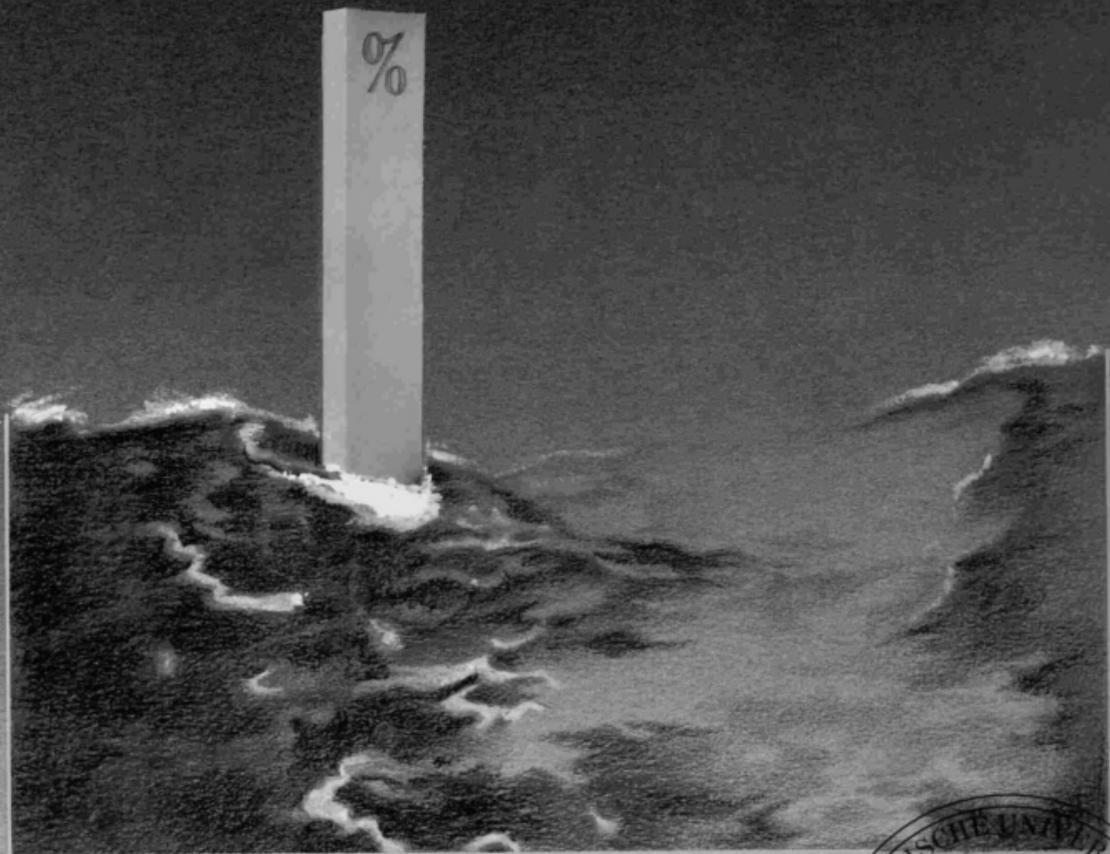


bild der wissenschaft

Sonderdruck



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

Fritz Ulmer



Wahlprognosen und Meinungsumfragen

Schon kurze Zeit nach Schließung der Wahllokale am 25. Januar wird die erste Projektion auf die Endergebnisse vorliegen und mit recht guter Genauigkeit erkennen lassen, zu wessen Gunsten die Bundestagswahl ausgegangen ist - ein Triumph der Hochrechnung und der Datentechnik.

Wie aber steht es mit den Prognosen, die von verschiedenen Instituten in den Monaten und Wochen vor der Wahl gestellt werden? Geben sie schon verlässliche Hinweise auf den Ausgang der Wahl? 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 Demoskopien, falls sie gelingen. 1983 etwa lag die Allensbacher Vorhersage wie bei den vorhergehenden Bundestagswahlen recht nahe am tatsächlichen Wahlergebnis:

CDU	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 am 10. März 1985 lag die Prognose völlig daneben, wie die Leiterin des Instituts, die Demoskopie-Pionierin Prof. Dr. Elisabeth Noelle-Neumann, in einem Rechtsstreit mit dem Spiegel offenlegen mußte:

CDU	47,0%	tatsächlich 37,3%
SPD	44,2%	tatsächlich 49,2%
FDP	1,9%	tatsächlich 10,0%
Grüne	6,5%	tatsächlich 2,5%

Die im Auftrag der saarländischen CDU erstellte Prognose war auch nicht für die Öffentlichkeit bestimmt und basierte auf nur 474 Interviews. Das einträgliche Alltagsgeschäft der Prognose-Industrie, kommerziell motivierte Meinungsumfragen, basiert in der Regel auf einigen hundert bis tausend 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, beispielsweise um einer kleinen Partei doch noch über die 5%-Hürde zu helfen. Doch so sehr Gründe wie diese von den Demoskopien 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.

Um dies ausführlicher zu erklären, ist es zweckmäßig, 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 weitverbreitete Unkenntnis der Bedeutung von Erst- und Zweitstimmen und der Auswirkung des Stimmen-Splittings führt zu Verzerrungen, die wegen der geringen Zahlen statistisch nicht mehr erfaßbar sind.

Die zweite Ebene ist rein mathematisch-statistischer Natur. Denn 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 Fehler-Spielraum.

Genau hier liegt der Kern des Problems, der alle gängigen Wahlprognosen und erst recht kommerziell motivierte Meinungsumfragen als pseudowissenschaftliche Vorspiegelung exakter Zahlen entlarvt.

Zunächst: 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.

Was ist ein repräsentativer Querschnitt?

Wie wird ein repräsentativer Querschnitt zusammengestellt, was kann man damit anfangen und was nicht? Erwartet man Unterschiede, wenn zwei unabhängige Stellen gleichzeitig einen repräsentativen Querschnitt erheben ?

Die Bezeichnung ruft eine klare und zwingende Vorstellung hervor: Ein "repräsentativer" Querschnitt von 1000 Wahlberechtigten sei eine Art Miniaturbild, welches die Gesamtheit der Wahlberechtigten in den wesentlichen Merkmalen - Alter, Geschlecht, Familienstand, Beruf, Einkommen, Wohngegend und was sonst noch in der amtlichen Statistik aufgeführt ist - widerspiegelt. Darum könnten auch die in diesem Miniaturbild ermittelten prozentualen Anteile auf alle Wahlberechtigten übertragen werden. Ein zum gleichen Zeitpunkt aus 1000 anderen Wahlberechtigten gebildeter repräsentativer Querschnitt müsse im wesentlichen dasselbe Bild ergeben.

Diese Vorstellung ist falsch. Ein repräsentativer Querschnitt wird mit einem Auswahlverfahren erstellt, das im Idealfall jeder Gruppe von 1000 Wahlberechtigten die gleiche Chance gibt, ausgewählt zu werden. Vereinfacht ausgedrückt: Ein repräsentativer Querschnitt wird per Lotterie zusammengestellt. Demographische Merkmale finden überhaupt keine Beachtung.

Das erscheint auf den ersten Blick widersinnig, denn man möchte doch ein Abbild der Gesamtheit finden und nicht einen zusammengewürfelten Haufen. Könnte man da nicht etwas völlig Atypisches erreichen ?

Könnte man nicht ausgerechnet 1000 Grüne auswählen? Oder 1000 CDU/CSU-, 1000 SPD- oder 1000 FDP-Wähler? Um von solch pathologischen repräsentativen Querschnitten nicht aus der Fassung gebracht zu werden, muß man sich vor Augen führen, daß die Anzahl aller möglichen repräsentativen Querschnitte unvorstellbar groß ist, nämlich rund $2820 \dots 0$, wobei die Punkte für 5030 Nullen stehen. Gemessen an dieser gigantischen Zahl gehen die pathologischen Fälle völlig unter.

Ein perfektes Zufallsverfahren zur Auswahl eines repräsentativen Querschnitts läßt sich in der Praxis nicht durchführen. Die Demoskopie-Institute sind aus Kostengründen darauf angewiesen, daß jeder Interviewer seine etwa acht bis zehn Befragungen in einem kleinen Bezirk durchführen kann. Das ADM-Verfahren, ein abgestimmtes Verfahren des Arbeitskreises Deutscher Marktforschungs-Institute, sucht deshalb von vornherein nur einen kleinen Teil der fast 50000 Stimmbezirke aus, beispielsweise 201. Das bedeutet aber, daß alle die Tausender-Gruppen von Wahlberechtigten, die aus mehr als 201 Stimmbezirken zusammengesetzt sind, überhaupt keine Chance bekommen, einen repräsentativen Querschnitt zu bilden. Dabei machen sie fast alle der $2,82 \times 10^{5034}$ theoretisch denkbaren Möglichkeiten aus, während die des ADM-Verfahrens völlig untergehen.

Die Wohnnachbarschaft hat häufig gemeinsame Partei-Präferenzen zur Folge. Es wäre empirisch zu untersuchen - anhand früherer Bundestagswahlen -, welche Auswirkungen sich dadurch ergeben.

Das ADM-Verfahren stellt einen Kompromiß zwischen Kostenaufwand und Informationsverlust dar. Die Reisekosten der Interviewer werden durch die Beschränkung auf einen Stimmbezirk drastisch reduziert, aber ebenso die gewonnenen Informationen. Was nutzen Interviews in einem Stimmbezirk, in welchem eine Partei dominiert? De facto bedeutet dies eine Reduktion des Querschnitt-Umfangs - und die Kostenersparnis erweist sich als Bumerang.

Technisch ausgedrückt: Das ADM-Verfahren beabsichtigt zwar, jedem Wahlberechtigten die gleiche Chance zu geben, beabsichtigt aber in keiner Weise 1000 statistisch unabhängige Wahlberechtigte auszuwählen. Sowohl Chancengleichheit, als auch statistische Unabhängigkeit sind aber unerläßliche Kriterien.

Diese Vorstellung ist falsch. Die Desinformation des Bürgers 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 den Medien, 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 Zufallsauswahl gilt als Qualitätsmerkmal einer Umfrage, je zufälliger, desto besser. Das Verfahren wurde vom Arbeitskreis Deutscher Marktforschungs-Institute (ADM) entwickelt und wird als ADM-Mastersample bezeichnet. Es arbeitet in drei Stufen:

- Ausgangspunkt sind die rund 50000 Stimmbezirke der Bundesrepublik Deutschland, die jeweils einem Wahllokal zugeordnet sind. Kleinere Bezirke werden mit anderen zusammgelegt, so daß jeder Bezirk mindestens 400 Wahlberechtigte enthält. 47976 Bezirke bleiben so übrig. Aus diesen werden beispielsweise 201 Bezirke 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. Demographische Gesichtspunkte aber spielen keine Rolle. Von den geplanten rund 1600 Interviews kommt gut ein Drittel nicht zustande, entweder weil die ausgelosten Wahlberechtigten nicht angetroffen werden oder weil sie die Befragung ablehnen.
- In jedem Haushalt sucht der Interviewer, wiederum nach einem Zufallsverfahren, einen Wahlberechtigten aus. Da der einzelne dadurch 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 soviel wie ein Vertreter eines Zwei-Wahlberechtigten-Haushalts.

Damit 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 Verfahrens und seine praktische Durchführung ist die Realität von diesem Ideal weit entfernt. Beispielsweise werden durch das Verfahren stets Gruppen von sieben bis zehn Wahlberechtigten ausgesucht, die in unmittelbarer Nachbarschaft wohnen und damit oft eine bestimmte Präferenz für die gleiche Partei aufweisen: Zum Beispiel gibt es Stimmbezirke mit sehr hohen CDU/CSU-Anteilen oder sehr 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.

Doch sehen wir von diesen Problemen ab und unterstellen ideale repräsentative Querschnitte. Die zentrale Frage lautet dann: Wie verlässlich sind Aussagen über eine 45-Millionen-Gruppe, die aus Querschnitten mit 1000 Mitgliedern gewonnen werden? Mit welchen Fehlern ist zu rechnen?

Das ist eine rein mathematische Frage. Die aus den soziopsychologischen Problemen resultierenden Fehler der Interviews kommen bei einer Wahlprognose dann additiv hinzu. Sie können weit größer sein - wie aus den eklatanten Fehlprognosen hervorgeht - und sind weder kontrollier- noch meßbar.

Gehen wir, um die Problemebenen klar zu trennen, von einem fiktiven, aber objektiven Sachverhalt aus: Die realen, am Wahltag abgegebenen Stimmzettel - rund 40 Millionen bei 90% Wahlbeteiligung und 45 Millionen Wahlberechtigten - seien in einer riesigen Trommel gut vermischt worden. 1000 davon werden nun nach einem Zufallsverfahren gezogen. Wie gut stimmt ein solcher Querschnitt mit dem Gesamtergebnis überein?

Um diese Frage mathematisch zu lösen, muß man zunächst einmal sämtliche Möglichkeiten überdenken, wie man aus 40 Millionen Stimmzetteln 1000 auswählen kann. Die entscheidende Frage lautet dann: Wie häufig wird ein Griff in die Original-Wahlstimmen-Trommel, ein repräsentativer Querschnitt also, mit dem wirklichen Ergebnis übereinstimmen, wie stark werden die Abweichungen sein?

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 der repräsentativen Querschnitte kommt der wahren Stimmverteilung wie nahe?

Betrachten wir ein konkretes Beispiel: Angenommen die FDP hat 4,5% der Stimmen erhalten, also 1800000 Stimmzettel. Gibt es unter den repräsentativen Querschnitten welche, die der FDP fälschlicherweise einen Einzug in den Bundestag prophezeien, also einen Anteil von 50 oder mehr FDP-Stimmzetteln unter den gezogenen 1000 Stimmzetteln enthalten und damit einen Anteil von 5% oder mehr? Wieviele sind dies?

Die Antwort (siehe Tabelle Scheitern an der 5%-Hürde): 24%. In Worten: Jeder vierte repräsentative Querschnitt täuscht fälschlicherweise 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 wahren Stimmzettel. Allein aufgrund der Zufallsauswahl 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 beispielsweise einen wahren Anteil von 5,5 %, also deutlich über der 5%-Grenze, dann behaupten etwa 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, wie sie in der Tabelle ebenfalls enthalten sind: Bei 500 "Interviews" sind die Ergebnisse vollends katastrophal. Um bei einem wahren Anteil von 4,9% das Verpassen der 5%-Hürde einigermaßen sicher zu diagnostizieren, müßte man fast 130000 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 handele sich um eine politisch motivierte Prognose, den Grünen auf Kosten von SPD-Zweitstimmen den Einzug in den Landtag zu ermöglichen. Demoskopie hat eben viele Anwendungsmöglichkeiten.

Was kann man mit einer Meinungsumfrage klären?

Wie die Tabellen zu verstehen sind

Voraussetzung für alle Beispiele ist eine ideale Meinungsumfrage. Damit ist gemeint, daß ein - fiktives - amtliches Wahlergebnis mit bestimmter Stimmenverteilung vorliegt, aus dessen Stimmzetteln nach einem perfekten Zufallsverfahren Stichproben von gegebenem Umfang gezogen werden. Die soziopsychologischen Probleme einer Befragung sollen also ausgeschlossen, das "wahre" Ergebnis bekannt sein.

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.

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. 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.

Beispiel 1: Die 5%-Hürde

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 würde. Hier werden also alle Querschnitte gezählt, die 5% oder mehr für die FDP ermitteln. Die einzelnen Spalten unterscheiden sich im Umfang der Stichproben.

Scheitern an der 5%-Hürde					
FDP-Anteil gemäß amtl. Wahlergebnis	Anzahl der Interviews				
	500	1000	2000	10000	127200
3,5	5,0	0,9	0,034	nahezu 0	nahezu 0
3,9	12,6	4,7	0,8	nahezu 0	nahezu 0
4,0	15,2	6,6	1,5	nahezu 0	nahezu 0
4,1	18,1	9,0	2,7	nahezu 0	nahezu 0
4,2	21,4	12,0	4,5	0,0057	nahezu 0
4,3	24,8	15,5	7,1	0,04	nahezu 0
4,4	28,5	19,6	10,7	0,22	nahezu 0
4,5	32,3	24,2	15,3	0,93	nahezu 0
4,6	36,4	29,3	21,0	3,1	nahezu 0
4,7	40,4	34,7	27,7	8,3	nahezu 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

Ein Beispiel: Wenn die FDP tatsächlich 4,7% (1. Spalte) der Stimmen erhalten hat, liefern 40,4% (2. Spalte) aller Querschnitte mit 500 Befragungen die falsche Antwort, daß nämlich die FDP die 5%-Hürde geschafft hätte. Selbst bei Querschnitten mit 10000 Befragungen stellen noch 8,3% die falsche Prognose.

Überwinden der 5%-Hürde					
FDP-Anteil gemäß amtl. Wahlergebnis	Anzahl der Interviews				
	500	1000	2000	10000	129000
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	nahezu 0
5,4	31,8	26,9	20,1	3,5	nahezu 0
5,5	28,5	22,5	15,1	1,25	nahezu 0
5,6	25,3	18,7	11,1	0,38	nahezu 0
5,7	22,4	15,3	7,9	0,10	nahezu 0
5,8	19,7	12,4	5,5	0,02	nahezu 0
5,9	17,2	9,9	3,7	nahezu 0	nahezu 0
6,0	14,9	7,8	2,4	nahezu 0	nahezu 0
6,1	12,9	6,1	1,6	nahezu 0	nahezu 0
6,2	11,1	4,7	0,98	nahezu 0	nahezu 0
6,3	9,5	3,6	0,60	nahezu 0	nahezu 0
6,4	8,1	2,7	0,36	nahezu 0	nahezu 0
6,5	6,9	2,0	0,21	nahezu 0	nahezu 0
6,9	3,4	0,56	0,02	nahezu 0	nahezu 0
7,0	2,8	0,40	0,01	nahezu 0	nahezu 0

Die zweite Tabelle ("Überwinden der 5%-Hürde") gibt entsprechend die Anteile an falschen Prognosen wieder, wenn das amtliche Ergebnis über 5% liegt. Fordert man, daß wenigstens 95% aller Querschnitte die richtige Antwort liefern, so ist aus den Tabellen abzulesen, daß bei amtlichen Ergebnissen zwischen 4,0% und 6,2% diese Forderung mit 1000 Befragungen nicht zu erreichen ist.

Beispiel 2: Die Mehrheit

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

Die beiden Tabellen zeigen wieder den prozentualen Anteil an repräsentativen Querschnitten, die ein falsches Ergebnis vortäuschen; die in der oberen Tabelle also die Koalition an der Regierung sehen, obwohl der unterstellte Stimmenanteil unter 50% liegt; ihr in der unteren Tabelle dagegen trotz Stimmenanteil über 50% keine Mehrheit zubilligen. Die Ergebnisse sind wiederum für verschiedene Umfänge der Querschnitte angegeben.

Verpassen der absoluten Mehrheit					
CDU/CSU/FDP-Anteil gemäß amt. Wahlergebnis	Anzahl der Interviews				
	500	1000	2000	10000	667000
45,0	1,4	0,08	nahezu 0	nahezu 0	nahezu 0
46,0	4,0	0,62	0,02	nahezu 0	nahezu 0
47,0	9,7	3,1	0,39	nahezu 0	nahezu 0
47,35	12,7	5,0	0,94	nahezu 0	nahezu 0
47,5	14,1	6,0	1,34	nahezu 0	nahezu 0
48,0	19,8	10,9	3,9	nahezu 0	nahezu 0
48,5	26,5	17,9	9,3	0,14	nahezu 0
49,0	34,4	27,4	19,2	2,3	nahezu 0
49,5	42,9	38,8	33,5	16,1	nahezu 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

Gewinnen der absoluten Mehrheit					
CDU/CSU/FDP-Anteil gemäß amt. Wahlergebnis	Anzahl der Interviews				
	500	1000	2000	10000	665000
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	nahezu 0
51,0	31,1	25,3	18,0	2,2	nahezu 0
51,5	23,7	16,3	8,6	0,13	nahezu 0
52,0	17,4	9,7	3,5	nahezu 0	nahezu 0
52,5	12,2	5,3	1,2	nahezu 0	nahezu 0
52,55	11,8	5,0	1,1	nahezu 0	nahezu 0
53,0	8,3	2,7	0,34	nahezu 0	nahezu 0
54,0	3,3	0,50	0,016	nahezu 0	nahezu 0
55,0	1,1	0,068	nahezu 0	nahezu 0	nahezu 0

Zum Beispiel: Bei 47,5% wahren Stimmenanteil und 1000 Interviews täuschen bereits 6% aller repräsentativen Querschnitte eine Mehrheit für die Koalition vor. Ähnlich bei 52,5%: Dort sind es 5,3% der Querschnitte, die der Koalition fälschlicherweise eine Wahlniederlage zuschreiben. Das bedeutet: Bei einem wahren Stimmenanteil zwischen 47,5% und 52,5% sind überhaupt keine verlässlichen Aussagen möglich, die wenigstens 19 von 20 Querschnitte mit 1000

Interviews richtig wiedergeben würden.

Wie die Tabellen berechnet wurden

Die Werte in den Tabellen wurden numerisch mit der Binomialverteilung berechnet. Zum Beispiel bei der Frage nach dem Scheitern an der 5%-Hürde und einem FDP-Anteil von p (unterhalb 5%) durch folgende Formel (n ist die Zahl der Interviews):

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

Die Tabellenwerte wurden auf einer Cyber 175 mit einer Computer-Simulation berechnet, zum Teil aber auch - in guter Übereinstimmung - mit der Multinomialverteilung. Die Programme schrieben Ullrich Tesche und Edith Achiman.

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 verlässlich beantworten können, wenn das Wahlergebnis für die FDP entweder unter 4,0% oder über 6,2% liegt. Nur dann liefern 95% der Prognosen, also 19 von 20, die richtige Antwort. Bei Ergebnissen zwischen 4,0% und 6,2% haben die Prognosen keine Aussagekraft für oder gegen einen Einzug in den Bundestag. Gerade da aber wären sie von Interesse.

Ist das vielleicht ein Spezifikum der 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 eine Mehrheit zu erringen - ungeachtet der 5%-Problematik?

Unterstellen wir, daß 51% der abgegebenen Stimmen auf diese Parteien entfallen, also 20400000 Stimmzettel für sie in der Trommel sind. Wie viele der denkbaren repräsentativen Querschnitte vom Umgang 1000 werden ein falsches Ergebnis, nämlich unterhalb 50% liefern, also maximal 499 Stimmzettel für die Koalition enthalten?

Die Tabelle Gewinnen der absoluten Mehrheit zeigt: Bei 1000 "Befragten" geben 23,5% der repräsentativen Querschnitte eine falsche Antwort. Jede vierte Wahlprognose muß bei diesem Wahlausgang allein aus statistischen Gründen 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 Stimmanteil kommen noch immer 10,9% der Querschnitte zum falschen Ergebnis, die Koalition hätte die absolute Mehrheit gewonnen.

Wollte man 49,9% von 50,0% unterscheiden können, müßte man den Umfang des Querschnitts auf über 650000 erhöhen, und auch dann wäre noch jedes zwanzigste Ergebnis falsch.

Bei 1000 Befragungen muß der wahre Stimmenanteil für eine 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.

Bei diesem Spielraum kann man sich ebensogut auf sein Gefühl oder den Kaffeesatz verlassen und die bis zu fünfstelligen Beträge für eine Wahlprognose sparen.

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

Begnügen sich die Wahlprognosen mit den Stimmen für die einzelnen Parteien, so versuchen alle Institute auch Trends über 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 ZDF erfaßt. Darum stütze ich mich bei den folgenden Analysen auf sie. Es hat nämlich wenig Sinn, mit "korrigierten" Rohdaten Statistik zu betreiben.

Im "Politbarometer" werden die Parteienstärken 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.

In den Monaten Mai bis August ermittelte das Politbarometer 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

Im Kommentar dazu hieß es: "Zum erstenmal seit Januar dieses Jahres lag die SPD mit 43%, wenn auch knapp, hinter der CDU 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.

Lassen wir wieder alle soziopsychologischen Probleme der Interviews außer acht und unterstellen, daß die obigen Zahlen die objektive Wahrheit seien, wie es das ZDF ja auch behauptet. Füllen wir für jeden der vier Monate eine Trommel mit entsprechend ausgefüllten Zetteln. Für den Mai beispielsweise sind das 15600000 CDU-, 17600000 SPD-, 2000000 FDP- und 4800000 Stimmen für die Grünen, zusammen 40000000.

Aus jeder der vier Trommeln ziehen wir nun nach einem Zufallsverfahren jeweils 1000 Stimmzettel als repräsentative Querschnitte. Wie viele dieser Querschnitte werden den behaupteten Trend reproduzieren, also die wahren Ergebnisse innerhalb von $\pm 0,5\%$ Genauigkeit liefern?

Das Ziehen der repräsentativen Querschnitte aus den Trommeln haben wir mit dem Computer simuliert. Die Ergebnisse sind im Kasten "Der zeitliche Trend der Parteienstärken" zusammengestellt.

Der zeitliche Trend der Parteienstärken

Im "Politbarometer" des ZDF werden monatlich die Parteienstärken durch einen repräsentativen Querschnitt ermittelt, der jedesmal neu zusammengestellt wird. Es sind also nicht dieselben Befragten. In den Monaten Mai bis August 1986 ergab sich folgendes Bild (alle Angaben in %):

	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 welchem Maße ist diese Befragung reproduzierbar? Was wäre herausgekommen, wenn man in denselben Monaten jeweils 1000 andere Wahlberechtigte, jeweils andere repräsentative Querschnitte befragt hätte? Diese Frage klingt zunächst hypothetisch. Die rein statistischen Effekte jedoch kann man simulieren.

Man nimmt an, die obigen Werte entsprechen exakt der Wirklichkeit, das heißt, die entsprechenden Anteile würden für die Gesamtheit aller Wahlberechtigten zutreffen, was im Politbarometer auch behauptet wird. Man nimmt also an, daß man von jedem Wahlberechtigten weiß, wie er bei einer hypothetischen Befragung durch das Politbarometer geantwortet hätte. Die Frage ist ja nur: Wie genau spiegeln sich diese Verhältnisse in beliebig herausgegriffenen repräsentativen Querschnitten wider?

Diese Querschnitte kann man vom Computer nach einem wirklichen Zufallsverfahren ziehen lassen. Hier sind die ersten drei Wiederholungen obiger Befragung, wie sie sich bei einem willkürlichen Start des Zufallsgenerators (Ranset von Cyber mit Startwert 0.12345) ergaben :

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

Die Abweichungen von den wahren Werten variieren zwischen -3,6% und +3,7%.

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

Die Abweichungen von den wahren Werten variieren zwischen -2,2% 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

Die Abweichungen von den wahren Werten variieren zwischen -2,1% und +1,9%.

Wie man sieht, liegen die Werte keineswegs innerhalb von $\pm 0,5\%$, wie es die Daten und vor allem ihre Analyse im ZDF suggerieren. Denn der Trend, daß beispielsweise die CDU die SPD überholt hätte, läßt sich nur dann erkennen, wenn man den Daten diese Genauigkeit unterstellt. In der ersten Wiederholung zum Beispiel scheint die SPD im August wieder deutlich die Führung übernommen zu haben.

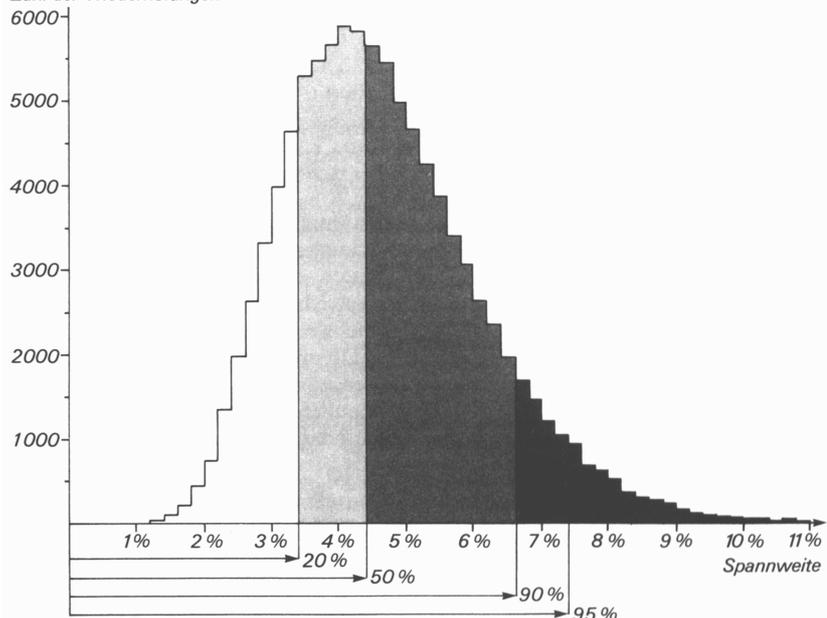
Oder sind die drei gezeigten Beispiele Ausreißer?

Wir haben die Ergebnisse einer Simulation von 100000 repräsentativen Querschnitten mit viermal 1000 Befragten analysiert. Bei jedem Querschnitt ermittelten wir zunächst die Spannweite, indem wir die größte Abweichung nach oben und die größte Abweichung nach unten zahlenmäßig addierten. Die obigen Beispiele haben Spannweiten von 7,3%, 3,7% und 4,0%.

Das Diagramm zeigt, wie die Spannweiten unter den 100000 Zufalls-Wiederholungen des "Politbarometers" verteilt sind. Zum Beispiel: Rund 2000 der 100000 Wiederholungen haben Spannweiten zwischen 2,4% und 2,6%. Die Hälfte aller Wiederholungen hat Spannweiten von mehr als 4,4%. Ein

100000 Wiederholungen des Politbarometers

Zahl der Wiederholungen



Zehntel aller Wiederholungen liegt über 6,6% und eine von 20 Wiederholungen hat sogar eine Spannweite von 7,5% und mehr.

Dieses Ergebnis macht deutlich, wie weit die statistische Qualität der "Politbarometer-Befragung" davon entfernt ist, den behaupteten Trend aufzeigen zu können.

Die durch Simulation und teilweise (Kommazahlen) auch durch numerische Berechnung mit der Multinomialverteilung gewonnen Zahlen in der Tabelle zeigen noch detaillierter, wie die Abweichungen bezogen auf die einzelnen Parteien einzuschätzen sind. Es könnte ja sein, daß die Spannweite ein zu schlechtes Resultat zeigt, weil sie immer die größte negative beziehungsweise positive Abweichung der Wiederholungen herauspicks. Zwei verschiedene Betrachtungsweisen sind dabei verwendet worden :

- Bei jeder Partei wurde derselbe Spielraum von beispielsweise $\pm 2,0\%$ für die sechzehn Werte der Parteienstärken zwischen Mai und August zugelassen. Aus den grauen Zeilen der Tabelle ist beispielsweise abzulesen, daß bei 1000 Interviews (2. Spalte) nicht einmal jede vierte Wiederholung (24,7%) innerhalb dieses Bereichs liegt. Dies bedeutet, daß in drei von vier Wiederholungen des Politbarometers bei mindestens einer Partei Abweichungen von mehr als 2% nach oben oder unten auftreten
- Es wurde berücksichtigt, daß für die kleinen Parteien ein Spielraum etwa von $\pm 3\%$ viel gravierender ist als für die großen Parteien. Entsprechend wurden die Spielräume gestaffelt, beispielsweise $\pm 3,0\%$ für CDU/CSU, $\pm 1,8\%$ für die Grünen und $\pm 1,2\%$ für die FDP. Die Zahlen ergeben sich aus dem Verhältnis der Standardabweichung.

Die Tabelle zeigt für diesen Fall (weiße Zeilen), daß nur 48,6% aller repräsentativen Querschnitte mit 1000 Interviews innerhalb dieses gestaffelten Spielraums liegen. Ausgangspunkt ist wiederum die Annahme, obiges "Politbarometer" trafe auf die gesamte Wahlbevölkerung zu.

Die Tabelle macht außerdem deutlich, daß Querschnitte mit größerem Umfang eine wesentlich bessere Qualität haben. Wegen der ungleich größeren Interview-Fehler hätten Umfragen mit 40000 oder gar 70000 Interviews allerdings wenig Sinn.

Erfolgs -Statistik von Wiederholungen des Politbarometers

Bandbreite in % bei				Anzahl der Interviews				
CDU/CSU	SPD	FDP	Grüne	1000	2000	10000	40000	70000
$\pm 0,5$	$\pm 0,5$	$\pm 0,5$	$\pm 0,5$	0,000912	0,0198	8	75	95
$\pm 0,75$	$\pm 0,75$	$\pm 0,3$	$\pm 0,45$	0,000651	0,0284	16	96	nahezu 100
$\pm 0,75$	$\pm 0,75$	$\pm 0,75$	$\pm 0,75$	0,022	0,635	42	99	nahezu 100
$\pm 1,0$	$\pm 1,0$	$\pm 0,4$	$\pm 0,6$	0,020	0,409	55	nahezu 100	nahezu 100
$\pm 1,0$	$\pm 1,0$	$\pm 1,0$	$\pm 1,0$	0,452	4,49	76	nahezu 100	nahezu 100
$\pm 1,5$	$\pm 1,5$	$\pm 0,6$	$\pm 0,9$	0,778	9,01	96	nahezu 100	nahezu 100
$\pm 1,5$	$\pm 1,5$	$\pm 1,5$	$\pm 1,5$	6,65	29,7	98	nahezu 100	nahezu 100
$\pm 2,0$	$\pm 2,0$	$\pm 0,8$	$\pm 1,2$	6,67	38,5	nahezu 100	nahezu 100	nahezu 100
$\pm 2,0$	$\pm 2,0$	$\pm 2,0$	$\pm 2,0$	24,7	63,3	nahezu 100	nahezu 100	nahezu 100
$\pm 2,5$	$\pm 2,5$	$\pm 1,0$	$\pm 1,5$	23,7	71,8	nahezu 100	nahezu 100	nahezu 100
$\pm 2,5$	$\pm 2,5$	$\pm 2,5$	$\pm 2,5$	49,0	85,5	nahezu 100	nahezu 100	nahezu 100
$\pm 3,0$	$\pm 3,0$	$\pm 1,2$	$\pm 1,8$	48,6	90,6	nahezu 100	nahezu 100	nahezu 100
$\pm 3,0$	$\pm 3,0$	$\pm 3,0$	$\pm 3,0$	70,7	96	nahezu 100	nahezu 100	nahezu 100
$\pm 3,5$	$\pm 3,5$	$\pm 1,4$	$\pm 2,1$	71,5	97,5	nahezu 100	nahezu 100	nahezu 100
$\pm 3,5$	$\pm 3,5$	$\pm 3,5$	$\pm 3,5$	85,4	99	nahezu 100	nahezu 100	nahezu 100
$\pm 4,0$	$\pm 4,0$	$\pm 1,6$	$\pm 2,4$	86,5	fast 100	nahezu 100	nahezu 100	nahezu 100
$\pm 4,0$	$\pm 4,0$	$\pm 4,0$	$\pm 4,0$	94	nahezu 100	nahezu 100	nahezu 100	nahezu 100
$\pm 5,0$	$\pm 5,0$	$\pm 2,0$	$\pm 3,0$	97,8	nahezu 100	nahezu 100	nahezu 100	nahezu 100
$\pm 5,0$	$\pm 5,0$	$\pm 5,0$	$\pm 5,0$	fast 100	nahezu 100	nahezu 100	nahezu 100	nahezu 100

Ergebnis: Lediglich einer von 100000 repräsentativen Querschnitten reproduziert alle sechzehn Einzelergebnisse mit $\pm 0,5\%$ Genauigkeit. Ein parallel durchgeführtes Politbarometer mit einem anderen Querschnitt wäre also zu einem ganz anderen Resultat gekommen.

Bei einem Spielraum von $\pm 2\%$ liegen noch immer 75,3% der repräsentativen Querschnitte außerhalb, das heißt bei drei von vier Politbarometern weichen einzelne der sechzehn angegebenen Parteistärken um mehr als $\pm 2\%$ von den wahren Werten ab.

Selbst bei einem Spielraum von $\pm 3\%$ für CDU/CSU und SPD beziehungsweise $\pm 1,8\%$ für die Grünen und $\pm 1,2\%$ für die FDP - entsprechend ihrer geringeren absoluten Anteile - gibt jedes zweite Politbarometer (51,4%) eine falsche Antwort, liegen also zumindest einzelne der sechzehn angegebenen Werte außerhalb des Spielraums.

Man müßte den Spielraum auf $\pm 4\%$ erhöhen, wenn man die übliche Sicherheit haben wollte, daß nur einer von 20 repräsentativen Querschnitten außerhalb des Spielraums liegt. Bei einem solchen Spielraum interessiert sich aber kein Mensch mehr für das Resultat, denn die Wählerbewegungen von Monat zu Monat gehen darin völlig unter.

Um den interessanten Spielraum von $\pm 0,5\%$ einhalten zu können, müßte man die Zahl der Befragten aber auf 70000 erhöhen.

Diese Zahlen machen deutlich, wie absurd es ist, eine zeitliche Entwicklung, die sich im Prozentbereich abspielt, mit einer Befragung von 4×1000 "repräsentativ" Ausgewählten analysieren zu wollen.

Offenbar setzen sich die Demoskopien über solch elementare, mathematische Grundlagen ihrer Arbeit schlicht hinweg. Sie suggerieren, durch angeblich repräsentative Befragungen aussagekräftige Meinungsbilder produzieren zu können, wo doch allein der statistische Fehler ihnen keine Chance dazu läßt.

Erst bei 10000 oder besser noch 100000 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 Interview-Fehler 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 nur das Eingeständnis: Die Demoskopie ist nicht in der Lage, Trendaussagen über vier Monate zu machen, wenn die Bewegungen im Prozentbereich liegen. 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.

Aufgrund von Fehlprognosen haben die Meinungsforschungs-Institute in den letzten Jahren zu einem sehr riskanten Mittel gegriffen: der "Kunst" der Gewichtung. Es werden nicht die tatsächlich in der Umfrage ermittelten Werte veröffentlicht, sondern diese werden rein rechnerisch erhöht oder erniedrigt und nur dieses modifizierte Resultat publiziert. Beispielsweise werden die erfragten Werte für die SPD und die Grünen erniedrigt, weil "erfahrungsgemäß" bis zum Wahlsonntag Wähler der SPD und der Grünen "ihre Meinung ändern". Oder es werden die ermittelten Werte für die FDP erhöht, weil sie "erfahrungsgemäß" am Wahltag Zweitstimmen von CDU-Wählern erhält.

Falls es sich um empirisch gesicherte Erkenntnisse und nicht um Stammtisch-Argumente handelt, sind solche Korrekturen zulässig. Sie können sogar nützlich sein. Aber sie müssen offen dargelegt und begründet werden, zusammen mit den Rohdaten. Das geschieht jedoch nicht, und die Abänderungen werden verschwiegen.

Die Meinungsforschungs-Institute schaffen sich hier einen Freiraum, von dessen Existenz die Öffentlichkeit keine Ahnung hat und der auch für Manipulationen mißbraucht werden kann.

Das Verschweigen der Gewichtungen kann handfeste politische Auswirkungen haben. Ist es zum Beispiel verantwortbar, 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 die 5%-Hürde. Wer garantiert aber, daß die Wähler am Wahltag für die "aus Erfahrung" hinzugerechneten 2% tatsächlich aufkommen? Etwa die Gewichtungs-Zauberlehrlinge?

Damit sollen die Anstrengungen der Demoskopien im sozialwissenschaftlichen Bereich keineswegs abgewertet werden, wo es beispielsweise um Fragetechniken geht, die dem Befragten nicht schon Antworten suggerieren.

Aber ein seriöses Institut sollte auch die statistische Verlässlichkeit seiner Umfragen nennen und auf die Interview-Fehler und das Gewichtungs-Dilemma hinweisen. Es sollte die Rohdaten offenlegen und nicht nur Zahlen, die es daraus durch eine geheimgehaltene Kosmetik gewinnt.

Die für die Öffentlichkeit bestimmten Wahlprognosen werden in aller Regel mit besonderem Aufwand betrieben, weil sie durch die Wahl selbst überprüft werden und ein Aushängeschild für die Qualität eines Instituts darstellen. Das einträgliche Alltagsgeschäft kommerzieller Umfragen aber könnte allenfalls durch Parallel-Aufträge überprüft werden, was aus Kostengründen kaum ein Auftraggeber tun wird.

Wehrlos allerdings ist ein Auftraggeber nicht. Er kann als Teil seines Auftrages verlangen, daß Wiederholungen der Befragung im Sinne dieses Beitrags simuliert werden. Bei den handelsüblichen Meinungsumfragen mit einigen hundert bis tausend Interviews werden dann viele der oft zu Dutzenden abgelieferten, angeblich so harten Zahlen wie Butter unter der Simulations-Sonne schmelzen.

Es ist einfach so, daß die Meinungsforschungs-Institute als Meßinstrument nur über eine Elle verfügen, nämlich über Tausender-Querschnitte, die nicht einmal nach einem sauberen Zufallsverfahren ausgewählt werden. Sie suggerieren aber, damit Millimeter-Bruchteile messen zu können, als hätten sie eine Schieblehre.

Daß mit großen Zahlen und demoskopischen Erkenntnissen verlässliche Aussagen möglich sind, werden die Hochrechnungen am Wahlabend beweisen. Sie basieren schon nach relativ kurzer Zeit auf Zehn- und Hunderttausenden von Stimmzetteln aus Wahlbezirken, deren demoskopische Eigenheiten bei der Auswertung berücksichtigt werden.