihnen leben müssen, ohne sie der "Lüge" bezichtigen zu können.

Vom Vorwurf des "Lügens" bleibt also nichts übrig. Er ist unsinnig oder zumin-

<sup>104</sup> Sie ist nicht die einzige Wissenschaft, bei der über wahr/falsch nicht mit den genannten Methoden entschieden werden kann. Man kann dies z.B. auch nicht in der Rechtswissenschaft, der aber niemand "Lügen" vorwirft.

<sup>105</sup> "The fact is that despite its mathematical base, statistics is as much an art as it is a science". D. Huff: How to lie with Statistics, Pelican Books 1973, S. 109.

<sup>106</sup> vgl. oben Abschn. 3b.

dest nicht durchdacht und er verkennt völlig den Charakter der Statistik als Wissenschaft. Er wird aber weiter der Statistik anhaften und es ist ein Phänomen, warum er sich so dauerhafter Beliebtheit erfreut und offenbar so vielen Menschen intuitiv überzeugend erscheint.

Ich kann hier nur eine Vermutung aussprechen: Das Lästern über Statistik ist m.E. der Frust eines enttäuschten Liebhabers. Die Statistik ist – wie eingangs dargestellt – mit ihrer "Quantifizierung angetreten, um den Streit mit Worten und über Worte durch Zahlen zu "objektiveren". Aber das, was "Descartes Traum" (vgl. Fußnote 17) von der Mathematisierung und Rechenbarkeit der Welt und vom Sieg der Vernunft über den Mythos (und die Ideologie) genannt wird, dass man nämlich durch eine unbezweifelbare Rechnung feststellen kann, dass z.B. das "Glück" jetzt 87,5 beträgt, und zwar so unbezweifelbar wie  $2 + 3 = 5$

ist, wo dann ja auch niemand kommt und meint, es könnte bei genauerer Betrachtung 6,3 sein, dieser Traum ist nicht in Erfüllung gegangen.

Er ist eine Illusion, die wohl unverwüstlich bleibt (und die der Statistik – wie schon eingangs gesagt – wohl zu ewigem Leben verhilft). Es gibt in der Politik nicht die ultimative Exaktheit, Objektivität und Gewissheit über "Fakten" und die Statistik kann uns nicht die Last nehmen, uns immer wieder mit Meinungen von Leuten auseinanderzusetzen, die wir für dumm und uneinsichtig halten. Der Streit mit Worten und über Worte hat trotz Statistik nicht aufgehört und man könnte auch sagen – womit man die Dinge aber vielleicht wohl zu trübe sieht – dass leider noch der Streit über Zahlen hinzugekommen ist.