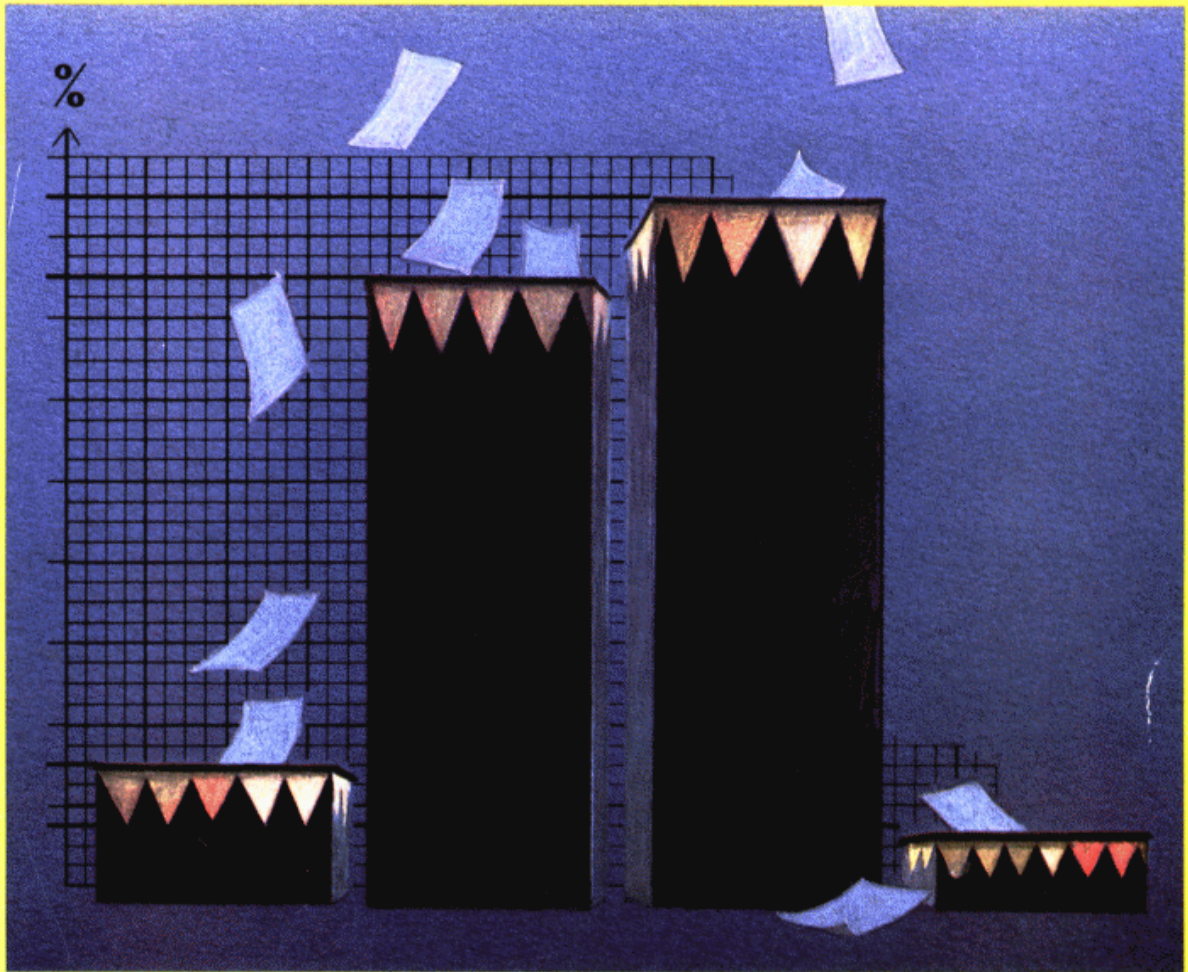


Der Lotteriecharakter des repräsentativen Querschnittes
Wahlprognosen und Meinungsumfragen

und der Abfahhandel mit Prozentzahlen

SONDERDRUCK AUS HEFT 30./31. JAHRGANG
DER ZEITSCHRIFT FÜR MARKT-, MEINUNGS- UND ZUKUNFTSFORSCHUNG

Fritz Ulmer



Wickert Institute Tübingen – Demokrit Verlag Tübingen

Der Lotterieverhalten des repräsentativen Querschnittes

Wahlprognosen und Meinungsumfragen und der Ablaßhandel mit Prozentzahlen

von Prof. Dr. Fritz Ulmer
Fachbereich Mathematik, Universität Wuppertal.

Eine Version dieses Artikels erschien erstmals vor Weihnachten 1986 in „Bild der Wissenschaft“, dann als Sonderdruck der Bergischen Universität und im Oktober 1987 als Sonderdruck in der Schriftenreihe „Zum Nachdenken: Lesenswerte Beiträge aus Zeitschriften und Büchern für politisch interessierte Bürger als Hilfe zur eigenen Urteilsbildung“, herausgegeben von der Landeszentrale für politische Bildung, einer staatlichen Stelle des Bundeslandes Hessen.

Daß nun auch im Kreis der Meinungsforscher Interesse an meiner Kritik gezeigt wird und ich von der „Zeitschrift für Markt-, Meinungs- und Zukunftsforschung“ um Abdruckerlaubnis gebeten wurde, hat mich sehr überrascht. Freundlicherweise gab „Bild der Wissenschaft“ die Genehmigung hierfür. Die unterschiedlichen Formate der beiden Zeitschriften machten Anpassungen bei den Tabellen und dem Fließtext erforderlich, was ich zum Anlaß nahm, verschiedene Teile zu überarbeiten und den Themenkreis zu erweitern. Die Umarbeitung führte schließlich zu einer Totalrevision, wobei das Datenmaterial von Mitte 1986 beibehalten wurde. In einigen Abschnitten sind neuere Daten verwendet worden, weil dies von der Thematik her geeigneter erschien. Die nun vorliegende Fassung¹ geht bezüglich Inhalt, Umfang und Kritik weit über die ursprüngliche Version hinaus. Satire - in Form von Parodie und Sarkasmus - ist Notwehr im Kampf gegen den grassierenden Ablaßhandel mit Prozentzahlen und die Zahlengläubigkeit. Mit Statistik allein ist diesem Krebsgeschwür nicht mehr beizukommen.

¹ Die Rechte für den hier erscheinenden Artikel bleiben beim Verfasser, die Rechte für die ursprüngliche Version bei Bild der Wissenschaft

Im Gedenken an meinen Vater

To J. and L. for their pressure to start this work
and to W. for the initial help

To my wife and children for enduring the production

Auch wenn ich in dieser Arbeit mit Spott über die Exzesse von Frau Noelle-Neumann und ihren Nachfahren im Prozentbereich herfalle und deren Auswirkungen auf die Gegenwartspolitik an den Pranger stelle, so strahlt dennoch ihr Werk - insbesondere die Schweigespirale - eine gewisse Faszination auf mich aus. Viele wichtige Dinge im Leben sind - auch wenn sie unzweifelhaft einen quantitativen Aspekt haben - nicht mit Zahlen begreif- und erfassbar. Beim Versuch, sie mit Prozentzahlen festzunageln, platzen sie leicht wie Seifenblasen und explodieren dem forschenden Untersucher als Knallfrösche ins Gesicht. Aber trotzdem sind sie wahr.

Inhaltsverzeichnis & Leitfaden

Abschnitt	Seite
I. Vorwort	5-14
II. Bild der Wissenschaft	15-20
Box: Dubiose Untersuchung	18-20
III. Der repräsentative Querschnitt	21-38
Box: Was ist ein Miniaturbild, was ist eine Zufallsauswahl?	29-33
Box: Simulation der Auslosung des repräsentativen Querschnittes	33-37
Box: Kommerzielle Grundlagen von Meinungsumfragen	37-38
IV. Die 5% Hürde	39-43
Box: 5% Hürde	41-43
V. Die Mehrheitsfrage	44-46
Box: Mehrheitsfrage	45-46
VI. Die Mehrheitsfrage unter Berücksichtigung der 5%-Hürde	47-49
Box: Auswirkung der 5%-Klausel auf die Mehrheitsbildung	48-49
VII. Die Vier-Parteien-Frage	50-57
Box: Die Auslosung von 100 repräsentativen Querschnitten	51
Box: Verteilungsfunktion der Vier-Parteien-Abweichungen	56
VIII. Pseudo-Trends	58-67
Box: Was der Zufall so alles aus eigener Kraft schafft	60-61
Box: Vergleich Pseudo-Trends und Trendmeldungen in DIE WELT	62
Box: Verteilungsfunktion der Pseudo-Trends	65
IX. Der zeitliche Trend über mehrere Monate	68-87
Box: Die ersten 35 Wiederholungen des Politbarometer	70-76
Box: 100.000 Wiederholungen des Politbarometers	82
Box: „Erfolgs“-Statistik von Wiederholungen des Politbarometers	86-87
X. Die Gewichtung	88-98
Box: Graphische Darstellung der Gewichtung des Politbarometers	91-93
Box: Wie das Politbarometer zur Gewichtung verdonnert wurde	94-96
Box: Spieglein, Spieglein an der Wand . . .	97-98
XI. Schlußfolgerungen	99-101
XII. Nachwort	102-107
Zusammenfassung	108-110
Summary	111-113

Es wird dem Leser empfohlen, sich bei seiner ersten Lektüre an den Fließtext zu halten – in Abschnitt III genügen sogar die ersten fünf Seiten – und sich nicht durch Einzelheiten in den zahlreichen Boxen aufhalten zu lassen. Bei einer zweiten Lektüre dürfte es leichter sein, diese etwas mehr technischen Details zu verstehen. Einige Druckfehler und etwas seltsame Redewendungen sollte man zum besseren Verständnis zweimal lesen.

I VORWORT

Anlaß für den Artikel in „Bild der Wissenschaft“ war die Bundestagswahl vom 25. Januar 1987. Während des Wahlkampfes ist die Öffentlichkeit einer Flut von „zuverlässigen“ Angaben über die aktuelle politische Stimmungslage ausgesetzt. In dieser Situation entsteht bei vielen das Bedürfnis nach grundsätzlicher Information: Wie werden Meinungsumfragen und Wahlprognosen gemacht? In welchem Maße sind sie wissenschaftlich begründet?

Ziel des Artikels war es, auf begrenztem Raum allgemeinverständlich Antworten auf diese und ähnliche Fragen zu geben und sie an konkreten Beispielen zu erläutern. Dabei waren Vereinfachungen und eine Stoffauswahl unumgänglich. Vieles von dem, was praktizierende Demoskopien und Wahlforscher zum eisernen Bestand zählen und was ihnen daher lieb und heilig ist, habe ich weggelassen, und dafür anderes, was ihnen unwesentlich erscheint, ins Zentrum gerückt, z.B. die Statistik. Diese Schwerpunktsetzung richtet sich danach, was meines Erachtens ein Demoskopie-Konsument wissen muß, um beim gegenwärtigen politischen Kräfteverhältnis den spekulativen Charakter von Wahlprognosen zu erkennen. Ihre Aussagekraft betreffend der wahlpolitisch relevanten Fragen – wie z.B. 5%-Hürde und Mehrheitsbildung sowie zeitlicher Trend der Parteistärken usw. – wird in der Presse, in Fernsehen und Radio, von Politikern und von Meinungsforschungsinstituten maßlos übertrieben. Hinter den so lautstark und oft in „wissenschaftlicher“ Verpackung feilgebotenen Zahlen verbirgt sich in Wirklichkeit nur ein Minimum an gesicherter Information.

Der „repräsentative Querschnitt“, auf dem die Legendenbildung beruht, wird – wie jeder Demoskopie-Kundige weiß – im Idealfall per Zufall erstellt, in die Umgangssprache übersetzt heißt dies: per Lotterie! Nur unter dieser Voraussetzung kann mit Hilfe von Statistik von den wenigen Befragten auf die ganze Population geschlossen werden. Die dem „repräsentativen Querschnitt“ entnommenen Daten haben daher zwangsläufig einen gewissen Grad an Lotteriencharakter. Das ist seit langem bekannt und vermag die Wahlforscher nicht mehr aufzuregen. Sie wännen diese Fehlerquelle – die lotteriebedingten Abweichungen – unter Kontrolle und halten sie im Vergleich zu den anderen Fehlern für unbedeutend. Daß dem nicht so ist, können die Demoskopien diesem Artikel entnehmen.

Die Fakten, die ich hier vorbringe, sind nicht neu², ich habe sie lediglich aus der „Verdrängung“ hervorgeholt und die seit langem praktizierten statistischen Alibiübungen und Milchmädchenrechnungen durch etwas realistischere Modelle und konkrete Berechnungen bzw. Simulationen ersetzt.

Wenn ein Demoskopiekonsument zum ersten Mal erfährt, daß der repräsentative Querschnitt im Prinzip per Lotterie erstellt wird, dann reagiert er bekanntlich höchst erstaunt: „Das ist doch Unsinn! Ein repräsentativer Querschnitt muß doch ein Spiegelbild der Bevölkerung oder zumindest ein Miniaturbild sein. Nur so kann der repräsentative Querschnitt die tatsächlichen Verhältnisse der ganzen Population widerspiegeln.“

Der Demoskopiekonsument empfindet die Gleichsetzung von „repräsentativ“ und „Lotterie“ als Irreführung. Und es hat durchaus seinen Grund, wenn die Demoskopie und die Vertreter ihrer Produkte in der Öffentlichkeit – die Medien, die Interessengruppen und die Politiker – grundsätzlich nur das Suggestivwort „repräsentativer Querschnitt“ verwenden und den Hinweis auf den Zufall konsequent unterdrücken. Man stelle sich vor, der Werbeslogan vom

² Man vergleiche z.B. Bulletin Nr. 60/1949 des Social Research Council "The Pre-Election Polls of 1948", nach dem Fiasko der Wahlprognosen bei der amerikanischen Präsidentenwahl.

„repräsentativen Querschnitt“ würde durch das „Kartellamt gegen unlauteren Wettbewerb“ aus dem Verkehr gezogen. Statt der irreführenden Floskel „Es wurden 1000 repräsentativ ausgewählte Personen befragt“ ist generell die Ausdrucksweise „Es wurden 1000 Personen ausgelost und dann befragt“ zu verwenden. Von nun an muß jedermann – von Helmut Kohl bis Jutta Ditfurth, von der FAZ über die Welt des Sterns und Spiegels bis hin zur Bilderzeitung, von der Handelskammer über den Kurverein bis Radio und Fernsehen, mit einem Wort: alles, was die Republik zusammenhält – seine demoskopischen Bedürfnisse mit Hilfe der Landeslotterie befriedigen. Ein erstes Beispiel:

Wegen eines Computerfehlers muß die Bundestagswahl annulliert werden. Da die Wiederholung mit riesigen Kosten verbunden wäre, werden die bewährten Methoden der modernen Demoskopie angewandt.

Die Nachwahl fand vom 2. bis 4. März 1987 unter dem Patronat der Landeslotterie statt. Teilnahmeberechtigt waren diejenigen Bürger, die am 25. Januar 1987 eine gültige Stimme abgegeben hatten. Von diesen wurden 1000 ausgelost und als repräsentativer Querschnitt deklariert. Die Auswertung der Stimmzettel ergab folgendes Resultat:

	CDU/CSU	SPD	FDP	Grüne	Rest
Nachwahl 2.3. bis 4.3.	40,3%	41,1%	8,6%	9,0%	1,0%
annulierte Wahl 25.1.1987	44,3%	37,0%	9,1%	8,3%	1,3%
Unterschied	-4,0%	+4,1%	-0,5%	+0,7%	-0,3%


Der repräsentative Querschnitt hat sein Machtwort gesprochen:
Die Koalition geht in die Opposition, das grünrote Bündnis wird Realität!


Das Machtwort des repräsentativen Querschnittes

Die Reaktion auf das Resultat der Nachwahl im meist betroffenen Blätterwald kann man sich leicht ausmalen:

- **DIE**  **WELT** verkündet den

Staatsstreich des repräsentativen Querschnittes

mit Entsetzen und ruft im Wirtschaftsteil und an den Finanzmärkten den Belagerungszustand aus. Aus Rücksicht auf **DIE**  **WELT** – Untergangsstimmung plädiert sie für eine vierjährige Staatstrauer oder Neuwahlen.

- Der  zeitung schlägt es den Rest der Sprache – und was sie weit härter trifft – auch die Bilder. In einer Sonderausgabe verpulvert sie ihre letzten Buchstaben zu einem ganzseitigen

PUTSCH!

Als Schadenersatz für den unblutigen Verlauf wird der Sturm auf die Bastille in einer Photomontage abgedruckt.

- Anders die **Frankfurter Allgemeine** (FAZ). Wie immer recherchieren ihre Leitartikler gründlich, und sie finden bald heraus, daß die Republik Anfang März vom Fasching heimgesucht worden war. In einem bewegten *Leid*artikel deckt der *Fromme Friedrich* auf, daß die Anhänger der Koalition auf Grund von empirisch/sozialen Unterschieden vermehrt unter den Anstrengungen vom Rosenmontag bis Aschermittwoch zu leiden hätten, was die Schweigespirale in Gang versetzt habe und diesen sonst nicht gerechtfertigten Vertrauensschwund auslöste. Er wirft die Frage auf, ob die bisherige Theorie der Gewichtung des repräsentativen Querschnittes solche soziologischen Traumata ausreichend berücksichtigt habe.

Die bisher gutgläubigen Konsumenten von Repräsentativumfragen werden wenig Verständnis für diese Erklärungen und Bemühungen zeigen. Sie kommen sich verschaukelt vor und behaupten mit Fug und Recht, daß vom Rosenmontag bis Aschermittwoch zwar allerlei Allotria, aber keine repräsentative Nachwahl möglich sei und daß das „Resultat“ nur das beweise, was man von einer Lotterie zu erwarten habe. Es sei doch seit den Wahlen nichts vorgefallen, was diesen politischen Erdrutsch erklären könne.

Ich kann dem aufgebrachten Demoskopie-Konsumenten an dieser Stelle nur versichern, daß die Nachwahl fair, korrekt und unter Einhaltung aller Kunstregeln der Demoskopie und der Statistik stattgefunden hat und daß es an dem Resultat nichts zu zweifeln und zu rütteln gibt. Die Nachwahl erfolgte so repräsentativ, wie ein Meinungsforscher nur träumen kann, sozusagen im demoskopischen Schlaraffenland : keine Verweigerer, keine Nicht-Angetroffenen, keine „Weiß nicht“, keine Noch-nicht-Entschiedenenen, sondern 100%-ige und wahrheitsgetreue Angaben – kurz: Photokopien der 1000 gültigen Stimmzettel, wie sie die Auszähler in der Wahlurne vorfanden. **Den Nachweis dafür, daß die Nachwahl repräsentativ war, sowie eine ausführliche Diskussion über den frommen, aber faulen Spruch vom repräsentativen Querschnitt findet der Leser in Abschnitt III.**

Die Retter der Republik: die Gewichtungskünstler

Im Gegensatz zu den Prognose-Endverbrauchern wird der Meinungsforscher die garantierte Zufallsauswahl der Landeslotterie und die 100%-ige Rücklaufquote erleichtert zur Kenntnis nehmen – hat man doch im demoskopischen Alltag mit dem Zufall und den Ausfällen so seine liebe Not.

Das Resultat der Nachwahl bereitet dem Wahlforscher wenig Kopfschmerzen. Er weiß seit langem, daß der „repräsentative Querschnitt“ nicht repräsentativ ist, solange er ihn nicht zurechtgebogen hat. Im Fachjargon heißt dies Kunst der Gewichtung, wobei – wie es bei bildenden Künsten üblich ist – jeder das Gewerbe auf seine Art und für seine Zwecke betreibt. Einen ersten Eindruck darüber vermittelt Frau Noelle-Neumann im Rheinischen Merkur Nr.37/1987:

„... nach der Theorie der Schweigespirale gewichten wir unsere Umfrageergebnisse um nach Meinungsklimadruck. Zwischen dem, was wir an Rohergebnissen erhalten, und dem, was wir als Prognose veröffentlichen, liegt manchmal eine Differenz von zehn oder elf Prozent. Denken Sie sich, wir würden die um zehn oder elf Prozent abweichenden Ergebnisse veröffentlichen!“

Veränderungen in diesem Ausmaß sind selbst für die Altmeisterin der Zunft immer wieder ein Ritt über den Bodensee:

„Ich bin jedesmal mit großer Sorge erfüllt.“

bekannt Frau Noelle-Neumann gegenüber der Wirtschaftswoche (Nr.1/2 1987). Es versteht sich von selbst, daß bei einer Umverteilung von 10% bis 11% genügend Freiraum bleibt, die „Theorie“ der Schweigespirale mit einem geeigneten Faschingsbonus auszuschnücken.

Zur politischen Gewichtung äußert sich die Wirtschaftswoche Nr.1/2 vom 2. Januar 1987 (Seite 28) ergänzend:

„Politische Gewichtung nennt sich der dramatische Höhepunkt **der demoskopischen Zirkuskunst, Recall-Verfahren die magische Formel**. Hauptgrundlage ist die neben der Sonntagsfrage gestellte Frage, hinter welche Partei der Interviewte bei der letzten Wahl sein Kreuzchen gemacht habe. Die so ermittelten Prozentwerte werden dann mit dem damaligen amtlichen Wahlergebnis verglichen.

Übliches Resultat – die Erinnerung stimmt nicht mit der Realität überein. In demselben Maße, wie die Erinnerung – der Recall – vom damaligen Wahlergebnis abweicht, werden die Angaben über die derzeitige Parteipräferenz umgerechnet. Behaupten etwa zu viele Befragte, sie hätten beim letztenmal sozialdemokratisch gewählt, so wird das aktuelle Umfrageergebnis für die SPD entsprechend heruntergewichtet. Liegt der Erinnerungswert für die FDP zu niedrig, werden die Liberalen entsprechend hochgewichtet: Deren aktueller Recall-Zuschlag liegt bei rund 100 Prozent.“

Das Resultat der Nachwahl bestätigt dem Wahlforscher also lediglich seine „Erfahrung“, daß bei Umfragen oft mehr Leute sagen, sie würden für die SPD und die Grünen stimmen, als es dann bei der Wahl tatsächlich der Fall ist. Das einzige, was ihn an der Nachwahl wirklich stört, ist der unerwartet hohe Anteil von 8,6% FDP-Stimmen. War der Verlust von 0,5% gegenüber der „annullierten“ Bundestagswahl zu klein? Nein, aber bei der FDP betrug die branchenübliche Profitmarge bis vor der Bundestagswahl rund 100% - d.h. man verdoppelte in etwa das Umfrageergebnis – was jetzt bei der Nachwahl nicht mehr so einfach geht, weil das Sparpotential bei den Grünen nicht genügend groß ist.

Als Wahlforscher beginnt man bei der Umgewichtung mit der wendigen FDP, denn wer diesen Faktor unter Kontrolle wähnt, hat fürs erste freie Fahrt in den Glauben, die politische Situation der Republik fest im Griff zu haben:

zweimal 8,6% gibt 17,2%

Das schluckt keiner mehr. 1983 erhielt die FDP bei den Bundestagswahlen 7,0% und 1987 bei der „annullierten“ Wahl 9,1%. Unter Berücksichtigung des Aufwärtstrends und der Tatsache, daß auch die FDP-Basis in letzter Zeit bei Umfragen vermehrt Farbe bekennt, läßt sich der wahre Anteil wie folgt mit 12% ermitteln:

Resultat der Repräsentativumfrage	8,6%
plus 50% historischer Mindestzuschlag für die Erfassung der klammheimlichen FDP-Wählerschaft	+4,3%
minus 10% für größere Bekenntnisfreudigkeit	-0,9%
<hr/>	
ergibt ein umgewichtetes Total für die FDP von	12,0%

Somit verbleiben 87% zur Verteilung für CDU/CSU, SPD und die Grünen (100% minus FDP minus einem Rest von 1%).

Nun zur Union: Man weiß – und viele teilen diese Ansicht -, daß es abwärts mit ihr geht. 1983 erhielt sie bei der Wende-Wahl 48,8%, aber 1985 und Anfang 1986 mußte man ihr bei den Umfragen oft kräftig unter die Arme greifen, um sie auf 45% hochzupumpen. Ihr historisches Tief bei der Bundestagswahl 1980 von 44,5% wird sie diesmal unterschreiten. Für die Mehrheit zusammen mit der FDP reichen 51% oder im schlimmsten Fall auch 50% oder sogar 49,5% (wenn man den Rest mit 1,1% ansetzt), folglich muß die Union 37,5% bis 39% auf die Beine bringen. Aber ein solches Resultat kann man gegenwärtig nicht veröffentlichen, denn unter Berücksichtigung historischer Perspektiven liegt die derzeitige untere Toleranzgrenze bei etwa 43%. Das würde eine rechnerische Mehrheit der Koalition von 55% ergeben. Kann man damit im heutigen politischen Umfeld auf den Markt gehen? Nein! Was soll man tun? Also 42% !...?

zum ersten ...
zum zweiten ...
und zum

Aber halt – in diesen schweren Zeiten und bei diesem harten Konkurrenzkampf will man sich als Geschäftsführer – Diplomkaufmann und Hobby-Psychosoziologe mit Volksschulabschluß in Statistik – möglichst wenig exponieren, denn mit einer zu auffälligen Fehlprognose könnte man leicht in den Ruch kommen, auch im lukrativen kommerziellen Alltagsgeschäft unzuverlässige Umfrageergebnisse zu liefern. So wird der Union schließlich bei 41% der Zuschlag erteilt. Aus Rücksicht auf das Koalitionsklima und die reichlich hohe Mehrheit von 53% wird zu guter Letzt der FDP-Anteil auf 11% reduziert, indem man entweder den oben auf 50% angesetzten Zuschlag für die Klammheimlichen auf 40% reduziert oder den Abzug für die Bekenntnisfreudigkeit von 10% auf 20% verdoppelt.

Damit verbleiben 47% für die SPD und die Grünen. Die Auslotung darüber könnte man entweder den nächsten grünroten Parteitag überlassen oder jetzt eine paritätische Aufteilung im Verhältnis 41 zu 9 vornehmen, welches sich bei der repräsentativen Nachwahl ergeben hat. Das würde 38,5% und 8,5% ergeben, aber weil erfahrungsgemäß in der Wahlkabine ein gewisser Anteil der Grünen trotz ihrer Neuen Heimat für die einstige politische Herkunft stimmt, wäre eine Aufteilung von 39% zu 8% oder gar 40% zu 7% realistischer. Aber die SPD bei 40%? Im Leben nie! Also bleibt nur die Möglichkeit von 39,5% zu 7,5%. Das amtlich-gewichtete Nachwahlresultat der Landeslotterie vom Rosenmontag bis Aschermittwoch lautet somit

CDU/CSU	SPD	FDP	GRÜNE	REST
41,0%	39,5%	11,0%	7,5%	1,0%

Wie der Demoskopie-Kundige sieht, haben wir nicht von dem ganzen Instrumentarium der Wahlforscher – Projektion, Recall, Feinjustierung und all dem Meinungsklimaschmu – Gebrauch gemacht, um die Republik wieder instand zu setzen. Normalerweise führen unsere Gewichtungskünstler einige weitere Additionen und Subtraktionen mit Prozentzahlen durch – die alle etwa von derselben Güte sind -, bis der repräsentative Querschnitt endlich „repräsentativ“ ist. Diese arithmetischen Verschiebungen werden aber nicht offen in der Buchhaltung als kurzfristige Verbindlichkeiten deklariert, sondern heimlich in der Bilanz aktiviert, als handle es sich um stille Reserven. Die Details der Gewichtungs-Arithmetik werden aus Sicherheitsgründen in einem wehrlosen Computerprogramm versteckt – nicht, weil sie jemand stehlen könnte – sondern weil es peinlich wäre, wenn sie ans Tageslicht kämen.

Die Wahlforscher sehen ihre Gewichtungskunst in nicht ganz so schwarzen Farben. So erklärte der Geschäftsführer von Basis-Research in der Wirtschaftswoche Nr.1/2 1987 optimistisch: „Eine fundierte theoretische, wissenschaftliche Grundlage dafür gibt es nicht. Das läuft rein pragmatisch ab.“ Solange eine reißende Nachfrage nach und der Glaube an Prozentzahlen und Prognosen besteht, werden uns diese

Pragmatiker, die mit beiden Füßen fest in der Luft stehen,

wohl-erhalten bleiben.

Nach diesem Exkurs in die demoskopische Zirkuskunst wollen wir das Geheimnis um die Nachwahl lüften, welche so tiefe Spuren in der Republik hinterlassen hat. Zu diesem Zweck führen wir erneut eine Wiederholung der „annullierten“ Bundestagswahl durch und lassen durch die Landeslotterie nochmals einen repräsentativen Querschnitt von 1000 Wahlberechtigten auslosen und ihre Stimmzettel auswerten. Weil das Resultat

CDU/CSU	SPD	FDP	GRÜNE	REST
41,1%	38,4%	8,7%	7,7%	1,1%

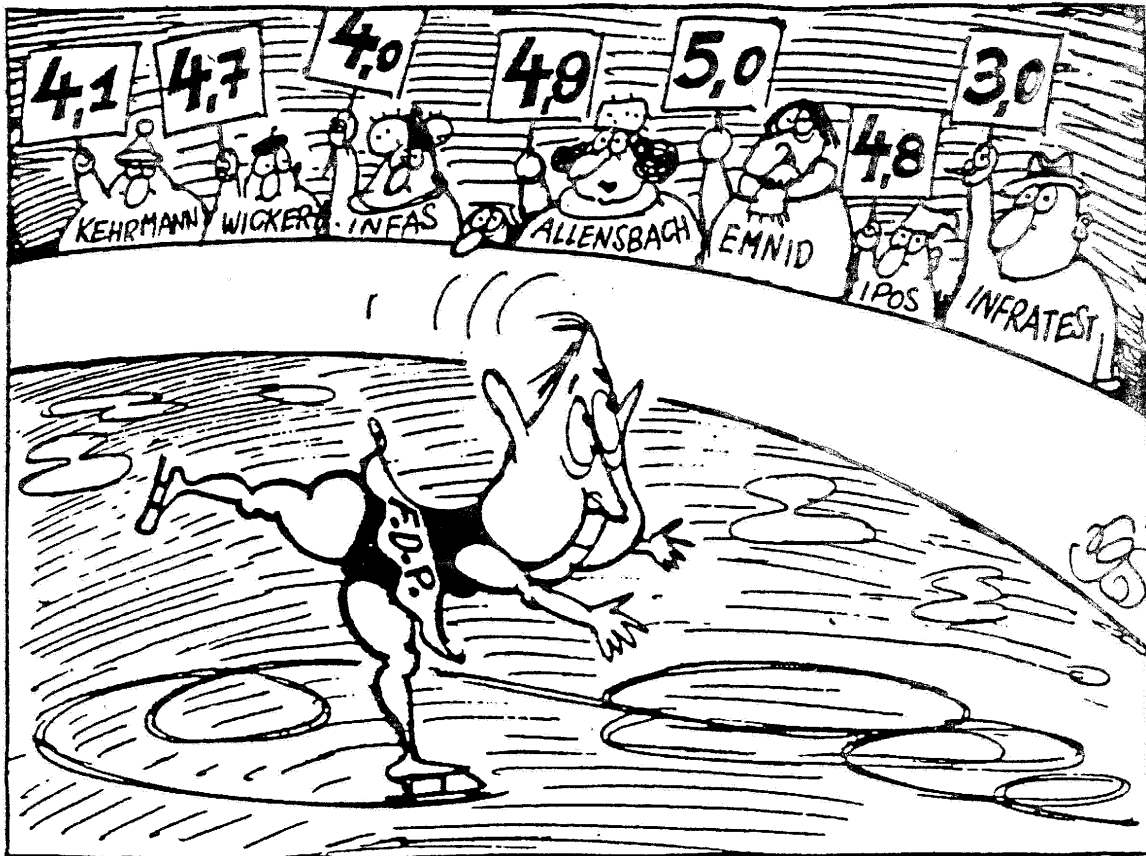
diesmal etwas anders aussieht, möchten uns die Gewichtungskünstler jetzt eine andere Geschichte erzählen. Aber geändert hat sich nichts! Wir haben nämlich – wie bei der ersten Nachwahl – den repräsentativen Querschnitt unter denjenigen Wahlberechtigten ausgelost, die am 25. Januar 1987 einen gültigen Stimmzettel in die Urne legten und die damals abgegebenen Stimmzettel ausgewertet. Die Wahlberechtigten der beiden repräsentativen Querschnitte für die Nachwahl und für ihre Wiederholung hatten somit gar keine Möglichkeit, ihre Meinung zu ändern. Nur die Launen des Zufalls und der „Meinungsklimadruck“ – die bei unseren Gewichtungskünstlern einen steten Leidensdruck bewirken – haben sich geändert! Aus diesem Grund haben wir die Gewichtungskünstler diesmal auf Kurzarbeit gesetzt und sie nur zum Stimmenzählen eingesetzt. Als verständig denkende Arbeitgeber haben wir sie beim Arbeitsamt zur Umschulung angemeldet. Wenn sie eine Computersprache erlernt haben – Basic genügt – dann werden sie auf dem Homecomputer ihrer Kinder – sei es ein Commodore C 64 oder ein Schneider CPC 464 – die Auslosung des repräsentativen Querschnittes mit Hilfe des eingebauten Zufallszahlengenerators selbst vornehmen können. Auf diese Weise werden sie in kürzester Zeit von ihren Zwangsvorstellungen betreffend Recallverfahren und Messung des „Meinungsklimadrucks“ geheilt werden.

Um den Leser mit den Launen des Zufalls vertrauter zu machen, haben wir auf dem Computer die Auslosung von einer Million „repräsentativer“ Querschnitte mit 1000 Befragten für die Bundestagswahl 1987 durchgeführt. Dieses Vorgehen bietet gegenüber den gängigen Methoden der Umfrageforschung einige Vorteile. Es kostet fast nichts, die Resultate liegen innerhalb einiger Stunden vor und jedermann merkt bald, welche Aussagefähigkeit „der“ repräsentative Querschnitt besitzt. Im folgenden sind die Resultate der ersten zehn Auslosungen des repräsentativen Querschnittes abgedruckt: Das amtliche Wahlergebnis lautete bekanntlich: CDU/CSU 44,3%, SPD 37,0%, FDP 9,1%, Grüne 8,3% und Rest 1,3%.

	CDU/CSU	SPD	FDP	Grüne	Rest
1. Auslosung	40,3%	41,1%	8,6%	9,0%	1,0%
2. Auslosung	44,1%	38,4%	8,7%	7,7%	1,1%
3. Auslosung	43,9%	38,5%	8,8%	8,2%	0,6%
4. Auslosung	47,1%	35,6%	7,8%	8,3%	1,2%
5. Auslosung	43,1%	37,4%	9,5%	8,8%	1,2%
6. Auslosung	47,0%	36,3%	8,4%	6,9%	1,4%
7. Auslosung	44,5%	34,3%	10,3%	9,4%	1,5%
8. Auslosung	43,5%	38,9%	8,4%	7,7%	1,5%
9. Auslosung	42,0%	37,9%	10,8%	8,1%	1,2%
10. Auslosung	43,7%	38,9%	8,7%	6,9%	1,8%

Diese 10 Auslosungen können wir uns als die Resultate von 10 unter idealen Bedingungen arbeitenden Meinungsforschungsinstituten vorstellen, denn diese Ergebnisse sind dadurch zustande gekommen, daß 10 mal 1000 Bürger unter denjenigen Wahlberechtigten ausgelost worden sind, die bei der Bundestagswahl eine gültige Stimme abgegeben haben. Es wurden die damals abgegebenen Stimmen ausgewertet. Wir ordnen daher jede Auslosung einem Meinungsforschungsinstitut zu, z.B. ist die 5. Auslosung das Resultat des 5. Institutes. Wie man sieht, reichen diese 10 Meinungsforschungsinstitute und ihre Resultate bereits aus, um jedermann glücklich zu machen. Die Union fühlt sich beim Institut Nr. 4 in guten Händen, denn es garantiert ihr 47,1% und hält gleichzeitig die FDP mit 7,8% in Schach. Die SPD wird vom Institut Nr. 1 mit dem Traumresultat von 41,1% beglückt und kann zusammen mit den Grünen (9,0%) versuchen, das grünrote Chaos in eine Regierungsmehrheit umzufunktionieren. Die Grünen, jedenfalls die Realos, halten sich ebenfalls an Institut Nr. 1, während die Fundis mit Institut Nr. 7 besser fahren, dort wird ihnen nicht nur 9,4% verbrieft, sondern es wird ihnen auch weis gemacht, daß die SPD mit 34,3% in den letzten Zügen liegt. Die FDP schließlich beruft sich auf Institut Nr. 9, nach welchem Strauß & Kohl mit 42,0% reif für Botschafterposten in Monaco und Liechtenstein sind, während Genscher mit 10,8% zum Quasi-Bundeskanzler und zur Integrationsfigur für den Rest der Union wird.

Man beachte, daß die 10 Institute weder gepfuscht noch geschummelt haben. Im Gegenteil, ihre Resultate – nämlich die 10 obigen Auslosungen – sind unter idealen Bedingungen zustande gekommen, es wurden alle Kunstregeln der Demoskopie und Statistik eingehalten. Die Unterschiede beruhen also nicht auf schlechter Arbeit oder gar Günstlingswirtschaft. **Vielmehr lassen sie sich grundsätzlich nicht vermeiden, sie sind durch die Methoden der Demoskopie bedingt, nämlich durch die Auslosung des repräsentativen Querschnittes.** Die folgende Karikatur (Spiegel Nr.8/1983) im Vorfeld der Bundestagswahl 1983, bei welcher es um das Überleben der FDP ging, macht klar, wie sehr der Leistungsanspruch an die Demoskopie und ihre tatsächlichen Möglichkeiten auseinanderklaffen.



Zitter-Kür

Mannheimer Morgen

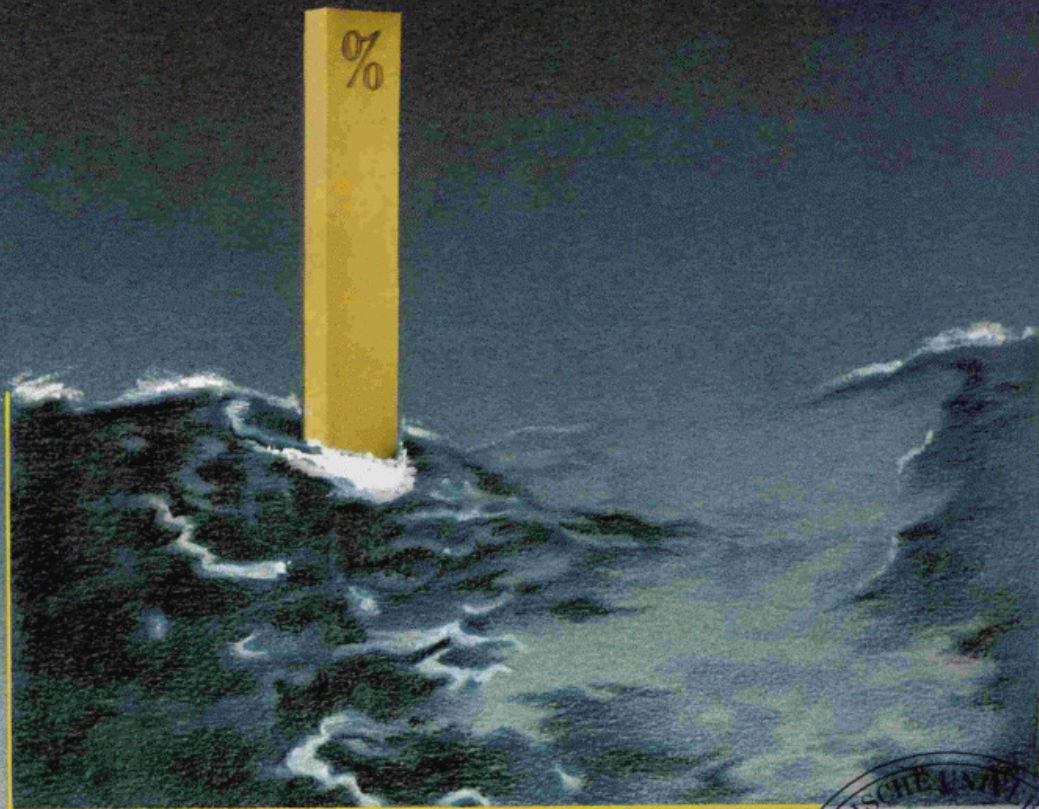
Die Karikatur und der Kommentar dazu gehen am Kern des Problems vorbei. Sie suggerieren, die unterschiedlichen Resultate seien auf subjektive Einflüsse bzw. ungenaues, nichtobjektives Messen zurückzuführen. In Wirklichkeit besteht der Kern des Problems in der unüberbrückbaren Kluft zwischen überzogenen Erwartungen einerseits und den realen Möglichkeiten der Demoskopie andererseits. Den Wunschvorstellungen der zahlengläubigen Prognosen-Endverbraucher stehen Meßmethoden der Demoskopie gegenüber, deren Genauigkeit und Zuverlässigkeit an Werkzeuge aus der Steinzeit erinnern. So wird die Nachfrage nach aussagekräftigen Daten von den Meinungsforschungsinstituten mit Pseudoinformation in Form von Prozentzahlen – nicht selten mit einer Nachkommastelle (wie z.B. bei der 5%-Hürde) – befriedigt. Die zur Debatte stehenden politischen Fragen können damit häufig nicht beantwortet werden, weil bereits die auslosungsbedingten Abweichungen zu groß sind. Hinzu kommen die Fehler, die bei der Befragung entstehen. Über das Ausmaß von beiden wird Stillschweigen gewahrt. Stattdessen wird durch die Verwendung des Suggestivwortes „repräsentativer Querschnitt“ beim ahnungslosen Demoskopiekonsumenten der Eindruck von Genauigkeit und Zuverlässigkeit erweckt.

Ein wesentliches Problem besteht also in der Weitergabe bzw. Überlassung von Prozentzahlen, mit denen der (politisch oder medienpolitisch motivierte) Auftraggeber anstellen kann, was er will. Wenn er von einem Dutzend Institute Resultate erhält, dann kann er sich aus dem Datenmaterial die ihm zweckdienlichen Zahlen auswählen und für seine Zwecke in der Öffentlichkeit einsetzen. Er braucht also nicht einmal Zahlen zu fälschen oder Suggestivfragen in Auftrag zu geben, um Meinungsumfragen zu manipulieren.

Auf Grund der Auslosung des repräsentativen Querschnitts sind die ermittelten Prozentzahlen zwangsläufig mit Fehlern behaftet, und die Chance, daß diese Fehler vernachlässigbar sind, ist – wie bereits die obigen zehn Auslosungen illustrieren – praktisch Null! Mit anderen Worten: Die von den Meinungsforschungsinstituten im „repräsentativen Querschnitt“ ermittelten Parteistärken treffen mit an Sicherheit grenzender Wahrscheinlichkeit nicht zu. Dies liegt nicht daran, daß die Institute ungenau und unzuverlässig arbeiten, sondern vielmehr lassen die Meßmethoden der Demoskopie grundsätzlich nur approximative Aussagen zu. Mit einer handelsüblichen Meinungsumfrage kann man nur Prozentbereiche für die Parteistärken angeben – z.B. CDU/CSU 42% bis 50%, SPD 33% bis 41%, FDP 6% bis 11%, Grüne 5% bis 10% (bei 1000 Interviews und der Standardsicherheit von 95%) - aber keine Prozentzahlen (wie z.B. CDU/CSU 46%, SPD 35%, FDP 8%, Grüne 10%), von einer Nachkommastelle ganz zu schweigen.

bild der wissenschaft

Sonderdruck



Der Orakelspruch
mit dem repräsentativen Querschnitt
**Wahlprognosen
und Meinungsumfragen**

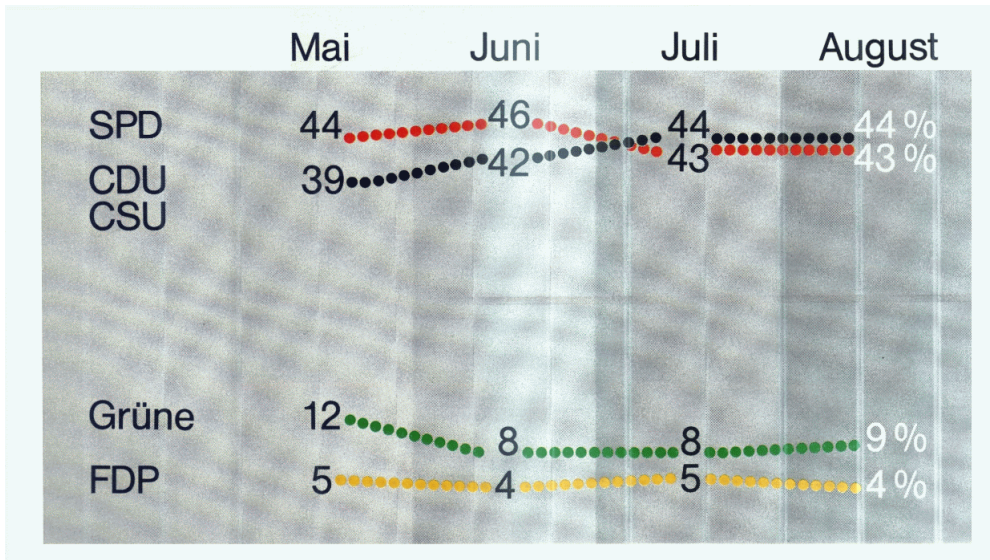
Fritz Ulmer



Politbarometer im Zweiten Deutschen Fernsehen:

„Welche Partei bevorzugen Sie im Augenblick?“

so fragen wir jeden Monat. Wir haben einen repräsentativen Querschnitt von über 1000 Wahlberechtigten befragt. Die Forschungsgruppe Wahlen hat das Ergebnis ausgewertet.



Vor der Bundestagswahl haben die Meinungsforscher wieder Hochkonjunktur. Mit repräsentativen Befragungen wollen sie den Wahlausgang voraussagen. Ist das Wissenschaft oder nur ein wissenschaftlich verbrämtes Orakel der Neuzeit? Unter einem repräsentativen Querschnitt verstehen die Demoskopon keineswegs ein verkleinertes Abbild der Bevölkerung, sondern eine per Lotterie getroffene Auswahl. Je zufälliger, desto besser. Sie berufen sich dabei auf die Statistik; doch was diese wirklich und seit langem sagt, geht bei der Prognose-Industrie oft verloren und wird durch Wunschenken ersetzt.

Schon kurze Zeit nach Schließung der Wahllokale am 25. Januar 1987 lag die erste Projektion auf die Endergebnisse vor und ließ mit recht guter Genauigkeit erkennen, wie die Wahl ausging. Wie aber steht es mit den Prognosen, die von verschiedenen Instituten in den Monaten und Wochen vor jeder Wahl gestellt werden? Geben sie schon verlässliche Hinweise auf den Ausgang? In der Vergangenheit war das keineswegs immer der Fall.

Zum Beispiel das Allensbacher Institut für Demoskopie: Es erstellte vor den Bundestagswahlen sehr aufwendige Prognosen, die sich auf bis zu 6000 Interviews stützten – sind doch die in der Öffentlichkeit vorgestellten Wahlprognosen das beste Werbemittel für die Demoskopen, falls sie gelingen. 1983 etwa lag die Allensbacher Vorhersage wie bei den vorhergehenden Bundestagswahlen recht nahe am tatsächlichen Wahlergebnis:

CDU/CSU	47,0%	(tatsächlich 48,8%)
SPD	40,0%	(tatsächlich 38,2%)
FDP	6,2%	(tatsächlich 7,0%)
Grüne	6,5%	(tatsächlich 5,6%)

Doch bei der Landtagswahl im Saarland 1985 lag die Prognose völlig daneben, wie die Leiterin des Instituts, die Demoskopie-Pionierin Frau Prof. Dr. Noelle-Neumann, in einem Rechtsstreit mit dem Spiegel offenlegen mußte:

CDU/CSU	47,0%	(tatsächlich 37,3%)
SPD	44,2%	(tatsächlich 49,2%)
FDP	11,9%	(tatsächlich 10,0%)
Grüne	6,5%	(tatsächlich 2,5%)

Die im Auftrag der saarländischen CDU erstellte Prognose basierte auf einer „Repräsentativ-Umfrage“ von nur 370 Interviews.³ Sie war nicht für die Öffentlichkeit bestimmt. Das einträgliche Alltagsgeschäft der Prognoseindustrie, kommerziell motivierte Meinungsumfragen, basiert in der Regel auf einigen hundert bis zweitausend Interviews. Bei solchen Umfragen schlägt keine Stunde der Wahrheit, gibt es keine Wahl. Wollte ein Auftraggeber das gelieferte Produkt überprüfen, hätte er nur die Möglichkeit, nochmals einen vier- bis fünfstelligen Betrag auf den Tisch zu legen und ein Konkurrenz-Institut mit der gleichen Umfrage zu beauftragen.

Es gibt viele Gründe, warum Wahlprognosen sich nicht erfüllen. Beispielsweise können sich die Wähler noch bis zur letzten Minute anders besinnen. Es ist sogar möglich, daß sie gerade aufgrund einer Prognose ihr Verhalten ändern, zum Beispiel um einer kleinen Partei doch noch über die 5%-Hürde zu helfen. Doch so sehr Gründe wie diese von den Demoskopen als Entschuldigung angeführt werden – und zweifellos auch eine Rolle spielen - : Es gibt ein Kernproblem, das damit nicht wegdiskutiert werden kann, und dies ist rein mathematischer, statistischer Natur.

³ Dies steht im Gegensatz zu dem, was Frau Noelle-Neumann in "Umfragen in der Massen gesellschaft" (S.152) als Überschrifts-Frage und sinngemäße Antwort formulierte:

"Sind weniger als 2000 auch repräsentativ?" Das hängt von der Genauigkeit ab.

Ihren Umgang mit der (statistischen) Ungenauigkeit ihrer Zahlen dürfte ihr der damalige saarländische CDU-Chef Zeyer kaum verzeihen, kam doch die daraus resultierende falsche Wahlkampfstrategie - der Griff nach der absoluten Mehrheit - als Bumerang auf ihn zurück. Die absolute Mehrheit kam zwar zustande, allerdings für die SPD. (Man vergleiche den ausführlichen Spiegel-Bericht in der auf der übernächsten Seite beginnenden Box).

Um dies ausführlicher zu erklären, ist zwischen zwei grundlegend verschiedenen Problem-Ebenen der Demoskopie zu unterscheiden:

Da sind zunächst die soziologischen und psychologischen Aspekte. Wie kann man jemanden so befragen, daß er die „Wahrheit“ antwortet? Selbst wenn man einen Original-Wahlzettel vorlegt und ankreuzen läßt, wird ein kleiner, vielleicht aber entscheidender Teil der Befragten sich am Wahlsonntag angesichts der Wahlurne doch anders entscheiden. Nicht nur, weil er inzwischen Gespräche mit anderen geführt hat oder zwischenzeitliche Ereignisse ihn umstimmen. Allein die Umgebung, die Situation des Interviews, der Zusammenhang der Fragen spielen eine Rolle. Der ungebetene Interviewer mit seinem seitenlangen Fragebogen kann die Entscheidung beeinflussen, ohne es zu beabsichtigen.

Die zweite Ebene ist rein mathematisch-statistischer Natur. Es kann schon aus Kostengründen nur ein kleiner Teil der Bevölkerung – einer von 40000 – befragt werden, ein sogenannter repräsentativer Querschnitt. Von diesem wird dann auf die Gesamtbevölkerung geschlossen. Das aber ist nur im statistischen Sinn möglich, enthält also einen bestimmten Fehlerspielraum. Genau hier liegt der Kern des Problems, der alle gängigen Wahlprognosen und erst recht kommerziell motivierte Meinungsumfragen als pseudowissenschaftliche Vorspiegelung exakter Zahlen entlarvt.

<p>Auszug aus dem SPIEGEL Nr. 43/1986 Seiten 146/147</p>	<p>Das Noelle-Institut ermittelte fünf Wochen vor der Wahl im Saarland am 10. März 1985:</p>	<p>eine Veröffentlichung zu verhindern, jeder auf eine andere Weise: Zeyer schwieg, und Frau Noelle-Neumann schwindelte. Sie bestritt die 47 Prozent und schickte einige frisierte Zahlen nach Hamburg, mit denen sie über das Desaster hinwegtäuschen wollte.</p>
<p>Meinungsforschung Dubiose Untersuchung Einen Rechtsstreit gegen den SPIEGEL glaubte Allensbach-Chefin Noelle-Neumann dadurch gewinnen zu können, daß sie eine bislang geheimgehaltene Umfrage vorlegte.</p>	<ul style="list-style-type: none"> • 47.0 Prozent für die CDU, die bei der Wahl fast zehn Prozent weniger erhielt • 44,2 Prozent für die SPD, die dann die absolute Mehrheit erkämpfte und einen Regierungswechsel erzwang. • 1,9 Prozent für die FDP, die ein zweistelliges Wahlergebnis erreichte. • 6,5 Prozent für die Grünen, die nur 2,5 Prozent bekamen. 	<p>Auch ohne Auskünfte der beiden Beteiligten und ohne Kenntnis der Umfrage-Tafelbände reichten die Informationen für einen Artikel (SPIEGEL 17/1985).</p>
<p>In allen (Bundes-) Ländern wurden Umfragen bekannt, die vor den jeweiligen Wahlen durchgeführt wurden. Oft taten sich die Institute schwer, realistische Zahlen zu ermitteln, wie sich auch bei den letzten Wahlen in Niedersachsen und Bayern zeigte.</p>	<p>Auftraggeber Werner Zeyer, zur Zeit der Umfrage CDU- und Regierungschef an der Saar, und Auftragnehmerin Noelle-Neumann versuchten</p>	<p>Die Allensbach-Chefin verlangte eine Gegendarstellung, die nach deutschem Presserecht unabhängig von ihrem Wahrheitsgehalt zu veröffentlichen ist. Als</p>

dies nur mit einem Zusatz geschah („Der SPIEGEL bleibt bei seiner Darstellung“), beantragte sie beim Hamburger Landgericht eine einstweilige Verfügung: Den SPIEGEL solle verboten werden, seinen Satz zu wiederholen. Ihre Saar-Unterlagen fügte sie bei.

Bei der Lektüre der 150 Seiten stellte sich heraus, daß die Untersuchung noch dubioser war, als die Recherchen für den ersten Artikel ergeben hatten.

Es handelt sich um zwei Umfragen. Für die erste lieferte das Institut zwei verschiedene Ergebnisse (siehe Graphik), weil in Allensbach der erste Bericht schon geschrieben wurde, als noch gar nicht alle Fragebögen im Institut eingetroffen und ausgewertet worden waren.

Bei der zweiten Umfrage wurden nur 474 Männer und Frauen befragt, von denen ein Fünftel die Auskunft über die Partei ihrer Wahl verweigerte. Mithin lagen den Angaben über die Stärke oder Schwäche der Saar-Parteien nur 370 Interviews zugrunde, jeweils 3,7 Befragte machten ein Prozent aus.

Ein Rechenbeispiel zeigt, auf welches Vabanquespiel sich ein Institut mit solch niedrigen Zahlen einläßt. Erklären 19 Befragte, sie würden FDP wählen, werden dieser Partei die lebens-

notwendigen fünf Prozent zugesprochen. Nennen 15 Befragte die Partei, zählen sie als vier Prozent. Es hat schon seinen Sinn, wenn Frau Noelle-Neumann – allerdings an ganz anderer Stelle – verkündet, daß „seriöse politische Umfragen erst bei 800 Befragten anfangen“.

Wollte man den Allensbacher Zahlen glauben, dann müßten sich in Saarland ohne irgendeinen politischen Grund zwei ganz verschiedene Entwicklungen vollzogen haben: in der Zeit zwischen der ersten und der zweiten Umfrage stark zugunsten der CDU, zwischen der zweiten Umfrage und den Wahlen noch weit stärker zugunsten der SPD.

Der Vergleich der ersten mit der zweiten Umfrage animierte die Allensbacher Wahlforscher, die Saar-CDU auf Sieg zu programmieren. Die Partei habe sich „klar vor die SPD gesetzt“: „Den Kampf um die absolute Mehrheit im Saarland zu führen, wird nach dieser Sachlage die Aufgabe der CDU sein müssen.“

Etwaige Zweifel wurden von vornherein erstickt: „Die Zunahme der Wahlabsicht für die CDU ist kein isoliertes Ergebnis, sondern sie wird abgestützt durch zahlreiche andere Ergebnisse.“ Da wundert man sich nicht, daß der leichtgläubige Zeyer als eine Art Allensbach-Echo

verkündete: „Wir liegen vor der SPD, und es wird deutlich, daß die CDU sich im Aufschwung befindet.“

Umgekehrt hätten die SPD-Genossen, wäre ihnen die Allensbach-Expertise zu Gesicht gekommen und hätten sie ihr vertraut, den Mut verlieren müssen. Die Situation dieser Partei, die einen in der Bundesrepublik fast beispiellosen Wahlsieg erkämpfte (nur zweimal gelang es einer Partei in einem anderen Bundesland, aus der Opposition heraus die absolute Mehrheit zu erkämpfen), schilderte das Noelle-Institut so, als stehe sie kurz vor ihrem Waterloo:

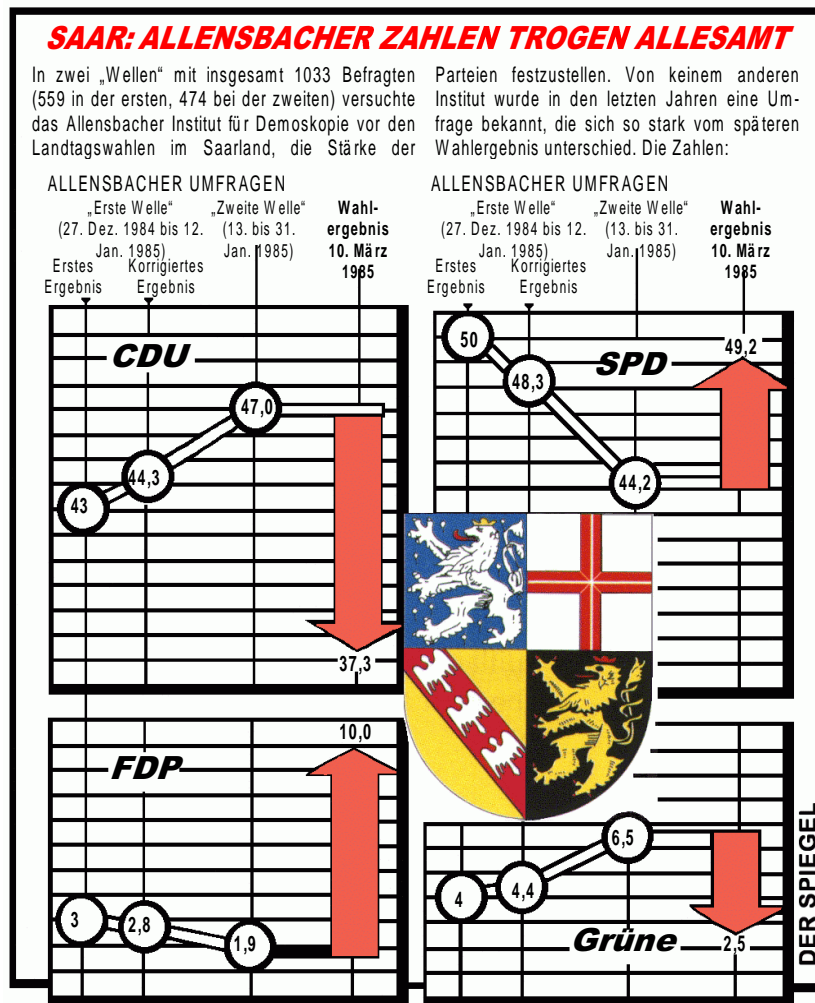
Sie sei „deutlich geschwächt“, auch habe „die Aktivität der SPD-Anhänger eher nachgelassen“. Es gebe eine „Abnahme in der Sympathie von Lafontaine“, und die SPD habe „an Vertrauen in ihre wirtschaftliche Kompetenz verloren“. Auch sei die „Erwartung, es werde nach der Landtagswahl zu einer SPD-geführten Regierung kommen, abgeflaut“.

Und, Schließlich: „für die SPD erscheint das Ziel, eine absolute Mehrheit im Saarland zu erreichen, gegenwärtig aussichtslos.“

Der Fehlprognosen noch immer nicht genug: es scheine „jetzt sicher, daß die Grünen die Fünf-Prozent-Hürde überwinden“.

Die Zivilkammer 24 des Landgerichtes Hamburg wies angesichts solcher Zahlen und Sprüche den Allensbacher Antrag auf Erlaß einer einstweiligen Verfügung zurück: Die Behauptung, die Frau Noelle-Neumann verbieten lassen wolle, müsse „als richtig angesehen werden“.

Und: Wenn Allensbach „einen Monat vor der Wahl von einer möglichen absoluten Mehrheit spricht und der tatsächliche Wahlausgang für die beratene Partei lediglich 37 Prozent ergibt, darf zu Recht von einem „Danebentappen“ gesprochen werden.



Anmerkung: Dieses Beispiel von „Repräsentativumfragen“ auf der Basis von 370 Interviews illustriert auch, daß bei Frau Noelle-Neumann Theorie und Praxis betreffend Pseudogenauigkeit von Prozentzahlen auseinanderklaffen. Man vergleiche ihre Ausführungen zu diesem Punkt in ihrem Buch „Umfragen in der Massengesellschaft“ (S. 223): „Obwohl es mathematisch-statistisch niemals falsch sein kann, Prozente auf Dezimale genau auszurechnen, können Dezimalzahlen psychologisch irreführen, indem sie eine größere Genauigkeit vortäuschen, als die ursprünglichen Zahlen für sich beanspruchen können. Sind die Prozentzahlen ... auf der Basis von 300 Fällen berechnet und nicht auf der Basis von 10 000 Fällen, so ist es viel besser, auf die Bedeutungslosigkeit solcher Unterschiede (in der Nachkommastelle) hinzuweisen, indem man die Dezimale wegläßt und auf- bzw. abrundet.“

III Grabrede für den repräsentativen Querschnitt

Was ist eigentlich ein repräsentativer Querschnitt von 1000 Wahlberechtigten? Die Wortwahl suggeriert, dies sei eine Art Miniaturbild der Bevölkerung, welches bezüglich wichtiger Merkmale wie Geschlecht, Alter, Familienstand, Beruf, Einkommen, Wohngegend (Stadt/Land) die tatsächlichen Verhältnisse widerspiegelt. Darum ergäbe eine Befragung dieser Gruppe ein ähnliches Resultat wie die Befragung der gesamten Bevölkerung, und wenn man einen anderen repräsentativen Querschnitt befragte, so würde man im wesentlichen dieselben Resultate bekommen.

Diese Vorstellung ist falsch. Die Desinformation der Öffentlichkeit ist einerseits darauf zurückzuführen, daß die präzise Definition begrifflich nicht einfach ist, andererseits liegt es nicht im Geschäftsinteresse der Prognose-Industrie und ihrer Kunden in Medien und Politik, Klarheit über diesen zentralen Begriff zu schaffen. Eine Klärung hätte nämlich zur Folge, daß viele Prognosen und die darauf aufgebauten – so populären – Geschichten in sich zusammenbrechen würden.

Tatsächlich ist das einzig repräsentative eines solchen Querschnitts die Absicht, daß jeder Wahlberechtigte die gleiche Chance hat, befragt zu werden. Die Chancengleichheit, d.h. die Zufallsauswahl, gilt als Qualitätsmerkmal einer Umfrage, je zufälliger, desto besser.

Der Leser wird sich fragen, weshalb die Meinungsforscher zu einem so dubiosen und unsinnig erscheinenden Verfahren wie der Zufallsauswahl greifen, um einen repräsentativen Querschnitt zu erstellen. Weshalb wird nicht versucht, ein Miniaturbild der Population zu erstellen?⁴ Die Antwort ist für einen Laien nicht einfach zu verstehen: Erstens ist dies mathematisch unmöglich und zweitens würde es nichts nutzen.

Der Glaube und die Vorstellung vom Miniaturbild oder vom verkleinerten Modell dienen ausschließlich den Werbezwecken der Prognoseindustrie: Durch diese Suggestivwörter wird bei Abnehmern und Konsumenten von „Repräsentativumfragen“ der Eindruck erweckt, die gelieferten Daten seien glaubwürdig und zuverlässig.

Der Versuch, aus 45 Millionen eine Gruppe von 1000 Personen auszuwählen, derart, daß diese bezüglich wichtiger Merkmale wie Alter, Geschlecht, Konfession, Familienstand, Schulbildung, Beruf, Einkommen⁵, Wohnortgröße, (Dorf,...,Großstadt), Bundesland usw. in etwa so

⁴ vergleiche Noelle-Neumann, Seite 109/110 in "Umfragen in der Massengesellschaft" : "Das verkleinerte Modell" "Das soziologisch-statistische Modell des repräsentativen Querschnitts, das der Umfrageforschung zugrunde liegt, muß man sich folgendermaßen vorstellen: Der gesamte interessierende Personenkreis, der vor der Anlage einer Untersuchung abgegrenzt werden muß, setzt sich aus Personen verschiedenen Alters, Berufs und so weiter in den einzelnen Bundesländern und Gemeindegrößenklassen zusammen. Es ist nun die Idee des repräsentativen Querschnitts, ein in der Zusammensetzung dieser Personengruppe gleichwertiges verkleinertes Modell zu erstellen. In dem verkleinerten Modell der Gesamtgruppe sind auch die zu untersuchenden Ansichten und Antwort-Kategorien in modellgerechtem Anteil vertreten, *wenn dieses Modell nach bestimmten Verfahren korrekt zusammengesetzt wurde.*"

Die Grabrede für "das verkleinerte Modell" verbirgt sich hinter dem letzten, kursiv geschriebenen Nachsatz.

⁵ In der Praxis dürfte es kaum gelingen, das zweifellos wichtige Merkmal "Einkommen" zu verwenden. Die Einkommensverteilung der ganzen Population läßt sich zwar leicht ermitteln, aber bei der Zusammenstellung des "repräsentativen Querschnittes" stellt sich das Problem, daß wohl nur wenige Bürger bereit sein dürften, dem Interviewer die Lohnsteuerkarte oder die Einkommensteuererklärung vorzulegen. Bei der Ermittlung der Merkmale Geschlecht, Familienstand, Konfession, Altersklasse, Wohngegend stellt sich dieses Problem nicht bzw. es spielt eine vergleichsweise untergeordnete Rolle.

zusammengesetzt ist wie die ganze Population, erweist sich, wie nun gezeigt wird, als undurchführbar. Man kann zwar die prozentualen Klassenanteile bezüglich der einzelnen Merkmale aus der amtlichen Statistik entnehmen, aber damit läßt sich der Begriff „repräsentativ“ nicht charakterisieren. Wenn beispielsweise eine Gruppe von 1000 Personen zu 50% aus Männern und zu 50% aus Frauen besteht und gleichzeitig aus 50% Katholiken und 50% Protestanten, dann kann sie trotzdem völlig unrepräsentativ betreffend Geschlecht und Konfession sein. Sie kann nämlich aus 500 katholischen Männern und 500 protestantischen Frauen bestehen. Um dies auszuschließen, muß man auch die Kombinationen männlich/katholisch, weiblich/katholisch, männlich/protestantisch und weiblich/protestantisch berücksichtigen und deren prozentuale Anteile ermitteln. Nur wenn auch diese Anteile die Verhältnisse der ganzen Population widerspiegeln, kann man von der Tausender-Gruppe sagen, sie sei betreffend Konfession und Geschlecht „repräsentativ“. Dasselbe Problem tritt für die anderen Merkmale auf. Die Anzahl der oben aufgeführten Merkmale und die zugehörigen Klasseneinteilungen (bezüglich Geschlecht, Konfession, Schulbildung, Alter, Beruf, Einkommen, Wohnortgröße) ist zwar relativ klein, aber die Anzahl der möglichen Kombinationen

- z.B. selbständig erwerbende Gärtner mit mittlerer Reife, in einer Großstadt wohnend, katholisch, verheiratet mit Kindern und einem Einkommen von DM 35000 bis 45000 –

geht in die Zehntausende und übertrifft den Umfang des geplanten „repräsentativen Querschnittes“ um ein mehrfaches. Unsere katholischen Großstadt-Gärtner von mittlerer Reife ... haben somit keine Aussicht, in der nächsten Zeit in den „repräsentativen Querschnitt“ zu kommen, denn dieser ist für Jahre ausgebucht

Dazu kommt, daß die Angaben über die prozentualen Anteile all dieser Kombinationen nicht verfügbar sind. Aber selbst wenn diese Probleme nicht existierten, wäre mit diesem Vorgehen nicht viel gewonnen. Denn offensichtlich wird das Wahlverhalten eines Bürgers nicht durch Merkmale wie Alter, Geschlecht, Beruf, Einkommen usw. bestimmt. Unsere katholischen Großstadt-Gärtner mit einem Einkommen von ... können sowohl für CDU/CSU, die FDP, die Grünen als auch die SPD stimmen. Es gibt keine Kausalität, sondern lediglich tendenzielle Aussagen – sogenannte Korrelationen – und die daraus resultierenden Unsicherheiten schließen geradezu aus, daß aus einem solchen Konstrukt je ein verlässlicher „repräsentativer Querschnitt“ betreffend Parteipräferenz und Wahlverhalten werden kann.

Auch Frau Noelle-Neumann bemerkt in ihrem Buch „Umfragen in der Massengesellschaft“ auf Seite 134, daß ein verkleinertes Modell in keiner Weise repräsentativ zu sein braucht:

„Die Repräsentanz der Quoten-Stichprobe wird nicht gesichert, indem bestimmte „quotierte“ Merkmale – Geschlecht, Alter usw. genauso verteilt sind wie in der Grundgesamtheit (=Population). Man kann sich zahlreiche Stichproben (=repräsentative Querschnitte) denken – beispielsweise Patienten von Krankenhäusern oder Reisende der Bundesbahn -, die in ihrer Zusammensetzung nach Geschlecht, Alter, Berufsgruppe, regionaler Verteilung den Proportionen der erwachsenen Bevölkerung der BRD genau entsprechen, ohne deshalb in irgendeiner Weise eine repräsentative Stichprobe der erwachsenen Bevölkerung zu bilden.“

So bleibt der Demoskopie nur das übrig, was ihr die Statistik als Ersatz für die Fata Morgana „repräsentativer Querschnitt“ alias „verkleinertes Modell“ anbietet, nämlich die Auswahl per Lotterie. Dieses Verfahren liefert viel bessere Resultate, als der Laie sich vorstellen kann, aber

bei weitem nicht so gute, wie die Demoskopie seit Jahrzehnten behauptet.

Ein per Lotterie erstellter „repräsentativer Querschnitt“ vom Umfang 1000 kann nicht für dutzende oder gar hunderte von Merkmalen in gleichem Maße „repräsentativ“ sein wie für ein einziges „Ja/Nein“-Merkmal. Auch hier gilt das Sprichwort, daß jedes Ding seinen Preis hat : Je umfangreicher und detaillierter die Information ist, die mit einer „Repräsentativumfrage“ ermittelt werden soll, desto mehr weicht der „repräsentative Querschnitt“ von einem Miniaturbild ab.(vgl. Abschn. VI-IX und die Box S. III/12 bis III/14).

Man kann eben nicht mit Sicherheit von 1000 auf 45 Millionen schließen, dies ist zwangsläufig mit einem Fehler und einem Risiko verbunden. Die Statistik macht nun Aussagen darüber, wie groß die Chance ist, daß die durch die Auslosung verursachten Abweichungen bei einem „Ja/Nein“-Merkmal innerhalb eines vom Demoskop frei wählbaren Spielraums liegen ⁶(3) (Dieser wird – allerdings unzutreffend – auch als statistischer Fehlerbereich oder Vertrauensintervall bezeichnet). Wählt der Demoskop den Spielraum für die durch Auslosung verursachten Abweichungen zu klein (z.B. $\pm 0,5\%$), dann wird die Repräsentativumfrage die tatsächlichen Verhältnisse mit großer Chance falsch widerspiegeln. Wählt der Demoskop umgekehrt einen großen Spielraum (z.B. $\pm 5\%$), dann wird zwar die Chance für die Richtigkeit der Repräsentativumfrage groß – und erreicht den in der Statistik üblichen Standard – aber dafür besteht die Gefahr, daß die (an sich richtigen) Aussagen inhaltlich nichtssagend sind. In der Wahlforschung hätte ein Spielraum von $\pm 5\%$ zur Folge, daß eine Prognose in der Form „die CDU/CSU liegt zwischen 40% und 50%“ bekannt gegeben werden müßte. Na und? Wer hat das nicht schon vorher gewußt?

Für die Beurteilung der Aussagekraft von Umfrageergebnissen ist die Kenntnis von Fehlerspielraum und Chance unerlässlich.

Aus optischen Gründen werden bei Meinungsumfragen und Wahlprognosen oft genaue Prozentzahlen (manchmal mit Nachkommastelle) angegeben, d.h. es werden die in der „Repräsentativumfrage“ ermittelten Prozentzahlen auf die ganze Population übertragen. Diese Prozentzahlen treffen – wie in diesem Aufsatz gezeigt wird – mit an Sicherheit grenzender Wahrscheinlichkeit nicht zu, sie sind pseudoexakt. Im demoskopischen Alltag wird die Problematik Spielraum – Chance dadurch gelöst, daß die Chance der Meinungsumfragen, richtig zu sein, überhaupt nicht bzw. inkorrekt angegeben wird.

Neben den durch die Auslosung (des repräsentativen Querschnittes) verursachten Abweichungen treten – wie vorhin erwähnt – bei der Befragung zusätzliche Fehler auf, die als Interviewfehler bezeichnet werden. Die im Interview gegebene Antwort stimmt nicht mit dem überein, was der Befragte dann auf dem Stimmzettel ankreuzt und in die Wahlurne legt. Zum Beispiel, weil er seine Meinung geändert hat, er dem Interviewer nicht die richtige Antwort geben wollte, er sich noch nicht entschieden hat, er nicht zur Urne gehen wollte und dann doch ging oder umgekehrt. Außerdem gibt fast ein Drittel der Ausgelosten keine Antwort oder kann vom Interviewer nicht

⁶ Im demoskopischen Alltag wird überhaupt nicht beachtet, daß die Aussagen der Statistik sich normalerweise auf nur ein "Ja/Nein"-Merkmal beziehen. In der Praxis wird so operiert, als würden diese Aussagen auch für dutzende oder hunderte von Merkmalen mit mehreren Ausprägungen zutreffen, was nicht stimmt. Bei dieser "Überforderung" des per Lotterie erstellten repräsentativen Querschnittes müssen zwangsläufig bei einzelnen Merkmalen erhebliche Verzerrungen auftreten. Aber man weiß nicht, bei welchen Merkmalen der repräsentative Querschnitt "verrückt" spielt (vergleiche Abschnitt VII und VIII).

erreicht werden. Ferner spielt die weitverbreitete Unkenntnis über die unterschiedliche Bedeutung von Erst- und Zweitstimme eine Rolle. Es kommt auch vor – allerdings nicht sehr häufig –, daß der Interviewer den Fragebogen selbst ausfüllt, weil die Bezahlung nicht gerade fürstlich ist und es so viel schneller geht. Der Interviewfehler kann weit größer sein als die durch die Auslosung verursachten Fehler und 5% übersteigen – wie aus den eklatanten Fehlprognosen hervorgeht – und er ist weder kontrollier- noch meßbar. Das potentielle Ausmaß des Interviewfehlers geht auch aus der folgenden Äußerung von Frau Noelle-Neumann im „Rheinischen Merkur (Extra)“ Nr.37, 1987 hervor:

„Zwischen dem, was wir an Rohergebnissen (=tatsächliche Umfrageergebnisse) erhalten und dem, was wir veröffentlichen, liegt manchmal eine Differenz von zehn bis elf Prozent.“

Entscheidend für die Qualität des „repräsentativen Querschnittes“ ist die Chancengleichheit des Auswahlverfahrens. Aus Kostengründen werden hier erhebliche Abstriche gemacht, was unkontrollierbare Auswirkungen zur Folge hat. Die praktische Durchführung der „Zufallsauswahl“ variiert erheblich. Im folgenden wird eine kurze Beschreibung der beiden wichtigsten Verfahren und der damit verbundenen Probleme gegeben. Eine vollständige Darstellung würde den Rahmen dieses Aufsatzes sprengen. Die Diskussion, inwieweit man von Zufallsauswahl sprechen kann, müßte allein vom Umfang her der Gegenstand einer eigenen Arbeit werden.

Das ADM-Verfahren wurde vom Arbeitskreis Deutscher Marktforschungs-Institute entwickelt. Es arbeitet in drei Stufen:

- Ausgangspunkt sind die rund 60000 Stimmbezirke der Bundesrepublik Deutschland, die jeweils einem Wahllokal zugeordnet sind. Kleinere Bezirke werden mit anderen zusammengelegt, so daß jeder Bezirk mindestens 400 Wahlberechtigte enthält. Rund 50000 Bezirke bleiben so übrig. Aus diesen werden beispielsweise 210 Bezirke – ein sogenanntes ADM-Netz – durch ein Zufallsprogramm ausgewählt, das die Chancen jedes Bezirks proportional zur Zahl seiner Haushalte steuert.
- In den ausgewählten Bezirken suchen Interviewer nach einem bestimmten Zufalls-Verfahren sieben bis zehn Haushalte aus. Dadurch soll gewährleistet werden, daß jeder bundesdeutsche Haushalt die gleiche Chance hat, in die Auswahl zu kommen.
- In jedem Haushalt sucht der Interviewer, wiederum nach einem Zufallsverfahren, einen Wahlberechtigten aus. Da der einzelne eine größere Chance bekommt teilzunehmen, wenn er in einem kleinen Haushalt lebt, werden die Angaben der Befragten mit der Größe des Haushalts gewichtet. Ein Vertreter eines „Vier-Wahlberechtigten-Haushalts“ zum Beispiel zählt also doppelt so viel wie ein Vertreter eines „Zwei-Wahlberechtigten-Haushalts“.
- Von den geplanten rund 1600 Interviews kommen gut ein Drittel nicht zustande, weil entweder die ausgelosten Wahlberechtigten nicht angetroffen werden oder sie die Befragung ablehnen.

Mit diesem Verfahren wird angestrebt, daß jeder Wahlberechtigte mit gleicher Wahrscheinlichkeit in den repräsentativen Querschnitt kommt. Im Idealfall sollte die Auswahl so zufällig sein, als würde man für jeden Wahlberechtigten ein Los in eine Trommel legen und daraus 1000 Lose ziehen. Doch durch die Dreistufigkeit des ADM-Verfahrens und seine

praktische Durchführung ist die Realität von diesem Ideal weit entfernt.

Einerseits wird aus Kostengründen die erste Stufe des Verfahrens – die Auswahl von 210 Stimmbezirken – nicht bei jeder Erstellung eines repräsentativen Querschnittes wiederholt, sondern ein Meinungsforschungsinstitut führt seine Umfragen stets in den gleichen (wenigen) ADM-Netzen durch, für die es die Unterlagen gekauft hat.⁷ Mit anderen Worten, die Grundgesamtheit der Wahlberechtigten wird durch diejenige der erworbenen ADM-Netze ersetzt. Der „Zufallsfehler“ auf der ersten Stufe wird damit zum systematischen Fehler auf der Institutsebene. Man kann für die ADM-Gemeinde nur hoffen, daß die Wahlberechtigten jedes ADM-Netzes bei jeder Bundestagswahl auf die Wunschvorstellungen der ADM-Produzenten Rücksicht nehmen werden. Als Meinungsforschungsinstitut xy und (un)glücklicher Besitzer von einigen wenigen ADM-Netzen würde ich allerdings nach jeder Bundestagswahl meine Schäfchen zählen und feststellen, in welchem Maß jedes meiner Netze das amtliche Wahlergebnis reflektiert, was durch Addition der Wahlergebnisse der 210 Stimmbezirke festgestellt werden kann. Denn offensichtlich können meine Wahlprognosen nur das widerspiegeln, was die Wahlberechtigten meines permanent ausgesaugten ADM-Netzes an politischer Pluralität offenbaren.

Andererseits werden durch das Auswahlverfahren stets Gruppen von sieben bis zehn Wahlberechtigten ausgesucht, die im gleichen Stimmbezirk und folglich nahe beieinander wohnen und damit nicht selten eine Präferenz für die gleiche Partei aufweisen.⁸ Zum Beispiel gibt es Stimmbezirke mit hohen CDU/CSU-Anteilen oder hohen SPD-Anteilen. Das führt zu unkontrollierbaren statistischen Abhängigkeiten mit der Folge, daß der repräsentative Querschnitt nicht die Information von 1000 unabhängigen Wahlberechtigten enthält, sondern von erheblich weniger.

Das ADM-Verfahren stellt einen Kompromiß zwischen Kostenaufwand und Informationsverlust dar. Die Nichtwiederholung der Auswahl auf der ersten Stufe bedeutet eine gigantische Kosteneinsparung, das Verfahren wäre ohne diese Vereinfachung nicht praktikabel. Dafür geht man das Risiko der Verewigung eines systematischen Fehlers ein. Die Reisekosten der Interviewer werden durch die Beschränkung auf einen Stimmbezirk reduziert, aber möglicherweise ebenso die gewonnenen Informationen. Was nutzen 7-10 Interviews in einem Stimmbezirk, in welchem eine Partei dominiert? De facto bedeutet dies eine Reduktion des Querschnittumfangs – und die Kostenersparnis erweist sich als Bumerang.⁹

⁷ Im Jahre 1982 wurden insgesamt 96 (!) verschiedene ADM-Netze von je 210 Stimmbezirken mit Hilfe einer sogenannten systematischen Zufallsauswahl "Für Alle Ewigkeit" kreiert. Wenn der Zufall zu seinem Recht gekommen ist, dann werden einige dieser ADM-Netze nicht ganz so repräsentativ sein, wie sich das die ADM-Produzenten gedacht haben. Aber wie es mit den Launen des Zufalls nun einmal ist, man weiß nicht, welche ADM-Netze nicht ganz stubenrein sind. Diese 96 ADM-Netze wurden insgesamt 12 Markt- und Meinungsforschungsinstituten in unterschiedlicher Anzahl - je nach Anforderung und gegen Entgelt - zugeteilt, wobei jedes ADM-Netz an höchstens ein Institut abgegeben wurde. Größter Abnehmer war der Umsatzriese der Prognoseindustrie, Infratest, mit 16 Netzen (vgl. z.B. Ute Hanefeld, Das Sozio-ökonomische Panel, S.146-151/S.171-175, Campus-Verlag 1987).

⁸ Es wäre empirisch - d.h. anhand früherer Bundestagswahlen - zu untersuchen, welche Auswirkungen sich dadurch ergeben.

⁹ Technisch ausgedrückt: Das ADM-Verfahren beabsichtigt zwar (tut es aber nicht), jedem Wahlberechtigten die gleiche Chance zu geben, es beabsichtigt aber in keiner Weise, 1000 statistisch unabhängige Wahlberechtigte auszuwählen. Sowohl Chancengleichheit, als auch statistische Unabhängigkeit sind aber unerläßliche Kriterien. Wie bereits erwähnt, wird durch das ADM-Verfahren die Chancengleichheit nicht gewährleistet. (Ein

In der Praxis wird hauptsächlich das kostengünstigere Quotenverfahren verwendet. Dieses ist viel älter als das ADM-Verfahren. Das ursprüngliche Konzept des Quotenverfahrens bestand darin, im „repräsentativen Querschnitt“ ein Miniaturbild der Population bezüglich einiger ausgewählter (=quotierter) Merkmale zu realisieren. Das Vorgehen war rein empirisch: Funktionierte etwas nicht – d.h. erhielt man schlechte Resultate –, dann wurde das Verfahren modifiziert, bis es wieder „brauchbare“ Resultate lieferte. Für eine theoretische Begründung schien wenig Bedarf bzw. die gegebenen Erklärungen betreffend Miniaturbild erwiesen sich als unhaltbar. Aus diesem Grund wird heute auch von den Verfechtern des Quotenverfahrens der Schwerpunkt auf die „Zufallsauswahl“ gelegt:

„Gleiche Chance für jeden“

herzustellen, sei der eigentliche Sinn und Zweck des Quotenverfahrens, schreibt Frau Noelle-Neumann auf Seite 135 ihres Buches „Umfragen in der Massengesellschaft“. Diese Bedingung herzustellen sei nicht einfach. Sie versucht, dies mit Quotenanweisungen betreffend der folgenden Merkmale zu erreichen: Wohnortgröße (4 Klassen), Geschlecht, Alter (5 Klassen), Berufstätigkeit (8 Klassen) und nicht Berufstätige (7 Klassen)¹⁰. Dazu führt sie auf den Seiten 132 bis 135 aus:

„Da der modellgerechte Miniaturschnitt die gleichen Proportionen besitzen soll, arbeitet man für jeden Interviewer, der an der Umfrage beteiligt ist, eine „Quote“ aus: wieviel Männer und Frauen er befragen soll, wieviel Angehörige der verschiedenen Altersgruppen, wieviel Berufstätige,

Auswahlverfahren auf einer Grundgesamtheit heißt bekanntlich "Realisierung der Gleichverteilung oder Chancengleichheit", wenn bei einer großen Anzahl von Wiederholungen des Verfahrens jedes Element der Grundgesamtheit mit ungefähr der gleichen relativen Häufigkeit ausgewählt wird.) Da bei der Erstellung eines repräsentativen Querschnittes der Auswahlvorgang auf der ersten Stufe - d.h. die Selektion der 210 Stimmbezirke - nicht wie derholt wird, findet eine "Zufallsauswahl" allenfalls innerhalb der Subpopulation statt, die durch die fest gewählten 210 Stimmbezirke des jeweiligen ADM-Netzes definiert wird. Ist diese Subpopulation repräsentativ für die ganze Population (im Sinne eines Miniaturbildes bezüglich der zu untersuchenden Merkmale), dann reicht diese eingeschränkte Zufallsauswahl aus. Es spielt dann keine Rolle, daß die Elemente außerhalb der Subpopulation keine Chance haben, in den repräsentativen Querschnitt zu gelangen.

Damit stellt sich die Frage, wie man überprüfen kann, ob diese Subpopulation ein Miniaturbild der ganzen Population darstellt. Ein ähnliches Problem stellt sich bei dem im folgenden beschriebenen Quotenverfahren betreffend der Subpopulation des "Bekanntenkreises" der eingesetzten Interviewertruppe. Die dort skizzierte "Lösung" ist auch hier anwendbar, so weit man überhaupt von einer Lösung sprechen kann. Meines Erachtens gibt es keinen anderen Ausweg. (Es sei denn, man wiederholt die Selektion der 210 Stimmbezirke bei jeder Erstellung des repräsentativen Querschnittes, was aus Kosten- und Praktikabilitätsgründen undenkbar ist.) Eine ausführlichere Diskussion findet man auf den Seiten III/12-15.

¹⁰ Die Anzahl der theoretisch möglichen Klassen beträgt demnach $4 \cdot 2 \cdot 5 \cdot 8 + 4 \cdot 2 \cdot 5 \cdot 7 = 600$ von denen einige praktisch nicht vorkommen können oder keine Rolle spielen. Was immer auch die exakte Anzahl der relevanten Kombinationen ist, so ist sie offensichtlich groß genug, um die Erstellung eines Miniaturbildes vom Umfang 2.000 oder 1.000 - von 370 wie bei der Saarwahl ganz zu schweigen - unmöglich zu machen, welches die relevanten Kombinationen in den richtigen Proportionen reflektiert (Wie groß diese Anteile wirklich sind, läßt sich nicht aus der amtlichen Statistik entnehmen). Frau Noelle-Neumann und andere Verfechter des Quotenverfahrens "helfen" sich dadurch, daß sie nur die Anteile innerhalb jedes der vier Merkmale (Wohnortgröße, Geschlecht, Alter, Berufstätigkeit) berücksichtigen und die Kombinationen ignorieren.

wieviel Personen aus den verschiedenen Berufsgruppen und wieviel nichtberufstätige Angehörige in den verschiedenen Berufskreisen. Werden diese ausgearbeiteten Quoten zusammengesetzt, so ergibt sich in allen diesen statistischen Proportionen das Bild der Grundgesamtheit. Die richtige Verteilung auf Ortsgröße, Länder oder Regierungsbezirke stellt sich in der Regel aus der Zahl der Aufträge an die Interviewer in den verschiedenen Gebieten und in den verschiedenen Ortsgrößen her.“

.
. .

„Die wirkliche Funktion der Quoten ist: sie sollen den Interviewer zu einer Zufallsauswahl veranlassen, bei der jedes Mitglied der Grundgesamtheit praktisch die gleiche Chance hat, in die Stichprobe zu gelangen; ohne das Wirken eines solchen Zufallsmechanismus ist die Bildung einer repräsentativen Stichprobe nicht möglich. Erst in zweiter Linie wirken die Quoten außerdem wie eine „Schichtung“ bei Random-Stichproben, indem sie in bestimmten Merkmalen die Übereinstimmung zwischen Stichprobe und Universum sichern.“

Dem Quotenverfahren wird häufig vorgeworfen, daß es keine Zufallsauswahl – d.h. gleiche Chance für jeden – garantiere¹¹. Dies ist zweifellos richtig – es ist leicht zu zeigen, daß das Quotenverfahren die Chancengleichheit nicht gewährleistet – aber das ist für das Funktionieren des Verfahrens nicht notwendig¹². Auf dem Papier erscheint das ADM-Verfahren als halbwegs praktikable Notlösung für den idealen Würfel, während das Quotenverfahren den Verdacht

¹¹ Vergleiche auch M. Kaase in Rheinischer Merkur (Extra) Nr.37 1987:

"Der entscheidende Unterschied zwischen sogenannten Quotenstichproben, bei denen der Interviewer nach Maßgabe bestimmter Quoten-Vorgaben (zum Beispiel ein 30-40jähriger männlicher Arbeiter in einer Großstadt) die konkrete Befragungsperson selber aussucht, und Zufallsstichproben, bei denen der Interviewer einen bestimmten Haushalt beziehungsweise eine bestimmte Person vorgegeben bekommt, die er dann befragen muß, besteht nur darin, daß nur bei Zufallsstichproben der Stichprobenfehler rechnerisch näher bestimmt werden kann."

In der demoskopischen Praxis ist die rechnerische Bestimmung des Stichprobenfehlers kaum mehr als ein frommer Spruch. Falls überhaupt ein Versuch in dieser Richtung gemacht wird, so handelt es sich meist um eine statistische Alibiübung oder Milchmädchenrechnung.

¹² Es reicht nämlich aus, daß durch die ausgewählten Interviewer und ihren "Bekanntenkreis" eine Subpopulation definiert wird, die "repräsentativ" für die zu untersuchenden Merkmale ist. Dies könnte unter Umständen dann eintreten, wenn die Auswahl der Interviewer genügend "zufällig" und ausreichend regional gestreut erfolgt und dem einzelnen Interviewer möglichst wenige Interviews zugeteilt werden. Durch eine geschickte Quotenvorschrift kann dann möglicherweise eine "Zufallsauswahl" innerhalb dieser Subpopulation erreicht werden. Auf diese Weise könnte erklärt werden, wie durch die "Verkettung von unglücklichen Umständen" etwas Brauchbares entstehen kann.

Eine Verifikation dieser spekulativen Hypothese kann nur auf der Basis von umfangreichem Datenmaterial erfolgen. Man kann bekanntlich nicht beweisen, daß ein Verfahren zu einer Zufallsauswahl führt, sondern lediglich mit Hilfe von Versuchsreihen gewisse Regelmäßigkeiten feststellen, und dann zeigen, daß die aufgetretenen Abweichungen (von den Regelmäßigkeiten) sich im Rahmen von Zufallsabweichungen bewegen. Unter diesem Gesichtspunkt sind die folgenden Ausführungen von Frau Noelle-Neumann in ihrem Buch "Umfragen in der Massengesellschaft" auf den Seiten 136 und 137 zu sehen:

"Ohne daß die Quotenanweisung an die Interviewer etwas über den Familienstand oder die Konfession oder die Größe des Haushalts, in dem die Befragten leben, aussagen, stellen sich die Proportionen, die die amtliche Statistik in diesen Merkmalen für die Bevölkerung ausweist, auch in der Quotenstichprobe her. Das gleiche gilt für den ebenfalls an Hand amtlicher Daten leicht zu kontrollierenden Flüchtlingsanteil. Zur Überprüfung der Methode werden in diesen Interviews laufend wechselnde Fragen nach Sachverhalten eingeschlossen, die sich an externen Daten kontrollieren lassen."

(Man vergleiche dazu auch ihre Fußnoten Nr.30,35,43 auf den Seiten 134-148). Diese Vorgehensweise scheint mir sehr vernünftig. Ich würde jedenfalls ähnlich wie sie vorgehen.

erweckt, aus Kostengründen würde eine Kokosnuß als Würfel verkleidet. Entscheidend ist allerdings weder der erste Eindruck noch die Papierform, sondern allein die gemachten Erfahrungen. Dabei könnte sich herausstellen, daß die Kokosnuß nicht schlechter ist als ein „idealer“ Würfel aus dem ADM-Kaufhaus. Die Frage nach dem Unterschied bzw. der statistischen Qualität von ADM- und Quotenverfahren kann letztlich nur durch lange Versuchsreihen unter gleichbleibenden Bedingungen beantwortet werden. Doch wer soll das bezahlen, wer hat soviel Geld?

Sehen wir von all diesen Problemen ab und unterstellen ideale repräsentative Querschnitte, d.h. es wird die (unrealistische) Annahme gemacht, daß erstens 1000 Wahlberechtigte nach einem perfekten Zufallsverfahren ausgelost werden und zweitens bei der Befragung keine Fehler entstehen (als würde der Interviewer von den Befragten die ausgefüllten und gültigen Stimmzettel erhalten, sofern sie am Wahltag einen solchen in die Urne legen). Die zentrale Frage lautet dann: Wie verlässlich sind Aussagen über eine 45-Millionen-Gruppe, die auf diese Weise gewonnen werden, wie wird sich das Resultat vom amtlichen Wahlergebnis unterscheiden? Mit welchen Fehlern ist zu rechnen?

Das ist eine rein mathematische Frage. Die aus der Befragung resultierenden Abweichungen – d.h. die Interviewfehler – kommen bei einer Wahlprognose dann additiv hinzu.

Gehen wir, um die Problemebenen klar zu trennen, von einem fiktiven, aber objektiven Sachverhalt aus: Wählen wir von rund 40 Millionen Wahlberechtigten, die am Wahltag eine gültige Stimme abgegeben haben, per Lotterie 1000 aus und ermitteln die Parteistärken in diesem „repräsentativen Querschnitt“. Wir können uns diesen Sachverhalt auch so vorstellen, daß die abgegebenen rund 40 Millionen gültigen Stimmzettel in einer riesigen Trommel gut vermischt werden und dann daraus 1000 mit einem perfekten Zufallsverfahren gezogen werden.

Man muß nun zunächst einmal sämtliche Möglichkeiten überdenken, wie man aus 40 Millionen Stimmzetteln 1000 auswählen kann. Die Zahl der möglichen Querschnitte ist zwar unvorstellbar groß, aber relativ einfach zu berechnen. Es ist eine Zahl, die mit der Ziffer 2 beginnt und 5034 Stellen hat.

Jeder dieser Querschnitte ist statistisch gleichberechtigt. Unter ihnen findet man exotische wie 1000 Stimmen für die FDP und 0 Stimmen für alle anderen Parteien. Doch solch extreme repräsentative Querschnitte kommen nur selten vor. Die meisten liegen näher am „wahren“ Stimmresultat. Wie nahe ist das aber? Oder präziser: Welcher Bruchteil kommt der wahren Stimmverteilung wie nahe?

Was ist ein Miniaturbild? Was ist Zufallsauswahl?

Was ist ein Miniaturbild?

Die Vorstellung eines Miniaturbildes beinhaltet den Wunsch, daß es sich um eine vergleichsweise kleine Teilpopulation der Grundgesamtheit handeln soll, die deren wesentliche Eigenschaften widerspiegelt. Aber mit diesem unpräzisen Wunsch kommt man nicht weiter. Es muß zuerst spezifiziert werden, was die „wesentlichen Merkmale“ sind. Insbesondere kann es sich nicht um Merkmale handeln, die man irgendwann einmal benötigt, aber jetzt noch nicht kennt.

Definition: Sei G eine Population und S eine Teilpopulation. In der Population seien eine Anzahl Merkmale spezifiziert, die sich als Teilmengen der Population beschreiben lassen. Diese Teilmengen werden mit M_1, M_2, \dots, M_n und ihre Anteile in der Population mit p_1, p_2, \dots, p_n bezeichnet.

Eine Teilpopulation S von G heißt Miniaturbild von G bezüglich der Merkmale – oder repräsentativ bezüglich M_1, M_2, \dots, M_n – falls diese in der Teilpopulation ebenfalls mit den Anteilen p_1, p_2, \dots, p_n auftreten.

Wenn man heute feststellt, daß eine Teilpopulation S ein Miniaturbild von G bezüglich bestimmter Eigenschaften ist, dann besteht a priori kein Grund dafür, daß dies in Zukunft auch noch der Fall sein wird oder daß S ein Miniaturbild von G bezüglich weiterer Merkmale ist.

Im allgemeinen wird es ein hoffnungsloses Unterfangen sein, festzustellen, ob eine Teilpopulation S einer Population G ein Miniaturbild bezüglich bestimmter Merkmale ist, es sei denn, man zählt die entsprechenden Teilmengen der Teilpopulation.

Man neigt gelegentlich zur Annahme, daß dies der Fall sein wird, wenn S groß genug ist. Das braucht aber nicht der Fall zu sein. Weder Bayern noch Nordrhein-Westfalen sind Miniaturbilder der Bundesrepublik bezüglich aller demoskopischen Merkmale. Für einige Merkmale besteht Repräsentativität, für andere nicht.

Was ist eine Zufallsauswahl in einer Population G ?

Es besteht bei vielen Anwendern und auch bei Statistikern der Glaube, eine Zufallsauswahl sei eine Teilpopulation, die man mit einem geeigneten Auswahlverfahren aus einer Population gewonnen habe. Diese Vorstellung ist jedoch falsch. Eine Zufallsauswahl ist im Gegensatz zu einem Miniaturbild keine Teilpopulation!

Eine Zufallsauswahl ist ein Auswahlverfahren in einer Population, welches im Prinzip eine Lotterie imitiert.

Wieso kann man nun nicht durch n -maliges Anwenden des Auswahlverfahrens eine Teilmenge der Population erhalten, die man als Zufallsauswahl betrachten kann? Das kann man schon tun, aber die Wahrscheinlichkeitsaussagen der Statistik beziehen sich auf das Auswahlverfahren und nicht auf die ausgewählte Teilmenge.

Es ist genauso sinnlos, von einer Teilmenge der Population zu sagen, sie sei eine Zufallsauswahl, wie von einer einmal gewürfelten Drei zu behaupten, sie sei eine Zufallszahl.

Es gibt viele Auswahlverfahren, jeder trifft täglich Entscheidungen und „wählt damit aus“. In der Umgangssprache wird der Begriff des Zufalls häufig mit „unerwartet“ oder „willkürlich“ in Zusammenhang gebracht. Aber dies hat nichts mit Zufall im statistischen Sinne zu tun. Obwohl die Idee der Lotterie seit Jahrhunderten bekannt ist – schon im alten Ägypten gab es Würfel (auch gefälschte) – wurde der Begriff der Zufallsauswahl erst zu Beginn dieses Jahrhunderts exakt definiert. Die Menschheit hat Statistik und Wahrscheinlichkeitsrechnung seit über drei Jahrhunderten mit unterschiedlichem Erfolg angewendet, ohne über eine präzise Definition des Begriffes „Zufallsauswahl“ zu verfügen. Viele falsche Anwendungen und Mißbräuche von Statistik – in diesem, im letzten, vorletzten und nächsten Jahrhundert – beruhen darauf, daß das Auswahlverfahren in der Luft schwebt.

Die Zufallsauswahl wird durch zwei Eigenschaften definiert, nämlich durch die der Chancengleichheit (=Gleichverteilung) und die der Regellosigkeit (=statistische Unabhängigkeit).

I) Die Chancengleichheit ist leicht zu verstehen: Wird das Auswahlverfahren in einer Population mit N Elementen genügend oft wiederholt, dann muß jedes Element mit ungefähr der relativen Häufigkeit $1/N$ ausgewählt werden. Die zugrunde liegende Vorstellung basiert auf dem „idealen Würfel“, von dem man erwartet, daß jede Seite mit einer relativen Häufigkeit von $1/6$ gewürfelt wird, sofern man genügend Ausdauer zeigt....

II) Die Eigenschaft der Regellosigkeit für Stichproben vom Umfang n kann man sich zwar bildlich noch relativ einfach vorstellen, die Definition ist jedoch nicht leicht zu verstehen: Jede Gruppe von n Elementen in der Population hat die gleiche Chance, ausgewählt zu werden, wenn das Auswahlverfahren n -mal angewendet wird, d.h. bei häufiger n -maliger Auslosung muß jede Gruppe von n Elementen mit einer relativen Häufigkeit von ungefähr $1/\binom{N}{n}$ ausgewählt werden.

Die Regellosigkeit der Auswahlverfahren beinhaltet unter anderem, daß ein einmal selektiertes Element der Population keinen Effekt auf das Auswahlverfahren haben darf, insbesondere darf dadurch nicht die Auswahl des nächsten, übernächsten ... Elementes beeinflußt werden.

Auch wenn man beim Würfeln schon zehnmal eine sechs gewürfelt hat, dann wird dadurch die Chance für eine erneute Sechs nicht kleiner und die Chancen für die anderen Seiten nicht größer. Diese Vorschrift geht gegen das (naive) Gefühl. Man neigt zum Glauben, die Chance für eine Sechs sei größer, wenn man lange keine Sechs mehr gewürfelt hat. Umgekehrt erwartet man nach zehnmalem Würfeln einer Sechs, daß beim nächsten Wurf bestimmt keine Sechs mehr auftreten wird, als sei eine ausgleichende Gerechtigkeit im Spiel.

Nach den Vorstellungen und Wünschen der Meinungsforscher soll die *Zufallsauswahl* – alias ADM & Quote – dafür geradestehen, daß die repräsentativen Querschnitte *Miniaturbilder* sind. Aber bezüglich welcher Merkmale? Natürlich für alle, die den Demoskopien je interessieren mögen! *Aber das hat die Statistik (bzw. das Gesetz der Großen Zahlen) der Demoskopie nie in Aussicht gestellt.* Die üblichen Aussagen beziehen sich auf ein einziges „Ja/Nein“-Merkmal. Dies sei an einem Beispiel erläutert: Angenommen, 52% der Wahlberechtigten seien Frauen. Dann werden bei häufiger Auslosung 95% der Tausender-Querschnitte einen Frauen-Anteil von 49% bis 55% aufweisen, in 5% der Fälle wird sich also ein Frauen-Anteil von über 55% oder unter

49% ergeben. Betrachten wir nun ein weiteres „Ja/Nein“-Merkmal, die Konfession. (Zur Vereinfachung wird angenommen, es gäbe nur Katholiken und Protestanten.) Liegt der Anteil der Katholiken bei 45%, so läßt sich eine analoge Aussage machen. Bei häufiger Auslosung der repräsentativen Querschnitte wird in 95% der Fälle der Katholiken-Anteil zwischen 42% und 48% liegen. Aber daraus kann man nicht schließen, daß bei häufiger Auslosung des repräsentativen Querschnittes in 95% aller Fälle ein Frauen-Anteil zwischen 49% und 55% und ein Katholiken-Anteil zwischen 42% und 48% eintreten wird. Das wird nur in etwa 90% aller Fälle eintreten.

Betrachtet man nun zehn „Ja/Nein“-Merkmale dieser Art in einem repräsentativen Querschnitt, dann wird sich obige Zahl auf 60% reduzieren und bei fünfzig „Ja/Nein“-Merkmalen sogar auf 8%. Mit anderen Worten, bei häufiger Auslosung des repräsentativen Querschnittes wird die gleichzeitige Einhaltung der 95%-Intervalle für die einzelnen Merkmale sehr unwahrscheinlich. Will man eine Sicherheitswahrscheinlichkeit von 95% für zehn bzw. fünfzig bzw. hundert „Ja/Nein“-Merkmale aufrechterhalten, so muß man die Spielräume um über 40% bzw. um fast 70% bzw. um fast 80% vergrößern. Bei einer Sicherheitswahrscheinlichkeit von 90% beträgt die notwendige Vergrößerung der Spielräume über 55% bzw. über 85% bzw. 100%.

Eine analoge Reduktion der Zuverlässigkeit des repräsentativen Querschnittes tritt ein, wenn ein Merkmal nicht zwei, sondern mehrere Ausprägungen hat (wie z.B. verschiedene Parteien, Altersklassen u.s.w.). In den Abschnitten VI bis IX wird darauf näher eingegangen. Zur Aufrechterhaltung einer Sicherheitswahrscheinlichkeit von 95% bei einem Merkmal mit 10 bzw. 50 Ausprägungen sind die Spielräume um über 40% bzw. um fast 70% - im Vergleich zu einem „Ja/Nein“-Merkmal - zu vergrößern. Bei vielen Merkmalen dieser Art ist eine entsprechende Vergrößerung der Spielräume erforderlich.

Die Moral von der Geschichte ist: Je mehr Merkmale in einer „Repräsentativumfrage“ untersucht werden und je mehr Ausprägungen die einzelnen Merkmale haben, desto mehr weicht der repräsentative Querschnitt von einem Miniaturbild ab. Man kann also den repräsentativen Querschnitt nicht ungestraft ausquetschen. Je mehr man aus ihm herauszuholen versucht, desto größer wird seine statistische Unzuverlässigkeit. In einer späteren Arbeit wird detailliert auf dieses Phänomen eingegangen.

Die direkte Überprüfung, ob ein Auswahlverfahren die Eigenschaften der Chancengleichheit und der Regellosigkeit aufweise, ist praktisch meist unmöglich. In der Demoskopie ist dies wegen der Größe der Population a priori unmöglich. Der Nachweis, daß ein Auswahlverfahren die Chancengleichheit und die Regellosigkeit verletzt, ist manchmal einfacher.

Die Schöpfer und Verfechter des ADM-Verfahrens und des Quotenverfahrens behaupten von ihren Auswahlverfahren, daß diese den Kriterien der Zufallsauswahl genügen würden. Das bedeutet, daß bei genügend häufiger Wiederholung bei beiden Verfahren in einer Population von ca. 45.000.000 jedes Element mit ungefähr der relativen Häufigkeit $1/45.000.000$ in einen repräsentativen Querschnitt gelangen müßte und außerdem jede 1000er-Gruppe gleich oft als repräsentativer Querschnitt ausgewählt werden müßte.

Was das ADM-Verfahren betrifft, so sieht es in der Praxis so aus, daß ein Meinungsforschungsinstitut ausschließlich in den (wenigen) von ihm erworbenen ADM-Netzen seine Befragungen durchführt, wobei aus Kostengründen zunächst einmal stets dasselbe Netz abgegrast wird. Dadurch wird eine Teilpopulation von durchschnittlich 190.000

Wahlberechtigten festgelegt, wobei eine beträchtliche Schwankungsbreite besteht. Bei Wiederholung des Auswahlverfahrens wird der repräsentative Querschnitt stets aus Mitgliedern *dieser* Teilpopulation zusammengestellt. Alle Wahlberechtigten *außerhalb* dieser Teilpopulation haben die Chance Null, in den repräsentativen Querschnitt zu gelangen.

Bei Quotenverfahren ist die Situation ähnlich: Die Sample-Points, das heißt die Orte der Befragung, werden einmal festgelegt und werden nicht *jedesmal* bei der Erstellung eines repräsentativen Querschnittes *neu* ausgewählt. Die Interviewer werden in den Sample-Points ausgewählt und führen die Interviews an ihrem Wohnort oder in der Umgebung durch (zwecks Kostenersparnis). Durch die Auswahl der Sample-Points und der Interviewer wird ein Wirkungskreis festgelegt, der zu einer Teilpopulation führt, die man – überspitzt formuliert – als „Bekanntenkreis“ der eingesetzten Interviewer bezeichnen kann. Bei Erstellung des repräsentativen Querschnittes können daher zwangsläufig nur Mitglieder dieser Teilpopulation selektiert werden. Wahlberechtigte *außerhalb dieser* Teilpopulation haben also die Chance Null, in den repräsentativen Querschnitt zu gelangen.

In beiden Fällen findet also – wenn überhaupt – eine „Lotterie“, das heißt eine Zufallsauswahl, lediglich innerhalb der durch das feste ADM-Netz bzw. den festen Interviewer-Wirkungskreis definierten Teilpopulation statt, nicht aber in der betrachteten gesamten Population. Die Substitution der Gesamtheit durch diese Teilpopulation ist aus praktischen (und finanziellen) Gründen eine Notwendigkeit. Historisch gesehen sind Meinungsumfragen stets auf diese Art durchgeführt worden.

Ob die Teilpopulation ein Miniaturbild für die gesamte Population bezüglich der in der Umfrage behandelten Merkmale darstellen, ob also die gefundenen Ergebnisse auf die gesamte Population übertragen werden können, kann nicht bewiesen werden. Als indirekter „Beweis“ sollte allerdings zumindest versucht werden, mit einer *ausreichenden* Anzahl *geeigneter* Kontrollmerkmale, deren Verteilung in der Gesamtpopulation bekannt ist, die „Miniaturbild“-Eigenschaft der Teilpopulation zu überprüfen.

Was die Chancengleichheit innerhalb der Teilpopulation betrifft, so scheint diese – jedenfalls auf dem Papier – beim ADM-Verfahren gewährleistet zu sein. Entscheidend sind die Details der Auswahlverfahren in den einzelnen Stimmbezirken (Begehungspläne). Diese werden durch das ADM-Verfahren nicht festgelegt und sind den Instituten überlassen.

Beim Quotenverfahren soll die Chancengleichheit innerhalb der Teilpopulation durch die Quotenvorschriften erreicht werden. Auf den ersten Blick scheint dies hochgradig suspekt. Falls die Ausführungen von Frau Noelle-Neumann betreffend den Kontrollmerkmalen zutreffend sind (Umfragen in der Massengesellschaft, S.136/137,146; Fußnote 30/43), so dürfte eine abgeschwächte Form der Chancengleichheit vorliegen, die statistisch ausreichend erscheint. (Auf die Details kann hier nicht eingegangen werden.) Wesentlich erscheint mir, daß die Anzahl der Befragungen pro Interviewer möglichst klein ist.

Wie sieht es nun in diesen Teilpopulationen mit der Regellosigkeit der Auswahlverfahren aus? Beeinflußt ein einmal ausgewähltes Element der Teilpopulation die Auswahl des nächsten, des übernächsten oder irgendeines anderen Elementes?

Hat wirklich jeder denkbare 1000er-Querschnitt innerhalb eines ADM-Netzes die gleiche

Chance, ausgewählt zu werden? Bei dem ADM-Verfahren werden in jedem (!) Stimmbezirk 7-10 Bürger befragt. Alle 1000er-Querschnitte, die sich nicht so zusammensetzen, haben also keine Chance, als repräsentativer Querschnitt ausgewählt zu werden. Dadurch wird die Regellosigkeit verletzt. Durch die Begehungspläne für die Stimmbezirke sind die Interviews statistisch abhängig. Praktikabilität und Kostenersparnis bewirken de facto fast immer eine Verletzung der Regellosigkeit.

Das Quotenverfahren erfüllt die Eigenschaft der Regellosigkeit ebenfalls nicht. De facto werden meist nur Bürger aus dem Wirkungskreis (=Bekanntenkreis) der Interviewer befragt. Durch eine möglichst kleine Anzahl von Befragungen pro Interviewer kann dieser Effekt reduziert werden.

Bei beiden Verfahren hat dies zur Folge, daß ein repräsentativer Querschnitt vom Umfang 1000 nicht wirklich die Information von 1000 unabhängigen Wahlberechtigten liefert, sondern von weniger. Je stärker die Regellosigkeit verletzt wird, desto größer wird der Informationsverlust. Daraus folgt, daß der auslosungsbedingte Spielraum größer wird.

Simulation der Auslosung des repräsentativen Querschnittes auf einem Computer

Die Simulation wird anhand eines Beispiels – der Bundestagswahl 1987 – erläutert.

An der Bundestagswahl nahmen 38.225.294 Wahlberechtigte teil, die 37.867.319 gültige Zweitstimmen abgegeben haben, davon

16.761.572	(= 44,3 %)	für die CDU/CSU,
14.025.763	(= 37,0 %)	für die SPD,
3.440.911	(= 9,1 %)	für die FDP,
3.126.256	(= 8,3 %)	für die Grünen und
512.817	(= 1,3 %)	für die Splitterparteien.

Zur Vereinfachung der Ausdrucksweise wird im folgenden der Term „*Wahlberechtigte, die bei der Wahl eine gültige Zweitstimme abgegeben haben*“ abgekürzt durch „*Wahlteilnehmer*“.

Der Zweck der Simulation einer Auslosung von 1000 Wahlteilnehmern besteht nicht darin, eine Schätzung für das Resultat der Bundestagswahl zu bekommen – diese wird ja als bekannt vorausgesetzt –, sondern eine Schätzung über die möglichen Abweichungen zu bekommen, die durch die Auslosung des repräsentativen Querschnittes verursacht werden. Im Prinzip könnte die Auslosung durch die Landeslotterie erfolgen, indem jeder Wahlteilnehmer genau ein Los zugeteilt erhält. Von den 37.867.319 Teilnehmern werden dann 1000 „Gewinner“ ausgelost und deren Stimmzettel ausgewertet. Praktisch ist dies nicht durchführbar, da alle Wahlteilnehmer erfaßt werden müßten.

Es ist aber erstaunlich einfach, den Auslosungseffekt auf einem Computer zu simulieren und dadurch Aufschluß über die möglichen Abweichungen eines „repräsentativen Querschnittes“ zu erhalten. Hierfür denken wir uns die Wahlteilnehmer von 1 bis 37.867.319 numeriert. Zuerst werden alle CDU/CSU-Wähler in irgendeiner Reihenfolge numeriert, anschließend alle SPD-

Wähler, dann die FDP-Wähler, dann alle Grün-Wähler und zuletzt alle Wähler von Splitterparteien. Der erste CDU/CSU-Wähler erhält die Nummer 1 und der letzte die Nummer 16.761.572. Somit erhält der erste SPD-Wähler die Nummer 16.761.573 und der letzte die Nummer 30.787.335. Entsprechend wird der erste FDP-Wähler mit der Nummer 30.787.336 versehen und der letzte mit der Nummer 34.228.246. Der erste Grün-Wähler erhält demnach die Nummer 34.228.247 und der letzte die Nummer 37.354.502. Schließlich erhält der erste Splitterparteiwähler die Nummer 37.354.503 und der letzte die Nummer 37.867.319. Dies ist natürlich die Anzahl aller Wahlteilnehmer.

In einem Computer ist ein sogenannter Zufallszahlengenerator eingebaut, der dazu verwendet werden kann, Nummern zwischen 1 und 37.867.319 auszulosen. Dieser Zufallszahlengenerator wird nun tausendmal hintereinander aktiviert und so erhält man die Nummern von 1000 Wahlteilnehmern, die als repräsentativer Querschnitt deklariert werden. Aufgrund der Anordnung der Parteien in der Numerierung – zuerst die Nummern der CDU/CSU-Wähler, dann diejenigen der SPD-Wähler usw. – ist es nun leicht, bereits anhand der ausgelosten Nummer zu bestimmen, für welche Partei der entsprechende Wahlteilnehmer gestimmt hat. Liegt die ausgeloste Nummer zwischen 1 und 16.761.572, dann hat der Wahlteilnehmer für die CDU/CSU gestimmt. Liegt die Nummer zwischen 16.761.573 und 30.787.335, dann hat der ausgeloste Wahlteilnehmer für die SPD gestimmt. Liegt die Nummer zwischen 30.787.336 und 34.228.246, dann hat der ausgeloste Wahlteilnehmer für die FDP gestimmt. Liegt die Nummer zwischen 34.228.247 und 37.354.502, dann hat der Wahlteilnehmer für die Grünen gestimmt. Liegt die Nummer zwischen 37.354.503 und 37.867.319, dann hat der Wahlteilnehmer für eine Splitterpartei gestimmt. Auf diese Art läßt sich leicht das Resultat für den ausgelosten repräsentativen Querschnitt bestimmen, und – was wichtiger ist – die Abweichungen vom amtlichen Resultat. Wie man sieht, ist es für die Ermittlung des Resultats der „Repräsentativumfrage“ und der Abweichungen gar nicht notwendig zu wissen, wie die Wahlteilnehmer heißen, die hinter den ausgelosten Nummern stehen.

Auf einem Homecomputer – z.B. einem SCHNEIDER CPC 464, der heute für weniger als 400 DM zu haben ist – dauert die Auslosung eines repräsentativen Querschnittes vom Umfang 1000 bei vier Parteien plus Splitterparteien inklusive der Auswertung der ausgelosten Stimmzettel weniger als 10 Sekunden, auf einem Großcomputer etwa 10 Millisekunden. Man kann daher leicht hunderttausend oder gar eine Million repräsentativer Querschnitte auslosen und die Abweichungen der Resultate vom tatsächlichen Wahlergebnis untersuchen.

Die im Vorwort I/Seite 9 angegebenen 10 Auslosungen des repräsentativen Querschnitts (und ebenso die 100 in Abschnitt VII/Seite 2) kann jeder Leser mit einem Homecomputer leicht selbst durchführen. Hierzu ist weiter unten ein Programm in BASIC angegeben. Je nach Zufallszahlengenerator bzw. Startwert desselben werden etwas andere Resultate herauskommen. Bei 100 Auslosungen (Dauer ca. 14 Minuten) werden die Angaben über auslosungsbedingte Spielräume der Tabelle in Abschnitt VII/Seite 9 größenordnungsmäßig erkennbar, bei 1000 Auslosungen (Dauer ca. 140 Minuten) schon relativ genau. Insbesondere wird klar, daß in über fünfzig Prozent der Auslosungen die Resultate nicht innerhalb folgender Spielräume liegen:

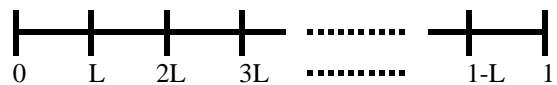
CDU/CSU	SPD	FDP	Grüne
42,3% - 46,3%	35,0% - 39,0%	8,1% - 10,1%	7,3% - 9,3%

Die Aussagen über auslosungsbedingte Abweichungen scheinen zunächst auf die Bundestagswahl 1987 beschränkt zu sein. Dies ist jedoch nicht der Fall, weil

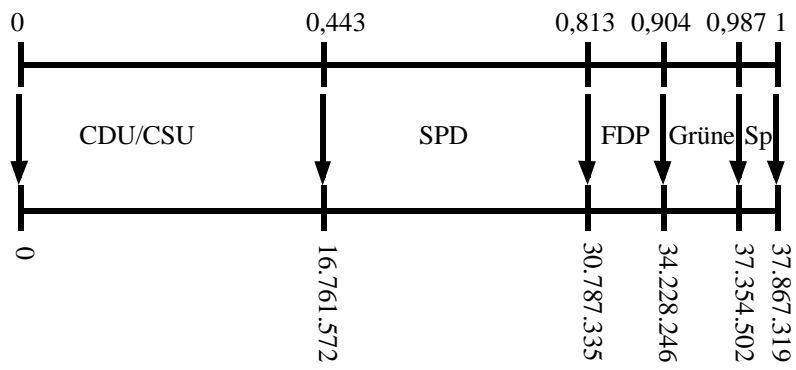
- 1) die Anzahl der Wahlteilnehmer so gut wie keine Rolle spielt, es können hunderttausend oder hundert Millionen sein.
- 2) eine genaue Kenntnis der Parteistärken nicht erforderlich ist, da die Abweichungen nicht von den Parteistärken selbst, sondern nur von deren Standardabweichungen abhängen. Es reicht deshalb aus, die Größenordnungen der Parteistärken zu kennen. Die CDU/CSU oder die SPD können zwischen 30% und 60% variieren und die FDP oder die Grünen zwischen 7% und 12%, ohne daß die simulierten Abweichungen wesentlich beeinflußt werden. Durch Standardisierung kann das Problem auf gleichstarke Parteien zurückgeführt werden.

Die Aussagen über die auslosungsbedingten (standardisierten) Abweichungen hängen jedoch wesentlich vom Umfang des repräsentativen Querschnitts sowie der Anzahl der Parteien ab. Je mehr Parteien, desto größer die Abweichungen. Je größer der Umfang des repräsentativen Querschnitts, desto kleiner die Abweichungen.

Für die effektive Durchführung der Simulation auf einem Computer wird nur die Anzahl der Parteien, ihre prozentualen Stärken und der Umfang des repräsentativen Querschnitts eingegeben, nicht aber die Anzahl der Wahlteilnehmer. Die Nummer eines Wahlteilnehmers wird durch ein kleines Intervall der Länge $L=1/37.867.319$ ersetzt, und diese werden der Reihe nach auf einer Zahlenachse mit Anfangspunkt 0 und Endpunkt 1 aufgetragen, wie dieses in folgendem Bild illustriert wird



Der im Computer eingebaute Zufallsgenerator ist ein Auswahlverfahren, welches Zahlen zwischen 0 und 1 auslost und welches die in Abschnitt III/Seite 12-13 erwähnten Kriterien der Chancengleichheit und der Regellosigkeit approximativ erfüllt. Bei jeder Aktivierung des Zufallszahlengenerators wird eine Zahl zwischen 0 und 1 ausgewählt, welche in genau eines der obigen Intervalle fällt. Dadurch wird die Nummer eines Wahlteilnehmers festgelegt, mit anderen Worten, dadurch ist ein Wahlteilnehmer ausgelost worden. Die Zuordnung Zufallszahl \rightarrow Nummer des Wahlteilnehmers wird durch folgendes Bild verdeutlicht:



Wie man sieht, entsprechen

den CDU/CSU	- Wählern die Zahlen zwischen	0	und	0,443 ,
den SPD	- Wählern die Zahlen zwischen	0,443	und	0,813 ,
den FDP	- Wählern die Zahlen zwischen	0,813	und	0,904 ,
den Grünen	- Wählern die Zahlen zwischen	0,904	und	0,987 und
den Splitter	- Wählern die Zahlen zwischen	0,987	und	1 .

Im auf der nächsten Seite abgedruckten BASIC-Programm wird die Auslosung von 1000 Wahlteilnehmern und die Auswertung ihrer Stimmzettel durchgeführt. Die Variablen CDU, SPD, FDP, Gruen und Splitter dienen als Zähler für die Stimmzettel der ausgelosten Wahlteilnehmer. Der Inhalt dieser Zählervariablen sowie ihre Abweichung vom Sollwert (= tatsächliches Resultat der Bundestagswahl 1987) wird nach 1000 Auslosungen auf den Bildschirm geschrieben. Dann beginnt eine neue Auslosung von 1000 Wahlteilnehmern und die Auswertung ihrer Stimmzettel. Wenn der Leser die Resultate und Abweichungen gesehen hat, kann er das Programm abbrechen. Der Leser kann das Programm auch verändern. Er kann andere Parteistärken eingeben und die Anzahl n der Interviews erhöhen oder erniedrigen. Die Modifikationen hierfür sind einfach. Man kann auch die Anzahl der Parteien ändern, indem zusätzliche Zählervariablen in das Programm eingebaut werden. Ferner können Wiederholungsbefragungen wie in Abschnitt VIII simuliert werden, um damit Pseudo-Trends sichtbar zu machen.

Anmerkungen zum Programm:

Einige der verwendeten Befehle sind spezifisch für den BASIC-Dialekt auf dem SCHNEIDER CPC 464, zum Beispiel TIME, RANDOMIZE, RND, der Aufbau der IF-THEN-Befehle und die Form der PRINT-Befehle.

Auf einem anderen Computer, etwa einem Commodore C64 sind diese Befehle entsprechend dem dort verwendeten BASIC-Dialekt zu verändern.

Für eventuelle Rückfragen steht Ullrich Tesche (Telefon: 0202/439-2533) zur Verfügung.

```

1  '  Ausdruck der Überschrift
2  '
3  CLS : PRINT
4  PRINT "   CDU/CSU           SPD           FDP           Grüne           Splitter"
5  '
10 RANDOMIZE(TIME)           'Setzen des Zufallszahlengenerators mit Hilfe
11                           'einer im Computer eingebauten Uhr ( TIME )
12 n=1000                    'Festlegen des Querschnittumfangs
13 CDU=0                     'Zähler der CDU/CSU – Stimmen wird 0 gesetzt
14 SPD=0                     'Zähler der   SPD   – Stimmen wird 0 gesetzt
15 FDP=0                     'Zähler der   FDP   – Stimmen wird 0 gesetzt
16 Gruen=0                   'Zähler der   Grünen – Stimmen wird 0 gesetzt
17 Splitter=0                'Zähler der   Splitter – Stimmen wird 0 gesetzt
18 '
20 FOR wtn=1 to n            'Schleife für die Auslosung des Querschnitts
21 z=RND                     'Erzeugen einer Zufallszahl
22 '
23 IF z<0,443 THEN CDU=CDU+1 : GOTO 37 'Abfrage nach CDU/CSU – Stimme und
24                                   'Erhoehen des CDU/CSU – Zählers, wenn
25                                   'z < 0,443
26 IF z<0,813 THEN SPD=SPD+1 : GOTO 37 'Abfrage nach SPD – Stimme und
27                                   'Erhoehen des SPD – Zählers, wenn
28                                   'z < 0,813
29 IF z<0,904 THEN FDP=FDP+1 : GOTO 37 'Abfrage nach FDP – Stimme und
30                                   'Erhoehen des FDP – Zählers, wenn
31                                   'z < 0,904

```

```

32 IF z<0,987 THEN Gruen=Gruen+1 : GOTO 37 `Abfrage nach Grünen – Stimme und
33 `Erhoehen des Grünen – Zählers, wenn
34 `z < 0,987
35 Splitter=Splitter+1 `Erhoehen des Splitter– Zählers, wenn
36 `z > 0,987
37 NEXT wnt
38 `
40 ` Ausgabe der Ergebnisse der Auslosung des repräsentativen Querschnitts
41 `
42 PRINT : PRINT“ “; : PRINT USING“##.#“;CDU/n*100; : PRINT“ % “;
43 PRINT USING“##.#“;SPD/n*100; : PRINT“ % “;
44 PRINT USING“##.#“;FDP/n*100; : PRINT“ % “;
45 PRINT USING“##.#“;Gruen/n*100; : PRINT“ % “;
46 PRINT USING“##.#“;Splitter/n*100; : PRINT“ % “;
47 `
50 ` Ausgabe der Abweichungen von den tatsächlichen Werten der Bundestagswahl
51 `
52 PRINT : PRINT“ “; : PRINT USING“##.#“;CDU/n*100-44.3; : PRINT“ % “;
53 PRINT USING“##.#“;SPD/n*100-37; : PRINT“ % “;
54 PRINT USING“##.#“;FDP/n*100-9.1; : PRINT“ % “;
55 PRINT USING“##.#“;Gruen/n*100-8.3; : PRINT“ % “;
56 PRINT USING“##.#“;Splitter/n*100-1.3; : PRINT“ % “;
57 `
58 GOTO 13

```

Die kommerziellen Grundlagen von Umfragen

Die Beschreibung der kommerziellen Grundlagen erfolgt hier nur in groben Umrissen.

Die Befragung eines einzigen repräsentativen Querschnittes von 1000 oder 2000 Personen ist mit hohen Fixkosten verbunden: Bekommt ein Interviewer pro Interview 20 bis 30 DM, so ergeben sich allein daraus fixe Kosten von 20.000 bis 30.000 DM bei 1000 durchgeführten Interviews – bei 2000 sogar 40.000 bis 60.000 DM. Hinzu kommen die Lohnkosten des festangestellten Personals des Meinungsforschungsinstitutes.

Nur in Ausnahmefällen wird eine Meinungsumfrage von einem einzigen Auftraggeber übernommen. Im Normalfall werden Fragen von mehreren Auftraggebern zu einer Umfrage zusammengefaßt (Omnibus-, Mehrthemenumfragen). Diese Vorgehensweise ergibt sich zwangsläufig aus den Preisen, die sich am freien Markt für Umfragen gebildet haben. Ein paar Beispiele (aus: context Nr.17/87):

- GFM-Getas, Mehrfachumfrage BEVÖLKERUNG ab 14 Jahren/Haushalte. 2000 Fälle Random Route. Preise: Geschlossene Frage DM 1400, offene Frage DM 1800. Lieferung nach 30 Arbeitstagen
- Marplan Forschungsgesellschaft, Mehrthemenumfrage BEVÖLKERUNG ab 14 Jahren. 2000 Fälle. Preise: Geschlossene Frage DM 1300, offene Frage DM 1750. Lieferung am 07.10.87

- Institut Schölzel GmbH, MARKETING-CONTAINER: Befragung bei 1000 Haushalten in der BRD und West-Berlin. Preise: Geschlossene Frage DM 750, offene Frage DM 1150. Dauer der Feldarbeit: Zwei bis drei Wochen.
- GfK-Marktforschung, Beteiligungs-Untersuchung BEVÖLKERUNG 14 bzw. 16 bis 69 Jahre, 2000 Fälle. Richtpreis: Geschlossene Frage DM 1400, offene Frage DM 1900. Lieferung Mitte Oktober.
- Ifak-Institut, Mehrthemenumfrage BEVÖLKERUNG ab 14 Jahren, 2000 Fälle, darunter 1000 PKW-Fahrer. Preise: Geschlossene Frage DM 1400, offene Frage DM 1750. Lieferung Anfang Oktober
- IJF Institut für Jugendforschung, Mehrthemenumfrage bei 7 BIS 15JÄHRIGEN, 1000 Fälle, Random. Preise: Geschlossene Frage DM 1900, offene Frage DM 2600. Lieferung im November
- Ifak-Institut, Mehrthemenumfrage BEVÖLKERUNG ab 14 Jahren, 2000. Preise: Geschlossene Frage DM 1400, offene Frage DM 1750. Lieferung Anfang Oktober
- Emnid bietet an: MEHRTHEMENUMFRAGE. Bevölkerung: jede Woche + Haushalte: jede Woche. PKW-Fahrer: alle 14 Tage. Rufen Sie Frau Glock an: Tel 0521/26001-30

Wie bei einer Fluggesellschaft, die einen Flug erst rentabel durchführen kann, wenn die Passagier-Einnahmen die Fixkosten decken, muß auch bei einer „Repräsentativumfrage“ die Kapazität zwecks Kostendeckung genügend ausgelastet sein. Die obere Grenze für die Anzahl der Fragen von Auftraggebern, die in einer Repräsentativumfrage untergebracht werden können, hängt dabei von mehreren Faktoren ab (maximal zumutbare Länge eines Interviews, „gleichmäßige“ Beschäftigung der Interviewer und Terminvorstellungen der Auftraggeber). Mit zunehmender Auslastung einer „Repräsentativumfrage“ sinkt aber die Qualität der Ergebnisse – nicht nur wegen Ermüdungserscheinungen bei den Interviewern und den Befragten, sondern aus statistischen Gründen. Je mehr Fragen gestellt werden und je umfangreicher diese sind, desto mehr nimmt die statistische Zuverlässigkeit ab, d.h. bei gleicher Sicherheitswahrscheinlichkeit werden die Vertrauensintervalle größer (vergleiche auch S. III/13-14 und Abschnitte VI bis IX)

IV Die 5% - Hürde

Betrachten wir ein konkretes Beispiel. Angenommen, die FDP hat 4,5% der Stimmen erhalten, also 1.800.000 Stimmzettel: Wie viele der repräsentativen 1000er-Querschnitte enthalten exakt 45 Stimmzettel für die FDP, reproduzieren also genau ihr Wahlergebnis von 4,5%? Diese Frage läßt sich sowohl durch theoretische Berechnung mit Hilfe der Binomialverteilung als auch durch Simulation - hinreichend häufige Auslosung des repräsentativen Querschnittes - beantworten.

Der relative Anteil der repräsentativen 1000er-Querschnitte, die genau 45 FDP-Stimmzettel enthalten, beträgt *nur* sechs Prozent, das heißt, bei häufiger Auslosung von repräsentativen Querschnitten wird im Durchschnitt nur etwa jeder siebzehnte repräsentative Querschnitt das Wahlergebnis von 4,5% für die FDP reproduzieren.

Ferner findet man jede Anzahl von FDP-Stimmzetteln zwischen 40 und 49 in ungefähr 5 bis 6% der ausgelosten repräsentativen 1000er-Querschnitte.

Wie man hieraus sieht, ist ein Querschnitt vom Umfang 1000 ein recht untaugliches Mittel zur Bestimmung des FDP-Anteils. Obwohl wir wissen, daß dieser 4,5% beträgt, werden uns die 1000er-Querschnitte jedes Resultat von 4,0 bis 4,9% mit einer etwa gleich geringen Häufigkeit bescheren.

Bei diesem diffusen Bild stellt sich die Frage, wie viele der repräsentativen Querschnitte der FDP fälschlicherweise einen Einzug in den Bundestag prophezeien, also 50 oder mehr FDP-Stimmzettel unter den ausgelosten 1000 enthalten und damit einen FDP-Anteil von 5% oder mehr vortäuschen. **Der Anteil der Querschnitte mit genau 50 FDP-Stimmen beträgt 4,4% und unterscheidet sich damit kaum von demjenigen mit 45 Stimmen für die FDP (dem tatsächlichen Wert).** Die relativen Anteile der repräsentativen Querschnitte, die zwischen 50 und 60 FDP-Stimmzettel enthalten, sind in der folgenden Tabelle aufgeführt. In der ersten Zeile ist die Anzahl der gefundenen FDP-Stimmzettel und in der zweiten Zeile der prozentuale Anteil der repräsentativen 1000er-Querschnitte mit dieser Anzahl angegeben:

50	51	52	53	54	55	56	57	58	59	60
4,4%	3,8%	3,3%	2,8%	2,3%	1,9%	1,5%	1,2%	0,9%	0,7%	0,5%

Summiert man die untere Zeile, so folgt, daß rund 23% aller repräsentativen Querschnitte vortäuschen, die FDP habe 5% bis 6% der Wählerstimmen erhalten und damit den Einzug in den Bundestag geschafft, obwohl der tatsächliche Stimmenanteil für die FDP bei 4,5% liegt! Zusätzlich enthalten über 1% aller 1000er-Querschnitte 61 oder mehr FDP-Stimmzettel und täuschen vor, daß die FDP über 6% der abgegebenen Stimmen erhalten hat.

Somit werden über 24% aller repräsentativen Querschnitte der FDP fälschlicherweise den Einzug in den Bundestag prophezeien (siehe Tabelle auf der folgenden Seite). Jeder vierte repräsentative Querschnitt täuscht also ein Überwinden der 5%-Hürde vor.

Man muß sich bei diesem Ergebnis noch einmal klar machen, daß der Ausgangspunkt die vollzogene Wahl ist : Jedes Prognose-Institut bekommt die zu seinem repräsentativen Querschnitt gehörenden ausgefüllten Stimmzettel. Allein aufgrund der Tatsache, daß 1000 Wahlberechtigte per Lotterie ausgewählt wurden, kommt es zu so starken Schwankungen, daß eine von vier Prognosen nicht einmal die wichtige Frage der 5%-Hürde richtig beantwortet, obwohl unterstellt ist, daß die FDP in „Wahrheit“ mit 4,5% relativ weit darunter liegt.

Je näher der tatsächliche Stimmenanteil an die 5%-Marke gerät, desto schlechter werden die

Chancen der Querschnitte, wie die Tabelle zeigt. Auch im umgekehrten Fall, daß die FDP in Wahrheit über 5% liegt, sieht es nicht besser aus (siehe Tabelle „Überwinden der 5%-Hürde“): Hat die FDP einen wahren Anteil von 5,5%, liegt sie also deutlich über der 5%-Grenze, dann behaupten 22,5% aller repräsentativen Querschnitte fälschlicherweise, die FDP hätte den Einzug in den Bundestag nicht geschafft.

Das Problem liegt einfach in der viel zu geringen Zahl 1000. Das wird deutlich, wenn man dieselbe Überlegung mit anderen Querschnittsumfängen durchführt (man vergleiche dazu die Tabelle auf der folgenden Seite).

Um bei einem wahren Anteil von 4,9% das Verpassen der 5%-Hürde einigermaßen sicher zu diagnostizieren, müßte man fast 130.000 Interviews führen. Es grenzt deshalb an Verhältnis-Schwachsinn, wenn ein Meinungsforschungsinstitut auf der Basis von 1075 Interviews den Grünen mit 4,9% den Nichteinzug in einen Landtag prophezeit - es sei denn, es handle sich um eine politisch motivierte Prognose, die die SPD dadurch zu schwächen versucht, daß ihr Zweitstimmen zur Rettung der Grünen abspenstig gemacht werden. Demoskopie hat eben viele Anwendungsmöglichkeiten.

Doch zurück zu den Wahlprognosen mit 1000 Stimmzetteln. Ein Blick in die Tabellen zeigt, daß die Prognosen die Frage nach der 5%-Hürde nur dann einigermaßen verlässlich beantworten können, wenn das Wahlergebnis für die FDP entweder unter 4% oder über 6,2% liegt. Nur dann liefern 95% der Prognosen, also 19 von 20, die richtige Antwort. **Bei Wahlergebnissen zwischen 4% und 6,2% haben die Prognosen keine Aussagekraft für oder gegen den Einzug in den Bundestag. Gerade da aber wären sie von Interesse.**

Was kann man mit einer Meinungsumfrage klären?

Wie die Tabellen zu verstehen sind

Voraussetzung für alle Beispiele ist eine ideale Meinungsumfrage (Man vergleiche dazu Abschnitt III, Seite 11). Da die Stichproben nur einen kleinen Teil der tatsächlichen Stimmen umfassen, können sie zufallsbedingt mehr oder weniger stark vom wahren Ergebnis abweichen. In den Beispielen wird jeweils eine einfache Frage betrachtet, beispielsweise ob die FDP die 5%-Hürde überwindet oder ob die Koalition CDU/CSU und FDP die Mehrheit gewinnt. Ein Teil der Stichproben wird die Frage in Übereinstimmung mit dem amtlichen Ergebnis beantworten, der Rest dagegen wird die falsche Antwort geben. Die Tabellen führen den jeweiligen Anteil der *falschen* Prognosen an. Durchgespielt werden sowohl verschiedene amtliche Ergebnisse, die in der ersten Spalte aufgeführt sind, als auch verschiedene Umfänge der Querschnitte, die in den weiteren Spalten zu finden sind.

Vereinfachend wurde außerdem angenommen, daß nur die vier großen Parteien angetreten und keine ungültigen Stimmen möglich sind.

Zu den interessantesten Fragen der Bundestagswahl zählt, ob die FDP die 5%-Hürde nehmen wird. In der ersten Spalte der Tabellen ist der im amtlichen Wahlergebnis unterstellte tatsächliche Stimmenanteil aufgeführt, bei der ersten Tabelle zwischen 3,5% und 4,9%. Rechts daneben steht der prozentuale Anteil der repräsentativen Querschnitte, die fälschlicherweise voraussagen, daß die FDP in den Bundestag einziehen wird. Hier werden also die Querschnitte gezählt, die 5% oder mehr für die FDP ermitteln. Die einzelnen Spalten unterscheiden sich im Umfang der Stichproben.

Die zweite Tabelle („Überwinden der 5%-Hürde“) gibt entsprechend die Anteile der falschen Prognosen wieder, wenn das amtliche Ergebnis über 5% liegt.

Scheitern an der 5%-Hürde

In der Tabelle sind die prozentualen Anteile der falsch positiven Querschnitte angeführt, d.h. es wird ein Überschreiten der 5%-Hürde vorgetäuscht, obwohl der wahre FDP-Anteil unter 5% liegt. Zum Beispiel: Wenn die FDP tatsächlich 4,7% (1. Spalte) der Stimmen erhalten hat, liefern 40,4% (2. Spalte) aller Querschnitte mit 500 Befragten die falsche Antwort, daß nämlich die FDP die 5%-Hürde geschafft hätte. Selbst bei Querschnitten mit 10 000 Befragten stellen noch 8,3% die falsche Prognose (alle Angaben in %).

FDP-Anteil gemäß amtli. Wahlergebnis	Anzahl Interviews 500	Anzahl Interviews 1000	Anzahl Interviews 2000	Anzahl Interviews 10 000	Anzahl Interviews 127 200
3,5	5,0	0,9	0,034	~ 0	~ 0
3,9	12,6	4,7	0,8	~ 0	~ 0
4,0	15,2	6,6	1,5	~ 0	~ 0
4,1	18,1	9,0	2,7	~ 0	~ 0
4,2	21,4	12,0	4,5	0,0057	~ 0
4,3	24,8	15,5	7,1	0,04	~ 0
4,4	28,5	19,6	10,7	0,22	~ 0
4,5	32,3	24,2	15,3	0,93	~ 0
4,6	36,4	29,3	21,0	3,1	~ 0
4,7	40,4	34,7	27,7	8,3	~ 0
4,8	44,6	40,4	35,2	18,0	0,046
4,9	48,8	46,2	43,2	32,8	5,0

Die Werte in der Tabelle wurden numerisch mit der Binomialverteilung berechnet. Eine exakte Berechnung mit Hilfe der hypergeometrischen Verteilung ergibt bei der angegebenen Stellenzahl in der Tabelle die gleichen Werte. Die hypergeometrische Verteilung wird also in den Fällen in der Tabelle gut durch die Binomialverteilung approximiert. Unter Zugrundelegung des Stichprobenumfangs n und des wahren FDP-Anteils p wird folgender Ausdruck berechnet:

$$\sum_{k=n \cdot 0,05}^n \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Ist z.B. der Stichprobenumfang n=1000 und der FDP-Anteil gemäß amtlichem Wahlergebnis p=4,5%, so berechnet man:

$$\sum_{k=50}^{1000} \binom{1000}{k} \cdot 0,045^k \cdot 0,955^{1000-k} = 0,242 \Rightarrow 24,2\%$$

Überwinden der 5%-Hürde

In der Tabelle sind die prozentualen Anteile der falsch negativen Querschnitte angeführt, d.h. es wird ein Überschreiten der 5%-Hürde vorgetäuscht, obwohl der wahre FDP-Anteil über 5% liegt. Der Anteil der falsch negativen Querschnitte hängt von der Anzahl n der Interviews und vom wahren FDP-Anteil p gemäß amtlichem Wahlergebnis ab (alle Angaben in %).

FDP-Anteil gemäß amtli. Wahlergebnis	Anzahl Interviews 500	Anzahl Interviews 1000	Anzahl Interviews 2000	Anzahl Interviews 10 000	Anzahl Interviews 129 000
5,0	47,1	48,0	48,6	49,4	49,8
5,1	43,1	42,3	40,5	31,8	5,0
5,2	39,2	36,8	33,0	17,8	0,055
5,3	35,4	31,6	26,1	8,6	~ 0
5,4	31,8	26,9	20,1	3,5	~ 0
5,5	28,5	22,5	15,1	1,25	~ 0
5,6	25,3	18,7	11,1	0,38	~ 0
5,7	22,4	15,3	7,9	0,10	~ 0
5,8	19,7	12,4	5,5	0,02	~ 0
5,9	17,2	9,9	3,7	~ 0	~ 0
6,0	14,9	7,8	2,4	~ 0	~ 0
6,1	12,9	6,1	1,6	~ 0	~ 0
6,2	11,1	4,7	0,98	~ 0	~ 0
6,3	9,5	3,6	0,60	~ 0	~ 0
6,4	8,1	2,7	0,36	~ 0	~ 0
6,5	6,9	2,0	0,21	~ 0	~ 0
6,9	3,4	0,56	0,02	~ 0	~ 0
7,0	2,8	0,40	0,01	~ 0	~ 0

Die Werte in der Tabelle wurden wie zuvor mit der Binomialverteilung berechnet. Unter Zugrundelegung des Stichprobenumfangs n und des wahren FDP-Anteils p wird der folgende Ausdruck berechnet:

$$\sum_{k=0}^{n \cdot 0,05 - 1} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Ist z.B. der Stichprobenumfang n=1000 und der FDP-Anteil gemäß amtlichem Wahlergebnis p=5,5%, so berechnet man:

$$\sum_{k=0}^{49} \binom{1000}{k} \cdot 0,055^k \cdot 0,945^{1000-k} = 0,225 \Rightarrow 22,5\%$$

Die Programme zur Bestimmung der Tabellenwerte schrieb Ullrich Tesche.

V Die Mehrheitsfrage

Ist das vielleicht ein Spezifikum einer kleinen Partei: Sind wenigstens die Zahlen für die großen Parteien verlässlich? Auch hierzu eine typische **Fragestellung**: Gelingt es der CDU/CSU, gemeinsam mit der FDP die Mehrheit zu erringen - **ungeachtet der 5%-Problematik**?

Unterstellen wir, daß 51% der abgegebenen Stimmen auf diese Parteien entfallen, also 20.400.000 Stimmzettel für sie in der Trommel sind : Wie viele der repräsentativen Querschnitte vom Umfang 1000 werden genau dieses Ergebnis liefern, also exakt 510 Stimmzettel für die Koalition enthalten? Auch diese Antwort läßt sich sowohl durch eine Berechnung mit der Binomialverteilung als auch durch Simulation finden : Lediglich 2,5% aller repräsentativen 1000er-Querschnitte enthalten genau 510 Koalitionsstimmzettel, d.h. **bei häufiger Auslosung von repräsentativen Querschnitten wird durchschnittlich nur jeder vierzigste repräsentative Querschnitt einen Koalitionsanteil von 51,0% aufweisen**. Es ist also sehr unwahrscheinlich, daß ein repräsentativer Querschnitt den Wert von 51,0% liefert. Vielmehr findet man jeden Wert von 49,9% bis 52,1% in 2% bis 2,5% aller Querschnitte.

Wie man hieraus sieht, ist ein Querschnitt vom Umfang 1000 ein recht untaugliches Mittel zur Bestimmung des genauen Koalitionsanteils. Obwohl wir wissen, daß dieser 51% beträgt, werden uns die repräsentativen 1000er-Querschnitte jedes Resultat von 49,9% bis 52,1% mit einer geringen Wahrscheinlichkeit von einem Vierzigstel bis einem Fünfzigstel bescheren.

Bei diesem diffusen Bild stellt sich die Frage, wie viele der repräsentativen Querschnitte vom Umfang 1000 der Koalition fälschlicherweise das Verpassen der Mehrheit prognostizieren, also maximal 499 Stimmzettel für die Koalition enthalten und damit einen Koalitionsanteil von 49,9% oder weniger vortäuschen.

Die folgende Tabelle zeigt nun die relativen Anteile der repräsentativen 1000er-Querschnitte, die Koalitionsanteile unterhalb von 50% enthalten und somit vortäuschen, daß die Koalition die absolute Mehrheit verpaßt hat, obwohl sie diese tatsächlich mit 51% der abgegebenen Stimmen erhielt. Dabei zeigt die obere Zeile die Anzahl der gefundenen Koalitionsstimmzettel und in der unteren Zeile ist der prozentuale Anteil der repräsentativen 1000er-Querschnitte mit dieser Anzahl angegeben:

499	498	497	496	495	494	493	492	491	490
2,0%	1,9%	1,8%	1,7%	1,6%	1,5%	1,4%	1,3%	1,2%	1,1%

Die Summe der unteren Zeile beträgt 15,5%. Obwohl der tatsächliche Koalitionsanteil 51% beträgt, täuschen also 15,5% aller repräsentativen Querschnitte einen Anteil zwischen 49,0% und 49,9% und damit einen Verlust der absoluten Mehrheit vor.

Zusätzlich täuschen über 7% der Querschnitte einen Koalitionsanteil zwischen 48,0% und 48,9% vor, wie aus der folgenden Tabelle hervorgeht:

489	488	487	486	485	484	483	482	481	480
1,0%	1,0%	0,9%	0,8%	0,7%	0,7%	0,6%	0,5%	0,5%	0,4%

Außerdem enthalten noch fast 3% aller repräsentativen Querschnitte weniger als 480 Koalitionsstimmzettel und täuschen somit vor, daß die Koalition **weniger als 48%** der

abgegebenen Stimmen erhalten hat.

Auf die Ausgangsfrage, wie viele der repräsentativen Querschnitte fälschlicherweise ein Verpassen der Mehrheit bei einem tatsächlichen Koalitionsanteil von 51% prophezeien, gibt es demnach folgende Antwort (siehe Tabelle „Gewinnen der absoluten Mehrheit“): 25,3%.

Jede vierte Wahlprognose muß bei diesem Wahlausgang allein aus **statistischen** Gründen, d.h. weil 1000 Wahlberechtigte per Lotterie ermittelt werden, zum qualitativ falschen Ergebnis kommen und den Regierungsparteien eine Wahlniederlage statt eines Sieges prophezeien.

Oder im umgekehrten Fall („Verlieren der absoluten Mehrheit“): Bei **48% tatsächlichem** Stimmenanteil kommen noch immer 10,9% der Querschnitte zu dem falschen Ergebnis, daß die Koalition die absolute Mehrheit gewonnen hätte.

Wollte man 49,9% von 50% unterscheiden können, müßte man den Umfang des Querschnitts auf über 650.000 erhöhen, und auch dann wäre noch das Ergebnis jeder zwanzigsten Umfrage falsch. Bei 1000 Befragungen muß der wahre Stimmenanteil für eine einigermaßen verlässliche Aussage über die Mehrheit entweder unter 47,5% liegen oder über 52,5%. Im Zwischenbereich dagegen sind keine verlässlichen Aussagen möglich, wenn man wieder verlangt, daß von 20 repräsentativen Querschnitten höchstens einer zum falschen Ergebnis kommt.

Was kann man mit einer Meinungsumfrage klären?

Die entscheidende Frage der Wahl lautet, ob CDU/CSU und FDP zusammen mehr als 50% der Stimmen gewinnen. Das Problem der 5%-Hürde sei dabei ausgeklammert.

Ein Teil der Stichproben wird die Frage in Übereinstimmung mit dem amtlichen Ergebnis beantworten, der Rest dagegen wird die umgekehrte Antwort geben. Die Tabellen führen den jeweiligen Anteil der *falschen* Prognosen an. In der ersten Spalte der Tabellen ist der unterstellte tatsächliche Stimmenanteil gemäß amtlichem Wahlergebnis aufgeführt, bei der ersten Tabelle zwischen 45,0% und 49,9% und bei der zweiten 50,0% bis 55,0%. Die einzelnen Spalten unterscheiden sich im Umfang der Stichproben

Verpassen der absoluten Mehrheit

In der Tabelle sind die prozentualen Anteile der falsch positiven Querschnitte angeführt. Beispielsweise täuschen bei 47,5% wahren Stimmenanteil und 1000 Interviews bereits 6% aller Querschnitte eine Mehrheit für die Koalition vor (alle Angaben in %).

CDU/CSU/ FDP-Anteil gemäß amtl. Wahlergebnis	Anzahl Interviews 500	Anzahl Interviews 1000	Anzahl Interviews 2000	Anzahl Interviews 10 000	Anzahl Interviews 667 000
45,0	1,4	0,08	~ 0	~ 0	~ 0
46,0	4,0	0,62	0,02	~ 0	~ 0
47,0	9,7	3,1	0,39	~ 0	~ 0
47,35	12,7	5,0	0,94	~ 0	~ 0
47,5	14,1	6,0	1,34	~ 0	~ 0
48,0	19,8	10,9	3,9	~ 0	~ 0
48,5	26,5	17,9	9,3	0,14	~ 0
49,0	34,4	27,4	19,2	2,3	~ 0
49,5	42,9	38,8	33,5	16,1	~ 0
49,8	48,2	46,2	43,8	34,8	0,05
49,9	50,0	48,7	47,3	42,4	5,0

Gewinn der absoluten Mehrheit

In der Tabelle sind die prozentualen Anteile der falsch negativen Querschnitte angeführt, d.h. es wird ein Verpassen der Mehrheit vorgetäuscht. Zum Beispiel täuschen bei 52,5% wahren Stimmenanteil und 1000 Interviews bereits 5,3% aller repräsentativen Querschnitte eine Niederlage der Koalition vor.

CDU/CSU/ FDP-Anteil gemäß amtl. Wahlergebnis	Anzahl Interviews 500	Anzahl Interviews 1000	Anzahl Interviews 2000	Anzahl Interviews 10 000	Anzahl Interviews 667 000
50,0	48,2	48,7	49,1	49,6	49,95
50,1	46,4	46,2	45,5	41,7	5,0
50,2	44,7	43,7	42,0	34,1	0,05
50,5	39,4	36,4	31,9	15,6	~ 0
51,0	31,1	25,3	18,0	2,2	~ 0
51,5	23,7	16,3	8,6	0,13	~ 0
52,0	17,4	9,7	3,5	~ 0	~ 0
52,5	12,2	5,3	1,2	~ 0	~ 0
52,55	11,8	5,0	1,1	~ 0	~ 0
53,0	8,3	2,7	0,34	~ 0	~ 0
54,0	3,3	0,50	0,016	~ 0	~ 0
55,0	1,1	0,068	~ 0	~ 0	~ 0

Die Berechnung erfolgte, wie bei der 5%-Hürde dargestellt, mit Hilfe der Binomialverteilung. Die Programme hierzu schrieb Ullrich Tesche.

VI Die Mehrheitsfrage unter Berücksichtigung der 5%-Klausel

Unter Berücksichtigung der 5%-Klausel verkompliziert sich die Frage nach der Mehrheitsbildung beträchtlich, weil gleichzeitig zwei voneinander abhängige Merkmale betrachtet werden müssen. Es reicht nämlich nicht mehr aus, daß wir nur unterstellen, daß die Koalition die Mehrheit - z.B. 51% - gewinnt, wenn sich diese Mehrheit beispielsweise aus einem CDU/CSU-Anteil von 47% und einem FDP-Anteil von 4% zusammensetzt. Deshalb müssen die beiden Merkmale, daß erstens CDU/CSU und FDP zusammen mehr als 50% der gültig abgegebenen Stimmen erhalten haben und zweitens die FDP mehr als 5% der Stimmen erhalten hat, in den Berechnungen berücksichtigt werden.

Angenommen, die CDU/CSU habe tatsächlich 45% und die FDP 5% der Stimmen erhalten. Dann stellt sich die Frage, wie viele der repräsentativen 1000er-Querschnitte genau 450 CDU/CSU- und 50 FDP-Stimmzettel enthalten, also genau das Wahlergebnis für beide Parteien reproduzieren.

Die Beantwortung dieser Frage ist sowohl durch theoretische Berechnung mit der Trinomialverteilung als auch durch Simulation möglich. Während nach Abschnitt IV rund 6% aller repräsentativen 1000er Querschnitte genau 50 FDP-Stimmen und nach Abschnitt V rund 2,5% genau 450 CDU/CSU-Stimmen aufweisen, haben nur 0,15% aller repräsentativen Querschnitte die Eigenschaft, daß sie sowohl exakt 50 FDP-Stimmen als auch genau 450 CDU/CSU-Stimmen enthalten. Bei häufiger Auslosung von repräsentativen Querschnitten wird im Durchschnitt demnach nur jeder 670-te das tatsächliche Ergebnis von 45,0% für die CDU/CSU und 5,0% für die FDP reproduzieren. Ferner wird fast jede Kombination von k_1 CDU/CSU-Stimmen und k_2 FDP-Stimmen, wobei

$$438 \leq k_1 \leq 462 \quad \text{und} \quad 45 \leq k_2 \leq 55$$

mit einer Wahrscheinlichkeit zwischen einem Tausendstel und $1/670$ auftreten. Eine Meinungsumfrage vom Umfang 1000 ist also ein extrem unzuverlässiges Mittel zur gleichzeitigen Bestimmung des exakten CDU/CSU- und FDP-Anteils. Wie man hieraus sieht, besteht nicht der geringste Grund, die per Meinungsumfrage ermittelten Werte für FDP und CDU/CSU als die tatsächlichen Ergebnisse hinzustellen, die gefundenen Werte treffen nämlich mit an Sicherheit grenzender Wahrscheinlichkeit nicht zu.

Auf die Ausgangsfrage bezogen kann man folgende Resultate herleiten¹³:

Unter der Annahme, daß die CDU/CSU tatsächlich 45% und die FDP 5% der Stimmen erhalten haben, täuschen 67% (!) aller repräsentativen Querschnitte eine Niederlage der Koalition vor (ohne Berücksichtigung der 5%-Klausel wären es nur 48,7% gewesen).

Haben CDU/CSU 46% und FDP 5% der Stimmen erhalten, so täuschen 50% der repräsentativen Querschnitte eine Niederlage der Koalitionsparteien vor (ohne Berücksichtigung der 5%-Klausel wären es nur 25,3% gewesen).

Setzt man voraus, daß CDU/CSU 47,4% und FDP 5,6% der Stimmen erhalten, so täuschen 10% der Querschnitte eine Niederlage der Koalitionsparteien vor (ohne Berücksichtigung der 5%-Klausel wären es nur 2,7% gewesen).

Die übliche statistische Sicherheit beträgt 95% : Bei dieser Problemstellung dürfen also nur 5%

¹³ Die Details der Berechnung der folgenden Fehlerraten sind in der Box "Auswirkung der 5%-Klausel auf die Mehrheitsbildung" zu finden.

der repräsentativen Querschnitte der Koalition fälschlicherweise eine Niederlage prognostizieren. Im folgenden sind einige Kombinationen von Parteianteilen aufgeführt, für die dies zutrifft.¹⁴

Anteil der falsch negativen repräsentativen Querschnitte:						
CDU/CSU	49,5%	49,0%	48,6%	47,2%	46,8%	46,1%
FDP	5,0%	5,3%	5,5%	6,0%	6,2%	6,6%
-----	-----	-----	-----	-----	-----	-----
KOALITION	54,5%	54,3%	54,1%	53,2%	53,0%	52,7%
bei Berücksichtigung der 5%-Klausel:	5 %	5 %	5 %	5 %	5 %	5 %
ohne Berücksichtigung der 5%-Klausel:	0,2%	0,3%	0,4%	2,0%	2,7%	4,1%
-----	-----	-----	-----	-----	-----	-----

Hat beispielsweise die CDU/CSU 49,5%, die FDP 5,0% der gültig abgegebenen Stimmen erhalten, so täuschen mit Berücksichtigung der 5%-Klausel 5%, ohne deren Berücksichtigung dagegen nur 0,2% der repräsentativen Querschnitte der Koalition eine Niederlage statt eines Sieges vor.

Bei den vorhin beschriebenen Fehlerraten kann man sich ebensogut auf sein Gefühl oder den Kaffeesatz verlassen und die fünf- bis sechststelligen Beträge für eine Wahlprognose sparen.

Generell gilt: Je detaillierter die Auskunft einer Prognose sein soll, desto schlechter wird ihre statistische Zuverlässigkeit.

Auswirkungen der 5%-Klausel auf die Mehrheitsbildung

Technische Erläuterungen

Die Ergebnisse werden mit Hilfe der Trinomialverteilung berechnet. Dabei wird vereinfachend angenommen, daß die Grünen den Einzug in den Bundestag schaffen und daß die Anteile der Splitterparteien auf die vier bisher im Bundestag vertretenen Parteien umgewichtet werden (Die Berechnung ohne diese Annahme erfolgt ähnlich, der Aufwand ist jedoch wegen der vielen Fallunterscheidungen erheblich größer.).

Zur Berechnung führen wir folgende Bezeichnungen ein:

- n: Anzahl der ausgelosten Stimmzettel
- p₁: tatsächlicher relativer Anteil der CDU/CSU-Stimmen
- p₂: tatsächlicher relativer Anteil der FDP-Stimmen
- p₃: tatsächlicher relativer Anteil der SPD bzw. Grünen-Stimmen
- k₁: Anzahl der ausgelosten CDU/CSU-Stimmzettel
- k₂: Anzahl der ausgelosten FDP-Stimmzettel
- k₃: Anzahl der ausgelosten SPD- bzw. Grünen-Stimmzettel

¹⁴ Im Gegensatz zu der Frage der 5%-Hürde bzw. Mehrheitsbildung läßt sich aus den obigen Fehlerraten nicht direkt ein zweidimensionaler Test für die Mehrheitsfrage bei vorgegebener Sicherheitswahrscheinlichkeit angeben. Die Konstruktion dieses Testes und seine praktische Implementierung ist nicht schwierig, sie würde jedoch den Rahmen dieser Arbeit sprengen.

Der relative Anteil der repräsentativen Querschnitte, die bei einem tatsächlichen Anteil von p_1 CDU/CSU-Stimmen, p_2 FDP-Stimmen und p_3 SPD- bzw. Grünen-Stimmen genau k_1 CDU/CSU-, k_2 FDP- und k_3 SPD- bzw. Grünen-Stimmzettel enthalten, beträgt:

$$\frac{n!}{k_1!k_2!k_3!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$

Im folgenden wird angenommen, die Koalition aus CDU/CSU und FDP habe die Mehrheit. Dies ist dann der Fall, wenn entweder die FDP an der 5%-Hürde scheitert, aber die Union stärker als SPD und Grüne zusammen ist – in Formeln: $p_2 < 0,05$ und $p_1 > p_3$ – oder die FDP die 5%-Hürde überschritten hat und gleichzeitig die Union und die FDP zusammen mehr als 50% haben – in Formeln: $p_2 \geq 0,05$ und $p_1 + p_2 > 0,5$.

Zur Bestimmung des Anteils der repräsentativen Querschnitte, die unter Berücksichtigung der 5%-Klausel fälschlicherweise behaupten, daß die Koalition der CDU/CSU und der FDP die Wahlen verloren hat, sind die folgenden Fälle zu betrachten:

ENTWEDER

Der Anteil der FDP-Stimmen ist kleiner als 5%, wobei gleichzeitig der CDU/CSU-Anteil kleiner als der Anteil von SPD und Grünen ist – d.h. für eine Anzahl der im repräsentativen Querschnitt gefundenen Stimmzettel gilt:

$$k_2 < n \cdot 0,05 \text{ und } k_1 < k_3$$

Man muß nun die relativen Anteile aller repräsentativen Querschnitte addieren, die diese einschränkenden Bedingungen erfüllen. Damit ergibt sich folgende Summe:

$$\sum_{\substack{k_1=0, k_2=0, k_3=0 \\ k_1+k_2+k_3=n, k_2 < n \cdot 0,05 \wedge k_1 < k_3}}^n \frac{n!}{k_1!k_2!k_3!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$

ODER

die Koalition hat weniger als die Hälfte der ausgelosten Stimmzettel erhalten, obwohl der Anteil der FDP-Stimmzettel größer als 5% ist – d.h. für die Anzahl der im repräsentativen Querschnitt gefundenen Stimmzettel gilt:

$$k_2 \geq n \cdot 0,05 \text{ und } k_1 + k_2 < n \cdot 0,5$$

Man muß nun die relativen Anteile aller repräsentativen Querschnitte addieren, die diese einschränkenden Bedingungen erfüllen. Damit ergibt sich folgende Summe:

$$\sum_{\substack{k_1=0, k_2=0, k_3=0 \\ k_1+k_2+k_3=n, k_2 \geq n \cdot 0,05 \wedge k_1+k_2 < n \cdot 0,5}}^n \frac{n!}{k_1!k_2!k_3!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$

Insgesamt bestimmt sich der gesuchte Anteil repräsentativer Querschnitte, die fälschlicherweise behaupten, daß die Koalition aus CDU/CSU und FDP die Wahl verloren hat, als Summe der beiden obigen Einzelergebnisse.

VII Die Vier-Parteien-Frage

In den Abschnitten IV und V wurde jeweils ein einfaches Problem betrachtet - in IV die 5%-Hürde und in V die Mehrheitsfrage. Diese lassen sich statistisch mit der Binomialverteilung behandeln. Allerdings setzt dies voraus, daß der repräsentative Querschnitt lediglich dazu verwendet wird, nur die gerade gestellte Frage zu beantworten, aber keine anderen. Insbesondere kann für die Beantwortung der Mehrheitsfrage nicht der gleiche repräsentative Querschnitt verwendet werden wie für die Beantwortung der Frage, ob die FDP die 5%-Hürde überwinden kann. Das Modell der Binomialverteilung läßt dies nicht zu. Will man beide Fragen mit Hilfe des gleichen repräsentativen Querschnitts beantworten, dann muß man das Modell der Trinomialverteilung verwenden, wie dies in Abschnitt VI gemacht wurde - was dazu führt, daß bei gleicher Chance die auslosungsbedingten Spielräume größer werden. Bei der Trinomialverteilung werden die SPD und die Grünen in einen Topf geworfen, was als statistische Vereinfachung legitim sein mag, aber demoskopisch unhaltbar ist, da von einem Wahlforscher erwartet wird, daß er in der Lage ist, getrennte Angaben über alle vier Parteien zu machen.

In der Praxis holt er weit mehr Informationen aus einem repräsentativen Querschnitt heraus, als nur über die vier Parteistärken.¹⁵ Würde er sich darauf beschränken, oder gar zwei verschiedene repräsentative Querschnitte für die Mehrheitsfrage und die 5%-Hürde verwenden, dann wäre er längst bankrott. (Man vergleiche die Box in Abschnitt III, Seite 20.)

Um die folgenden Ausführungen zu veranschaulichen, sind auf der übernächsten Seite die Resultate von 100 „Wiederholungen der Bundestagswahl“ aufgeführt. D.h. es wurde hundertmal ein Querschnitt vom Umfang 1000 ausgelost - und zwar unter denjenigen Wahlberechtigten, die bei der Wahl im Januar 1987 eine gültige Stimme abgaben - und die damaligen Stimmzettel ausgewertet. Man soll sich diese Resultate als die Ergebnisse von 100 verschiedenen Instituten vorstellen. Dadurch erhält man einen Eindruck davon, zu welchen Eskapaden „der“ repräsentative Querschnitt fähig ist.

¹⁵ In den demoskopischen Standardwerken und statistischen Lehrbüchern wird für die Behandlung von Meinungsumfragen das Binomial- (oder hypergeometrische) Modell verwendet. Damit wird für die Berechnung von Konfidenzintervallen und Sicherheitswahrscheinlichkeiten stillschweigend vorausgesetzt, daß eine Meinungsumfrage aus einer einzigen Frage besteht, die mit "Ja" oder "Nein" zu beantworten ist. Obwohl diese Annahme offensichtlich absurd ist, wird sie immer gemacht. Genauer, es wird einfach nicht darüber gesprochen, und es wird so getan, als ob das binomiale Weltbild in Ordnung wäre.

Bundestagswahl 1987

CDU/CSU 44,3% SPD 37,0% FDP 9,1% Grüne 8,3% Rest 1,3%

Was der Zufall so alles aus eigener Kraft schafft...

Die Resultate von 100 ausgelosten „repräsentativen“ Querschnitten

	CDU/CSU	SPD	FDP	Grüne	Rest		CDU/CSU	SPD	FDP	Grüne	Rest
1.Wdh.	40,3%	41,1%	8,6%	9,0%	1,0%	51.Wdh.	43,2%	37,3%	9,6%	7,9%	2,0%
2.Wdh.	44,1%	38,4%	8,7%	7,7%	1,1%	52.Wdh.	42,7%	38,5%	9,6%	7,9%	1,3%
3.Wdh.	43,9%	38,5%	8,8%	8,2%	0,6%	53.Wdh.	45,7%	36,0%	9,8%	7,1%	1,4%
4.Wdh.	47,1%	35,6%	7,8%	8,3%	1,2%	54.Wdh.	45,7%	37,7%	8,3%	7,5%	0,8%
5.Wdh.	43,1%	37,4%	9,5%	8,8%	1,2%	55.Wdh.	44,2%	36,2%	9,4%	8,9%	1,3%
6.Wdh.	47,0%	36,3%	8,4%	6,9%	1,4%	56.Wdh.	43,5%	37,7%	8,3%	9,2%	1,3%
7.Wdh.	44,5%	34,3%	10,3%	9,4%	1,5%	57.Wdh.	45,0%	36,2%	10,8%	7,1%	0,9%
8.Wdh.	43,5%	38,9%	8,4%	7,7%	1,5%	58.Wdh.	42,4%	38,6%	10,9%	6,9%	1,2%
9.Wdh.	42,0%	37,9%	10,8%	8,1%	1,2%	59.Wdh.	47,3%	34,7%	7,7%	9,2%	1,1%
10.Wdh.	43,7%	38,9%	8,7%	6,9%	1,8%	60.Wdh.	42,8%	39,1%	9,4%	7,6%	1,1%
11.Wdh.	44,6%	36,2%	9,5%	7,8%	1,9%	61.Wdh.	46,3%	36,2%	7,3%	8,4%	1,8%
12.Wdh.	42,8%	38,0%	8,8%	8,4%	2,0%	62.Wdh.	45,5%	33,4%	10,6%	8,2%	2,3%
13.Wdh.	43,9%	38,3%	9,1%	7,7%	1,0%	63.Wdh.	45,2%	34,5%	9,6%	8,9%	1,8%
14.Wdh.	43,6%	38,3%	8,5%	8,2%	1,4%	64.Wdh.	44,7%	36,3%	9,5%	8,3%	1,2%
15.Wdh.	46,2%	34,8%	10,4%	6,9%	1,7%	65.Wdh.	43,2%	37,1%	10,7%	7,6%	1,4%
16.Wdh.	43,3%	40,3%	8,4%	7,0%	1,0%	66.Wdh.	43,6%	40,5%	7,5%	7,0%	1,4%
17.Wdh.	44,7%	36,9%	8,9%	8,7%	0,8%	67.Wdh.	44,8%	34,9%	10,4%	8,3%	1,6%
18.Wdh.	46,0%	35,4%	8,8%	8,5%	1,3%	68.Wdh.	42,9%	36,7%	9,1%	10,2%	1,1%
19.Wdh.	42,0%	38,1%	9,5%	8,8%	1,6%	69.Wdh.	45,9%	34,0%	7,9%	10,2%	2,0%
20.Wdh.	43,3%	37,1%	9,4%	8,5%	1,7%	70.Wdh.	44,0%	35,9%	10,4%	8,5%	1,2%
21.Wdh.	44,4%	36,7%	9,0%	8,5%	1,4%	71.Wdh.	45,9%	37,1%	8,3%	7,4%	1,3%
22.Wdh.	45,2%	36,8%	9,6%	6,9%	1,5%	72.Wdh.	42,4%	37,4%	9,3%	9,9%	1,0%
23.Wdh.	45,7%	35,3%	11,0%	6,6%	1,4%	73.Wdh.	42,4%	37,6%	8,4%	10,9%	0,7%
24.Wdh.	45,5%	35,0%	9,8%	8,5%	1,2%	74.Wdh.	44,9%	36,2%	9,1%	8,7%	1,1%
25.Wdh.	44,7%	36,5%	9,5%	8,2%	1,1%	75.Wdh.	44,2%	37,0%	8,4%	9,1%	1,3%
26.Wdh.	42,8%	38,8%	9,5%	7,9%	1,0%	76.Wdh.	44,8%	36,8%	9,3%	8,2%	0,9%
27.Wdh.	41,9%	39,7%	8,8%	8,4%	1,2%	77.Wdh.	44,3%	37,7%	8,8%	7,5%	1,7%
28.Wdh.	43,7%	37,0%	9,1%	8,0%	2,2%	78.Wdh.	44,7%	34,3%	10,1%	9,5%	1,4%
29.Wdh.	44,4%	37,3%	9,4%	7,8%	1,1%	79.Wdh.	46,1%	36,1%	9,3%	7,2%	1,3%
30.Wdh.	45,4%	34,3%	9,5%	9,3%	1,5%	80.Wdh.	44,4%	37,7%	8,5%	8,0%	1,4%
31.Wdh.	43,5%	37,7%	9,8%	7,8%	1,2%	81.Wdh.	43,9%	37,6%	8,7%	8,2%	1,6%
32.Wdh.	45,4%	37,0%	9,9%	7,2%	0,5%	82.Wdh.	41,2%	36,9%	12,1%	9,3%	0,5%
33.Wdh.	43,6%	35,9%	10,5%	8,2%	1,8%	83.Wdh.	42,6%	38,7%	8,8%	9,5%	0,4%
34.Wdh.	43,8%	36,8%	8,1%	9,8%	1,5%	84.Wdh.	46,9%	35,8%	8,0%	8,2%	1,1%
35.Wdh.	43,0%	37,8%	9,1%	8,8%	1,3%	85.Wdh.	43,4%	39,5%	7,6%	8,4%	1,1%
36.Wdh.	42,5%	36,2%	9,9%	10,1%	1,3%	86.Wdh.	43,1%	37,2%	9,9%	8,5%	1,3%
37.Wdh.	44,8%	36,3%	9,1%	7,5%	2,3%	87.Wdh.	42,7%	38,8%	7,6%	9,3%	1,6%
38.Wdh.	48,3%	35,3%	7,8%	7,7%	0,9%	88.Wdh.	45,7%	36,3%	8,6%	7,9%	1,5%
39.Wdh.	43,7%	34,7%	11,1%	9,5%	1,0%	89.Wdh.	44,3%	37,3%	9,8%	6,9%	1,7%
40.Wdh.	44,3%	39,0%	8,0%	7,2%	1,5%	90.Wdh.	43,9%	37,4%	8,0%	9,2%	1,5%
41.Wdh.	45,6%	34,7%	9,1%	8,6%	2,0%	91.Wdh.	45,1%	37,4%	8,2%	8,0%	1,3%
42.Wdh.	43,1%	36,8%	9,8%	8,9%	1,4%	92.Wdh.	42,8%	36,4%	11,0%	8,2%	1,6%
43.Wdh.	47,0%	32,9%	9,3%	9,7%	1,1%	93.Wdh.	43,6%	35,7%	9,7%	9,5%	1,5%
44.Wdh.	41,1%	39,7%	9,4%	8,6%	1,2%	94.Wdh.	42,1%	38,7%	8,9%	9,7%	0,6%
45.Wdh.	43,4%	37,0%	9,2%	9,5%	0,9%	95.Wdh.	44,3%	37,3%	9,1%	7,9%	1,4%
46.Wdh.	42,8%	38,1%	9,3%	8,6%	1,2%	96.Wdh.	43,3%	36,4%	9,1%	10,4%	0,8%
47.Wdh.	40,8%	38,5%	10,1%	9,1%	1,5%	97.Wdh.	42,8%	36,9%	9,8%	8,8%	1,7%
48.Wdh.	45,8%	38,5%	8,1%	6,2%	1,4%	98.Wdh.	43,7%	36,7%	10,3%	7,9%	1,4%
49.Wdh.	44,0%	37,7%	9,2%	7,8%	1,3%	99.Wdh.	41,7%	38,4%	10,4%	7,9%	1,6%
50.Wdh.	46,9%	34,0%	9,2%	8,3%	1,6%	100.Wdh.	45,0%	36,3%	9,6%	7,9%	1,2%

Ein Blick auf die obige Tabelle zeigt, daß die „meisten“ der 100 idealen repräsentativen Querschnitte das tatsächliche Resultat für **eine** Partei relativ gut, aber gleichzeitig für **eine andere** ziemlich schlecht approximieren. Wie man sieht, sind die guten und schlechten Schätzungen ziemlich gleichmäßig über alle vier Parteien verteilt. Folglich kommt man damit nicht weiter. Die Statistik kann nun Aussagen darüber machen, wie groß die Chance ist, daß die durch die Auslosung des repräsentativen Querschnittes verursachten Abweichungen für eine, zwei ... oder alle Parteien in einem **vom Meinungsforscher frei wählbaren Spielraum** liegen. Diese Chance kann in einfachen Fällen explizit berechnet werden. Meist ist dies jedoch nicht möglich und sie muß durch Simulation (d.h. Probieren) ermittelt werden. Man führt hierzu eine große Anzahl von Auslosungen durch und prüft bei jeder nach, ob die Abweichungen innerhalb des vorgegebenen Spielraums liegen. Der prozentuale Anteil der Auslosungen, bei dem dies der Fall ist, stellt dann die Chance dar.

Bei dem gegenwärtigen politischen Kräfteverhältnis ist ein repräsentativer Querschnitt offensichtlich nur dann aussagekräftig, wenn er gleichzeitig für alle vier Parteien hinreichend genaue und zuverlässige Resultate liefert.¹⁶

Wählt der Demoskop einen Spielraum von 1% für die beiden großen und 0,5% für die beiden kleinen Parteien, so hat dies zur Folge, daß er die Resultate des ersten und zweiten repräsentativen Querschnittes in der Form

	CDU/CSU	SPD	FDP	Grüne
1.Wdh.	39,8% - 40,8%	40,6% - 41,6%	8,4% - 8,9%	8,8% - 9,3%
2.Wdh.	43,6% - 44,6%	37,9% - 38,9%	8,5% - 9,0%	7,5% - 8,0%

bekanntgeben müßte, und natürlich ebenso die weiteren. (Auf die Wahl eines Spielraums bei den Splitterparteien kann aus den vorhin erwähnten Gründen verzichtet werden.) Führt er diese Prozedur mit den Resultaten aller 100 repräsentativen Querschnitte durch, so merkt er bald, daß er mit **diesem Spielraum** der Wahrheit nicht näher kommt. Eine Überprüfung zeigt, **daß lediglich zwei** der 100 repräsentativen Querschnitte, nämlich Nr. 21 und Nr. 76 :

	CDU/CSU	SPD	FDP	Grüne
21.Wdh.	43,9% - 44,9%	36,2% - 37,2%	8,8% - 9,3%	8,3% - 8,8%
76.Wdh.	44,3% - 45,3%	36,3% - 37,3%	9,1% - 9,6%	8,0% - 8,5%

die Eigenschaft haben, daß sich das tatsächliche Wahlergebnis

CDU/CSU	SPD	FDP	Grüne
44,3%	37,0%	9,1%	8,3%

für **alle vier** Parteien innerhalb der gewählten Spielräume befindet und daß das Umfrageergebnis in der **präsentierten Form** somit zutreffend ist. Man kann mit 100 ausgelosten repräsentativen Querschnitten das Problem natürlich nur illustrieren, aber nicht daraus schließen, daß die Chance zwei Prozent beträgt. Anders verhält sich dies bei einer Million Auslosungen, bei welchen sich

¹⁶ Was den Spielraum betrifft, können die im Rest zusammengefaßten Splitterparteien ignoriert werden. Die Chance, daß sie über die 5%-Hürde rutschen, ist nämlich äußerst gering. Das Wahlgesetz hat bekanntlich zur Folge, daß der Anteil der Splitterparteien proportional auf diejenigen Parteien verteilt wird, welche die 5%-Hürde überwinden.

eine Chance von etwas mehr als 1,2% ergibt. Dieser Wert stimmt gut mit dem Resultat (=1,25%) einer theoretischen Berechnung mit der Multinomialverteilung überein. Somit ergibt sich: Im Durchschnitt liegen **nur bei einem von 80** repräsentativen Querschnitten die wahren Werte für alle vier Parteien innerhalb der obigen Spielräume. **Gibt der Demoskop das Resultat mit diesen Toleranzgrenzen an, dann müßte er seinem Auftraggeber mitteilen, daß das Umfrageergebnis - allein wegen der Auslosung des repräsentativen Querschnittes - mit einer Wahrscheinlichkeit von fast 99% falsch ist, wobei der Interviewfehler nicht berücksichtigt ist.**¹⁷

Wählt der Demoskop bei den großen Parteien einen Spielraum von 2% und bei den kleinen einen solchen von 1%, dann hat er die Resultate des ersten und zweiten repräsentativen Querschnittes in der Form

	CDU/CSU	SPD	FDP	Grüne
1. Wdh.	39,3% - 41,3%	40,1% - 42,1%	8,1% - 9,1%	8,5% - 9,5%
2. Wdh.	43,1% - 45,1%	37,4% - 39,4%	8,2% - 9,2%	7,2% - 8,2%

anzugeben und entsprechend die weiteren. Eine Überprüfung der obigen Tabelle zeigt, daß trotz dieses nicht gerade bescheidenen Spielraums *lediglich* 14 der 100 repräsentativen Querschnitte die Eigenschaft besitzen, daß das Wahlergebnis für alle vier Parteien sich innerhalb der gewählten Toleranzgrenzen befindet. Wie vorhin besitzt diese Anzahl lediglich indikativen Wert. Der genaue Wert läßt sich mit einer theoretischen Berechnung bzw. Simulation ermitteln: Die Chance beträgt 9,8%. Der Demoskop müßte also bei der Bekanntgabe des Resultates in der obigen Form dem Auftraggeber mitteilen, daß das Umfrageergebnis mit einer Wahrscheinlichkeit von über 90% falsch ist.¹⁸

Greift der langsam verzweifelnde Demoskop zu einem Spielraum von 4% für die großen Parteien und zu einem solchen von 2% für die kleinen, dann hat er die Ergebnisse der beiden ersten repräsentativen Querschnitte in der Form

	CDU/CSU	SPD	FDP	Grüne
1. Wdh.	38,3% - 42,3%	39,1% - 43,1%	7,6% - 9,6%	8,0% - 10,0%
2. Wdh.	42,1% - 46,1%	36,4% - 40,4%	7,7% - 9,7%	6,7% - 8,7%

bekanntzugeben und ebenso die weiteren. Der Auftraggeber wird ihm entgegenhalten - sei es ein Mediengewaltiger oder ein Politiker - daß Umfrageergebnisse in dieser Form keine politisch relevante Information mehr beinhalten, weil der Spielraum viel zu groß ist. Was soll der arme Demoskop darauf erwidern? Er könnte seinen Auftraggeber bitten, er möge doch mit ihm zusammen die Schäfchen in der obigen Tabelle zählen. Dann könne er sich selber überzeugen, daß trotz des peinlich großen Spielraums nicht einmal die Hälfte der 100 repräsentativen Querschnitte die Eigenschaft besitzen, daß das Wahlergebnis für alle vier Parteien sich innerhalb der Toleranzgrenzen befindet (nur 43 haben diese Eigenschaft). Wir können dem Demoskop

¹⁷ Zu den Abweichungen, die von der Auslosung herrühren, kommen noch die Fehler hinzu, die bei der Befragung entstehen. Der Interviewfehler wurde in Abschnitt III erörtert.

¹⁸ Da hat Bundeskanzler Kohl ganz andere Vorstellungen und Ansprüche betreffend Genauigkeit von Prognosen. Ein Spielraum von 2% ist für ihn indiskutabel. Mit 46% hat er gerechnet, genau 44,3% erhielt er. Einen Verlust in dieser Größenordnung habe niemand vorausgesehen, erklärte er am Wahlabend im ZDF. Er führte das schlechte Resultat für die Union nicht etwa auf den Prognosen-Spielraum, sondern auf die um 5% geringere Wahlbeteiligung zurück.

und seinem Auftraggeber versichern, daß die exakte theoretische Berechnung (bzw. Simulation) auf der Basis des gewählten Spielraums eine Chance von 44,5% ergibt. Dem Auftraggeber bleibt der Trost, daß die nichtssagenden Umfrageergebnisse wenigstens mit einer Wahrscheinlichkeit von über 55% falsch sind - wobei der Interviewfehler noch nicht berücksichtigt ist - und dem Demoskop Selbstkosten in fünfstelliger Höhe bescheren.

Wen wundert es bei dieser Sachlage, daß der Demoskop und einige seiner Auftraggeber zum Schluß kommen, es sei für beide besser - wenn auch aus unterschiedlichen Motiven - die Flucht nach vorne anzutreten. Die Toleranzen gehen über Bord und in Striptease-Manier erscheinen die Umfrageergebnisse splinternackt auf dem Markt. Als Feigenblatt wird der fromme Spruch vom repräsentativen Querschnitt zelebriert. Spielraum und Chance mögen ihre akademische Berechtigung haben, aber sie sind nun einmal praxisfremd, weil sie dem Demoskop und seinem Medien-Auftraggeber die Meistbegünstigten-Klausel vorenthalten. Das Betriebsrisiko ist beim Demoskopie-Endverbraucher billiger aufgehoben und außerdem bleibt die Illusion von exakten Zahlen erhalten.

Ein Spielverderber wird den Demoskop und seinen Auftraggeber daran erinnern, daß in den demoskopischen Lehrbüchern üblicherweise das Hohe Lied von einer Sicherheitswahrscheinlichkeit von 95 und noch mehr Prozent gesungen wird. Leute, die nur mit 90% operieren, gelten schon fast als Nestbeschmutzer.

Nimmt man 95% ins Visier, so ist - wie gleich gezeigt wird - der Spielraum für die großen Parteien auf 8,2% und für die kleinen auf 4,2% anzusetzen. Die Resultate des ersten und zweiten repräsentativen Querschnittes der obigen Tabelle sind dann in der Form

	CDU/CSU	SPD	FDP	Grüne
1. Wdh.	36,2% - 44,4%	37,0% - 45,2%	6,5% - 10,7%	6,9% - 11,1%
2. Wdh.	40,0% - 48,2%	34,3% - 42,5%	6,6% - 10,8%	5,6% - 9,8%

anzugeben und ebenso die weiteren. Der Demoskop wird erleichtert feststellen, daß mit dieser Präsentation der Umfrageergebnisse nun eine wahre Aussage gemacht wird, denn das Wahlergebnis befindet sich erstmals innerhalb der gewählten Toleranzgrenzen, was bisher nicht der Fall war. Diese Eigenschaft weisen insgesamt 98 der 100 obigen Wiederholungen auf, nur die Nr. 73 und Nr. 82 tanzen noch aus der Reihe. Natürlich kann man daraus nicht schließen, daß die Fehlerrate bei diesem Spielraum lediglich 2% beträgt. Die exakte Berechnung mit der Multinomialverteilung (bzw. mit einer hinreichend großen) Simulation ergibt - wie oben angekündigt - eine Chance von 95,6% dafür, daß die durch die Auslosung verursachten Abweichungen für die großen Parteien höchstens $\pm 4,1\%$ und für die kleinen höchstens $\pm 2,1\%$ betragen. Es versteht sich von selbst, daß bei solchen Abweichungen kein Mensch mehr an Meinungsumfragen und Prognosen interessiert ist. Die Streuung von Stammtischschätzungen wird erheblich geringer sein.

In ihrem Buch „Umfragen in der Massengesellschaft“ pflegt Frau Noelle-Neumann in Sachen Spielraum die Vorstellung

„Es haben nicht alle Werte der Toleranzspanne (=Spielraum) die gleiche Wahrscheinlichkeit, bei der Stichprobenziehung (=Auslosung des repräsentativen Querschnittes) gezogen zu werden, sondern die dicht am wahren Wert liegenden haben eine größere Chance, Infolgedessen kann man sagen: der in der Stichprobenziehung gefundene Wert hat eine größere Wahrscheinlichkeit als andere Werte, innerhalb der Toleranzspanne auch der wahre Wert zu sein.“

Das klingt wie eine posthume Rechtfertigung ihrer Gewohnheit, die Prognose-Zahlen mit einer Nachkommastelle zu verschnörkeln, anstatt sie mit Anstands-Toleranzgrenzen zu relativieren. Es ist evident, daß die Angabe von einem Spielraum bei vielen Demoskopie-Endverbrauchern den Verdacht von Ungenauigkeit und Nichtwissen erwecken würde, was keine ersprießliche Geschäftsgrundlage darstellt.

Frau Noelle-Neumann legt ihren Wunschvorstellungen betreffend Spielraum die Standard-Normalverteilung zugrunde, welche bei vielen Anwendern den Glauben nährt, daß kleine Fehler häufig auftreten, große hingegen selten. Dies trifft jedoch für die auslosungsbedingten Abweichungen des repräsentativen Querschnittes nicht zu. Das Gegenteil ist der Fall. Kleine Abweichungen bei allen vier Parteien treten praktisch mit der Wahrscheinlichkeit Null auf. Die häufigsten (Vier-Parteien-)Abweichungen liegen bei etwa $\pm 2\%$ für die großen und entsprechend bei etwa $\pm 1\%$ für die kleinen Parteien. Die Verteilungsfunktion (der Vier-Parteien-Abweichungen) ist also zweibucklig und sie weist in der Mitte (um die Null) ein Loch auf. In der graphischen Darstellung auf der folgenden Seite ist die Verteilungsfunktion mit einer durchgezogenen Linie dargestellt; zum Vergleich ist die Standard-Normalverteilung punktiert angegeben. Während bei der letzteren bekanntlich rund 68% bzw. 95% der Werte innerhalb der einfachen bzw. doppelten Standardabweichung liegen, so befinden sich bei der Verteilungsfunktion (der Vier-Parteien-Abweichungen) lediglich 24% bzw. 76% innerhalb des gleichen Spielraums. Der Kurvenverlauf wurde mit einer Computersimulation ermittelt. Es wurden eine halbe Million Auslosungen von repräsentativen Querschnitten vom Umfang 10000 (Zehntausend) vorgenommen - und die Resultate auf repräsentative Querschnitte vom Umfang 1000 umgerechnet. Dadurch konnte der Diskretisierungseffekt reduziert werden, so daß eine optisch ins Gewicht fallende Glättung nicht mehr erforderlich war.

Verteilungsfunktion der Vier-Partei-Abweichungen

Wie werden die Vier-Partei-Abweichungen auf der Skala eingetragen?

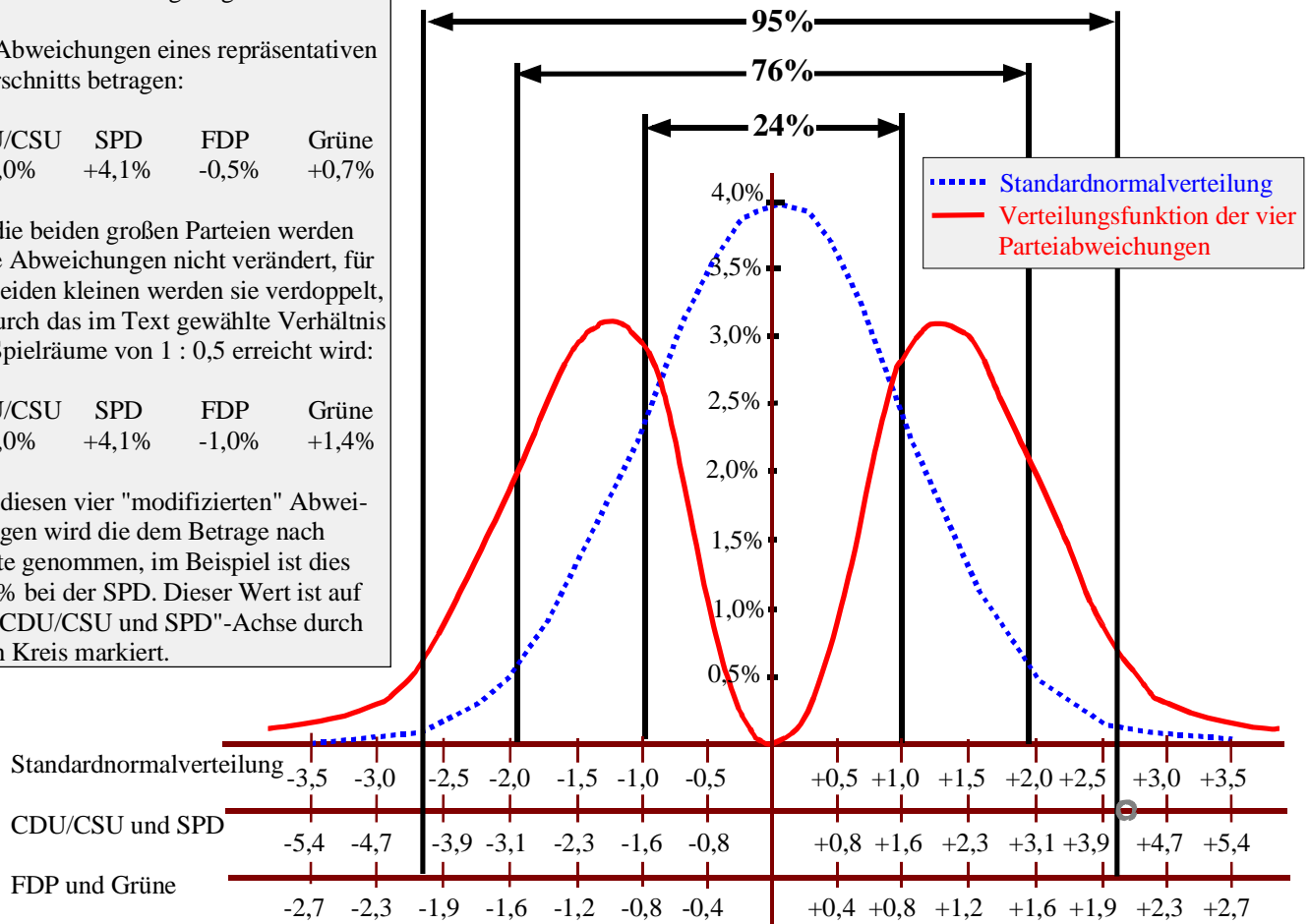
Die Abweichungen eines repräsentativen Querschnitts betragen:

CDU/CSU	SPD	FDP	Grüne
-4,0%	+4,1%	-0,5%	+0,7%

Für die beiden großen Parteien werden diese Abweichungen nicht verändert, für die beiden kleinen werden sie verdoppelt, wodurch das im Text gewählte Verhältnis der Spielräume von 1 : 0,5 erreicht wird:

CDU/CSU	SPD	FDP	Grüne
-4,0%	+4,1%	-1,0%	+1,4%

Von diesen vier "modifizierten" Abweichungen wird die dem Betrage nach größte genommen, im Beispiel ist dies +4,1% bei der SPD. Dieser Wert ist auf der "CDU/CSU und SPD"-Achse durch einen Kreis markiert.



In der folgenden Tabelle ist der Zusammenhang zwischen Spielraum und Chance dargestellt. In der ersten und zweiten Spalte sind die gewählten Spielräume für die großen und kleinen Parteien angeführt. In der dritten Spalte ist die Chance dafür angegeben, daß das Wahlergebnis für alle vier Parteien sich innerhalb der gewählten Spielräume befindet.

Gewählter Spielraum ¹⁹ CDU/CSU & SPD	FDP & Grüne	Chance, ²⁰ daß der gewählte Spielraum eingehalten wird
±0,5%	±0,25%	1,25 Prozent
±1,0%	±0,5%	9,8 Prozent
±2,0%	±1,0%	44,5 Prozent
±2,1%	±1,1%	50,8 Prozent
±3,0%	±1,5%	78,2 Prozent
±3,6%	±1,8%	89,8 Prozent
±4,0%	±2,0%	94,2 Prozent
±4,1%	±2,1%	95,6 Prozent
±5,0%	±2,5%	98,9 Prozent

Eine „Repräsentativumfrage“ ohne gleichzeitige Angabe von Fehlerspielraum und Chance ist so irreführend wie die Reklame einer Landeslotterie, durch den Kauf von Losen werde man Millionär.

Aber diese entscheidende Information zur Beurteilung der Aussagefähigkeit von Umfrageergebnissen erreicht den Demoskopiekonsumenten nie. Vor allem im politischen Bereich fällt der Spielraum einer stillen Symbiose von Politikern, Medien und Meinungsforschungsinstituten schmerzlos zum Opfer. Wer würde eine Prognose mit einem Spielraum von 8% für die großen und 4% für die kleinen Parteien veröffentlichen? Etwa der Spiegel, der Stern, die Welt, die FAZ oder gar die Bildzeitung? Etwa das Politbarometer des ZDF oder der Meinungsreport der ARD? An Einschaltquote Null sind weder ZDF noch ARD interessiert. Die Bildzeitung, der Stern, die Welt und auch der Spiegel wollen ihre Auflage verkaufen und nicht einstampfen.

¹⁹ Für die großen Parteien wurde ein doppelt so großer Spielraum wie für die kleinen gewählt. Der Grund hierfür liegt darin, daß eine lotterieberingte Abweichung von z.B. ±3% bei den großen Parteien weit weniger ins Gewicht fällt als bei den kleinen. Das Verhältnis der Spielräume kann man in einem gewissen Rahmen frei wählen, wobei inhaltliche und statistische Überlegungen eine Rolle spielen. Bei den gegenwärtigen Parteistärken scheint aus statistischer Sicht ein Verhältnis von etwa 1:0,6 "optimal".

²⁰ Berechnungsgrundlage : Multinomialverteilung für vier Merkmale

VIII Pseudo-Trends

Wie wirkt sich der Zufall im Wahlkampf aus? Jedermann will wissen, ob und wie sich die haltlosen Versprechen, die Fernsehdebatten usw. in Prozente umsetzen. Frau Noelle-Neumann hat den Wahlkampf mit Hilfe von zehn „Repräsentativumfragen“ verfolgt - es wurden jeweils 1000 Wahlberechtigte befragt - und in DIE WELT in ganzseitigen Reportagen darüber berichtet:

exklusiv WELT - Wahlbarometer

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Befragung: 29.10 - 10.11.86	46,5%	36,6%	6,4%	10,0%	0,5%
2.Befragung: 14.11 - 17.11.86	45,2%	34,6%	8,2%	11,7%	0,3%
3.Befragung: 21.11 - 25.11.86	47,2%	34,7%	6,5%	11,2%	0,4%
4.Befragung: 28.11 - 02.12.86	48,0%	35,0%	6,9%	10,0%	0,1%
5.Befragung: 06.12 - 09.12.86	46,9%	33,4%	8,5%	10,8%	0,4%
6.Befragung: 13.12 - 18.12.86	48,7%	32,4%	7,8%	10,7%	0,4%
7.Befragung: 23.12 - 30.12.86	46,8%	34,8%	6,0%	11,7%	0,7%
8.Befragung: 03.01 - 07.01.87	44,8%	36,7%	8,5%	9,5%	0,5%
9.Befragung: 10.01 - 14.01.87	44,0%	35,3%	9,6%	10,4%	0,7%
10.Befragung: 15.01 - 19.01.87	44,6%	37,3%	8,9%	8,3%	0,9%
11.Befragung: 15.01 - 23.01.87	46,0%	36,0%	8,5%	9,0%	0,5%

Zeitliches Auf und Ab der Parteistärken²¹

	CDU/CSU	SPD	FDP	Grüne	Rest
Trend von 1. zu 2. Befragung	-1,3%	-2,0%	+1,8%	+1,7%	-0,2%
Trend von 2. zu 3. Befragung	+2,0%	+0,1%	-1,7%	-0,5%	+0,1%
Trend von 3. zu 4. Befragung	+0,8%	+0,3%	+0,4%	-1,2%	-0,3%
Trend von 4. zu 5. Befragung	-1,1%	-1,6%	+1,6%	+0,8%	+0,3%
Trend von 5. zu 6. Befragung	+1,8%	-1,0%	-0,7%	-0,1%	0,0%
Trend von 6. zu 7. Befragung	-1,9%	+2,4%	-1,8%	+1,0%	+0,3%
Trend von 7. zu 8. Befragung	-2,0%	+1,9%	+2,5%	-2,2%	-0,2%
Trend von 8. zu 9. Befragung	-0,8%	-1,4%	+1,1%	+0,9%	+0,2%
Trend von 9. zu 10. Befragung	+0,6%	+2,0%	-0,7%	-2,1%	+0,2%
Trend von 10. zu 11. Befragung	+1,4%	-1,3%	-0,4%	+0,7%	-0,4%

Ihre Berichte über die „momentanen“ Absichten der Wahlberechtigten von Ende Oktober 1986 bis Mitte Januar 1987 sollen den Verlauf des Wahlkampfes widerspiegeln und nicht irgendeine skurrilen Lotteriefekte demonstrieren. Allerdings gibt Frau Noelle-Neumann ihre Zahlen wie üblich in nackter Form - d.h. ohne irgendwelche Spielräume - an. Ein Vergleich mit den 10 repräsentativen Querschnitten in Abschnitt I, Seite 9, die alle aus der gleichen Grundgesamtheit stammen, zeigt, daß ihre Trendmeldungen, gar nicht so aufregend sind. Es ist

²¹ Das Resultat der letzten Befragung hat Frau Noelle-Neumann nicht in die Welt veröffentlicht, sondern als offizielle Wahlprognose am Wahlabend um 1800 im Satellitenfernsehen SAT1 bekanntgegeben. Wir haben ihr Resultat der Vollständigkeit halber angeführt. Nach ihren Angaben in SAT1 basierte ihre offizielle Prognose auf 2000 Interviews in drei Umfragen - und nicht auf 1000 wie zuvor im Welt-Wahlbarometer. Diese ist daher nicht direkt vergleichbar mit den vorangehenden "Stimmungsbildern" in DIE WELT.

zu beachten, daß die Umfragen von Frau Noelle-Neumann über einen Zeitraum von fast drei Monaten und daher in verschiedenen Grundgesamtheiten stattgefunden haben. Im WELT-Wahlbarometer variierte die CDU/CSU zwischen 44,0% und 48,7%, die SPD zwischen 32,4% und 37,3%, die FDP zwischen 6,0% und 9,6% und die Grünen zwischen 8,3% und 11,7%. Bei den aus der gleichen Grundgesamtheit ausgelosten repräsentativen Querschnitten (Vorwort S.I/9) hingegen variierte die CDU/CSU zwischen 40,3% und 47,1%, die SPD zwischen 34,3% und 41,1%, die FDP zwischen 7,8% und 10,8% und die Grünen zwischen 6,9% und 9,4%. Es scheint also, daß „der repräsentative Querschnitt“ auf seine Erstellung per Lotterie mindestens so empfindlich reagiert, wie auf das Gepolter eines Franz Josef Strauß oder den Kassensturz bei der Neuen Heimat. Wir stellen uns daher die Frage, welche Trendmeldungen herauskommen werden, wenn sich die Parteistärken tatsächlich nicht geändert haben, wenn also in Wirklichkeit kein Trend vorliegt. Falls Frau Noelle-Neumanns Ausführungen im Welt-Wahlbarometer überhaupt einen Sinn ergeben sollen, muß man offensichtlich unterstellen, daß sich in diesem Fall ein Bild von praktisch bedeutungslosen Trends ergibt. Trifft dies wirklich zu? Zu welchen Trend-Kapriolen ist der „repräsentative Querschnitt“ fähig, weil er per Lotterie erstellt wird? Dazu gehen wir vom Ergebnis der Bundestagswahl 1987 aus: CDU/CSU 44,3%, SPD 37,0%, FDP 9,1%, Grüne 8,3% und Rest 1,3%. Wir erstellen nun per Lotterie Paare von repräsentativen Querschnitten von jeweils 1000 Wahlberechtigten, die bei der Bundestagswahl eine gültige Stimme abgegeben haben, und werten die damals abgegebenen Stimmzettel aus. Dies ist auf einem Computer ganz einfach durchzuführen. Wie der Leser auf den folgenden Seiten sehen wird, entpuppen sich die Trends, die in etwa Null sein sollten, als wilde Pseudo-Trends. Auf der nächsten Seite sind die Resultate von 10 simulierten Paarumfragen abgedruckt, wobei das Ergebnis der ersten und zweiten Umfrage sowie der daraus resultierende Pseudo-Trend angegeben. Auf der übernächsten Seite sind die Resultate von 100 simulierten Paarumfragen in abgekürzter Form aufgeführt, d.h. es werden nur noch die Pseudo-Trends angeführt. Dann folgt ein Vergleich.

Was der Zufall so alles aus eigener Kraft schafft...
Pseudo-Trends bei Wiederholungsbefragungen

1. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	44,2	35,7	8,9	9,3	1,9
2.Umfrage	40,3	41,1	8,6	9,0	1,0
Pseudotrend	-3,9	+5,4	-0,3	-0,3	-0,9

2. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	45,5	35,8	7,9	9,5	1,3
2.Umfrage	44,1	38,4	8,7	7,7	1,1
Pseudotrend	-1,4	+2,6	+0,8	-1,8	-0,2

3. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	44,7	36,7	8,9	8,6	1,1
2.Umfrage	43,9	38,5	8,8	8,2	0,6
Pseudotrend	-0,8	+1,8	-0,1	-0,4	-0,5

4. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	45,4	35,8	10,2	7,3	1,3
2.Umfrage	47,1	35,6	7,8	8,3	1,2
Pseudotrend	+1,7	-0,2	-2,4	+1,0	-0,1

5. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	42,2	39,1	9,8	6,7	2,2
2.Umfrage	43,1	37,4	9,5	8,8	1,2
Pseudotrend	+0,9	-1,7	-0,3	+2,1	-1,0

6. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	43,3	37,4	9,1	8,6	1,6
2.Umfrage	47,0	36,3	8,4	6,9	1,4
Pseudotrend	+3,7	-1,1	-0,7	-1,7	-0,2

7. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	43,8	35,7	9,3	10,0	1,2
2.Umfrage	44,5	34,3	10,3	9,4	1,5
Pseudotrend	+0,7	-1,4	+1,0	-0,6	+0,3

8. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	44,7	37,8	9,5	7,0	1,0
2.Umfrage	43,5	38,9	8,4	7,7	1,5
Pseudotrend	-1,2	+1,1	-1,1	+0,7	+0,5

9. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	45,6	35,5	9,1	8,0	1,8
2.Umfrage	42,0	37,9	10,8	8,1	1,2
Pseudotrend	-3,6	+2,4	+1,7	+0,1	-0,6

10. Wiederholung

	CDU/CSU	SPD	FDP	Grüne	Rest
1.Umfrage	46,8	35,1	8,4	8,3	1,4
2.Umfrage	43,7	38,9	8,7	6,9	1,8
Pseudotrend	-3,1	+3,8	+0,3	-1,4	+0,4

Was der Zufall so alles aus eigener Kraft schafft...

Pseudo-Trends bei Wiederholungsbefragungen

	CDU/CSU	SPD	FDP	Grüne	Sonst		CDU/CSU	SPD	FDP	Grüne	Sonst
1. Wdh.	-3,9	+5,4	-0,3	-0,3	-0,9	51. Wdh.	-1,4	+0,8	+0,3	-1,1	+1,4
2. Wdh.	-1,4	+2,6	+0,8	-1,8	-0,2	52. Wdh.	+0,8	-0,2	+0,9	-1,4	-0,1
3. Wdh.	-0,8	+1,8	-0,1	-0,4	-0,5	53. Wdh.	+0,9	+0,1	+0,8	-2,3	+0,5
4. Wdh.	+1,7	-0,2	-2,4	+1,0	-0,1	54. Wdh.	+1,5	0,0	-1,6	-0,2	+0,3
5. Wdh.	+0,9	-1,7	-0,3	+2,1	-1,0	55. Wdh.	-0,9	+2,1	-0,7	-0,6	+0,1
6. Wdh.	+3,7	-1,1	-0,7	-1,7	-0,2	56. Wdh.	-1,7	+1,2	-1,3	+1,0	+0,8
7. Wdh.	+0,7	-1,4	+1,0	-0,6	+0,3	57. Wdh.	-0,4	+0,3	+2,4	-1,7	-0,6
8. Wdh.	-1,2	+1,1	-1,1	+0,7	+0,5	58. Wdh.	+1,1	-1,4	+1,6	-1,0	-0,3
9. Wdh.	-3,6	+2,4	+1,7	+0,1	-0,6	59. Wdh.	+2,3	-2,6	-1,7	+1,6	+0,4
10. Wdh.	-3,1	+3,8	+0,3	-1,4	+0,4	60. Wdh.	-4,5	+5,0	+0,2	-0,7	0,0
11. Wdh.	+0,7	-0,3	-0,6	-0,8	+1,0	61. Wdh.	+6,1	-4,7	-2,7	+0,7	+0,6
12. Wdh.	-1,0	-0,4	+0,1	-0,2	+1,5	62. Wdh.	+3,1	-4,4	+0,9	-0,7	+1,1
13. Wdh.	-1,8	+1,8	+0,7	-0,5	-0,2	63. Wdh.	-1,3	-1,5	+1,7	+0,6	+0,5
14. Wdh.	-0,5	-0,7	+0,2	+0,9	+0,1	64. Wdh.	-1,2	+1,1	-0,7	+1,0	-0,2
15. Wdh.	+3,0	-1,3	+0,4	-2,4	+0,3	65. Wdh.	-5,7	+2,4	+3,1	+0,8	-0,6
16. Wdh.	-1,0	+3,1	-1,1	-0,7	-0,3	66. Wdh.	+0,1	+3,2	-2,7	-1,3	+0,7
17. Wdh.	-0,9	+1,8	+0,6	-1,5	0,0	67. Wdh.	+0,2	-0,8	+1,8	-1,6	+0,4
18. Wdh.	-1,2	+0,3	+0,1	+0,5	+0,3	68. Wdh.	-1,6	-0,7	+0,8	+2,4	-0,9
19. Wdh.	-2,1	-1,2	0,0	+2,3	+1,0	69. Wdh.	+1,6	-2,9	-1,0	+0,9	+1,4
20. Wdh.	-2,4	+1,0	+1,9	-0,6	+0,1	70. Wdh.	-0,3	-1,9	+2,2	+0,3	-0,3
21. Wdh.	-0,4	+2,4	-0,9	-0,7	-0,4	71. Wdh.	+2,3	-0,7	+0,2	-1,6	-0,2
22. Wdh.	-2,7	+1,9	+0,8	-0,4	+0,4	72. Wdh.	-3,8	+1,4	+1,5	+1,1	-0,2
23. Wdh.	+1,1	-0,1	+2,2	-2,7	-0,5	73. Wdh.	-2,3	-1,1	+1,2	+2,6	-0,4
24. Wdh.	+3,7	-2,8	+0,1	-0,5	-0,5	74. Wdh.	+1,6	-1,7	-0,7	+1,1	-0,3
25. Wdh.	+1,6	-1,5	-1,1	+1,6	-0,6	75. Wdh.	+2,0	-1,8	-0,1	+0,1	-0,2
26. Wdh.	-0,7	+1,8	-0,2	-0,8	-0,1	76. Wdh.	-0,6	+0,3	+0,1	+0,3	-0,1
27. Wdh.	-0,9	+0,7	-0,3	-0,3	+0,8	77. Wdh.	+0,5	-1,8	+0,1	+0,4	+0,8
28. Wdh.	-0,8	+0,5	+0,4	-0,7	+0,6	78. Wdh.	-1,2	-1,7	+1,3	+1,4	+0,2
29. Wdh.	+3,0	-1,7	0,0	-0,7	-0,6	79. Wdh.	+1,3	-0,8	+0,3	-0,8	0,0
30. Wdh.	-1,5	-2,8	+1,8	+1,6	+0,9	80. Wdh.	-1,5	+1,6	+0,1	-0,8	+0,6
31. Wdh.	0,0	+0,6	+0,6	-0,6	-0,6	81. Wdh.	+0,7	-0,6	-0,7	+0,3	+0,3
32. Wdh.	+4,0	-1,4	+0,4	-1,7	-1,3	82. Wdh.	-2,5	-0,4	+2,9	+0,7	-0,7
33. Wdh.	-1,7	-0,2	+1,4	-0,2	+0,7	83. Wdh.	-2,4	+3,1	-1,7	+1,5	-0,5
34. Wdh.	-0,9	+0,1	-1,7	+1,8	+0,7	84. Wdh.	+3,3	-1,7	-1,0	-0,6	0,0
35. Wdh.	-1,4	+1,0	+0,3	+0,4	-0,3	85. Wdh.	-2,2	+2,2	+0,2	-0,2	0,0
36. Wdh.	-0,3	-0,9	+0,2	+1,6	-0,6	86. Wdh.	-2,3	+0,3	+0,4	+1,8	-0,2
37. Wdh.	-0,6	-0,9	+0,5	-0,5	+1,5	87. Wdh.	+1,8	-0,8	-2,4	+1,0	+0,4
38. Wdh.	+6,8	-2,5	-2,4	-1,2	-0,7	88. Wdh.	+1,2	-1,8	+2,2	-1,9	+0,3
39. Wdh.	-1,5	-1,9	+2,8	+0,5	+0,1	89. Wdh.	-1,9	+2,2	+0,5	-1,3	+0,5
40. Wdh.	0,0	+2,6	-0,8	-1,9	+0,1	90. Wdh.	-0,8	-1,1	-0,6	+2,2	+0,3
41. Wdh.	-2,0	-0,2	-0,2	+1,9	+0,5	91. Wdh.	+1,4	+2,5	-2,9	-1,2	+0,2
42. Wdh.	-0,7	+0,2	+1,1	-1,1	+0,5	92. Wdh.	+0,3	-2,3	+3,2	-1,2	0,0
43. Wdh.	+2,5	-4,7	+0,1	+2,2	-0,1	93. Wdh.	+1,0	-3,4	+1,0	+1,1	+0,3
44. Wdh.	-4,4	+3,1	+1,6	+0,4	-0,7	94. Wdh.	-4,7	+1,4	+0,7	+2,9	-0,3
45. Wdh.	+1,8	-0,6	-0,6	-0,4	-0,2	95. Wdh.	-1,1	+2,9	-0,2	-1,5	-0,1
46. Wdh.	-3,3	+0,9	+0,7	+1,5	+0,2	96. Wdh.	-0,2	-3,5	+1,0	+3,3	-0,6
47. Wdh.	-5,9	+2,8	+3,8	-0,3	-0,4	97. Wdh.	-0,6	-1,5	+1,8	+0,2	+0,1
48. Wdh.	+2,8	-1,9	-0,2	-0,7	0,0	98. Wdh.	-0,3	-2,4	+2,7	+0,7	-0,7
49. Wdh.	-0,9	+2,1	-0,4	-0,4	-0,4	99. Wdh.	-1,5	+0,3	+1,4	-0,2	0,0
50. Wdh.	+1,9	-2,5	+1,0	-0,6	+0,2	100. Wdh.	+2,4	-2,9	+0,3	+0,3	-0,1

Vergleiche zwischen den Pseudo-Trends und den Trendaussagen in DIE WELT

Um die 100 Pseudo-Trends der vorigen Seite mit den Trendmeldungen von Frau Noelle-Neumann im WELT-Wahlbarometer zu vergleichen, sind im folgenden zuerst die 10 kleinsten Pseudo-Trends, dann ihre Trendmeldungen und schließlich die 10 größten Pseudo-Trends aufgeführt:

Die 10 kleinsten Pseudo-Trends:

	CDU/CSU	SPD	FDP	Grüne	Rest
Pseudo-Trend der 76. Wdh.	-0,6%	+0,3%	+0,1%	+0,3%	-0,1%
Pseudo-Trend der 27. Wdh.	-0,9%	+0,7%	-0,3%	-0,3%	+0,8%
Pseudo-Trend der 37. Wdh.	-0,6%	-0,9%	+0,5%	-0,5%	+1,5%
Pseudo-Trend der 12. Wdh.	-1,0%	-0,4%	+0,1%	-0,2%	+1,5%
Pseudo-Trend der 31. Wdh.	0,0%	+0,6%	+0,6%	-0,6%	-0,6%
Pseudo-Trend der 18. Wdh.	-1,2%	+0,3%	+0,1%	+0,5%	+0,3%
Pseudo-Trend der 81. Wdh.	+0,7%	-0,6%	-0,7%	+0,3%	+0,3%
Pseudo-Trend der 28. Wdh.	-0,8%	+0,5%	+0,4%	-0,7%	+0,6%
Pseudo-Trend der 11. Wdh.	+0,7%	-0,3%	-0,6%	-0,8%	+1,0%
Pseudo-Trend der 79. Wdh.	+1,3%	-0,8%	+0,3%	-0,8%	0,0%

Zeitliches Auf und Ab der Parteistärken im WELT-Wahlbarometer

	CDU/CSU	SPD	FDP	Grüne	Rest
Trend von 1. zu 2. Befragung	-1,3%	-2,0%	+1,8%	+1,7%	-0,2%
Trend von 2. zu 3. Befragung	+2,0%	+0,1%	-1,7%	-0,5%	+0,1%
Trend von 3. zu 4. Befragung	+0,8%	+0,3%	+0,4%	-1,2%	-0,3%
Trend von 4. zu 5. Befragung	-1,1%	-1,6%	+1,6%	+0,8%	+0,3%
Trend von 5. zu 6. Befragung	+1,8%	-1,0%	-0,7%	-0,1%	0,0%
Trend von 6. zu 7. Befragung	-1,9%	+2,4%	-1,8%	+1,0%	+0,3%
Trend von 7. zu 8. Befragung	-2,0%	+1,9%	+2,5%	-2,2%	-0,2%
Trend von 8. zu 9. Befragung	-0,8%	-1,4%	+1,1%	+0,9%	+0,2%
Trend von 9. zu 10. Befragung	+0,6%	+2,0%	-0,7%	-2,1%	+0,2%
Trend von 10. zu 11. Befragung	+1,4%	-1,3%	-0,4%	+0,7%	-0,4%

Die 10 größten Pseudo-Trends:

	CDU/CSU	SPD	FDP	Grüne	Rest
Pseudo-Trend der 38. Wdh.	+6,8%	-2,5%	-2,4%	-1,2%	-0,7%
Pseudo-Trend der 47. Wdh.	-5,9%	+2,8%	+3,8%	-0,3%	-0,4%
Pseudo-Trend der 61. Wdh.	+6,1%	-4,7%	-2,7%	+0,7%	+0,6%
Pseudo-Trend der 96. Wdh.	-0,2%	-3,5%	+1,0%	+3,3%	-0,6%
Pseudo-Trend der 65. Wdh.	-5,7%	+2,4%	+3,1%	+0,8%	-0,6%
Pseudo-Trend der 92. Wdh.	+0,3%	-2,3%	+3,2%	-1,2%	0,0%
Pseudo-Trend der 1. Wdh.	-3,9%	+5,4%	-0,3%	-0,3%	-0,9%
Pseudo-Trend der 82. Wdh.	-2,5%	-0,4%	+2,9%	+0,7%	-0,7%
Pseudo-Trend der 91. Wdh.	+1,4%	+2,5%	-2,9%	-1,2%	+0,2%
Pseudo-Trend der 94. Wdh.	-4,7%	+1,4%	+0,7%	+2,9%	-0,3%

Man braucht nichts von Statistik zu verstehen, um aus dem obigen Vergleich der 100 Pseudo-Trends mit den Trendmeldungen von Frau Noelle-Neumann den Schluß zu ziehen, daß ihre Wahlkampfverlauf-Geschichten im Welt-Wahlbarometer eine gewisse Ähnlichkeit mit Hoffmanns Erzählungen aufweisen. Offensichtlich reagiert der „repräsentative Querschnitt“ auf die Auswahl per Lotterie genau so spontan und heftig wie auf das Kasperle-Theater zwischen Strauß und Genscher und die Pannen und Querelen innerhalb der grünroten Des-Organisation.

Das erstaunliche ist, daß Frau Noelle-Neumann von ihren Trend-Geschichten überzeugt ist, wie sie am Wahlabend im Fernsehen (SAT1) in eindrücklicher Weise demonstrierte. Nachdem sie, wie bereits erwähnt, um 1800 ihre Prognose verlesen hatte - CDU/CSU 46%, SPD 36%, FDP 8,5%, Grüne 9,0% -, wurden um 1830 die ersten Hochrechnungen für die CDU/CSU im ZDF (45%) und in der ARD (41%) bekannt gegeben, die sich völlig widersprachen. Zu dieser Situation befragt, entwickelte sich der folgende Dialog:

Reporter von SAT1:

„Es geht ja wohl ein bißchen drunter und drüber. Die einen sehen die CDU bei 45%, die anderen bei 41%. Wie kann das denn nun noch ausgehen?“

Frau Noelle-Neumann:

„Also ich kann nur sagen, daß wir ja gesehen haben, daß die CDU in der Weihnachtspause von ihrem Hoch herunterfiel bis auf 44%.“

Während sie sprach, griff sie zur Illustration ihrer Ausführungen zu einer Graphik, in welcher der CDU/CSU-Anteil von Oktober 1986 bis Januar 1987 gemäß ihren WELT-Wahlbarometer-Daten dargestellt war. Daß diese Verlaufskurve mit der gestellten Frage betreffend der sich widersprechenden Hochrechnungen nichts zu tun hatte, schien weder sie noch den Reporter zu stören, und sie fuhr fort:

„Ich kann mir denken, daß das bis 44% tatsächlich noch immer runtergegangen ist. Aber ich rechne damit, daß es auf 45% mindestens wieder hochgegangen ist und daß das zusammen mit der FDP, das ist die wichtige Aussage, eben die Mehrheit der Sitze sein wird.

...

Also jedenfalls augenblicklich muß man sagen, daß die CDU deutlich durch die Weihnachtspause verloren hat an einer Führung, die damals bei 48% lag.“

In ihrem Buch „Umfragen in der Massengesellschaft“ legt Frau Noelle-Neumann ein uneingeschränktes Plädoyer für die Auswahl des repräsentativen Querschnittes per Lotterie ab, wobei sie das Synonym „Zufallsauswahl“ vorzieht. Als begeisterte Anhängerin und Verfechterin der Zufallsauswahl bleibt ihr nur noch die (statistisch gesehen) kleine Hoffnung, daß die obigen 100 Pseudo-Trends zufällig völlig aus der Reihe tanzten und deshalb nicht „repräsentativ“ seien. Um diesem (im Prinzip berechtigten) Einwand zu begegnen, wurden eine halbe Million Auslosungen von Querschnitten vom Umfang 10000 vorgenommen - und die Resultate auf Querschnitte vom Umfang 1000 umgerechnet. Die Resultate sind auf der folgenden Seite graphisch dargestellt. (Eine optisch ins Gewicht fallende Glättung war nicht erforderlich.) Wie man sieht, zeigt die Verteilungsfunktion der Pseudo-Trends einen ähnlichen Verlauf wie die in Abschnitt VII betrachteten Vier-Parteien-Abweichungen. Sie ist zweibucklig, hat in der Mitte ein Loch, aber die Buckel sind im Vergleich zu VII/8 weiter nach außen gerutscht und das Loch in der Mitte ist größer. Die häufigsten Pseudo-Trends liegen für die großen Parteien bei etwa $\pm 3\%$ und für die kleinen bei etwa $\pm 1,8\%$. Während bei der Standard-Normalverteilung rund 68% bzw.

95% innerhalb der einfachen bzw. doppelten Standardabweichung liegen, befinden sich lediglich etwa 12% bzw. 56% der Pseudo-Trends innerhalb der gleichen Spielräume. In der Graphik auf der folgenden Seite sind die obigen 10 kleinsten Pseudo-Trends auf den Achsen als Kreise o eingezeichnet. Wie zu erwarten, kuscheln sie sich in der Mitte, im Loch der Verteilung. Die obigen 10 größten Pseudo-Trends sind als Quadrate mit Diagonalen \boxtimes markiert, sie sind weit außen in der Verteilung zu finden. Zwischen den zehn größten und kleinsten Pseudo-Trends liegen die „Trendmeldungen“ von Frau Noelle-Neumann im WELT-Wahlbarometer, die als schwarze Punkte (gekennzeichnet sind.²² Wie man sieht, liegen ihre Trendmeldungen unter den beiden Buckeln der Verteilungsfunktion, sie erfreuen sich also der Gesellschaft der am häufigsten vorkommenden Pseudo-Trends.

²² Man beachte, daß die Trendmeldungen von Frau Noelle-Neumann nicht zur Grundgesamtheit der Pseudo-Trends gehören. Sie sind lediglich zu Vergleichszwecken in die Graphik aufgenommen worden. Ihre Trendmeldungen stellen eine Überlagerung der drei Komponenten Pseudotrends, tatsächliche Änderungen der Parteistärken (von Oktober 1986 bis Januar 1987) und Gewichtung (vergleiche Abschnitt X) dar. Mit dem Vergleich soll illustriert werden, daß aus ihren Trendmeldungen nichts über die tatsächlichen Änderungen ausgesagt werden kann. Die Pseudo-Trends sind nämlich oft größer als die tatsächlichen Änderungen. Meßbar ist nur eine Überlagerung von Pseudo-Trends und tatsächlichen Änderungen. Eine Aufteilung auf die beiden Komponenten ist aber offensichtlich nicht möglich.

Verteilungsfunktionen von Pseudotrends

Wie werden die Pseudo-Trends auf der Skala eingetragen?

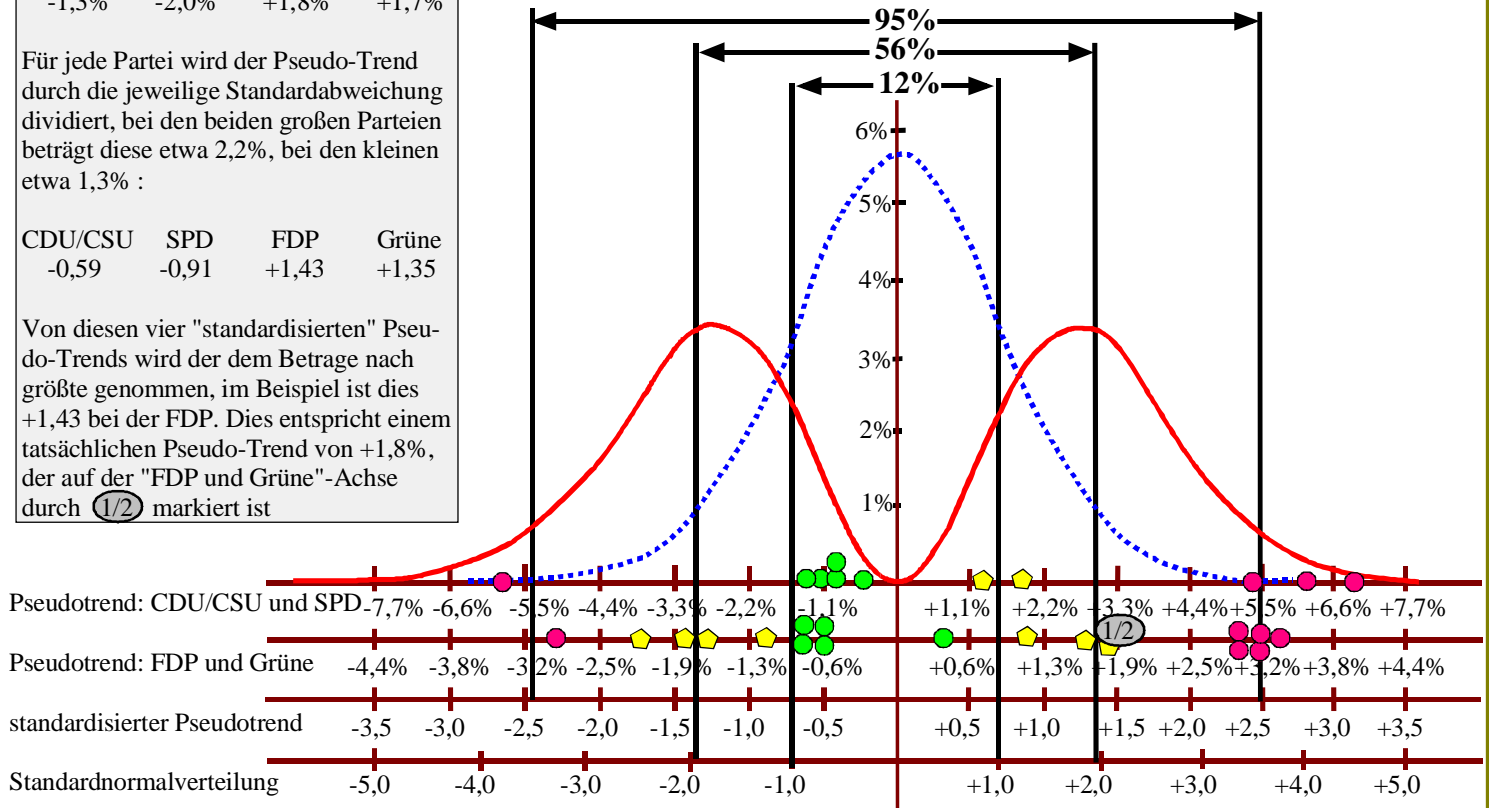
Die Pseudo-Trends betragen:

CDU/CSU	SPD	FDP	Grüne
-1,3%	-2,0%	+1,8%	+1,7%

Für jede Partei wird der Pseudo-Trend durch die jeweilige Standardabweichung dividiert, bei den beiden großen Parteien beträgt diese etwa 2,2%, bei den kleinen etwa 1,3% :

CDU/CSU	SPD	FDP	Grüne
-0,59	-0,91	+1,43	+1,35

Von diesen vier "standardisierten" Pseudo-Trends wird der dem Betrage nach größte genommen, im Beispiel ist dies +1,43 bei der FDP. Dies entspricht einem tatsächlichen Pseudo-Trend von +1,8%, der auf der "FDP und Grüne"-Achse durch $\frac{1}{2}$ markiert ist



- ◆ Trendmeldungen von Frau Noelle-Neumann von Nov. 1986 bis Jan. 1987 in exklusiv Welt Wahlbarometer
- Die zehn kleinsten Pseudotrends
- ⋯ Standardnormalverteilung
- Die zehn größten Pseudotrends
- Verteilungsfunktion der Pseudotrends

In der folgenden Tabelle sind einige Werte der obigen Verteilungsfunktion aufgeführt. In der ersten und zweiten Spalte sind die gewählten Grenzen aufgeführt. In der dritten Spalte ist die Chance (=Wahrscheinlichkeit) dafür angegeben, daß die Pseudo-Trends für alle vier Parteien die gewählten Grenzen nicht überschreiten.

Gewählter Spielraum²³		Chance, daß der gewählte Spielraum eingehalten wird²⁴
CDU/CSU & SPD	FDP & Grüne	
bis $\pm 0,5\%$	bis $\pm 0,3\%$	0,25%
bis $\pm 1,0\%$	bis $\pm 0,6\%$	3,1 %
bis $\pm 1,7\%$	bis $\pm 1,0\%$	14,4 %
bis $\pm 2,0\%$	bis $\pm 1,1\%$	19,9 %
bis $\pm 3,0\%$	bis $\pm 1,7\%$	50,4 %
bis $\pm 3,5\%$	bis $\pm 2,0\%$	64,9 %
bis $\pm 4,0\%$	bis $\pm 2,3\%$	76,8 %
bis $\pm 5,0\%$	bis $\pm 2,9\%$	91,7 %
bis $\pm 5,2\%$	bis $\pm 3,0\%$	93,2 %
bis $\pm 5,5\%$	bis $\pm 3,2\%$	95,4 %
bis $\pm 6,0\%$	bis $\pm 3,4\%$	97,3 %

Wie kann ein Wahlforscher auf der Basis von Meinungsumfragen von einem „Trend“ reden, wenn er diesen nicht von einem Pseudo-Trend unterscheiden kann? Aus der obigen Tabelle geht nämlich hervor : Haben sich die Parteistärken nicht geändert, dann täuschen zwei aufeinanderfolgende Meinungsumfragen von je 1000 Interviews mit einer Wahrscheinlichkeit von fast 50% vor, daß für mindestens eine Partei eine markante Veränderung eingetreten ist: Laut Wiederholungsbefragung ist eine der beiden großen Parteien (CDU/CSU oder SPD) um mindestens 3% oder eine der beiden kleinen (FDP oder Grüne) um mindestens 1.7% gestiegen/gefallen, während sich in Wirklichkeit nichts geändert hat!

Will ein Wahlforscher mit der üblichen statistischen Sicherheit von 95% ausschließen, daß der von ihm „ermittelte“ Vier-Parteien-Trend nicht für mindestens eine der vier Parteien eine Laune des Zufalls darstellt, dann muß er gegenwärtig für eine der großen Parteien eine Änderung von mindestens $\pm 5,5\%$ oder für eine der kleinen eine solche von mindestens $\pm 3,2\%$ nachweisen (bei je 1000 Interviews)²⁵.

Um es boshaft auszudrücken: Die „Gunst“ des Zufalls beschert den Wahlforschern für

²³ Der Spielraum für die Pseudo-Trends wurde für die kleinen Parteien um gut 40% kleiner gewählt als für die großen. Das Verhältnis der Spielräume kann man in einem gewissen Rahmen frei wählen, wobei inhaltliche und statistische Betrachtungen eine Rolle spielen. Bei den Parteistärken, wie sie sich bei der Bundestagswahl 1987 ergeben haben, scheint aus statistischer Sicht ein Verhältnis von etwa 1:0,6 bis 1:0,55 "optimal". (Technisch ausgedrückt: Die Pseudo-Trends werden mit den Standardabweichungen der Parteistärken gewichtet. Da die Standardabweichungen für die SPD und Union praktisch gleich sind - und ebenso für die FDP und die Grünen - führt dies zu einer Gewichtung der Pseudo-Trends im Verhältnis von etwa 1:0,57)

²⁴ Die Werte geben das Resultat einer Simulation von 1.000.000 Auslosungen an. Aufgrund der Simulationsungenauigkeit ist die Stelle nach dem Komma nicht gesichert, die tatsächlichen Werte können sich um etwa $\pm 0,1\%$ von den angegebenen Werten unterscheiden, beim ersten Wert (0,25%) beträgt die Simulationsungenauigkeit etwa $\pm 0,01\%$.

²⁵ Haben sich nämlich die Parteistärken nicht geändert, dann haben immer noch 5% aller Doppelumfragen vom Umfang 1000 die Eigenschaft, daß für mindestens eine der großen Parteien eine Änderung von mehr als $\pm 5,5\%$ oder für eine der kleinen eine solche von mehr als $\pm 3,2\%$ eingetreten ist.

mindestens eine Partei laufend Neuigkeiten und Schlagzeilen. Was sich wirklich abspielt, das weiß kein Mensch, und die von den Meinungsforschungsinstituten gelieferten Daten über das zeitliche Auf und Ab der Parteistärken - z.B. CDU/CSU -2.8%, SPD +1,9%, FDP +1.3%, Grüne - 0.4% - haben reinen „Horoskopcharakter“. Sie täuschen Veränderungen vor, die quantitativ mit an Sicherheit grenzender Wahrscheinlichkeit falsch sind und die häufig nicht einmal qualitativ richtig sind.

IX Der zeitliche Trend über mehrere Monate

Begnügen sich die Wahlprognosen mit den Stimmen für die einzelnen Parteien, so versuchen alle Institute auch Trends über mehrere Monate hinweg aufzuzeigen.

Meines Wissens legt allein die Forschungsgruppe Wahlen e.V. auch Rohdaten - also die bei der Umfrage tatsächlich ermittelten Zahlen - offen dar, die sie für das „Politbarometer“ des Zweiten Deutschen Fernsehens (ZDF) erfaßt. Darum stütze ich mich bei den folgenden Analysen auf sie. Es hat nämlich wenig Sinn, mit willkürlich „korrigierten“ Rohdaten Statistik zu betreiben.

Im „Politbarometer“ werden die Parteistärken über jeweils vier Monate hinweg miteinander verglichen. Bezugspunkt ist aber nicht etwa derselbe repräsentative Querschnitt, der immer wieder befragt wird, sondern es wird jedesmal ein neuer zusammengestellt, es werden also jeden Monat 1000 Wahlberechtigte neu ausgelost und befragt.

In den Monaten Mai bis August 1986 ermittelte das Politbarometer des ZDF folgende Zahlen (in Prozent)

	Mai	Juni	Juli	August
SPD	44,0	46,0	43,0	43,0
CDU/CSU	39,0	42,0	44,0	44,0
FDP	5,0	4,0	5,0	4,0
Grüne	12,0	8,0	8,0	9,0

In den Kommentaren dazu hieß es:

„Zum erstenmal seit Januar dieses Jahres lag die SPD (mit 43%), wenn auch knapp, hinter der Union (mit 44%). ...

FDP: schwankend zwischen 5% und 4%, derzeit 4%.“

Diese Aussage und die graphische Trenddarstellung suggerieren eine Genauigkeit von $\pm 0,5\%$, denn nur dann haben sie einen Sinn.

In welchem Maße ist diese Befragung reproduzierbar? Was wäre herausgekommen, wenn man in denselben Monaten jeweils 1000 andere Wahlberechtigte ausgelost, d.h. andere sogenannte „repräsentative Querschnitte“ befragt hätte?

Um den rein statistischen Aspekt des Problems zu behandeln, nehmen wir wieder an, daß die ausgelosten Wahlberechtigten nicht befragt werden, sondern wie in einer gut organisierten Volksdemokratie den ausgefüllten Stimmzettel für den Interviewer bereithalten; insbesondere auch jenes Drittel von Wahlberechtigten, das normalerweise dem Interviewer die Aussage verweigert oder von ihm nicht aufzuspüren ist. Wir wollen der Demoskopie damit nicht etwa die Existenzgrundlage entziehen, sondern lediglich den statistischen Kern des Problems klar herauschälen. Wir füllen also für jeden der Monate Mai bis August 1986 eine Trommel mit Stimmzetteln gemäß den Angaben des ZDF, beispielsweise für den Mai sind das 15.600.000 CDU/CSU-, 17.600.000 SPD-, 2.000.000 FDP- und 4.800.000 Grüne-Stimmen, zusammen 40.000.000. Das Ziehen der repräsentativen Querschnitte aus den Trommeln mit den „Stimmzetteln“ - d.h. die Auslosung von viermal 1000 Wahlberechtigten in den Monaten Mai bis August und die Auswertung ihrer Stimmzettel - haben wir mit einem perfekten Auswahlverfahren auf dem Computer durchgeführt (simuliert).

Wie viele dieser Querschnitte werden den behaupteten Trend reproduzieren, also die wahren Ergebnisse innerhalb von $\pm 0,5\%$ Genauigkeit liefern?

In den Tabellen auf den folgenden sieben Seiten sind die Resultate der ersten 35 Auslosungen von viermal 1000 Wahlberechtigten (bzw. von abgegebenen gültigen Stimmzetteln) abgedruckt, wie sie sich mit dem Startwert 0.12345 für den Zufallszahlengenerator der Rechenanlage Cyber 175 ergaben. Die Auslosungen sind mit Nummern von 1 bis 35 versehen und werden auch als Wiederholungen (der Auslosung) bezeichnet. Die 35 Wiederholungen können wir uns als Resultate von unter idealen Bedingungen arbeitenden Meinungsforschungsinstituten vorstellen, denn diese Ergebnisse sind dadurch zustande gekommen, daß 35 mal viermal 1000 Wahlberechtigte in den Monaten Mai, Juni, Juli und August 1986 ausgelost und „befragt“ wurden. Dabei wird vorausgesetzt, daß die vom ZDF ermittelten Umfrageergebnisse von Mai bis August 1986 zutreffen und daß für das ZDF der Interviewfehler im Sinne von Abschnitt III ausgeschaltet werden kann. Das letztere ist zwar völlig unrealistisch, aber selbst unter dieser idealisierenden Annahme werden die 35 Auslosungen des repräsentativen Querschnittes Resultate zeigen, die das ZDF aus Rücksicht auf die Einschaltquote nicht öffentlich eingestehen kann und will.

Die ersten 35 Wiederholungen des Politbarometers

1. Wiederholung

	Mai	Juni	Juli	August
SPD	43,9	49,7	41,9	44,0
CDU/CSU	40,4	38,4	46,0	41,9
FDP	3,7	4,2	4,7	4,6
Grüne	12,0	7,7	7,4	9,5

Abweichungen bei der 1. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,1	+3,7	-1,1	+1,0
CDU/CSU	+1,4	-3,6	+2,0	-2,1
FDP	-1,3	+0,2	-0,3	+0,6
Grüne	0,0	-0,3	-0,6	+0,5

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,6% und +3,7%, bei der FDP und den Grünen zwischen -1,3% und +0,6%

2. Wiederholung

	Mai	Juni	Juli	August
SPD	43,5	46,3	41,8	40,8
CDU/CSU	40,2	40,5	44,5	45,1
FDP	5,6	5,5	5,1	4,0
Grüne	10,7	7,7	8,6	10,1

Abweichungen bei der 2. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,5	+0,3	-1,2	-2,2
CDU/CSU	+1,2	-1,5	+0,5	+1,1
FDP	+0,6	+1,5	+0,1	0,0
Grüne	-1,3	-0,3	+0,6	+1,1

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,2% und +1,2%, bei der FDP und den Grünen zwischen -1,3% und +1,5%

3. Wiederholung

	Mai	Juni	Juli	August
SPD	45,9	46,6	43,9	40,9
CDU/CSU	37,6	41,9	44,2	44,5
FDP	4,4	3,7	5,0	4,4
Grüne	12,1	7,8	6,9	10,2

Abweichungen bei der 3. Wiederholung

	Mai	Juni	Juli	August
SPD	+1,9	+0,6	+0,9	-2,1
CDU/CSU	-1,4	-0,1	+0,2	+0,5
FDP	-0,6	-0,3	0,0	+0,4
Grüne	+0,1	-0,2	-1,1	+1,2

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,1% und +1,9%, bei der FDP und den Grünen zwischen -1,1% und +1,2%

4. Wiederholung

	Mai	Juni	Juli	August
SPD	44,7	45,7	42,7	44,4
CDU/CSU	40,0	43,9	42,7	41,9
FDP	4,2	3,7	5,4	4,4
Grüne	11,1	6,7	9,2	9,3

Abweichungen bei der 4. Wiederholung

	Mai	Juni	Juli	August
SPD	+0,7	-0,3	-0,3	+1,4
CDU/CSU	+1,0	+1,9	-1,3	-2,1
FDP	-0,8	-0,3	+0,4	+0,4
Grüne	-0,9	-1,3	+1,2	+0,3

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,1% und +1,9%, bei der FDP und den Grünen zwischen -1,3% und +1,2%

5. Wiederholung

	Mai	Juni	Juli	August
SPD	42,7	48,0	40,2	43,5
CDU/CSU	40,0	39,5	46,9	42,0
FDP	4,1	3,8	5,1	3,6
Grüne	13,2	8,7	7,8	10,9

Abweichungen bei der 5. Wiederholung

	Mai	Juni	Juli	August
SPD	-1,3	+2,0	-2,8	+0,5
CDU/CSU	+1,0	-2,5	+2,9	-2,0
FDP	-0,9	-0,2	+0,1	-0,4
Grüne	+1,2	+0,7	-0,2	+1,9

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,8% und +2,9%, bei der FDP und den Grünen zwischen -0,9% und +1,9%

6. Wiederholung

	Mai	Juni	Juli	August
SPD	44,2	45,9	43,8	44,2
CDU/CSU	38,6	42,0	42,1	44,0
FDP	5,3	3,7	5,8	3,9
Grüne	11,9	8,4	8,3	7,9

Abweichungen bei der 6. Wiederholung

	Mai	Juni	Juli	August
SPD	+0,2	-0,1	+0,8	+1,2
CDU/CSU	-0,4	0,0	-1,9	0,0
FDP	+0,3	-0,3	+0,8	-0,1
Grüne	-0,1	+0,4	+0,3	-1,1

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,9% und +1,2%, bei der FDP und den Grünen zwischen -1,1% und +0,8%

7. Wiederholung

	Mai	Juni	Juli	August
SPD	42,6	46,2	42,9	44,0
CDU/CSU	39,1	41,2	43,5	42,4
FDP	5,4	4,2	5,6	4,0
Grüne	12,9	8,4	8,0	9,6

Abweichungen bei der 7. Wiederholung

	Mai	Juni	Juli	August
SPD	-1,4	+0,2	-0,1	+1,0
CDU/CSU	+0,1	-0,8	-0,5	-1,6
FDP	+0,4	+0,2	+0,6	0,0
Grüne	+0,9	+0,4	0,0	+0,6

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,6% und +1,0%, bei der FDP und den Grünen zwischen 0,0% und +0,9%

8. Wiederholung

	Mai	Juni	Juli	August
SPD	45,0	43,7	42,6	43,7
CDU/CSU	39,3	43,3	43,7	41,8
FDP	4,7	4,0	5,4	4,5
Grüne	11,0	9,0	8,3	10,0

Abweichungen bei der 8. Wiederholung

	Mai	Juni	Juli	August
SPD	+1,0	-2,3	-0,4	+0,7
CDU/CSU	+0,3	+1,3	-0,3	-2,2
FDP	-0,3	0,0	+0,4	+0,5
Grüne	-1,0	+1,0	+0,3	+1,0

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,3% und +1,3%, bei der FDP und den Grünen zwischen -1,0% und +1,0%

9. Wiederholung

	Mai	Juni	Juli	August
SPD	42,8	45,6	39,8	41,5
CDU/CSU	40,6	41,4	46,5	46,2
FDP	5,1	4,1	4,9	3,9
Grüne	11,5	8,9	8,8	8,4

Abweichungen bei der 9. Wiederholung

	Mai	Juni	Juli	August
SPD	-1,2	-0,4	-3,2	-1,5
CDU/CSU	+1,6	-0,6	+2,5	+2,2
FDP	+0,1	+0,1	-0,1	-0,1
Grüne	-0,5	+0,9	+0,8	-0,6

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,2% und +2,5%, bei der FDP und den Grünen zwischen -0,6% und +0,9%

10. Wiederholung

	Mai	Juni	Juli	August
SPD	44,0	47,5	42,2	43,4
CDU/CSU	36,4	41,2	45,3	43,7
FDP	6,5	3,9	4,6	4,4
Grüne	13,1	7,4	7,9	8,5

Abweichungen bei der 10. Wiederholung

	Mai	Juni	Juli	August
SPD	0,0	+1,5	-0,8	+0,4
CDU/CSU	-2,6	-0,8	+1,3	-0,3
FDP	+1,5	-0,1	-0,4	+0,4
Grüne	+1,1	-0,6	-0,1	-0,5

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,6% und +1,5%, bei der FDP und den Grünen zwischen -0,6% und +1,5%

11. Wiederholung

	Mai	Juni	Juli	August
SPD	43,9	45,7	39,4	41,9
CDU/CSU	41,6	43,0	46,7	44,0
FDP	4,6	4,4	5,3	3,7
Grüne	9,9	6,9	8,6	10,4

Abweichungen bei der 11. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,1	-0,3	-3,6	-1,1
CDU/CSU	+2,6	+1,0	+2,7	0,0
FDP	-0,4	+0,4	+0,3	-0,3
Grüne	-2,1	-1,1	+0,6	+1,4

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,6% und +2,7%, bei der FDP und den Grünen zwischen -2,1% und +1,4%

12. Wiederholung

	Mai	Juni	Juli	August
SPD	43,5	44,8	45,4	42,4
CDU/CSU	40,6	42,6	43,2	45,0
FDP	3,1	4,2	3,7	4,3
Grüne	12,8	8,4	7,7	9,3

Abweichungen bei der 12. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,5	-1,2	+2,4	-0,6
CDU/CSU	+1,6	+0,6	-0,8	+1,0
FDP	-1,9	+0,2	-1,3	-0,7
Grüne	+0,8	+0,4	-0,3	+0,3

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,2% und +2,4%, bei der FDP und den Grünen zwischen -1,9% und +0,8%

13. Wiederholung

	Mai	Juni	Juli	August
SPD	45,3	45,7	43,7	45,1
CDU/CSU	37,5	42,2	44,2	40,8
FDP	5,7	4,5	5,7	4,2
Grüne	11,5	7,6	6,4	9,9

Abweichungen bei der 13. Wiederholung

	Mai	Juni	Juli	August
SPD	+1,3	-0,3	+0,7	+2,1
CDU/CSU	-1,5	+0,2	+0,2	-3,2
FDP	+0,7	+0,5	+0,7	+0,2
Grüne	-0,5	-0,4	-1,6	+0,9

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,2% und +2,1%, bei der FDP und den Grünen zwischen -1,6% und +0,9%

14. Wiederholung

	Mai	Juni	Juli	August
SPD	45,7	47,7	42,6	43,9
CDU/CSU	37,4	40,2	44,7	43,9
FDP	5,9	4,1	4,8	3,4
Grüne	11,0	8,0	7,9	8,8

Abweichungen bei der 14. Wiederholung

	Mai	Juni	Juli	August
SPD	+1,7	+1,7	-0,4	+0,9
CDU/CSU	-1,6	-1,8	+0,7	-0,1
FDP	+0,9	+0,1	-0,2	-0,6
Grüne	-1,0	0,0	-0,1	-0,2

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,8% und +1,7%, bei der FDP und den Grünen zwischen -1,0% und +0,9%

15. Wiederholung

	Mai	Juni	Juli	August
SPD	46,9	45,5	39,8	42,2
CDU/CSU	36,7	41,8	45,2	46,1
FDP	4,6	4,5	5,6	3,3
Grüne	11,8	8,2	9,4	8,4

Abweichungen bei der 15. Wiederholung

	Mai	Juni	Juli	August
SPD	+2,9	-0,5	-3,2	-0,8
CDU/CSU	-2,3	-0,2	+1,2	+2,1
FDP	-0,4	+0,5	+0,6	-0,7
Grüne	-0,2	+0,2	+1,4	-0,6

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,2% und +2,9%, bei der FDP und den Grünen zwischen -0,6% und +1,4%

16. Wiederholung

	Mai	Juni	Juli	August
SPD	44,8	46,1	46,0	41,9
CDU/CSU	39,3	41,3	43,0	43,3
FDP	4,2	4,1	4,3	4,5
Grüne	11,7	8,5	6,7	10,3

Abweichungen bei der 16. Wiederholung

	Mai	Juni	Juli	August
SPD	+0,8	+0,1	+3,0	-1,1
CDU/CSU	+0,3	-0,7	-1,0	-0,7
FDP	-0,8	+0,1	-0,7	+0,5
Grüne	-0,3	+0,5	-1,3	+1,3

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,1% und +3,0%, bei der FDP und den Grünen zwischen -1,3% und +1,3%

17. Wiederholung

	Mai	Juni	Juli	August
SPD	43,1	46,7	42,8	43,4
CDU/CSU	40,2	42,4	43,4	44,9
FDP	5,5	3,7	5,6	4,1
Grüne	11,2	7,2	8,2	7,6

Abweichungen bei der 17. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,9	+0,7	-0,2	+0,4
CDU/CSU	+1,2	+0,4	-0,6	+0,9
FDP	+0,5	-0,3	+0,6	+0,1
Grüne	-0,8	-0,8	+0,2	-1,4

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -0,9% und +1,2%, bei der FDP und den Grünen zwischen -1,4% und +0,6%

18. Wiederholung

	Mai	Juni	Juli	August
SPD	44,0	47,5	44,1	44,4
CDU/CSU	39,2	41,2	43,6	44,5
FDP	4,5	4,5	3,9	2,3
Grüne	12,3	6,8	8,4	8,8

Abweichungen bei der 18. Wiederholung

	Mai	Juni	Juli	August
SPD	0,0	+1,5	+1,1	+1,4
CDU/CSU	+0,2	-0,8	-0,4	+0,5
FDP	-0,5	+0,5	-1,1	-1,7
Grüne	+0,3	-1,2	+0,4	-0,2

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -0,8% und +1,5%, bei der FDP und den Grünen zwischen -1,7% und +0,5%

19. Wiederholung

	Mai	Juni	Juli	August
SPD	42,8	45,6	45,5	39,3
CDU/CSU	39,3	42,9	43,6	45,5
FDP	6,6	3,1	3,9	3,9
Grüne	11,3	8,4	7,0	11,3

Abweichungen bei der 19. Wiederholung

	Mai	Juni	Juli	August
SPD	-1,2	-0,4	+2,5	-3,7
CDU/CSU	+0,3	+0,9	-0,4	+1,5
FDP	+1,6	-0,9	-1,1	-0,1
Grüne	-0,7	+0,4	-1,0	+2,3

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,7% und +2,5%, bei der FDP und den Grünen zwischen -1,1% und +2,3%

20. Wiederholung

	Mai	Juni	Juli	August
SPD	43,1	46,8	43,7	42,6
CDU/CSU	39,2	42,3	43,9	43,8
FDP	5,3	3,3	4,6	4,6
Grüne	12,4	7,6	7,8	9,0

Abweichungen bei der 20. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,9	+0,8	+0,7	-0,4
CDU/CSU	+0,2	+0,3	-0,1	-0,2
FDP	+0,3	-0,7	-0,4	+0,6
Grüne	+0,4	-0,4	-0,2	0,0

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -0,9% und +0,8%, bei der FDP und den Grünen zwischen -0,7% und +0,6%

21. Wiederholung

	Mai	Juni	Juli	August
SPD	40,5	45,4	43,2	44,0
CDU/CSU	41,0	44,2	44,5	43,7
FDP	6,0	4,0	4,5	4,2
Grüne	12,5	6,4	7,8	8,1

Abweichungen bei der 21. Wiederholung

	Mai	Juni	Juli	August
SPD	-3,5	-0,6	+0,2	+1,0
CDU/CSU	+2,0	+2,2	+0,5	-0,3
FDP	+1,0	0,0	-0,5	+0,2
Grüne	+0,5	-1,6	-0,2	-0,9

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,5% und +2,2%, bei der FDP und den Grünen zwischen -1,6% und +1,1%

22. Wiederholung

	Mai	Juni	Juli	August
SPD	43,7	43,5	41,0	46,1
CDU/CSU	38,9	45,0	45,0	40,6
FDP	5,4	4,2	4,6	3,9
Grüne	12,0	7,3	9,4	9,4

Abweichungen bei der 22. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,3	-2,5	-2,0	+3,1
CDU/CSU	-0,1	+3,0	+1,0	-3,4
FDP	+0,4	+0,2	-0,4	-0,1
Grüne	0,0	-0,7	+1,4	+0,4

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,4% und +3,1%, bei der FDP und den Grünen zwischen -0,7% und +1,4%

23. Wiederholung

	Mai	Juni	Juli	August
SPD	46,8	46,1	41,3	44,6
CDU/CSU	36,9	43,0	43,5	42,8
FDP	4,8	4,1	5,6	4,5
Grüne	11,5	6,8	9,6	8,1

Abweichungen bei der 23. Wiederholung

	Mai	Juni	Juli	August
SPD	+2,8	+0,1	-1,7	+1,6
CDU/CSU	-2,1	+1,0	-0,5	-1,2
FDP	-0,2	+0,1	+0,6	+0,5
Grüne	-0,5	-1,2	+1,6	-0,9

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,1% und +2,8%, bei der FDP und den Grünen zwischen -1,2% und +1,6%

24. Wiederholung

	Mai	Juni	Juli	August
SPD	41,5	49,8	43,9	40,8
CDU/CSU	41,1	39,7	41,2	44,4
FDP	4,2	4,0	5,6	3,6
Grüne	13,2	6,5	9,3	11,2

Abweichungen bei der 24. Wiederholung

	Mai	Juni	Juli	August
SPD	-2,5	+3,8	+0,9	-2,2
CDU/CSU	+2,1	-2,3	-2,8	+0,4
FDP	-0,8	0,0	+0,6	-0,4
Grüne	+1,2	-1,5	+1,3	+2,2

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,8% und +3,8%, bei der FDP und den Grünen zwischen -1,5% und +2,2%

25. Wiederholung

	Mai	Juni	Juli	August
SPD	43,1	46,5	42,0	40,9
CDU/CSU	41,1	41,1	45,7	46,5
FDP	3,7	4,3	4,7	3,5
Grüne	11,9	8,1	7,6	9,1

Abweichungen bei der 25. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,7	+0,5	-1,0	-2,1
CDU/CSU	+2,1	-0,9	+1,7	+2,5
FDP	-1,3	+0,3	-0,3	-0,5
Grüne	-0,1	+0,1	-0,4	+0,1

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,1% und +2,5%, bei der FDP und den Grünen zwischen -1,3% und +0,3%

26. Wiederholung

	Mai	Juni	Juli	August
SPD	43,7	47,2	45,1	44,1
CDU/CSU	39,1	41,6	43,2	42,7
FDP	5,2	3,5	5,1	4,2
Grüne	12,0	7,7	6,6	9,0

Abweichungen bei der 26. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,3	+1,2	+2,1	+1,1
CDU/CSU	+0,1	-0,4	-0,8	-1,3
FDP	+0,2	-0,5	+0,1	+0,2
Grüne	0,0	-0,3	-1,4	0,0

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,3% und +2,1%, bei der FDP und den Grünen zwischen -1,4% und +0,2%

27. Wiederholung

	Mai	Juni	Juli	August
SPD	43,4	44,2	43,3	42,0
CDU/CSU	40,8	42,5	42,7	43,3
FDP	5,2	4,3	6,0	5,0
Grüne	10,6	9,0	8,0	9,7

Abweichungen bei der 27. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,6	-1,8	+0,3	-1,0
CDU/CSU	+1,8	+0,5	-1,3	-0,7
FDP	+0,2	+0,3	+1,0	+1,0
Grüne	-1,4	+1,0	0,0	+0,7

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,8% und +1,8%, bei der FDP und den Grünen zwischen -1,4% und +1,0%

28. Wiederholung

	Mai	Juni	Juli	August
SPD	43,4	45,2	42,2	44,1
CDU/CSU	41,9	43,7	43,3	42,9
FDP	4,3	3,8	6,2	3,8
Grüne	10,4	7,3	8,3	9,2

Abweichungen bei der 28. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,6	-0,8	-0,8	+1,1
CDU/CSU	+2,9	+1,7	-0,7	-1,1
FDP	-0,7	-0,2	+1,2	-0,2
Grüne	-1,6	-0,7	+0,3	+0,2

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,1% und +2,9%, bei der FDP und den Grünen zwischen -1,6% und +1,2%

29. Wiederholung

	Mai	Juni	Juli	August
SPD	42,7	45,1	45,7	45,1
CDU/CSU	40,4	42,2	40,6	41,3
FDP	5,1	3,5	4,6	4,5
Grüne	11,8	9,2	9,1	9,1

Abweichungen bei der 29. Wiederholung

	Mai	Juni	Juli	August
SPD	-1,3	-0,9	+2,7	+2,1
CDU/CSU	+1,4	+0,2	-3,4	-2,7
FDP	+0,1	-0,5	-0,4	+0,5
Grüne	-0,2	+1,2	+1,1	+0,1

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,4% und +2,7%, bei der FDP und den Grünen zwischen -0,5% und +1,2%

30. Wiederholung

	Mai	Juni	Juli	August
SPD	43,0	43,2	39,8	44,4
CDU/CSU	39,0	44,9	47,1	42,0
FDP	5,8	3,8	5,2	4,2
Grüne	12,2	8,1	7,9	9,4

Abweichungen bei der 30. Wiederholung

	Mai	Juni	Juli	August
SPD	-1,0	-2,8	-3,2	+1,4
CDU/CSU	0,0	+2,9	+3,1	-2,0
FDP	+0,8	-0,2	+0,2	+0,2
Grüne	+0,2	+0,1	-0,1	+0,4

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -3,2% und +3,1%, bei der FDP und den Grünen zwischen -0,2% und +0,8%

31. Wiederholung

	Mai	Juni	Juli	August
SPD	47,9	44,3	45,1	42,0
CDU/CSU	35,0	44,2	41,9	47,2
FDP	4,8	3,3	4,6	3,3
Grüne	12,3	8,2	8,4	7,5

Abweichungen bei der 31. Wiederholung

	Mai	Juni	Juli	August
SPD	+3,9	-1,7	+2,1	-1,0
CDU/CSU	-4,0	+2,2	-2,1	+3,2
FDP	-0,2	-0,7	-0,4	-0,7
Grüne	+0,3	+0,2	+0,4	-1,5

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,2% und +1,9%, bei der FDP und den Grünen zwischen -0,6% und +2,0%

32. Wiederholung

	Mai	Juni	Juli	August
SPD	41,8	45,0	41,8	42,5
CDU/CSU	39,8	43,9	44,8	43,7
FDP	4,4	3,6	5,0	4,1
Grüne	14,0	7,5	8,4	9,7

Abweichungen bei der 32. Wiederholung

	Mai	Juni	Juli	August
SPD	-2,2	-1,0	-1,2	-0,5
CDU/CSU	+0,8	+1,9	+0,8	-0,3
FDP	-0,6	-0,4	0,0	+0,1
Grüne	+2,0	-0,5	+0,4	+0,7

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,2% und +1,9%, bei der FDP und den Grünen zwischen -0,6% und +2,0%

33. Wiederholung

	Mai	Juni	Juli	August
SPD	39,3	47,6	43,8	44,3
CDU/CSU	42,2	39,9	42,8	40,3
FDP	4,8	4,2	5,7	4,0
Grüne	13,7	8,3	7,7	11,4

Abweichungen bei der 33. Wiederholung

	Mai	Juni	Juli	August
SPD	-4,7	+1,6	+0,8	+1,3
CDU/CSU	+3,2	-2,1	-1,2	-3,7
FDP	-0,2	+0,2	+0,7	0,0
Grüne	+1,7	+0,3	-0,3	+2,4

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -4,7% und +3,2%, bei der FDP und den Grünen zwischen -0,3% und +2,4%

34. Wiederholung

	Mai	Juni	Juli	August
SPD	43,4	44,9	42,6	44,7
CDU/CSU	40,7	43,8	44,6	44,4
FDP	4,8	3,9	4,5	3,4
Grüne	11,1	7,4	8,3	7,5

Abweichungen bei der 34. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,6	-1,1	-0,4	+1,7
CDU/CSU	+1,7	+1,8	+0,6	+0,4
FDP	-0,2	-0,1	-0,5	-0,6
Grüne	-0,9	-0,6	+0,3	-1,5

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -1,1% und +1,8%, bei der FDP und den Grünen zwischen -1,5% und +0,3%

35. Wiederholung

	Mai	Juni	Juli	August
SPD	44,1	44,0	43,5	43,3
CDU/CSU	39,1	46,2	43,7	43,9
FDP	4,7	2,8	5,0	3,2
Grüne	12,1	7,0	7,8	9,6

Abweichungen bei der 35. Wiederholung

	Mai	Juni	Juli	August
SPD	+0,1	-2,0	+0,5	+0,3
CDU/CSU	+0,1	+4,2	-0,3	-0,1
FDP	-0,3	-1,2	0,0	-0,8
Grüne	+0,1	-1,0	-0,2	+0,6

Die Abweichungen von den wahren Werten variieren bei der SPD und der CDU/CSU zwischen -2,0% und +4,2%, bei der FDP und den Grünen zwischen -1,2% und +0,6%

In der folgenden Aufstellung wurden diese 35 Auslosungen ihrer Güte nach geordnet. In der ersten Spalte sind die „besseren“ Auslosungen aufgelistet und in der zweiten die „schlechteren“ (dabei bedeutet

20.W: von -0.9% bis +0.8%

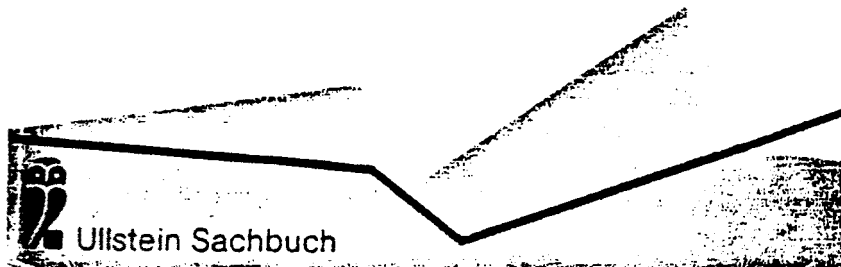
daß bei der 20. Auslosung Abweichungen von -0.9% bis +0.8% aufgetreten sind):

	Abweichungen		Abweichungen
20.W:	von -0,9% bis +0,8%	25.W:	von -2,1% bis +2,5%
17.W:	von -1,4% bis +1,2%	23.W:	von -2,1% bis +2,8%
7.W:	von -1,6% bis +1,0%	13.W:	von -3,2% bis +2,1%
6.W:	von -1,9% bis +1,2%	21.W:	von -3,5% bis +2,2%
18.W:	von -1,7% bis +1,5%	9.W:	von -3,2% bis +2,5%
34.W:	von -1,5% bis +1,8%	5.W:	von -2,8% bis +2,9%
14.W:	von -1,8% bis +1,7%	29.W:	von -3,4% bis +2,7%
26.W:	von -1,4% bis +2,1%	15.W:	von -3,2% bis +2,9%
8.W:	von -2,3% bis +1,3%	19.W:	von -3,7% bis +2,5%
27.W:	von -1,8% bis +1,8%	35.W:	von -2,0% bis +4,2%
2.W:	von -2,2% bis +1,5%	30.W:	von -3,2% bis +3,1%
4.W:	von -2,1% bis +1,9%	11.W:	von -3,6% bis +2,7%
3.W:	von -2,1% bis +1,9%	22.W:	von -3,4% bis +3,1%
10.W:	von -2,6% bis +1,5%	24.W:	von -2,8% bis +3,8%
32.W:	von -2,2% bis +2,0%	1.W:	von -3,6% bis +3,7%
16.W:	von -1,3% bis +3,0%	31.W:	von -4,0% bis +3,9%
12.W:	von -1,9% bis +2,4%	33.W:	von -4,7% bis +3,2%
28.W:	von -1,6% bis +2,9%		

Die beste der Auslosungen weist bei den Parteistärken Abweichungen von -0,9% bis +0,8% auf und die schlechteste solche von -4,7% bis +3,2%. Die meisten Auslosungen weisen Abweichungen in den Größenordnungen von $\pm 2\%$ auf, aber es treten auch Fälle von $\pm 3\%$ und mehr auf. **Wie man sieht, liegen die Werte keiner einzigen Auslosung innerhalb von $\pm 0,5\%$, wie es die Daten und vor allem ihre Analyse im ZDF suggerieren.**

Die 35 Auslosungen des Politbarometers mit jeweils anderen „repräsentativen Querschnitten“ stellen eine wahre Fundgrube für den „motivierten“ Berichterstatter dar. Je nach Standort läßt sich mancher Traum realisieren und entsprechend ausschmücken :

**Elisabeth
Noelle-
Neumann
Die
Schweige-
spirale**
Öffentliche Meinung –
unsere soziale Haut



Demo-Wischer und die Schweigespirale

von Carell-Hildebrandt

Variationen auf ein Thema von Elisabeth Noelle-Neumann:

Wasche die **Öffentliche Meinung** –
aber mach **unsere soziale Haut** *nicht naß!*

Nach einem Studium der 35 Wiederholungen und jeweils geschickter Wahl des repräsentativen Querschnittes gibt es folgendes zu berichten:

- ❖ Der Aufwärtstrend der Union hält seit Monaten an. Selbst der Traum von der absoluten Mehrheit findet seine Bestätigung im „repräsentativen Querschnitt“: Von 35,0% um Mai über 44,2% im Juni und einem Zwischentief im Juli wird im August der Rekordstand von 47,2% erreicht. Da die FDP mit 4% im Keller liegt – und damit sicher aus dem Bundestag ausscheidet – erhöht sich der Unionsanteil de facto auf 49,2%. Parallel zum Aufwärtstrend der Union hält der Abwärtstrend der SPD unvermindert an. Auch der neue *Ober-Vogel LaFontaine* vermag daran nichts zu ändern. Was kann dieser

linke Stürmer und Dränger
Bundeskanzler *Helmut Toll* schon entgegenstellen?
Leere Kassen bei Freibier & Saarstahl!

- ❖ Seit Mai liegt die SPD deutlich in Führung. Im Frühsommer überschritt die SPD mit 49,8% sogar die absolute Mehrheit, da die FDP mit 4% ausschied. Der Krebsgang der nicht mehr so *Jungen Union* setzt sich während des ganzen Sommers fort. Was kann

dieser unser *Kleinmut Hohl*
im Kanzleramt dem Jakobiner aus Fontainebleau schon entgegensetzen?
Steuer-Harmonie und soziale Gerechtigkeit für Seinesgleichen!

- ❖ Blutbad bei der FDP: War die FDP im Mai mit fast 7% sicher im Bundestag vertreten, so verlor sie innerhalb eines Monats über 50% ihrer Wählerstimmen (Juni: 3,1%). Trotz einer leichten Erholung im Juli auf 3,9% - die wahrscheinlich auf einen Mitleidseffekt zurückzuführen war – stellt sich die *Bange Frage*: Was hat den Zusammenbruch der FDP herbeigeführt?

Mangel- oder Böllermann?

- ❖ FDP seit Mai konstant über der 5%-Hürde, abgesehen von einem Früh-Sommerloch mit 4,3%. Darin spiegelt sich der Konsens wieder, daß eine un-

abhängige Kraft zwischen den Blöcken
erhalten bleiben muß.

- ❖ Die Grünen haben ihren Tschernobyl-Rekord vom Mai mit 9,9% im August mit 10,4% übertroffen. Wird die Abschaltung der *Kernkraftwerke* damit näherrücken

und der Kahl-Brüter endlich Demo-liert?

Werden unsere Kinder nun wieder cäsiumfreie Muttermilch und Nahrung bekommen?

- ❖ Tschernobyl erwies sich als Eintagsfliege für die Grünen. Nach 13,1% im Mai gewannen Vernunft und Sorge um die *Sicherheit der Kernkraftarbeitsplätze* wieder Oberhand. Im Juni fielen die Grünen auf 7,4% zurück,

was als Restrisiko tragbar ist.

Wahrscheinlich ließe sich das grüne Restrisiko weiter reduzieren, wenn man die Eisenbahnladung von gutem deutschen Molkepulver – welches wegen seiner Radio-aktiven Spritzfahrt quer durch die BRD zu einer parapsychologischen Belastung geworden ist – per Bundesnotwehr in die Schweiz abschieben würde. *Zwecks Endverbrauch*. Dort hat man bekanntlich mit der *natürlichen Strahlungskraft* von Milch und Mineralwasser seit Jahrhunderten gute Erfahrungen gemacht, und der *radioaktive Nährwert* wird traditionsgemäß als Qualitätsmerkmal auf den Flaschenetiketten festgehalten. ^(*)

^(*) bis die grün inspirierte Schweigespirale diesen unternehmerischen Leistungsnachweis sogar in die Schweiz zu hintertreiben vermochte.

Wie man sieht, ist die Beliebtheit des repräsentativen Querschnitts leicht verständlich: Niemand geht leer aus, jedermann erhält seinen Wunsch erfüllt. Man kann seinen Gefühlen und Phantasien freien Lauf lassen und als Erklärung für die „gefundenen“ Zahlen hinstellen. Prozentzahlen sind Waffen, die man für seine Ziele und bei Bedarf gegen den politischen Gegner einsetzen kann.

Die Forschungsgruppe Wahlen - als Betreiber des Politbarometers - könnte einwenden, daß diese 35 Auslosungen zufällig extreme Abweichungen aufweisen. Es wurden allerdings nicht nur 35, sondern 100.000 Auslosungen des Politbarometers simuliert, d.h. es wurden die Ergebnisse von 100.000 repräsentativen Querschnitten mit viermal 1000 Befragten analysiert - aber aus Platzgründen haben wir hier „nur“ die ersten 35 Auslosungen abgedruckt und die anderen auf verschiedene Arten statistisch ausgewertet:

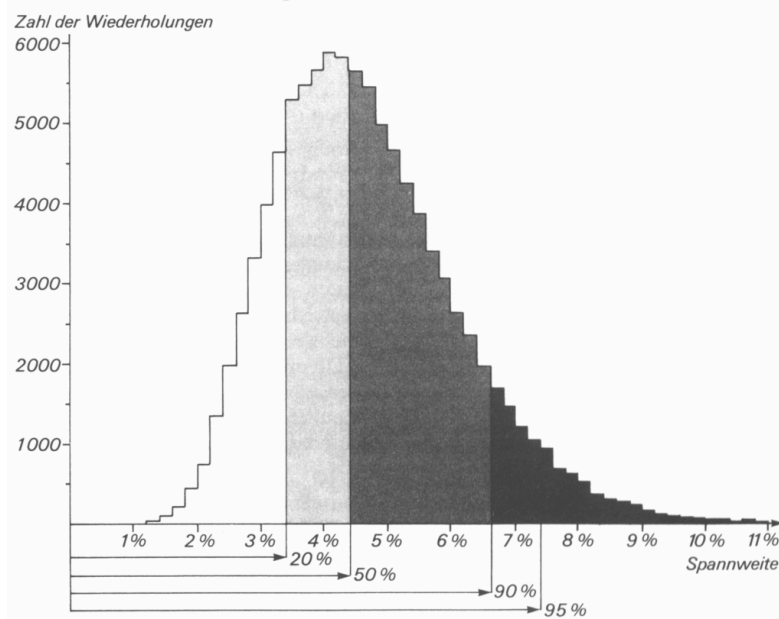
Zunächst betrachteten wir die weiter unten beschriebene Spannweite. Diese ist zwar sehr anschaulich, aber im Hinblick auf die Fragestellung nicht sehr effizient, weil die auslosungsbedingten Abweichungen der großen Parteien diejenigen der kleinen verschlucken. Die Spannweite vermittelt daher ein etwas zu günstiges Bild. Aus diesem Grund wurden die bei Politbarometer-Auslosungen auftretenden Abweichungen auch mit Hilfe von gestaffelten Spielräumen analysiert, wobei die unterschiedliche Auswirkung der Parteistärken auf die Standardabweichung berücksichtigt wurde.

Für die Berechnung der Spannweite einer Politbarometer-Auslosung ermittelten wir zunächst die größte Abweichung einer Parteistärke nach oben und die größte nach unten und addierten dann diese Zahlen betragsmäßig, d.h. ohne Berücksichtigung des Vorzeichens. Diese Summe der extremen Abweichungen wird als Spannweite bezeichnet.

Bei der ersten Auslosung (siehe Tabellen auf den vorangehenden Seiten) beträgt sie 7,3% (= 3,7% + 3,6%), bei der zweiten 3,7% (= 1,5% + 2,2%), bei der dritten 4,0% (= 1,9% + 2,1%). Das Diagramm auf der folgenden Seite zeigt die Verteilung der Spannweiten der 100.000 Auslosungen, soweit sie von 1,2% bis 11,0% variierten.²⁶ Die Hälfte aller Auslosungen hat eine Spannweite von mehr als 4,4%. Nur jede fünfte Auslosung zeigt eine Spannweite unter 3,3%, aber jede zehnte weist eine solche von über 6,6% auf und jede zwanzigste von über 7,5%.

²⁶ Ein verschwindend kleiner Anteil - nämlich 0,6 Promille - wies Spannweiten auf, die entweder von 1,0% bis 1,1% (3 Stück) oder von 11% bis 14,4% (57 Stück) variierten. Diese konnten wegen der geringen Anzahl im Diagramm graphisch nicht dargestellt werden.

100.000 Wiederholungen des Politbarometers



Wie ist das Diagramm zu verstehen? Es wurden je 100.000 Wiederholungen der vier Politbarometer (Mai bis August 1986) auf dem Computer simuliert. Zum Beispiel:

1. Wiederholung

	Mai	Juni	Juli	August
SPD	43,9	49,7	41,9	44,0
CDU/CSU	40,4	38,4	46,0	41,9
FDP	3,7	4,2	4,7	4,6
Grüne	12,0	7,7	7,4	9,5

Abweichungen bei der 1. Wiederholung

	Mai	Juni	Juli	August
SPD	-0,1	+3,7	-1,1	+1,0
CDU/CSU	+1,4	-3,6	+2,0	-2,1
FDP	-1,3	+0,2	-0,3	+0,6
Grüne	0,0	-0,3	-0,6	+0,5

Die Abweichungen beziehen sich auf die tatsächlichen (vom ZDF angegebenen) Werte.

Von den obigen 16 Abweichungen wurden ohne Berücksichtigung des Vorzeichens der größte negative und der größte positive Wert addiert, in unserem Beispiel also $-3,6\%$ und $+3,7\%$. Das Ergebnis dieser Rechnung wurde als „Spannweite“ der betrachteten Wiederholung bezeichnet, im Beispiel erhalten wir $7,3\%$. Für jede der 100.000 simulierten Wiederholungen wurde die Spannweite auf diese Art berechnet.

Das obige Diagramm zeigt, wie oft einzelne Spannweiten auftraten: Die waagerechte Achse ist in Abstände von $0,2\%$ unterteilt und deckt den Bereich der Spannweiten zwischen 0% und 11% ab. Die senkrechte Achse zeigt die Anzahl derjenigen Wiederholungen, die Spannweiten im entsprechenden Bereich aufwiesen.

Beispiel: Spannweiten zwischen $4,8\%$ und $5,0\%$ traten bei etwa 5000 der Wiederholungen auf und solche zwischen $2,4\%$ und $2,6\%$ bei circa 2000 der Wiederholungen.

Im Histogramm ist durch die Pfeile unter der waagerechten Achse und gleichzeitig durch die Farbgebung der Fläche zusätzlich folgendes angedeutet:

20% der Wiederholungen hatten eine Spannweite unter $3,3\%$

90% der Wiederholungen hatten eine Spannweite unter $6,6\%$

50% der Wiederholungen hatten eine Spannweite unter $4,4\%$

95% der Wiederholungen hatten eine Spannweite unter $7,4\%$

Das Histogramm der 100.000 Politbarometer-Auslosungen zeigt, daß die in den Tabellen auf den Seiten IX/3 bis IX/9 abgedruckten 35 Auslosungen durchaus keine Ausreißer darstellen, sondern den Normalfall illustrieren. Der einzige „Ausreißer“ unter den in den Tabellen abgedruckten Auslosungen ist die Nummer 20 mit einer Spannweite von 1,7%. Nur eine von 500 Auslosungen weist eine so geringe oder noch kleinere Spannweite auf. Mit der gleichen Häufigkeit treten Auslosungen mit einer Spannweite von über 10,2% auf!

Diese Beispiele machen deutlich, wie weit die Politbarometerbefragungen davon entfernt sind, den behaupteten Trend aufzeigen zu können. Hätte man in den vier Monaten andere Wahlberechtigte ausgelost - also einen anderen repräsentativen Querschnitt verwendet - dann wäre man mit ziemlicher Sicherheit zu einem ganz anderen Resultat gekommen.

Kommt es überhaupt vor, daß eine Auslosung des Politbarometers die Parteistärken mit einer Genauigkeit von $\pm 0,5\%$ reproduziert, d.h. eine Spannweite von nur 1% aufweist? Von unseren 100.000 Auslosungen des Politbarometers hat keine einzige dieses Kunststück geschafft. Dieses experimentell gewonnene Ergebnis stimmt mit dem Ergebnis einer theoretischen Berechnung mit Hilfe der Multinomialverteilung überein (siehe Tabelle „Erfolgsstatistik von Auslosungen des Politbarometers“): Lediglich 0,000912% aller Auslosungen - d.h. knapp eine von 100.000 - liegen innerhalb eines Spielraumes von $\pm 0,5\%$! Um sich ein Bild von dieser geringen Wahrscheinlichkeit zu machen: Die von der Forschungsgruppe Wahlen im Politbarometer des Zweiten Deutschen Fernsehens stillschweigend angenommene Genauigkeit von $\pm 0,5\%$ ist - wie eine einfache Wahrscheinlichkeitsrechnung zeigt - so unrealistisch wie der Glaube eines Lottospielers, mit einer (halben) Mark Einsatz fünf Richtige zu treffen. Der Unterschied besteht darin, daß unser Lottospieler seinen Einsatz mit einer Wahrscheinlichkeit von 99,9991% in den Wind schreiben kann, während die „Forsche Truppe Wahlen“ im Polit-Lotto-Meter des ZDF in gleicher Weise pokern kann. Ungestraft und nicht schlecht bezahlt! Das Risiko tragen die armen Zuschauer, die den wissenschaftlich verpackten Politbarometer-Prognosen in der präsentierten Form Glauben schenken.²⁷

Gibt es denn überhaupt Informationen zur aktuellen politischen Situation, denen der Zuschauer Glauben schenken kann und für deren Gehalt es sich lohnt, am Montagabend bis 22.00 Uhr vor dem Fernsehschirm für das Politbarometer auszuharren? Da eine Genauigkeit von $\pm 0,5\%$ bei Parteistärken und deren Trend *utopisch* ist, so ist der Spielraum für die Abweichungen - welche durch die Auslosung des repräsentativen Querschnittes hervorgerufen werden - massiv zu vergrößern, was natürlich den Informationsgehalt der „Repräsentativumfrage“ entsprechend reduziert. Die Resultate sind für Spielräume von $\pm 1\%$ bis $\pm 5\%$ in der oberen Tabelle auf Seite IX/20 zusammengestellt.

Bei einer Genauigkeit von $\pm 2\%$ für die Parteistärken liegen 75,3% der Auslosungen des Politbarometers außerhalb dieses Spielraums, das heißt bei drei von vier Politbarometern weichen einzelne der sechzehn **angegebenen Parteistärken um mehr als $\pm 2\%$ von den wahren Werten ab**. Selbst wenn man die Genauigkeit reduziert, indem man den Spielraum auf $\pm 2,5\%$ vergrößert, so vermögen noch immer die Hälfte der Politbarometer-Auslosungen diese Toleranz nicht einzuhalten. Man müßte den Spielraum auf $\pm 4\%$ erhöhen, wenn man die übliche statistische

²⁷ P.S. Im Januar 1988 hat das Politbarometer seine "Genauigkeit" weiter gesteigert. Die Parteistärken wurden auf halbe Prozent angegeben, was einer Genauigkeit von $\pm 0,25\%$ entspricht. Die "Forschungsgruppe Wahlen" sollte das Geheimnis ihrer Meßmethoden beim Patentamt sicherstellen lassen. Die zugrundeliegenden Prinzipien werden eines Tages den Bau von Überschall-Segelflugzeugen ermöglichen.

Sicherheit von 95% haben wollte. **Bei einem solchen Spielraum gehen die tatsächlichen Wählerbewegungen von Monat zu Monat völlig unter.**

Um bei einer statistischen Sicherheit von 95% den interessanten Spielraum von $\pm 0,5\%$ einhalten zu können, müßte man die Zahl der Befragten auf **70.000** erhöhen.

Bei der bisherigen Betrachtungsweise blieb unberücksichtigt, daß die durch die Auslosung des repräsentativen Querschnittes verursachten Abweichungen für die kleinen Parteien weit gravierender sind als für die großen. Eine große Partei kann $\pm 3\%$ - wenn auch schmerzhaft - noch verkraften. Bei einer kleinen Partei hingegen geht es bei solchen Abweichungen aufgrund der 5%-Klausel um Sein oder Nichtsein. Daher wurden die Spielräume auch nach Parteistärken gestaffelt²⁸, wie in der unteren Tabelle auf Seite IX/20 dargestellt ist.

Einen Spielraum von $\pm 2\%$ für CDU/CSU und SPD beziehungsweise $\pm 1,2\%$ für die Grünen und $\pm 0,8\%$ für die FDP halten nicht einmal sieben Prozent (!) von Politbarometerauslosungen ein. Ein Spielraum von $\pm 3\%$ für CDU/CSU und SPD beziehungsweise $\pm 1,8\%$ für die Grünen und $\pm 1,2\%$ für die FDP führt zu einer Trendmeldung der Gestalt

	Mai	Juni	Juli	August
SPD	41,0 - 47,0	43,0 - 49,0	40,0 - 46,0	40,0 - 46,0
CDU/CSU	36,0 - 42,0	39,0 - 45,0	41,0 - 47,0	41,0 - 47,0
FDP	3,8 - 6,2	2,8 - 5,2	3,8 - 6,2	2,8 - 5,2
Grüne	10,2 - 13,8	6,2 - 9,8	6,2 - 9,8	7,2 - 10,8

Soll sich der Fernsehzuschauer vielleicht damit trösten, daß selbst diese nichtssagende Trendmeldung mit einer Wahrscheinlichkeit von 50% falsch ist? Weil der repräsentative Querschnitt per Lotterie erstellt wird, schafft es nämlich nicht einmal jede zweite Auslosung des Politbarometers, diese Bandbreiten einzuhalten.

Man müßte den Spielraum für CDU/CSU und SPD auf $\pm 4,6\%$, für die Grünen auf $\pm 2,8\%$ und für die FDP auf $\pm 1,8\%$ erhöhen, wenn man die übliche statistische Sicherheit von 95% haben wollte. **Man führe sich konkret vor Augen, was die „Forschungs“-Gruppe Wahlen und der Politbarometer-Moderator Klaus Bresser auf dieser Grundlage dem Fernsehzuschauer effektiv an Trendinformation offerieren:**

	Mai	Juni	Juli	August
SPD	39,4 - 48,6	41,4 - 50,6	38,4 - 47,6	38,4 - 47,6
CDU/CSU	34,4 - 43,6	37,4 - 46,6	39,4 - 48,6	39,4 - 48,6
FDP	3,2 - 6,8	2,2 - 5,8	3,2 - 6,8	2,2 - 5,8
Grüne	9,2 - 14,8	5,2 - 10,8	5,2 - 10,8	6,2 - 11,8

Bei einem solchen Spielraum interessiert sich offensichtlich kein Mensch mehr für das Resultat. Diese Zahlen machen deutlich, wie absurd es ist, eine zeitliche Entwicklung, die sich im Prozentbereich abspielt, mit einer Befragung von viermal 1000 „repräsentativ

²⁸ Von der Fragestellung her wäre eine Staffelung wünschenswert, bei welcher der relative Fehler für alle Parteien ungefähr gleich wäre. Dies ist zwar möglich, aber daraus resultieren für die Parteien höchst unterschiedliche Sicherheitswahrscheinlichkeiten, was zu einer inhaltlich und statistisch unsinnigen Schätzung führt. Als Kompromiß wurde die Staffelung gewählt, die sich aus dem Verhältnis der Standardabweichungen der Parteistärken CDU/CSU (bzw. SPD) Grüne und FDP ergibt. Normiert man die Standardabweichung der CDU/CSU und SPD - welche praktisch gleich sind - auf 1, so ergibt sich etwa das Verhältnis 1 : 0,6 : 0,4.

Ausgewählten“ analysieren zu wollen.

Offenbar setzen sich die Demoskopien über solch elementare mathematische Grundlagen ihrer Arbeit schlicht hinweg. Durch den Mißbrauch des Begriffes „repräsentativer Querschnitt“ geben sie vor, aussagekräftige Meinungsbilder produzieren zu können, obwohl ihnen doch allein der statistische Fehler, der durch die Auslosung von 1000 Wahlberechtigten verursacht wird, keine Chance dazu läßt.

Erst bei 10.000 oder besser noch 100.000 Befragten könnten sie die Wahrheit einigermaßen sicher diagnostizieren. Doch wer sollte ihnen den entsprechenden Aufwand bezahlen? Woher sollten sie die Heerscharen ausgebildeter Interviewer nehmen, wo doch bekannt ist, daß sich deren Objektivität nach acht oder höchstens zehn Befragungen erschöpft?

Es hätte allerdings wenig Sinn, derart umfangreiche Befragungen durchzuführen. Denn der ungleich größere Interviewfehler wäre damit nicht aus der Welt geschafft. Man hätte lediglich das „letzte“ statistische Haar aus der Suppe entfernt, der Balken im Auge aber bliebe erhalten.

Es bleibt das Eingeständnis: Die Demoskopie ist nicht in der Lage, Trendaussagen zu machen, wenn die Bewegungen im Prozentbereich liegen. Was sich in Wirklichkeit abspielt, das weiß kein Mensch und die von den Meinungsforschungsinstituten gelieferten Daten über das zeitliche Auf und Ab der Parteistärken haben reinen „Horoskopcharakter“. Sie täuschen Veränderungen vor, die quantitativ mit an Sicherheit grenzender Wahrscheinlichkeit falsch sind und die häufig nicht einmal qualitativ richtig sind.

Man könnte noch mehrere Selbstüberschätzungen der Demoskopie aufzählen, zum Beispiel die beliebten Popularitätswerte von Politikern. Über weite Bereiche könnte man sie, vor allem ihr zeitliches Auf und Ab, ebenso gut mit dem Würfel bestimmen. Das wäre erstens billiger, und zweitens wäre klargestellt, um was es geht: um ein Unterhaltungsspiel im Fernsehen.

„Erfolgs“-Statistik von Wiederholungen des Politbarometers

Gleiche Spielräume für alle Parteien

Bei jeder Partei wurde derselbe Spielraum für die Abweichungen der Parteistärken von SPD, CDU/CSU, FDP und Grüne von Mai bis August 1986 von den wahren Werten zugelassen.

Spielräume in % bei				Anzahl der Interviews				
CDU/CSU	SPD	FDP	Grüne	1000	2000	10 000	40 000	70 000
±0,5	±0,5	±0,5	±0,5	0,000912	0,0198	8	75	95
±0,75	±0,75	±0,75	±0,75	0,022	0,635	42	99	~100
±1,0	±1,0	±1,0	±1,0	0,452	4,49	76	~100	~100
±1,5	±1,5	±1,5	±1,5	6,65	29,7	98	~100	~100
±2,0	±2,0	±2,0	±2,0	24,7	63,3	~100	~100	~100
±2,5	±2,5	±2,5	±2,5	49,0	85,5	~100	~100	~100
±3,0	±3,0	±3,0	±3,0	70,7	96	~100	~100	~100
±3,5	±3,5	±3,5	±3,5	85,4	99	~100	~100	~100
±4,0	±4,0	±4,0	±4,0	94	~100	~100	~100	~100
±5,0	±5,0	±5,0	±5,0	~100	~100	~100	~100	~100

Zum Beispiel: Wird für jede der vier Parteien ein Spielraum von $\pm 2,0\%$ für die Werte der Parteistärken zwischen Mai und August zugelassen, dann ist aus der Spalte (1000 Interviews) abzulesen, daß nicht einmal jede vierte Wiederholung (24,7%) innerhalb dieses Bereiches liegt. Dies bedeutet, daß in drei von vier Wiederholungen des Politbarometers bei mindestens einer Partei Abweichungen von mehr als 2% nach oben und unten auftreten.

Nach Parteistärken abgestufte Spielräume

Es wurde berücksichtigt, daß derselbe Spielraum für die kleinen Parteien viel gravierender ist als für die großen. Entsprechend wurden die Spielräume gestaffelt, beispielsweise $\pm 3,0\%$ für die CDU/CSU und die SPD, $\pm 1,8\%$ für die Grünen und $\pm 1,2\%$ für die FDP.

Spielräume in % bei				Anzahl der Interviews				
CDU/CSU	SPD	FDP	Grüne	1000	2000	10 000	40 000	70 000
±0,75	±0,75	±0,3	±0,45	0,000651	0,0284	16	96	~100
±1,0	±1,0	±0,4	±0,6	0,020	0,409	55	~100	~100
±1,5	±1,5	±0,6	±0,9	0,778	9,01	96	~100	~100
±2,0	±2,0	±0,8	±1,2	6,67	38,5	~100	~100	~100
±2,5	±2,5	±1,0	±1,5	23,4	71,8	~100	~100	~100
±3,0	±3,0	±1,2	±1,8	48,6	90,6	~100	~100	~100
±3,25	±3,25	±1,3	±1,95	58,8	95,1	~100	~100	~100
±3,3	±3,3	±1,3	±2,0	62,8	95,6	~100	~100	~100
±3,5	±3,5	±1,4	±2,1	71,5	97,5	~100	~100	~100
±4,0	±4,0	±1,6	±2,4	86,5	~100	~100	~100	~100
±4,6	±4,6	±1,8	±2,8	95,1	~100	~100	~100	~100
±5,0	±5,0	±2,0	±3,0	97,8	~100	~100	~100	~100

Zum Beispiel: Wird bei CDU/CSU und SPD ein Spielraum von $\pm 3,0\%$ zugelassen, bei den Grünen ein solcher von $\pm 1,8\%$ und bei der FDP $\pm 1,2\%$, dann ist aus der 1. Spalte (1000 Interviews) abzulesen, daß nur 48,6% aller Politbarometer-Wiederholungen innerhalb dieses gestaffelten Spielraums liegen. Dies bedeutet, daß in jeder zweiten Wiederholung des

Politbarometers für mindestens eine Partei eine Zahl außerhalb dieses Spielraumes ermittelt wird.

Berechnungsgrundlage zur Tabelle „Erfolgsstatistik von Wiederholungen des Politbarometers“

Die Werte in der Tabelle wurden teils numerisch mit der Multinomialverteilung, teils mit Simulation berechnet. In einigen Fällen wurden beide Methoden verwendet, wobei eine gute Übereinstimmung erreicht wurde. Die numerisch ermittelten Werte sind dadurch gekennzeichnet, daß Nachkommastellen angegeben sind.

Die Programme schrieben Edith Achiman und Ullrich Tesche.

Zur Multinomialverteilung:

Für die vier Monate Mai, Juni, Juli und August 1986 sind die Parteistärken wie folgt gegeben:

	Mai	Juni	Juli	August
SPD	$p_{11} = 0,44$	$p_{12} = 0,46$	$p_{13} = 0,43$	$p_{14} = 0,43$
CDU/CSU	$p_{21} = 0,39$	$p_{22} = 0,42$	$p_{23} = 0,44$	$p_{24} = 0,44$
FDP	$p_{31} = 0,05$	$p_{32} = 0,04$	$p_{33} = 0,05$	$p_{34} = 0,04$
Grüne	$p_{41} = 0,12$	$p_{42} = 0,08$	$p_{43} = 0,08$	$p_{44} = 0,09$

Da in den Monaten Mai, Juni, Juli und August der repräsentative Querschnitt jeweils neu erstellt worden ist, sind die Anteile der einzelnen Monate statistisch unabhängig. Innerhalb eines Monats folgen die Parteistärken der Multinomialverteilung (m=4) und nicht der Binomialverteilung, wie in den Standardtexten der Demoskopie angenommen wird.

Unter Zugrundelegung des Stichprobenumfangs n und den Spielräumen $\delta_1, \delta_2, \delta_3, \delta_4$ gemäß Tabelle ist der folgende Ausdruck getrennt für jeden Monat zu berechnen, wobei j = 1, 2, 3, 4 den Monat bezeichnet:

$$M_j = \sum_{\substack{n(p_{1j}+\delta_1), n(p_{2j}+\delta_2), n(p_{3j}+\delta_3), n(p_{4j}+\delta_4) \\ k_1=n(p_{1j}-\delta_1), k_2=n(p_{2j}-\delta_2), k_3=n(p_{3j}-\delta_3), k_4=n(p_{4j}-\delta_4) \\ k_1+k_2+k_3+k_4=n}} n! \frac{p_{1j}^{k_1} \cdot p_{2j}^{k_2} \cdot p_{3j}^{k_3} \cdot p_{4j}^{k_4}}{k_1! \cdot k_2! \cdot k_3! \cdot k_4!}$$

Beispiel: Monat Mai 1986

$$p_{11} = 0,44, \quad p_{21} = 0,39, \quad p_{31} = 0,05, \quad p_{41} = 0,12$$

$$\delta_1 = 0,02, \quad \delta_2 = 0,02, \quad \delta_3 = 0,008, \quad \delta_4 = 0,012$$

Für den Stichprobenumfang n=1000 berechnet man die folgende Summe:

$$M_1 = \sum_{\substack{460, 410, 58, 32 \\ k_1=420, k_2=370, k_3=42, k_4=108 \\ k_1+k_2+k_3+k_4=1000}} 1000! \frac{0,44^{k_1} \cdot 0,39^{k_2} \cdot 0,05^{k_3} \cdot 0,12^{k_4}}{k_1! \cdot k_2! \cdot k_3! \cdot k_4!} = 0,466$$

Dies führt man ebenso für die Monate Juni, Juli und August durch.

Für die obigen Werte n=1000 und $\delta_1=0,02, \delta_2=0,02, \delta_3=0,008, \delta_4=0,012$ erhält man die folgenden Ergebnisse:

$$\text{Mai: } M_1=0,466, \quad \text{Juni: } M_2=0,537, \quad \text{Juli: } M_3=0,508, \quad \text{August: } M_4=0,542.$$

Das Produkt dieser vier Zahlen $M_1 \cdot M_2 \cdot M_3 \cdot M_4 = 0,466 \cdot 0,537 \cdot 0,508 \cdot 0,524 = 0,0667 = 6,67\%$

findet man in der Tabelle in der Kolonne n=1000 und den Spielräumen CDU/CSU: $\pm 2\%$, SPD $\pm 2\%$, FDP $\pm 0,8\%$, Grüne $\pm 1,2\%$.

X Die Gewichtung

Aufgrund von Fehlprognosen haben die Meinungsforschungs-Institute in den letzten Jahren zu einem ebenso riskanten wie dubiosen Mittel gegriffen; der „Kunst“ der Gewichtung. Es werden nicht die tatsächlichen Umfrageergebnisse veröffentlicht, sondern diese werden rein rechnerisch erhöht oder erniedrigt, und nur dieses modifizierte Resultat wird publiziert.²⁹

M. Kaase führte dazu im „Rheinischen Merkur“ Nr.37/1987 aus:

„Wegen Verzerrungsmöglichkeiten hat die kommerzielle Wahlforschung das Verfahren der „politischen Gewichtung“ erfunden. Bei diesem höchst umstrittenen und theoretisch völlig unfundierten Ansatz macht man sich paradoxerweise gerade die Abweichung der Befragtenangaben zu ihrer zurückliegenden Wahlentscheidung (der sogenannten Recall-Frage) vom ja zweifelsfrei bekannten Ergebnis einer zurückliegenden Vergleichswahl zunutze, um die aktuelle „Prognose“, die an sich auf den Angaben der Befragten zu ihrer Wahlabsicht beruhen müßte, zu korrigieren. Solche politisch gewichteten „Prognosen“ haben die Wirkung, diese mehr in die Nähe der traditionell zu erwartenden Wahlergebnisse zu rücken und verfälschen damit gerade die eigentlich interessante Nachricht : Wie nämlich zu einem bestimmten Zeitpunkt die Stimmung der Bevölkerung bezüglich der politischen Parteien aussieht.“

Ein typisches Beispiel für die Umverteilung von Prozenten, wie sie vor der Bundestagswahl praktiziert wurde, war in der Wirtschaftswoche Nr. 1/2 vom 2. Januar 1987 dargestellt:

„Behaupten etwa zu viele Befragte, sie hätten beim letztenmal sozialdemokratisch gewählt, so wird das aktuelle Umfrageergebnis für die SPD entsprechend heruntergewichtet. Liegt der Erinnerungswert für die FDP zu niedrig, werden die Liberalen entsprechend hochgewichtet: Deren aktueller Recall-Zuschlag liegt bei rund 100 Prozent.“

Es wird oft behauptet, die erfragten Werte für die SPD und die Grünen müßten erniedrigt und diejenigen für die FDP erhöht werden. Als Gründe dafür werden angeführt, daß ein Teil der SPD- und Grüne-Wähler „erfahrungsgemäß“ bis zum Wahlsonntag „ihre Meinung ändern“ oder daß diese auf Grund von systematischen Verzerrungen in der Repräsentativumfrage überproportional vertreten seien. Von der FDP wird behauptet, sie würde „erfahrungsgemäß“ am Wahltag Zweitstimmen von CDU/CSU-Wählern erhalten und sie sei in Repräsentativumfragen systematisch untervertreten.

Falls es sich um empirisch gesicherte Erkenntnisse und nicht um plausible Stammtisch-Argumente handelt, können solche Modifikationen sinnvoll und nützlich sein, soweit sie der Korrektur systematischer Fehler dienen. Aber es ist absurd, auslosungsbedingte Abweichungen, die einen großen Spielraum unvermeidlich machen (vergleiche Abschnitte IV bis IX), durch Gewichtung kompensieren zu wollen. Dies wird aber gerade in der Praxis routinemäßig gemacht.

Die Größenordnung der Umgewichtung ist atemraubend. Nach Noelle-Neumann werden die

²⁹ Die Methode der Gewichtung von Umfrageergebnissen ist an sich fast so alt wie die "moderne" Demoskopie (vergleiche : The Pre-Election Polls of 1948, Bulletin Nr. 60 des Social Research Council Inc., New York 1949). Neu ist die Art und das Ausmaß der Gewichtung.

tatsächlich ermittelten Ergebnisse bis zu **10% oder 11%** abgeändert.^{30,31} Wie vorhin erwähnt, betrug im Januar 1987 der „aktuelle“ Recall-Zuschlag für die FDP rund 100%, d.h. das Umfrageergebnis wurde verdoppelt. In den Graphiken auf den Seiten X/5 bis X/7 ist die von der Forschungsgruppe Wahlen im Politbarometer des ZDF vorgenommene Umgewichtung in den Jahren 1986/87 detailliert dargestellt.

Die Meinungsforschungs-Institute schaffen sich durch die Gewichtung einen Freiraum, von dessen Existenz und Ausmaß die Öffentlichkeit keine Ahnung hat.

Ein seriöses Institut sollte auf das Gewichtungsdilemma und die Interview-Fehler hinweisen und wenigstens den auslosungsbedingten Spielraum der Umfragergebnisse angeben. Es sollte die Rohdaten offenlegen und nicht nur Zahlen, die es daraus durch eine geheimgehaltene Kosmetik gewinnt.

Das geschieht jedoch nicht. Die einzige Ausnahme ist meines Wissens die Forschungsgruppe Wahlen - d.h. das Politbarometer im Zweiten Deutschen Fernsehen - wo die Existenz und das Ausmaß der Abänderung von Umfrageergebnissen noch in Umrissen erkennbar ist. Die Forschungsgruppe Wahlen hatte sich stets gegen die Gewichtung von Umfragergebnissen gewehrt. Anfang 1986 wurde sie durch politischen Druck gezwungen, diese dubiosen Praktiken ebenfalls einzuführen und zu kaschieren. Die Details dieser Aktion und die Begründung für diese angebliche „Erweiterung“ der Berichterstattung sind in der Box auf Seite X/8 zu lesen.

Für die folgenden Ausführungen halte man sich die Tabellen auf den Seiten X/5 bis X/7 vor Augen, in welchen die Umgewichtungen graphisch dargestellt sind. Die Einführung der Gewichtung beim Politbarometer im März 1986 führte zu einer Erhöhung der Prozentzahlen für die Union bzw. die FDP um durchschnittlich etwa 1,5% bzw. 3% im Jahre 1986. Die Umfrageergebnisse für die SPD bzw. die Grünen wurden um durchschnittlich etwa 3% bzw. 2% nach unten korrigiert. Bei diesen Zahlen handelt es sich um Durchschnittswerte, die Bandbreite der Umgewichtung ist beträchtlich. Die Gewichtung zeigte bei den großen Parteien einen deutlichen zeitlichen Trend. Anfang des Jahres wurde die SPD massiv beschnitten, während die Union ein kräftiges „face lifting“ erhielt: im März -9% für die SPD und +6% für die Union. Im Sommer - Juni bis September - reduzierte sich der Aderlaß der SPD auf 2% bis 4%, während die CDU/CSU zunehmend „selbständiger“ wurde. Im Oktober stand die Union auf eigenen Füßen, und die SPD mußte nur noch die bescheidene Abgabe von einem Prozent entrichten. Im Oktober/November erfolgte dann eine eigentliche Kehrtwendung bei der Gewichtung : Man erinnere sich an die damalige Situation. Die November-Umfrageergebnisse des Politbarometers, welche Anfang Dezember die Wahlkampfarena erreichten, ergaben 49% für die Union, 37% für die SPD, 4% für die FDP und 10% für die Grünen. Damit wäre die FDP wie üblich - d.h. wie bei den meisten Politbarometerumfragen 1986 - an der 5%-Hürde gescheitert und die Union hätte folglich mit 49% die absolute Mehrheit errungen. Bekanntlich erhitzte diese Möglichkeit die Gemüter erheblich. Die CDU/CSU lief Sturm gegen das „Gerede von der absoluten Mehrheit“, gleichzeitig aber führte sie eine vehemente Zweitstimmenkampagne gegen die FDP mit der

³⁰ Man vergleiche dazu die Seiten I/4 und I/5.

³¹ Wirtschaftswoche Nr. 1/2 vom 2. Januar 1987:

"Solch gewaltige Veränderungen sind selbst für die Nestorin des Gewerbes stets wieder ein Ritt über den Bodensee. Elisabeth Noelle-Neumann bekennt: "Ich bin jedesmal wieder von großer Sorge erfüllt." Mit welchen Methoden der Justierung sie die groben Recallgewichtungen noch verfeinert, weiß dabei selbst in Allensbach nur ein kleiner Zirkel von Eingeweihten. Auch die anderen Institute verbergen die Details ihrer Computerprogramme, in die neben den Recallgewichten noch Faktoren wie Veränderungen der Bevölkerungsstruktur oder des Stimmungsklimas und der politischen Prioritäten eingehen."

Begründung, die FDP benötige keine Leihstimmen der Union, denn sie sei sicher über der 5%-Hürde.

In dieser Situation änderte das Politbarometer seine bisherige Gewichtungs-Praxis bei den großen Parteien. Zum ersten Mal wurden Prozente von der CDU/CSU zur SPD hin verschoben; der Anteil der Union wurde von 49% auf 47% reduziert, während die SPD eine Mutspritze von 37% auf 38% erhielt. Der Anteil der FDP wurde von 6% im Vormonat auf 7% erhöht, obwohl das zugrundeliegende Umfrageergebnis unverändert bei 4% lag. Durch diese Änderung der bisherigen Gewichtungs-Praxis lieferte das Politbarometer - was zweifellos unbeabsichtigt war - den „wissenschaftlichen Beweis“ für die Wahlkampfbehauptungen der Union.

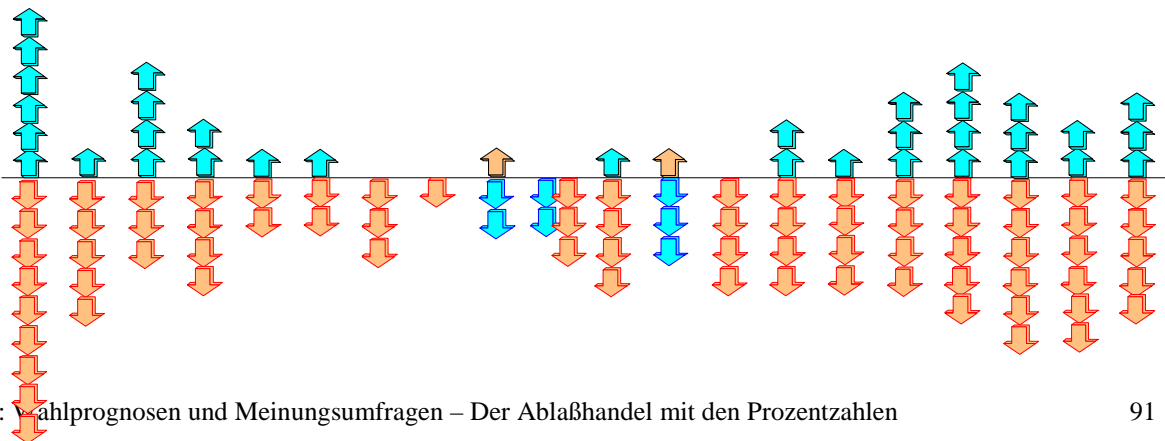
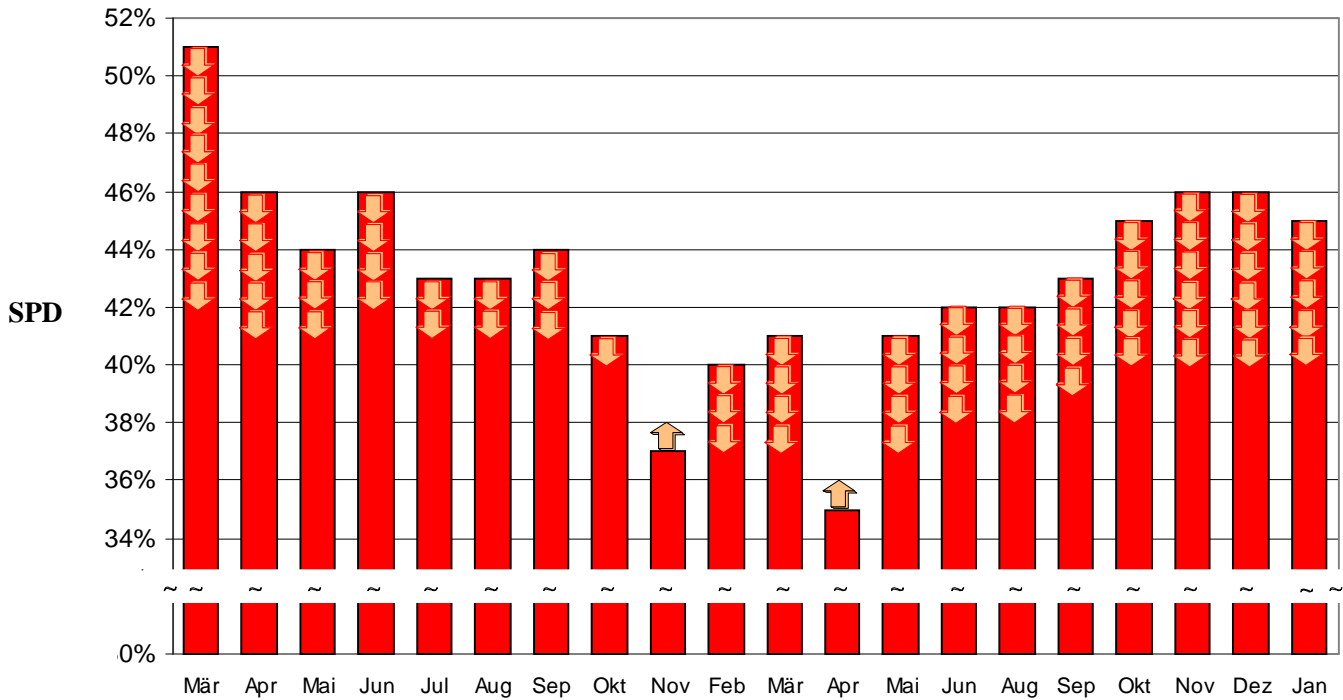
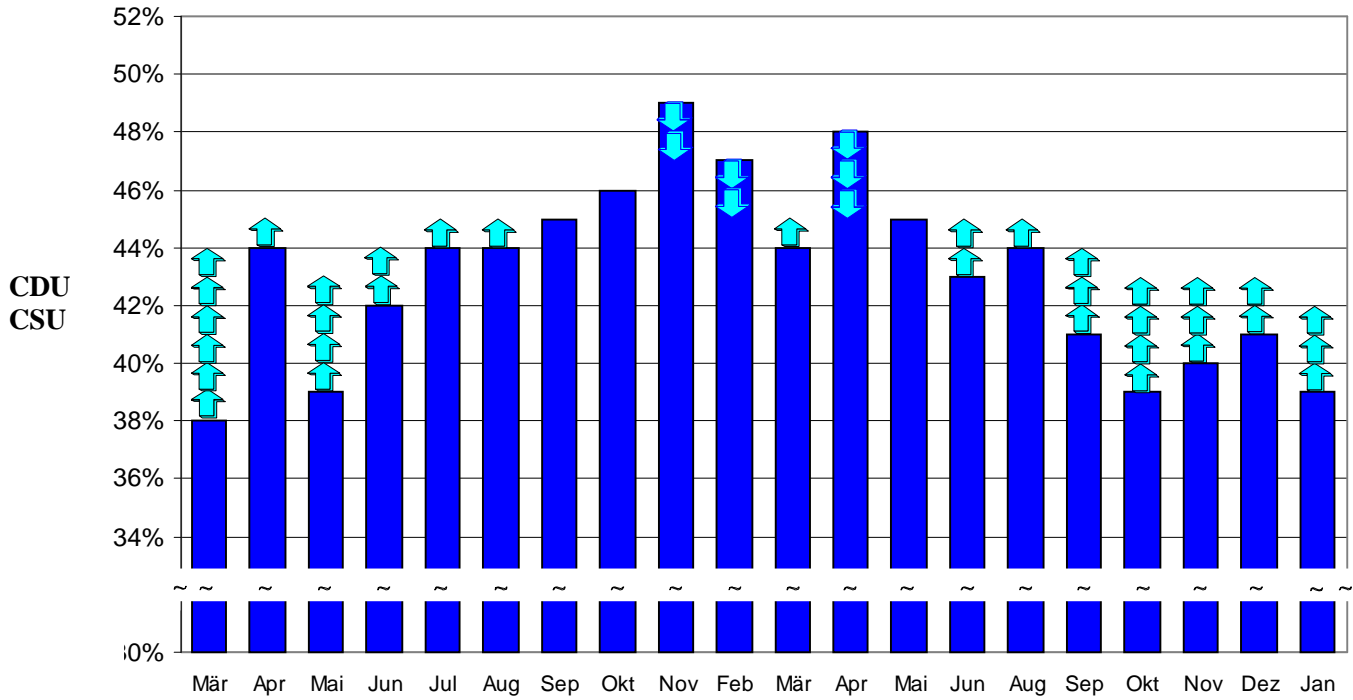
Das Verschweigen der Gewichtungen kann handfeste politische Auswirkungen haben. Ist es zum Beispiel zu verantworten, der FDP Monat für Monat 6% bis 7% auf den Bildschirm zu projizieren, obwohl ihr bei den laufenden Umfragen ermittelter Zweitstimmenanteil deutlich unter 5% liegt?

Mit diesen Zahlen wird doch dem Wähler suggeriert, die FDP sei sicher über der 5%-Hürde. Wer garantiert aber, daß die Wähler am Wahlsonntag für die „aus Erfahrung“ hinzugerechneten 2% tatsächlich aufkommen? Etwa die Gewichtungs-Zauberlehrlinge?

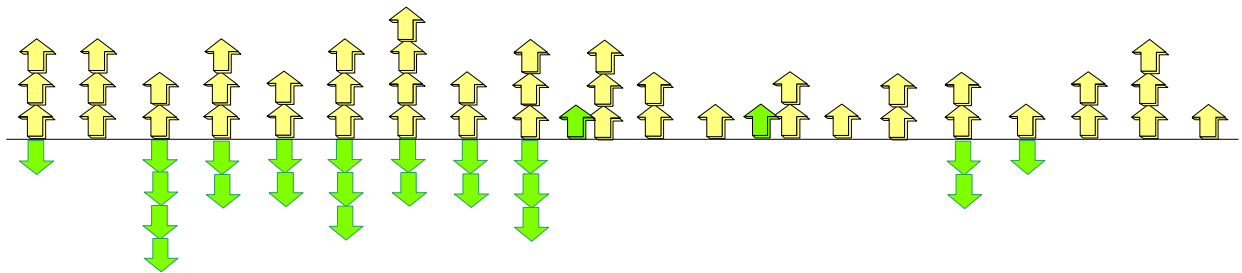
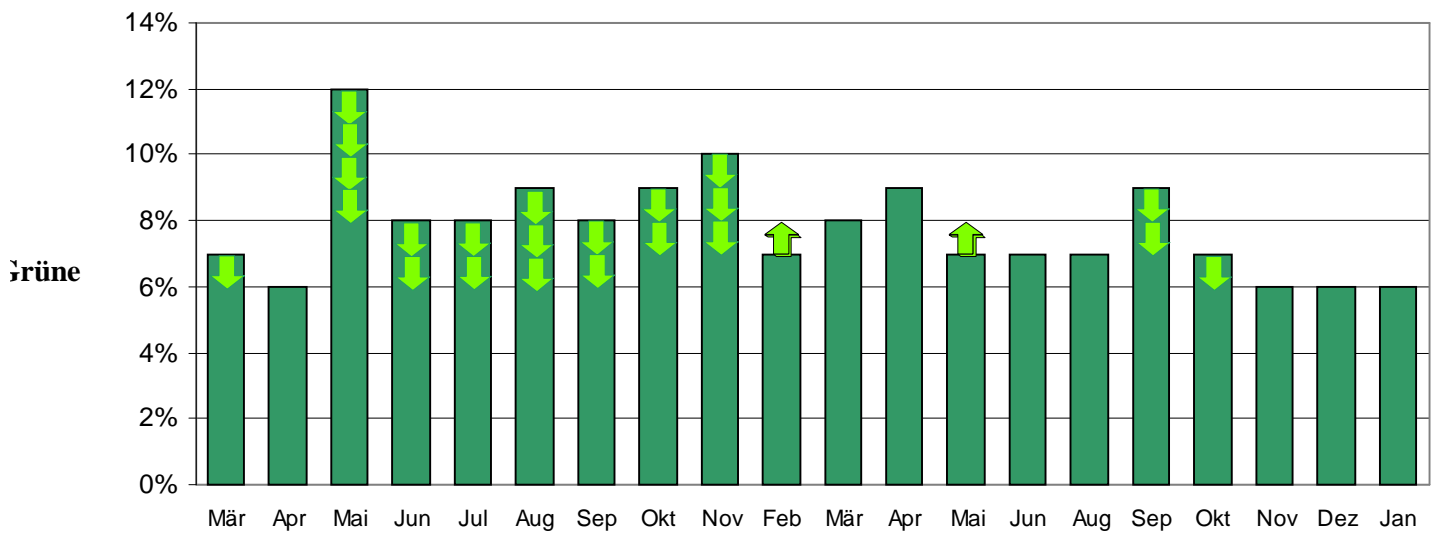
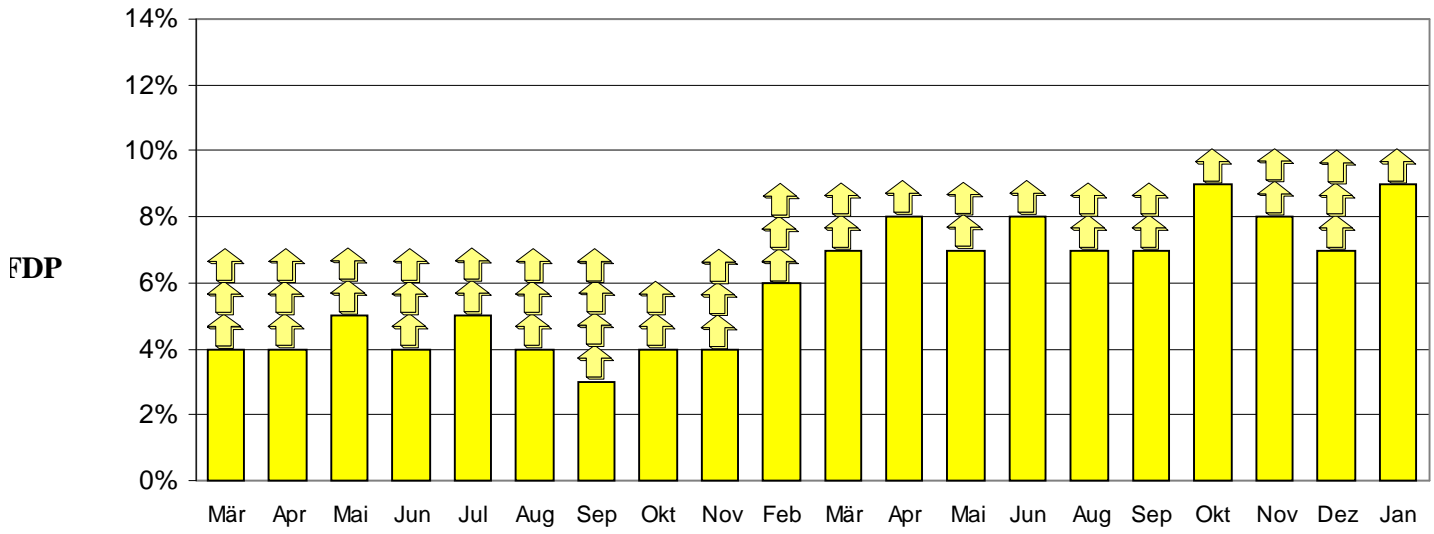
Es mag sein, daß die Stärke der FDP und der CDU/CSU mit Umfrageergebnissen in der Vergangenheit nicht selten unterschätzt worden sind. Ein Wahlforscher hat das Recht, wahrscheinlich sogar die Pflicht, darauf hinzuweisen.

Aber hat er das Recht, der Öffentlichkeit die Unzulänglichkeiten seiner Meßmethoden mit gefälschten Zahlen zu verheimlichen?

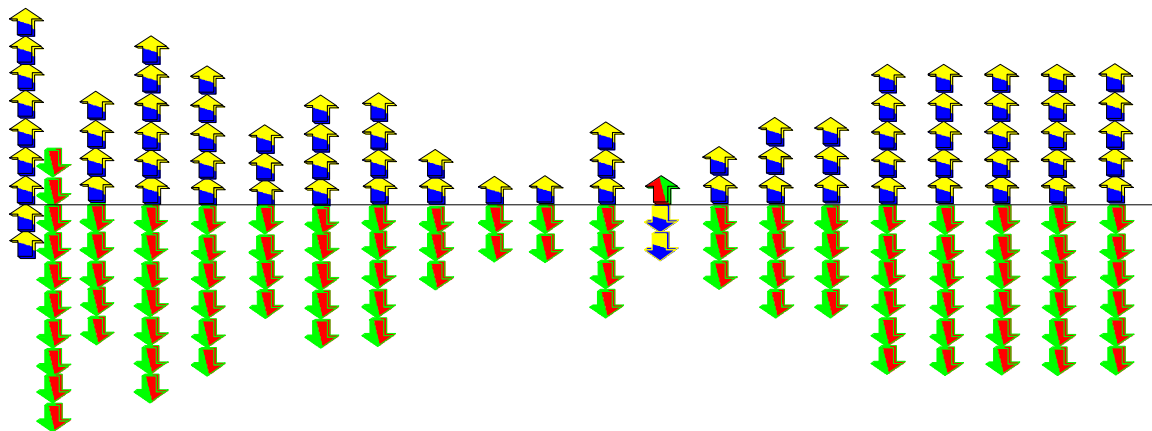
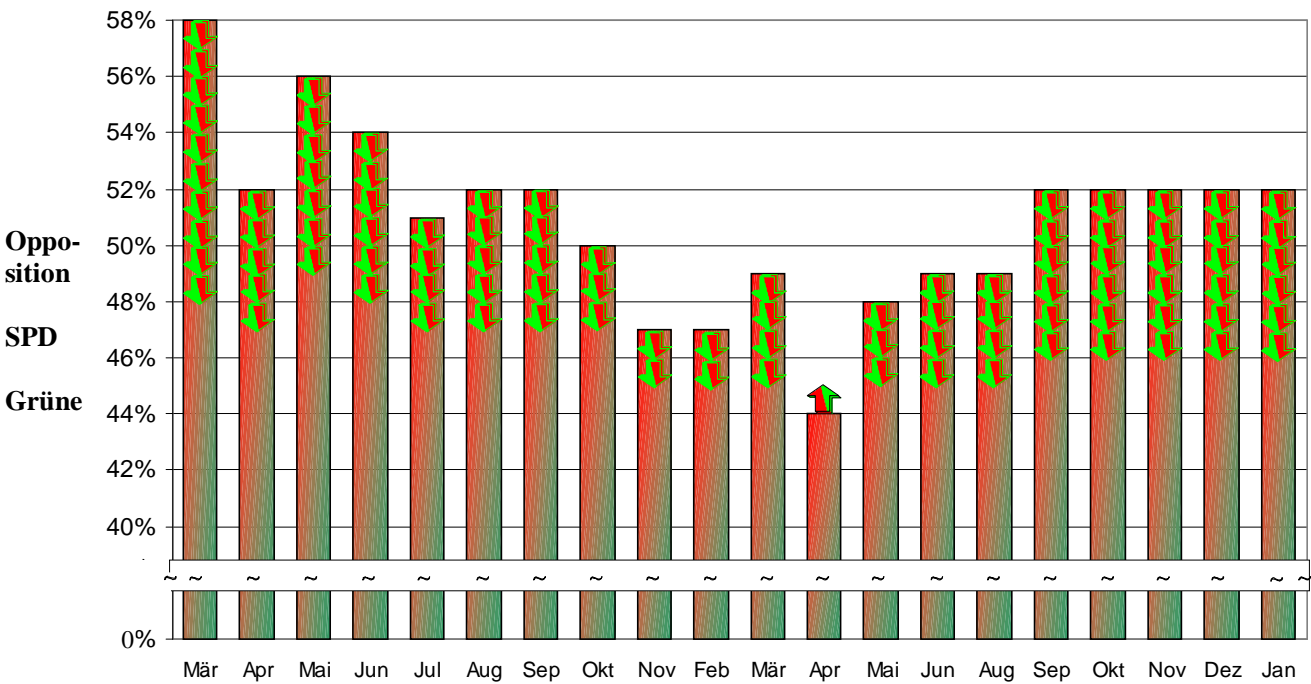
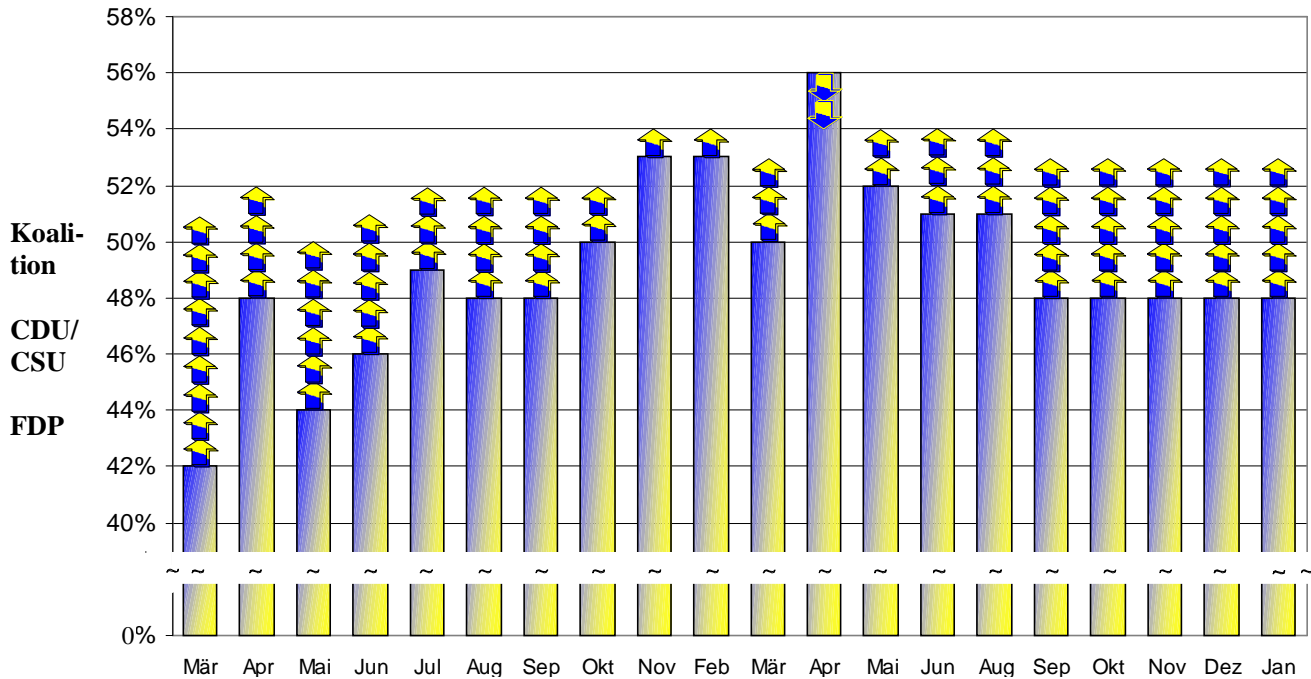
Politbarometer März 1986 bis Januar 1988



Politbarometer März 1986 bis Januar 1988



Politbarometer März 1986 bis Januar 1988



Wie das Politbarometer zur Gewichtung verdonnert wurde

Erwin K. Scheuch, Rheinischer Merkur Nr. 37/1987 (Merkur Extra):

„Wie wichtig Parteien die Veröffentlichung solcher Wahlvoraussagen nehmen, wird an einer Kontroverse über das sogenannte Politbarometer des ZDF deutlich. Diese Sendung berichtet nicht nur über die Meinung der Wähler, sondern pflegt als prominenteste Mitteilung das Ergebnis einer Frage nach der Parteisympathie vorzustellen. Daneben stellen Meinungsforschungsinstitute, die auch im Fernsehen zu Wort kamen und kommen, die sogenannte Sonntagsfrage, die wissen will, was er am nächsten Sonntag täte, gäbe es dann etwas zu wählen. Die Ergebnisse der Sonntagsfrage und der Sympathiefrage zeigten immer eine erhebliche Abweichung. Das kann nicht verwundern, weil eben verschiedene Dinge erfragt wurden. Aber nicht nur das Fragen war verschieden, sondern auch das Behandeln der Antworten, bevor sie publiziert wurden. Kein Umfrageinstitut veröffentlicht nämlich solche Antworten, ohne sie zu verändern, zu gewichten. Damit sollen Fehler ausgeglichen werden, die sich aus Antwortverweigerungen und anderen Störfaktoren ergeben können. Beim Politbarometer wurde das anders gewichtet als bei den übrigen Umfrageinstituten mit ihrer Sonntagsfrage, was an sich nur Fachleute zu interessieren braucht. Ich erlaubte mir in einem Aufsatz darauf zu verweisen, daß bei der Sympathiefrage die CDU 1986 mehrere Prozentpunkte –meist um drei bis vier – unter den Werten der anderen Institute lag, die SPD aber wesentlich stärker erschien als bei den Berichten über die Sonntagsfrage. „Zuschauer mißverstehen das Barometer aber als Aussage über die Wahlchancen“, überschrieb ich meinen Beitrag.

Das las die CDU als eine Bestätigung ihrer Vermutung, das Politbarometer betreibe mit seinen Zahlen Stimmungsmache zu Gunsten der SPD, und setzte entsprechend beim ZDF durch, daß in Zukunft die Forschungsgruppe Wahlen neben den Ergebnissen für die Sympathiefrage auch die Ergebnisse für die Sonntagsfrage vorzustellen habe. Deren bis dahin im ZDF veröffentlichten Ergebnisse sahen bei der Forschungsgruppe Wahlen nämlich auch nicht anders aus als bei den anderen Forschungsinstituten. Wie wichtig aber auch die SPD die Suggestivwirkung guter Ergebnisse für sich selber nahm, zeigt sich an den wütenden Angriffen im „Vorwärts“ auf diese Erweiterung der Berichterstattung.“

Wie das Politbarometer zur Gewichtung verdonnert wurde (Fortsetzung)

Kommentar: Beim Lesen des obigen Artikels von E.K. Scheuch entsteht der Eindruck, die Sympathiefrage und die Sonntagsfrage seien zwei völlig verschiedene Dinge, welche beide auf Umfrageergebnissen beruhen, und das Politbarometer habe vor März 1986 nur die Resultate der Sympathiefrage auf den Bildschirm gebracht, nicht aber diejenigen der Sonntagsfrage.

Das Politbarometer stellt die Sympathiefrage und die Sonntagsfrage. Nach den mir vorliegenden Unterlagen über 21 Monate von Januar 1986 bis November 1987 sowie September/Okttober 1985 – es gelang mir nicht, die Unterlagen für die Monate November und Dezember 1985 zu erhalten – stimmen die Antworten auf die beiden Fragen erstaunlich gut überein. Die Abweichungen liegen für die vier Parteien meist unter einem Prozent. Für jede Partei sind die Abweichungen sowohl positiv als auch negativ und sie heben sich im Durchschnitt praktisch auf.

Wenn das Politbarometer seine Berichterstattung wie gefordert erweitert hätte, dann hätte es meist zweimal die gleichen bzw. nur unwesentlich verschiedene Zahlen auf den Bildschirm bringen müssen. Dies war nicht der Sinn der Aktion. Das Politbarometer sollte die tatsächlichen Ergebnisse umgewichten (d.h. verfälschen), wie dies in der Prognoseindustrie üblich ist. **Der Vorschlag, die Meßergebnisse der Union um 3% bis 4% zu erhöhen, wurde durch die Ereignisse 1986/87 ad absurdum geführt.** Dies hätte vom September 1986 bis Mai 1987 meist

zu einer absoluten Mehrheit für die CDU/CSU geführt. Von November 1986 bis April 1987 wurden die Umfrageergebnisse für die CDU/CSU um durchschnittlich 2% heruntergewichtet, was das Schreckgespenst einer absoluten Mehrheit vom Fenster räumte.

Wie aus den Graphiken auf den vorangehenden Seiten hervorgeht, ergab die Gewichtung 1986/87 keinen erkennbaren SINN außer dem der Glättung **zwecks Kompensation auslosungsbedingter Abweichungen** und der Vortäuschung eines kontinuierlichen zeitlichen Trends. Die Gewichtung stellte sicher, daß die CDU/CSU nie „zu wenige“, aber auch nie „zu viele“ Prozente bekam. Zusammen mit der FDP wurde eine stetige Mehrheit von 50 bis 54% sichergestellt, egal wie die tatsächlichen Umfrageergebnisse lauteten.

Das Politbarometer zu zwingen, seine Umfrageergebnisse so zu fälschen, wie es in der Prognoseindustrie üblich ist, betrachte ich nicht als eine „sinnvolle Erweiterung“ der Berichterstattung. Eine solche würde allenfalls darin bestehen, daß das Politbarometer die Chance seiner Prognose – d.h. den auslosungsbedingten Spielraum – bekannt gibt, wie es auf Seite VII/9 illustriert wird. Dann würde sich das Problem der Gewichtung von selbst lösen ...

Bei Abänderungen der Rohergebnisse in den dargelegten Dimensionen stellt sich die Frage, weshalb überhaupt noch Meinungsumfragen für Wahlprognosen durchgeführt werden. Weshalb legt das Meinungsforschungsinstitut xy nicht einfach die Hand für einen Unionsanteil von x% ins Feuer? Seit den fünfziger Jahren schwankte die CDU/CSU bei den Bundestagswahlen bekanntlich zwischen etwa 45% und 50%, und dazwischen liegen nicht allzu viele Prozentzahlen, die als Prognosewert in Frage kommen Wie in Abschnitt VII gezeigt wurde, ist der auslosungsbedingte Spielraum größer.

Daß es auch „richtige“ Prognosen gibt, liegt zu einem guten Teil an der politischen Stabilität der Bundesrepublik und an den statistisch unvermeidlichen „Zufallstreffern“. Wenn die politische Landschaft in Bewegung geriete, würden den Wahlforschern die Felle sehr bald für jedermann sichtbar davonschwimmen. Einen Vorgeschmack für dieses Szenario lieferten etwa die Landtagswahlen in Bremen im September 1987: Die Form der im Zweiten Deutschen Fernsehen präsentierten Prognose,

CDU unter 30%

drückt so grotesk wie nur möglich das Eingeständnis aus, daß in einer *neuen* Situation nur nichtssagende Prognosen „sicher“ sind. Man traut den eigenen Daten nicht mehr, weil sie den bisherigen Rahmen sprengen. Das Wahlergebnis - 23,4% für die CDU, ein Verlust von 10% gegenüber 1983 - bestätigte die dumpfen Ahnungen der Forschungsgruppe Wahlen.

Bei großen Veränderungen - z.B. $\pm 10\%$ bis $\pm 20\%$ - werden die Karten neu verteilt. Dann müßten völlig unbekannte Größen gemessen werden, für deren Abschätzung es keine Präzedenzfälle gibt. Da reicht es nicht mehr aus, vom letzten Wahlergebnis Pi mal Daumen ein paar Prozente abzuziehen oder zu addieren. Nun muß der Wahlforscher eine Aussage machen, von der er weiß, daß er sich damit wahrscheinlich in die Nesseln setzen wird³².

In einer solchen Situation werden wir Zeuge einer bisher in der Zoologie nicht beobachteten Metamorphose: Unsere Politwetterfrösche werden aufgeschreckt aus ihrem seichten Teich hüpfen und als ratlos schnatternde Enten über dem Sumpf der Demoskopie flattern. Jeder fühlt sich allein - und von seinen Daten im Stich gelassen. Die Statistik hilft ihm nicht, sich aus der Schlinge zu

³² Bei einer normalen Umfrage im kommerziellen Alltagsgeschäft spürt der Meinungsforscher nicht die Faust der bevorstehenden Wahl im Nacken. Da ist er frei, seine Daten sprechen zu lassen, was immer sie auch sagen mögen....

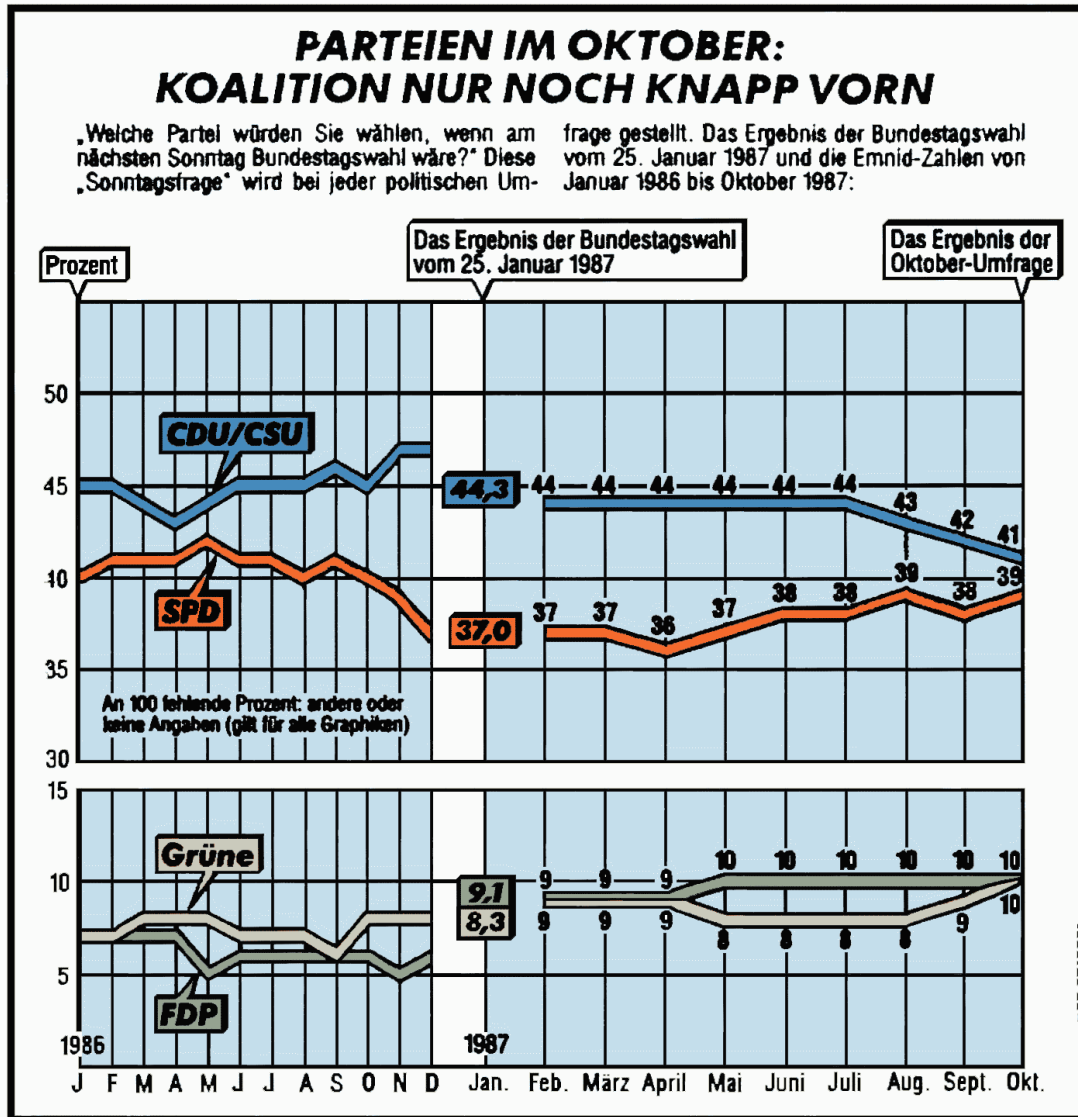
ziehen, in die er sich durch unhaltbare Versprechen verstrickt hat. Denn es treten - wie in den Abschnitten IV bis IX gezeigt wurde - als Folge der Auslosung des repräsentativen Querschnittes Abweichungen in einer Größenordnung auf, die der Meinungsforscher vor der Öffentlichkeit nicht eingestehen kann und will. So bleibt ihm als einziger Orientierungspunkt das Spektrum der „Prognosen“ der Konkurrenz. Die treibende Kraft ist die Angst, möglicherweise als einziger völlig daneben zu liegen. Zum Glück gibt es da die Gewichtung, mit deren Hilfe es möglich wird, die eigene Prognose im Mittelfeld der „Gerüchtebörse“, d.h. der „Konkurrenz“-Prognosen, anzusiedeln³³. Wenn die Sache schiefgeht, dann war man wenigstens nicht allein, was für das umsatzmäßig dominante und daher langfristig lebenswichtige kommerzielle Alltagsgeschäft von Bedeutung ist.

Eine Konsequenz dieser Angleichung (d.h. Koppelung) von Umfrageergebnissen - die manchmal auch bei normalen Änderungen des politischen Kräfteverhältnisses zu beobachten ist - besteht darin, daß die Fehler in der gleichen Richtung auftreten. Beispielsweise schätzen „alle“ Institute die CDU/CSU zu hoch oder „alle“ zu tief ein. Wenn der Zufall zu seinem Recht kommen würde, dann wären Fehler dieser Art extrem unwahrscheinlich.

Auch beim zeitlichen Verlauf der Parteistärken wird der Zufall vergewaltigt. Wenn z.B. im „Spiegel“ Nr. 45/1987 eine Graphik von Emnid abgedruckt ist, in welcher der CDU/CSU von Februar bis Juli 1987 konstant 44% zugeschrieben wird - und ebenso der FDP konstant 10% von Mai bis Oktober 1987 - dann handelt es sich um getürkte (d.h. geglättete bzw. gewichtete) Zahlen. Der Zufall mag ja seine Launen haben, aber so etwas schafft er nicht! (siehe folgende Seite)

³³ Eine neue Anwendung der Schweigespirale für Frau Noelle-Neumann?

Spieglein, Spieglein an der Wand,...
ist Augstein auch der Sauberste im ganzen Land?
Spiegel Märkte



Am Hofe König Rudolfs wird die Demoskopie seit langem als Kunst gepflegt: In monatlichen fünf- bis zehseitigen Spieglein-Berichten werden die Untertanen über die Lage im Reiche gründlich informiert. Objektiv! Da kodifiziert nicht einer der unzähligen Schreiber die vorherrschende Meinung a Hofe – nein, all die schönen Geschichtlein werden mit Zahlen belegt! Die Moral fließt förmlich aus den Zahlen, und kein verantwortungsbewußter Zahlengläubiger kann sich ihrem Sog entziehen. Von auslosungsbedingtem Spielraum und von Chance kann man von offizieller Seite in König Rudolfs Reich nichts erfahren. Nicht etwa, weil dann viele der schönen Demagogie-Geschichtlein wie Autos mit geplatzen Reifen davonholpern würden – nein, *Ehrenwort*, davon hat man am Hofe noch nie etwas gehört. Im Jahre *Barschels* und zu Ehren *Waldheims* haben wir deshalb ganz legal eine gutmütige Wanze beim Hof-Demagogen EMNID von König Rudolf installiert. Hören wir, was sie Pfeiffert:

Dies bedeutet in statistischem Sinne mit 90 % Wahrscheinlichkeit ($t = 1,64$), dass das tatsächliche Wahlergebnis für die einzelnen Parteien innerhalb folgender Fehlertoleranzgrenzen liegt:

CDU/CSU	= 42,3 - 49,2 %
SPD	= 32,9 - 41,1 %
F.D.P.	= 5,8 - 10,2 %
Grüne	= 5,8 - 10,2 %

EMNID - Institut GmbH & Co. KG

Tacke

F. Tacke

Urkundenrolle Nr. 5 / Jahrgang 1987

B e s c h e i n i g u n g

Hiermit wird der Emnid-Institut GmbH & Co. KG, Bielefeld, bescheinigt, daß die vorstehende Wahlprognose 1987 (Bundestagswahl am 25. Januar 1987) vom 24. Januar 1987 zu notariellem Protokoll - UR Nr. 4/87 - am 24. Januar 1987 in meine Verwahrung gegeben worden ist. Den versiegelten Umschlag habe ich heute, am 26. Januar 1987 geöffnet.

Bielefeld, 26. Januar 1987



H. Vitor
W. V. V.

XI Schlußfolgerungen

Die statistischen Grundlagen von Wahlprognosen und Meinungsumfragen basieren u.a. auf dem sogenannten schwachen Gesetz der großen Zahlen, vor allem aber auf dem unersättlichen Bedarf der zahlgläubigen Demoskopie-Endverbraucher.

Als Meßinstrument dient den Meinungsforschungs-Instituten eine *Elle*, nämlich die „*schwach repräsentativen*“ Tausender-Querschnitte, die mit einem nicht gerade stubenreinen Zufallsverfahren erstellt werden. Sie suggerieren aber durch die Präsentation ihrer Resultate, sie könnten damit *Millimeter-Bruchteile* messen, als hätten sie eine *Schieblehre*.

Nach den Gesetzen der Statistik ist jedes Meßresultat auch ein potentielles Miß-Resultat. Weil

- 1) die Ausfälle und Verzerrungen unkontrollierbar sind,
- 2) die Elle gegenüber den Schwächen der Interviewer und des Fragebogens immun ist,
- 3) bei der Zufallsauswahl aus Kostengründen erhebliche Abstriche gemacht werden,
- 4) kein Demoskop das Ausmaß des auslosungsbedingten Spielraums eingestehen will und kann,

wird aus Furcht vor dem Mißresultat der untaugliche Versuch gemacht, das Meßresultat mit dem *Gewichtungs-Vorschlaghammer* ins Lot zu rücken. Wen wundert es dann, wenn *diese „Hau den Lukas“*- Prognosen nicht immer auf dieser Welt landen?

Wenn ein Wahlforscher oder ein Meinungsforscher eine Prozentzahl nennt - mit oder ohne Nachkommastelle - oder eine solche dem Auftraggeber zum freien öffentlichen Gebrauch überläßt, dann ist dies unlauterer Wettbewerb, wahrscheinlich sogar Betrug. Er weiß doch ganz genau, daß er mit seinen Methoden grundsätzlich nicht in der Lage ist - auch unter idealen Voraussetzungen nicht -, eine gesicherte Prozentzahl zu ermitteln. Er hat Bereiche für die auslosungsbedingten Abweichungen seiner Ergebnisse und die zugehörige Chance anzugeben und auf die anderen Fehlerquellen hinzuweisen. Das Meßverfahren ist ausreichend zu beschreiben, damit eine Beurteilung der Resultate möglich wird.

Die von der Gewichtung und dem Mißbrauch des Begriffes „repräsentativer Querschnitt“ ausgehende Gefahr für die Öffentlichkeit liegt nicht so sehr in der Manipulation, denn die bevorstehenden Wahlen und die Rückwirkung von Fehlprognosen auf das umsatzmäßig dominante kommerzielle Alltagsgeschäft wirken ausreichend als Korrektiv, oft sogar schon als Über-Korrektiv.

Die eigentliche Gefahr besteht in der nicht beabsichtigten Irreführung durch Falsch-Informationen als Folge einer mißglückten Gewichtung³⁴ und des Verschweigens der unvermeidlichen Fehler, die durch das Interview und die Auslosung des repräsentativen Querschnittes verursacht werden. Man gibt vor, etwas wissenschaftlich gemessen zu haben, während man in Wirklichkeit nur spekuliert hat.

³⁴ So meinte der Geschäftsführer Jung des Meinungsforschungsinstitutes Basis Research GmbH gegenüber der Wirtschaftswoche (Nr.1/2 vom 2.Januar 1987), daß die Gewichtung in letzter Zeit versagt habe: "Die Rohzahlen waren oft genauer als die gewichteten Zahlen."

Besondere Bedeutung kommt der Tatsache zu, daß durch die Gewichtung der Umfrageergebnisse die unabhängige Berichterstattung der verschiedenen Meinungsforschungsinstitute zerstört wird. Welches Institut wagt es schon, auf die Dauer Umfrageergebnisse herauszugeben, die sich wesentlich vom Gros der anderen Institute unterscheiden? Die *Rohergebnisse* sind und bleiben zwar „*unabhängige Zeugen*“, aber durch den Konformitätsdruck werden diese Zeugen mit der Gewichtung oft mundtot gemacht. Die Ausschaltung der unabhängigen Berichterstattung hinterläßt keine Spuren, weil das Ausmaß der Abänderung verschwiegen wird und damit die Symptome des Angleichungsprozesses unterdrückt werden. Meines Erachtens gehen die Kollektiv-Fehlprognosen zum Teil auf diesen „*Meinungsklimadruck*“ unter den Instituten zurück. *Die Rohdaten möglichst vieler Institute über einen längeren Zeitraum stellen die bestmögliche Information dar.*

Die für die Öffentlichkeit bestimmten Wahlprognosen werden meist mit besonderem Aufwand betrieben, weil sie durch die Wahl überprüft werden und der Ruf des Instituts dabei auf dem Spiel steht. Um beim gegenwärtigen Kräfteverhältnis in der Bundesrepublik politisch interessante Aussagen betreffend der Stärken der vier relevanten Parteien machen zu können, ist eine Genauigkeit erforderlich, die die Möglichkeiten der Demoskopie weit übersteigt. Diese Genauigkeit läßt sich, wie in diesem Artikel dargelegt wurde, auch mit erhöhten Aufwand nicht erreichen, denn der erforderliche Stichprobenumfang ist praktisch nicht realisierbar und der Interviewfehler wird dadurch nicht aus der Welt geschafft. Wegen des auslosungsbedingten Spielraums vermögen Meinungsumfragen kaum mehr an gesicherter Information zu liefern, als was man aufgrund der relativen politischen Stabilität in der Bundesrepublik ohnehin schon weiß. Graphische Darstellungen des zeitlichen Verlaufs von Parteistärken, Popularitätswerten usw. sollte man im allgemeinen als Zeichentricksfilme oder „*Comic strips*“ betrachten.

Der Unterschied zwischen den Meinungsforschungsinstituten in der Wahlforschung besteht hauptsächlich darin, daß die einen mit mehr Glück und einem etwas besseren Riecher als die anderen spekulieren, wobei nicht zu bestreiten ist, daß Frau Noelle-Neumann der Konkurrenz manchmal um eine Nasenbreite voraus ist. Aber Spekulation bleibt es allemal.

Entscheidend für Erfolg bzw. Mißerfolg von Prognosen sind die politische Stabilität, der gute Riecher bzw. etwas Glück bei der Gewichtung und die Launen des Zufalls. Mit Wissenschaft hat das Ganze nichts zu tun, die „*Wissenschaft*“ dient lediglich als Aushängeschild.

Bei kommerziellen Umfragen schlägt keine Stunde der Wahrheit, gibt es keine Wahl. Wehrlos ist ein Auftraggeber allerdings nicht. Er kann als Teil seines Auftrages verlangen, daß Wiederholungen der Befragung simuliert werden, wie dies auf S. III/16-19 dargestellt wurde. Bei den handelsüblichen Meinungsumfragen mit einigen hundert bis 2000 Interviews werden dann viele der oft zu Hunderten abgelieferten, angeblich so harten Zahlen wie Butter unter der Simulations-Sonne schmelzen. Dann werden die durch die Auslosung des „repräsentativen Querschnittes“ verursachten Abweichungen in einer Weise zu Tage treten, welche die in den Abschnitten VI bis IX angeführten Spielräume als vergleichsweise „mustergültig“ erscheinen lassen. **Je mehr Fragen in einer „Repräsentativumfrage“ gestellt werden und je umfangreicher die Antwortmöglichkeiten sind, desto mehr weicht der „repräsentative Querschnitt“ von einem Miniaturbild ab.** Nicht nur weil die Länge des Fragebogens bei den Interviewern und Befragten Erschöpfungssymptome auslöst, sondern weil die statistische Qualität ausgehöhlt wird. Aufgrund vorläufiger Resultate ist mit einer Vergrößerung der Konfidenzintervalle (= lotteriebdingte Spielräume) um bis zu einem Faktor zwei und mehr zu

rechnen. Dies wird demnächst ausführlich behandelt werden.

Als Auftraggeber würde ich kommerzielle Umfragen grundsätzlich bei zwei verschiedenen Instituten in Auftrag geben - um *grobe* Fehler eher diagnostizieren zu können - und dabei sicherstellen, daß die Institute, was die Meßmethoden betrifft, *unabhängig* voneinander arbeiten. Unterschiede von ein bis zehn Prozent sind - je nach Anzahl und Umfang der gestellten Fragen und Größe des repräsentativen Querschnittes (1000-2000) - nicht als *signifikant*, sondern als *methodenbedingt* zu betrachten.

XII NACHWORT

Die Stunde der Wahrheit - die Wahl - begeht der Wahlforscher feierlich, nachdenklich und offen, etwas abgespannt durch die Strapazen des Wahlkampfes, aber innerlich gestärkt durch den erhöhten Umsatz. Das Resultat wird er gelehrt und aus sicherer Distanz „erklärend“ kommentieren. Hat er - wie es in der letzten Zeit häufiger passierte - mit Dichtung und Gewichtung Pech gehabt, dann wird er an der Gewichtungsschraube weiterdrehen. Ist er mit der zuletzt verwendeten Gewichtung ausnahmsweise nicht auf die Nase gefallen, so wird er dies der neuen „Methode“ zuschreiben. Weshalb sollte er sich die Freude mit dem Gedanken vergällen, daß der Erfolg wohl eher darauf beruht, daß der Zufalls- und Interviewfehler „versehentlich“ weggewichtet wurden? Auch ein blindes Huhn findet mal ein Korn...

So oder so: Die Launen des Zufalls werden dafür Sorge tragen, daß die Gewichtungszauberlehrlinge nicht zur Ruhe kommen. „Die Rohzahlen waren oft genauer als die gewichteten Zahlen“, klagte der Geschäftsführer von Basis-Research der Wirtschaftswoche Nr.1/2 Januar 1987.

Die Wahl wird man überstehen, der Konkurrenz ging es auch nicht viel besser. Ein Gang nach Canossa? Schon eher ein Einkaufsbummel ins nahegelegene Florenz. Und dann eine Erholungsreise an den Golf von Korinth, verbunden mit einem Kulturtrip zu den Tempeln in Delphi.

Auf dem Flug von Mailand nach Athen schlägt man noch schnell im Reiseführer das Wissenswerte zum „Orakel von Delphi“ nach:

„Delphi (7.Jh.v.Chr.) liegt am Südhang des Parnaß (2459m), welcher im Altertum als Sitz des Gottes Apollo galt. Das Orakel Apollos war das berühmteste Orakel des Altertums. Seine große politische Bedeutung fällt in die archaische Zeit. Damals wirkte das Orakel auf die Verfassung der Städte, die Kolonisation und viele politische Unternehmen ein. Noch Platon erkannte Delphis Weisungen eine hohe Bedeutung für die staatliche Ordnung zu.“

Die Ausführungen über den großen politischen Einfluß des Orakels bringen unsere Wahlforscher so richtig auf den Geschmack. Daß selbst Platon die Demoskopie und ihre Weisungen für das Wohl der Gemeinschaft und die Aufrechterhaltung der politischen Ordnung als Notwendigkeit betrachtete, macht unseren Wahlforschern wieder neuen Mut. Welch gute alte Zeit! Anstatt sich wie bisher jede Woche zwecks Entspannung den Glauben an die eigenen Prognosen beim Nervenklempner auf der Couch wieder ausreden zu lassen, könnte man es doch einmal mit dem Orakel versuchen, das seine Kunst offensichtlich viel besser beherrscht.

Im Reiseführer erfährt man, wie eine Konsultation beim Orakel abläuft:

Das Medium von Delphi war eine ältere Frau: die Pythia
Nach einer Opfergabe nahmen die Pythia und ihre ratsuchenden Klienten ein gemeinsames Bad in dem kastalischen Springbrunnen, wo ihr die Probleme vorgelegt wurden. Nach diesem Ritual ging sie in den Tempel und begann mit der Meditation. In ihrem veränderten Bewußtseinszustand fing die Pythia zu sprechen an. Ihre Wahrsprüche wurden den Ratsuchenden nicht direkt mitgeteilt, sondern sie wurden von Priestern interpretiert und in kunstvollen Versen niedergeschrieben, deren Auslegung nicht immer einfach war.

Auf der Fahrt nach Delphi beraten unsere Wallfahrer, was sie der Orakelpriesterin Pythia als Opfergabe bieten könnten. Sie einigen sich auf ein Wertpapier und legen ihre letzte Prognose auf den Altar. Vom Schatzmeister des Tempels nach Nominal- und Kurswert befragt, meinen sie, solche Papiere würden selten zum Nennwert genommen und die Kotierung an der Gerüchtebörse sei erst kürzlich erfolgt. Entscheidend sei aber der innere Wert, und der Selbstkostenpreis habe fast 100.000,- DM betragen. Der Schatzmeister erwidert sachlich aber höflich, der Altar sei keine Endlagerungsstätte und fünf Mark in bar wären ihm lieber. Nach dieser kalten Dusche verspüren unsere Wahlkampfmanager keine Lust mehr auf ein Bad im Nymphen-Springbrunnen. Auf ihre Bitte hin verlegt die Pythia die Reinigungsprozedur in die Sauna. Dort tragen sie ihr mit bewegten Worten die Klage über die Krankheit des repräsentativen Querschnittes vor:

seine **Wankelmütigkeit** !

Den Wahrspruch der Pythia am nächste Morgen

- Achte das Einmaleins der Statistik -

haben unsere Wahlforscher bis heute nicht enträtselt. Als Datenhändler hat man nicht viel Zeit zur Muße, auch werden sie den Verdacht nicht los, daß die Pythia sie im Dampfbad mit der Konkurrenz aus „Wetten daß...“ verwechselt hat. Wie dem auch sei, solange unsere Polit-Wetterfrösche für ihre gewichteten Lottozahlen dank der guten Tarnung reißenden Absatz finden und jedermann - von Politikern bis hin zur Regenbogenpresse - diesem Zahlen-Horoskop hörig ist, verhallt der Wahrspruch der Pythia ungehört. Man kann von unseren Datenhändlern - Kreuzritter für freies Unternehmertum - auch nicht erwarten, daß sie sich freiwillig in den eigenen Speck schneiden. Das wäre gänzlich neu in einer Narren-freien Marktwirtschaft.

In Meinungsforschungskreisen ist die „Pythia vom Bodensee“ eine wohlbekanntes und ein wohlbeachtetes Phänomen.

Bei dieser Institution handelt es sich aber nicht um eine Reinkarnation der Pythia in unserem Jahrhundert - allenfalls um eine von der Frankfurter Allgemeinen (FAZ) bestellte Thronräuberin - sondern um den Kosenamen von Frau Prof. Dr. Dr. h. c. Noelle-Neumann im eingeweihten Blätterwald.³⁵

Daß zwischen den beiden Damen Welten liegen, oder, wie wir in der Schweiz zu sagen pflegen, sie ihr Heu nicht auf der gleichen Bühne haben, geht auch aus dem „Wahr“-spruch der „Pythia vom Bodensee“

Die Schweigespirale, Öffentliche Meinung - unsere soziale Haut hervor, der mit dem Einmaleins der Statistik auf Kriegsfuß steht, soweit es die Anwendung auf die Gegenwartspolitik betrifft.

In den Auswirkungen auf die Staatspolitik ist jedoch eine gewisse Parallelität zwischen den beiden Damen unverkennbar. Ich kann mir allerdings nicht vorstellen, daß Platon den Prozentzahlen der „Pythia vom Bodensee“ so leichtfüßig gefolgt wäre wie Helmut Kohl³⁶.

³⁵ Den Hinweis auf die "Pythia vom Bodensee" verdanke ich Thomas Hochkirchen. Die Bezeichnung "Pythia vom Bodensee" ist seit langem im Umlauf und geht offenbar auf die FAZ zurück.

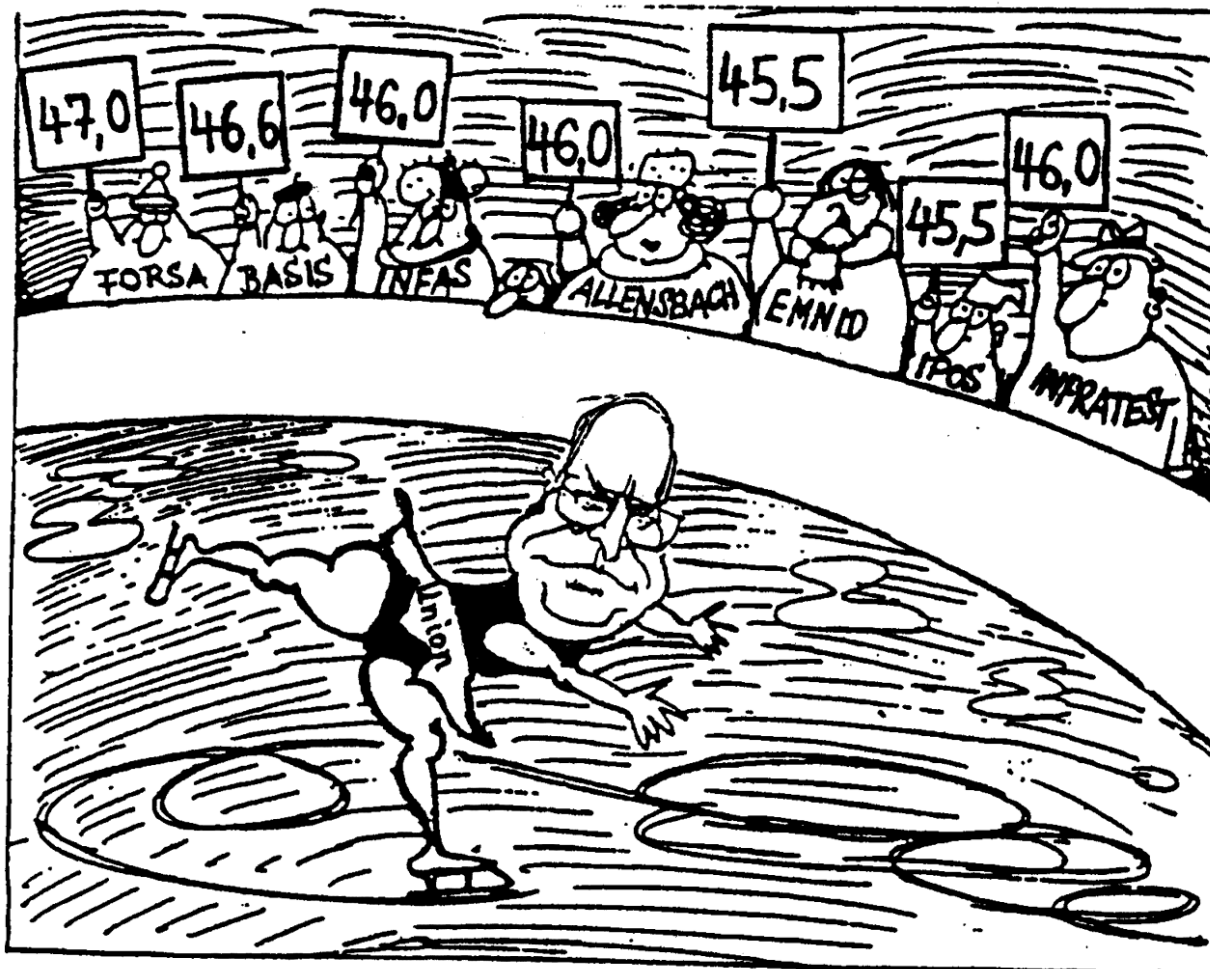
³⁶ und einige seiner Vorgänger im Amt mit der Ausnahme von H. Schmidt: "Ich hasse Demoskopie" erklärte dieser

abrupt in einem Gespräch mit einem Journalisten, in welchem er Dinge aufzählte, die ihm Freude bereiten.

Kohls Prognosen- und Zahlengläubigkeit wird illustriert durch seine erste Stellungnahme zu den Bundestagswahlen vom 25. Januar 1987. Nachdem sich bei den Hochrechnungen am Wahlabend ein Resultat von knapp 44% für die Union abzeichnete - also ein Verlust von etwa 5% - , zeigte er sich im ZDF enttäuscht:

"Wir in der Union haben unser Ziel nicht erreicht. Ich selbst habe gerechnet in einer Größenordnung von 46%, wir sind deutlich unter dieser Zahl geblieben. Das ist betrüblich.
Wir haben Verluste erlitten, Verluste, mit denen niemand in diesem Umfang gerechnet hat."

Aus dieser Aussage geht hervor, daß er bzw. seine demoskopische Ratgeberin die Vorstellung haben, der Unterschied von 46% und 44% sei so eklatant, daß er einem Demoskop unendlich entgehen könne. Als Erklärung für den Unterschied führte er die gegenüber 1983 um 5% geringere Wahlbeteiligung an. Auf den folgenden Seiten ist ein Statistik-Kurs in Bildform für das Bundeskanzleramt abgedruckt.

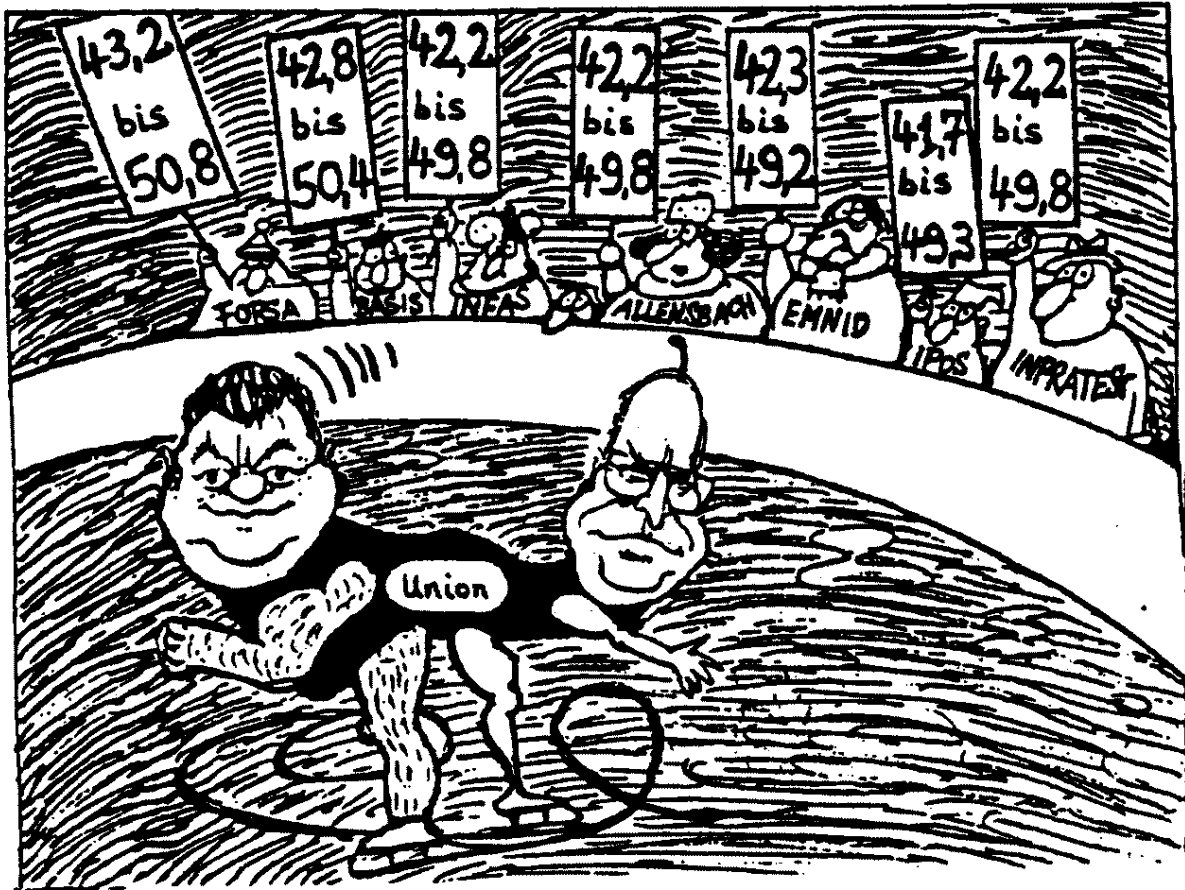


frei nach Mannheimer Morgen

Diagnose : Die Mitglieder der Jury haben geschummelt. Sie haben ihre Umfrageergebnisse solange umgewichtet, bis die vorherrschende Meinung herauskam.^(*)

^(*) Auf der Basis von 1000 Interviews besteht nämlich eine Chance von weniger als 1%, daß die Resultate von sieben *unabhängigen* Meinungsforschungsinstituten einen Spielraum von lediglich 45,5% bis 47% aufweisen. Selbst bei 2000 Interviews liegt diese Chance noch deutlich unter 5%

Erste-Hilfe-Kurs in Statistik



Siamesische Kür

frei nach Mannheimer Morgen

Danksagung

Diese Arbeit wurde durch eine Anfrage der Zeitschrift für Markt-, Meinungs- und Zukunftsforschung angeregt. Die Durchführung wurde durch ein vorzeitig gewährtes Forschungs-Freisemester ermöglicht. Die Arbeit stellt das Resultat einer Teamarbeit unter meiner Anleitung dar. Mein besonderer Dank geht an *Edith Achiman* und *Ullrich Tesche*.

Für Korrekturen und redaktionelle Vorschläge danke ich Thomas Hochkirchen und Frank Sauder. Bernd Wallaschkowski bemühte sich um die Textverarbeitung mit seinem privaten PC.

Ohne die zahlreichen Gespräche mit Nichtstatistikern - u. a. mit W. Huncke, G. Pinkernell und meiner Frau - die mich mit ihren Fragen oft zur Verzweiflung trieben, hätte ich nicht versucht, den statistischen Jargon über Bord zu werfen und die Auslosung des repräsentativen Querschnittes zum Mittelpunkt meiner Darstellung zu machen.

Ich danke meinen Kollegen und Studenten G. Heindl, H. Hoffmann, H. Scheid, E. Scholz, D. Vogt und A. Zieschang sowie meiner Frau und Herrn M. Kroemer für die kritische Durchsicht der verschiedenen Fassungen des Manuskriptes. Durch ihre Fragen und Vorschläge wurde die Gestaltung der Arbeit verbessert.

Ich danke Herrn Witte vom Uni-Rechenzentrum für seine Kooperation bei der Durchführung der vielen Simulationen.

Ich danke Yolanda Aguado, Ruth Horine und Martine Mérigeau sowie Agustin Blanco-Roiz und Gilbert Liébray für die Übersetzung der Zusammenfassung ins Französische und Spanische. Durch ihre Rückfragen und die daraus resultierenden Gespräche wurde die Gestaltung wesentlich verbessert.

Mein Dank geht auch an den Demokrit Verlag und an Herrn G. Wickert für die gewährte Unterstützung und für ihre Engelsgeduld.

Der Lotteriebertrag des repräsentativen Querschnittes

Wahlprognosen und Meinungsumfragen und der Ablaßhandel mit Prozentzahlen

Zusammenfassung

Das Ziel dieser Arbeit besteht darin, dem Nicht-Statistiker mit Hilfe von frappanten Beispielen und zugespitzten Formulierungen klar zu machen, daß die Meinungsforscher in der Bundesrepublik ihre hochtrabenden Versprechen betreffend Wahlprognosen seit über dreißig Jahren aus statistischen Gründen nicht einhalten können. Zunächst wird gezeigt, daß der fromme Spruch vom „repräsentativen“ Querschnitt - im Sinne eines Spiegel- oder Miniaturbildes der Bevölkerung - ein reiner Etikettenschwindel ist. Es ist aus mathematischen Gründen unmöglich, Miniaturbilder zu realisieren, die aus 1000 oder 2000 Personen bestehen. Sogar mit 10000 Personen geht es nicht, aber dann wären Meinungsumfragen wenigstens „unbezahlbar“, und die meisten Probleme würden sich erst gar nicht stellen. Selbst wenn Miniaturbilder eines Tages durch höhere Gewalt auf die Welt gebracht würden, könnten sie die völlig unrealistischen Vorstellungen und Erwartungen der Meinungsforscher und ihrer Kunden nicht erfüllen.

Durch das Suggestivwort „repräsentativer“ Querschnitt wird bei den Konsumenten und Auftraggebern von Meinungsumfragen der Eindruck erweckt, die gelieferten Daten seien glaubwürdig und zuverlässig. Die Autorität des „repräsentativen“ Querschnittes löst bei vielen eine Art Verpflichtung und Resignation aus, sich mit der gelieferten Information abzufinden, als hätte die große Mehrheit der Bevölkerung ganz persönlich zu ihnen gesprochen. Jeder Branchenkundige weiß jedoch, daß es die Zufallsstichprobe ist - d.h. eine per Lotterie getroffene Auswahl - welche die mit dem Werbeslogan „Repräsentativumfrage“ gemachten Versprechen einlösen soll (von den Quotenstichproben sehen wir fürs erste ab ...). **Die Auswahl per Lotterie liefert zwar viel bessere Resultate, als der Laie sich vorstellen kann, aber bei weitem nicht so gute, wie die Prognoseindustrie und ihre stille, aber äußerst geschickte Lobby seit Jahrzehnten behaupten.** Mit Computersimulationen läßt sich zeigen, daß die Auswahl per Lotterie in „repräsentativen“ Querschnitten größere Abweichungen verursacht, als bisher angenommen wurde. Dies macht es praktisch unmöglich, auf die Fragen nach den politischen Kräfteverhältnissen in der Bundesrepublik verbindliche Antworten zu geben, weil relativ geringe Verschiebungen die politische Situation grundlegend verändern können.

Im Gegensatz zu anderen Ländern bewegen sich hier die Änderungen der relevanten Parteistärken in engen Grenzen. Seit 1965 lagen beispielsweise die Resultate für die CDU/CSU zwischen 44,3% und 48,8%, für die SPD zwischen 37,0% und 45,8% und für die FDP zwischen 5,8% und 10,6%. Zählt man ab 1983 bei der SPD die Grünen hinzu, die zu einem guten Teil von der SPD „abgewandert“ sind, dann reduziert sich der Bereich der SPD auf 39,3% bis 45,8%. Im Vergleich dazu weisen die üblichen Vertrauensintervalle - deren Aufgabe bekanntlich darin besteht, Aufschluß darüber zu geben, in welchem Umfang der „repräsentative“ Querschnitt durch auslosungsbedingte Abweichungen verzerrt sein kann - ein Ausmaß von $\pm 2\%$ bis $\pm 4\%$ auf. Wie man sieht, liegen die Vertrauensintervalle in der Größenordnung der historischen Schwankungsbreiten der Parteistärken und übersteigen diese sogar gelegentlich. **Der „repräsentative“ Querschnitt kann also wegen der auslosungsbedingten Fehler nicht mehr Information liefern als die vorangehenden Wahlresultate, oft sogar weniger.**

Bei dieser Sachlage bleibt den Wahlforschern „nur“ der Ausweg, die auslosungsbedingten Abweichungen unter den Teppich zu kehren. Die Bekanntgabe von aktuellen Wahlprognosen in statistisch vertretbarer Form, d.h. unter Angabe der auslosungsbedingten Abweichungen wie z.B.

CDU/CSU 42% - 50%, SPD 35% - 43%, FDP 5% - 10%, Grüne 4% - 9%

würde so lächerlich wirken, daß sie nur in Witzblättern oder als Comic-strip möglich wäre. Wer würde für eine solche Banalität 50000 Mark hinblättern? Noch ärgerlicher wäre der laufende Einnahmeausfall, denn für den gleichen faulen Witz kann man nicht jede Woche abkassieren.

Also tun die Wahlforscher so, als hätten sie das statistische Ei des Columbus entdeckt: Eine magische Formel, welche die auslosungsbedingten Fehler in nichts auflöst. Mit Hilfe einer Umgewichtungszprozedur, deren alchimistischer Gehalt unverkennbar ist, wird den Vertrauensintervallen die überschüssige Luft abgelassen, und alsdann schlüpfen „exakte Zahlen“ aus der Retorte, bis hin zur Stelle nach dem Komma. Dieses Vabanquespiel basiert auf einer Umgewichtung der aktuellen Umfrageergebnisse in Abhängigkeit vom früheren Wahlverhalten (sogenannte Recall-Frage). Der statistische Gehalt dieser „Einkommens-umverteilung“ im Prozentbereich beruht auf dem folgenden „Gedankengang“: Falls die Wähler die Absicht haben, für die gleiche Partei zu stimmen wie bei der letzten Wahl vor vier Jahren, und falls sie dies dem Interviewer anvertrauen - unsere Wahlforscher hoffen also, daß die Wähler ihre Gedächtnislücken selbst „reparieren“, auch diejenigen, die damals noch nicht wahlberechtigt waren oder seither gestorben sind - ja, dann funktioniert die Umgewichtung statistisch tadellos. Abgesehen davon braucht man unter diesen Annahmen weder Meinungsumfragen noch eine Wahl durchzuführen, um Resultate zu erhalten. Man kopiere einfach die Resultate der letzten Wahl

Tragischerweise entpuppt sich das Ei des Columbus unserer Wahlforscher in der Praxis als Kuckucksei. Die alchimistische Recall-Formel produziert häufig surrealistische Resultate, die nicht zu vermarkten sind. So sind weitere Meinungsklima-Kuren vonnöten, deren phantasievolle Nomenklatur den höchsten Ansprüchen genügt. Ein Schönheitsfehler mag darin bestehen, daß das Endresultat ganz anders ermittelt wird: **Eine Prognose wird einfach so gemacht, daß das letzte Wahlresultat in Abhängigkeit von der vermeintlichen politischen Windrichtung Pimal Daumen ein bißchen nach oben oder unten korrigiert wird.** Die aktuellen Umfrageergebnisse spielen dabei eine untergeordnete Rolle. Sie dienen allenfalls dazu, die vermeintliche politische Windrichtung zu ermitteln. In Anbetracht der unkontrollierbaren auslosungsbedingten Abweichungen und Interviewfehler erfordert dieses Abenteuer jene Mischung von grenzenlosem Selbstvertrauen und unerschütterlicher Ahnungslosigkeit, welche sonst zum Rüstzeug von Politikern gehören. **Die laufend feilgebotenen Prognosezahlen für die sogenannte Sonntagsfrage**

„Wie würden Sie wählen, wenn am nächsten Sonntag Bundestagswahl wäre?“

reflektieren also ausschließlich die Spekulationen der Wahlforscher, oder, wie sie es salbungsvoll ausdrücken: ihre Gewichtungskunst.

Bei dieser Sachlage versteht es sich von selbst, daß Trendangaben keinen Bezug zur Realität haben, weil sie durch auslosungsbedingte Pseudo-Trends bis zur Unkenntlichkeit entstellt sind. Um es boshaft auszudrücken: Die „Gunst“ des Zufalls beschert den Wahlforschern laufend Neuigkeiten und Schlagzeilen. Was sich wirklich abspielt, das weiß kein Mensch, und die von den Meinungsforschungsinstituten gelieferten Daten über das zeitliche Auf und Ab der

Parteistärken - z.B. CDU/CSU -2,8%, SPD +1,9%, FDP +1,3%, Grüne -0,4% - haben ungetrübten Horoskop-Charakter. Sie täuschen Veränderungen vor, die quantitativ mit an Sicherheit grenzender Wahrscheinlichkeit falsch sind und die häufig nicht einmal qualitativ richtig sind.

Diese Probleme werden dadurch potenziert, daß in der Meinungsforschung ein falsches statistisches Modell zur Anwendung kommt: die Binomialverteilung und ihre Approximation - die Normalverteilung. Beide sind bei den empirischen Sozialforschern auf der ganzen Welt populär. Diese Verteilungen belohnen den glücklichen Gläubigen mit den kleinstmöglichen Vertrauensintervallen - die von der Auslosung des „repräsentativen“ Querschnittes verursachten Abweichungen werden also minimal. Was will er mehr? Er wird nicht belohnt, wenn er der Tatsache ins Auge sieht, daß dieses Modell in keiner Weise eine statistisch legitime Beschreibung seiner Meinungsumfrage darstellt, sondern der Ausdruck von reinem Wunschdenken ist, das von einer Generation von Lehrbüchern auf die nächste vererbt wird, wie ein Fluch bis ins dritte und vierte Geschlecht. Das Modell unseres Gläubigen setzt voraus, daß die von ihm organisierte Meinungsumfrage aus einer einzigen Frage besteht, die mit „ja“ oder „nein“ zu beantworten ist. Jede Hausfrau weiß, daß es nicht genügt, einmal „nein“ zu sagen, wenn der Interviewer vor der Tür steht. Es ist ihr sonnenklar, daß der ungebetene Gast - der sich an einem schönen Nachmittag unvermittelt in ihrem Wohnzimmer festsetzt - ihr mit seinem seitenlangen Fragebogen ein Loch in den Bauch fragen wird und daß sie vor lauter Möglichkeiten oft nicht weiß, welche Antwort sie geben soll. Statistisch bedeutet dies, daß das korrekte Modell für eine Meinungsumfrage nicht aus einer Binomialverteilung besteht, sondern aus dutzenden von Multinomialverteilungen mit verschiedenen Freiheitsgraden, die gleichzeitig betrachtet werden müssen, was - wenn überhaupt - meist nur mit einer Computersimulation möglich ist. Die Berechnung von Vertrauensintervallen bei Umfragen mit Hilfe der Normalverteilung erinnert an jenen Metzger, der eine Wurst in ein Rudel von Hunden wirft und dann jedem Besitzer den Schmaus in Rechnung stellt. Gewiß kann kein Besitzer beweisen, daß es nicht sein Hund war, der die Wurst erwischte. Dennoch ist der Metzger kein Modell für Integrität. Solche „Vertrauensintervalle“ haben die Funktion von statistischen Alibi-Übungen und Milchmädchenrechnungen.

Computersimulationen von Zufallsstichproben zeigen, daß solche Vertrauensintervalle bis zu 100% oder mehr vergrößert werden müssen, damit die gewählte statistische Sicherheit eingehalten werden kann. Je umfangreicher die Meinungsumfrage, desto größer der Zuschlag. Die Größe der tatsächlichen Vertrauensintervalle hängt von der Anzahl der Fragen und der Antwortmöglichkeiten ab und nicht allein von der Anzahl der Interviews und der statistischen Sicherheit, wie bisher angenommen wurde. Auf den Punkt gebracht: Unter der Glut der Simulations-Sonne schmelzen Berge von angeblich so harten Daten wie Butter.

Wenn der berühmte Psychologe Carl Gustav Jung - der Entdecker und Erforscher des kollektiven Unbewußten - noch lebte, würde er sich wohl gezwungen sehen, seine lange Liste kollektiver Inhalte durch einen weiteren Archetyp zu ergänzen: Zahlengläubigkeit.

Die Alchimisten vergangener Zeiten würden ihre heutigen Kollegen wohl beneiden. Was jenen während Jahrhunderten mangels geeigneter Stoffe und Know-how versagt blieb, ist für unsere Meinungsforscher zur Routine geworden:

aus Prozentzahlen und Aberglauben eine Goldgrube zu machen.

The lottery aspect of the representative sample

Election forecasts and opinion polls

The traffic with percentage points

The purpose of this paper consists in demonstrating to non-statisticians by means of striking examples that for statistical reasons pollsters in the Federal Republic of Germany cannot deliver the goods they have promised the public for the last thirty years. The pious wish of a “representative” sample in the sense of a mirror image or miniature image is shown to be a public relations gimmick. For mathematical reasons it is impossible to construct such images based on 1000 or 2000 people. With 10000 it won't work either, but at least opinion polls would become so costly that most problems would not exist in the first place. Even if miniature images were to be created by an act of God, they could not fulfil the totally unrealistic ideas and expectations of the pollsters and their customers.

The highly suggestive term “representative” sample sounds so reassuring. It brainwashes people into believing that election forecasts and opinion polls must be accurate and reliable. The moral authority of the “representative” sample provokes in many people a feeling of obligation and resignation. They swallow the numbers with which they have been fed, as if the silent majority had spoken to them in person. Everybody in the business knows however that it is random sampling - i.e. the selection by lottery - that is supposed to back up the promise behind “representative” sample, whatever the latter may mean. (As for quota samples, we prefer to look the other way.) **No doubt the lottery principle leads to much better results than a layman expects, but they are not nearly as good as the forecasting industry keeps claiming ever since it “incorporated” as a pressure group.** By means of computer simulations it can be shown that random sampling causes larger errors than has been assumed. These errors make it virtually impossible to provide politically relevant information in the Federal Republic of Germany by means of “representative” samples, because relatively minor changes in party strengths can lead to a completely different political situation.

For in Germany, unlike other countries, the strengths of the dominating parties have traditionally varied in narrow ranges of a few percentage points. For instance, since 1965 the CDU/CSU has varied between 44,3% and 48,8%, the SPD between 37,0% and 45,8% and the swing party FDP between 5,8% and 10,6%. If, as of 1983, one adds the Greens to the SPD (one may consider the former to a large extent as a “spin off” from the latter), then the range of SPD narrows to 6,5% (from 39,3% to 45,8%). By comparison the customary confidence intervals amount to $\pm 2\%$ to $\pm 4\%$. (Recall that their job is to keep you posted on the extent to which the information you get from your sample is distorted by lottery effects.) From this it is obvious that confidence intervals have the same order of magnitude as the historical variations of party strengths and sometimes they are even bigger. **This means that “representative” samples provide no more information than the previous elections and often even less.**

For this reason the “only” way out for the forecasters is to sweep the lottery effect under the rug. For if they were to divulge their forecasts in a form which could be backed up statistically, i.e. by clearly indicating the errors caused by random sampling, for example

CDU/CSU 42% - 50%, SPD 35% - 43%, FDP 5% - 10%, Greens 4% - 9%,

then their numbers would be eligible for the funny pages. Who would fork out thirty thousand dollars for such a banality? Even worse would be the continual loss of revenue because you can't cash in every week for the same bad joke.

So the pollsters act as if they had discovered the magic formula in statistics which makes random errors evaporate: By means of a weighting procedure, the alchemical origins of which are unmistakable, they first let out the surplus air from the confidence intervals, whereupon "exact figures" down to decimal points hatch out of the retort. Their gamble is based on modifying current voter preferences according to how people voted in the previous election. The statistical gist of their "reallocation" program is the following: If the voters plan to cast their ballots in the same way as they did in the election four years previously and if they confide their intention to the interviewer - our forecasters thus hope that the voters "make up" for their memory lapses, including those who were too young to vote then or have died since - , yes, then everything is perfect in the best of all possible worlds ... and even the statistics of their reallocation system work impeccably. Never mind the fact that under these assumptions you neither need polls nor an election to get results. Just copy the ones from the previous elections ... and there you are.

Tragically the pollsters magic formula frequently produces surrealistic results which they don't dare to sell because they just don't look right. To cover up the further doctoring of half-baked results our pollsters have developed a highfalutin nomenclature which exceeds any standard one might ever have. The minor blemish is that they arrive at the final results in a completely different way: **To produce a forecast they merely adjust the results of the last election a little bit upwards or downwards according to the alleged direction of the political trend.** They largely ignore current 1poll results and use them, if at all, to determine the alleged direction of the trend. Considering the uncontrollable nature of lottery and interview errors, this adventure requires that blend of unlimited self-confidence and unshakeable ignorance with which usually only politicians are endowed. **As for the famous "Sunday-question":**

"For whom would you vote if elections were held next Sunday?",

the upshot is that pollsters' results merely reflect their speculation, or, as they put it unctuously - their art of weighting.

From this it is obvious that the trends they report have no bearing on reality whatsoever. For the real trends are so distorted by pseudo-trends resulting from the lottery effect that they become unrecognisable. To put it sarcastically: The favours of chance continually provide our pollsters with news and headlines. In reality nobody knows what is going on. The numbers the pollsters keep dishing out - for instance CDU/CSU -2,8%, SPD +1,9%, FDP +1,3%, Greens -0,4% - contain no more information than a horoscope. They fake changes which are almost surely quantitatively false and are often qualitatively incorrect as well.

These problems are magnified by a wrong statistical model used in demoscropy: the binomial distribution and its substitute, the normal distribution. Both are popular among social scientists throughout the world. They reward the happy believer with the smallest possible confidence intervals - thus the errors due to random sampling become minimal. What more does he want? He gets no rewards for facing the fact that his model is not a legitimate statistical description of the poll he conducts. Instead it is merely a product of wishful thinking which is being handed down from one generation of textbooks to the next like a curse "unto the third and fourth generation".

For this model assumes that an opinion poll consists of a single question which is to be answered by either “yes” or “no”. Every housewife knows that it is not enough to say “no” once, when the interviewer knocks at her door. It is very clear to her that the uninvited guest is going to install himself comfortably in her living room for the afternoon, bombarding her with dozens of multiple choice questions so that she often is at a loss for an answer. Statistically this means that the correct model for an opinion poll does not consist of a single binomial distribution but of dozens of multinomial distributions with different degrees of freedom which have to be taken into account simultaneously. If at all, this can be done only by computer simulations. Computing confidence intervals for opinion polls by means of the normal distribution is reminiscent of that butcher who throws one sausage into a pack of dogs and then charges each owner for the “feast”. No owner can prove that it was not his dog that devoured the sausage. Still the butcher is not a model of integrity. Such “confidence” intervals merely serve as a statistical alibi. They have no bearing on reality whatsoever.

Computer simulations of random sampling show that confidence intervals based on the normal distribution have to be increased by 100% or more in order to maintain the chosen probability level. The longer the questionnaire, the greater the increase. The real confidence intervals depend on the number of questions asked and their multiple choice complexity, and not only on the number of people polled and the chosen probability level as has been assumed. In short: under the heat of the simulation-sun vast quantities of allegedly hard data melt down like butter.

If the famous psychologist, Carl Gustav Jung - the discoverer and explorer of the collective unconscious - were still alive, he would feel compelled to extend his long list of collective contents by another archetype:

the happy believer in numbers.

The alchemists of past times would envy their colleagues of today. What they tried to do in vain for centuries for lack of suitable resources and know-how has become routine for our pollsters:

to transform percentage points and superstition into a goldmine.