

Genauigkeit von Meinungsumfragen bleibt eine Utopie

Eine Antwort auf Professor Ronges Replik in BB 7/87 vom 11. April 1987

Mit „pseudoexakten Zahlen“ führten die Demoskopien, speziell die Wahlforscher, seit Jahren die Öffentlichkeit an der Nase herum. So die Kernaussage des Wuppertaler Demoskopie-Kritikers Professor Dr. Friedrich Ulmer in „Bild der Wissenschaft“, die er vor der Bundestagswahl den Demoskopien ins Stammbuch geschrieben und im Frühjahr in einem Interview mit den „Bergischen Blättern“ erläutert hat. „Denunziation der Meinungsforschung“ warf daraufhin der Wuppertaler Soziologe und ehemalige Infratest-Geschäftsführer Professor Dr. Volker Ronge dem Mathematiker vor: Statt auf die wirklich relevanten Probleme der Wahlforschung einzugehen, so Ronge, verbreite Ulmer „Banalitäten in schlechter Mixtur“. Harter Tobak auf beiden Seiten - im heutigen Beitrag geht Professor Ulmer auf Ronges Reaktion ein und faßt seine Kritik an den Wahlforschern nochmals zusammen.

Demoskopie beruht bekanntlich auf „Repräsentativumfragen“. Aus Kostengründen werden statt der ganzen Bevölkerung nur einige Auserwählte befragt, ein sogenannt „repräsentativer Querschnitt“, der in der Regel aus 500 bis 2000 Personen besteht. Wie eine „Repräsentativumfrage“ durchgeführt wird, wissen nur wenige. Aber jedermann weiß, was die Meinungsforscher und ihre Auftraggeber damit machen: Sie schließen vom „repräsentativen Querschnitt“ auf die ganze Bevölkerung, als handle es sich um ein Miniaturbild derselben.

Aber: Ein „repräsentativer Querschnitt“ ist durchaus nicht das, was man gemeinhin unter diesem Begriff versteht. Kein Miniaturbild der Bevölkerung, von einem Spiegelbild ganz zu schweigen. Tatsächlich ist das einzig Repräsentative einer „Repräsentativumfrage“ die Absicht, daß jeder Wahlberechtigte die gleiche Chance hat, befragt zu werden. Nur unter dieser Voraussetzung ist ein Schluß von den wenigen Befragten auf die ganze Bevölkerung mit Hilfe von Statistik möglich. Um es etwas salopp zu sagen: Jeder erhält von der Landeslotterie genau ein Los und dann werden 500 bis 2000 Gewinner ausgelost und als „repräsentativer Querschnitt“ von Interviewern befragt. Weil die Vorstellung einer Lotterie bei „Repräsentativumfragen“ wie die Faust aufs Auge paßt, läuft Herr Kollege Ronge Sturm dagegen und behauptet „Die Charakterisierung der Stichprobenziehung als Lotterie dient freilich nur der – wohl beabsichtigten – Verunklarung und Entwertung eines durchaus seriösen Vorgangs. Abweichungen solcher Querschnitte sind zwar

erwartbar, aber eben in berechenbaren Grenzen und nicht ‚irgendwie‘...“

Meine Beschreibung ist allgemeinverständlich und mathematisch-statistisch präzise. Wenn eine „Verunklarung“ vorliegt, dann liegt sie in Ronges Vorstellungen von Statistik. Seine Behauptung, die lotterieberingten Abweichungen würden in berechenbaren Grenzen liegen, ist offensichtlich widersinnig, denn bei einer Lotterie ist im Prinzip alles möglich. Die Statistik macht nun Aussagen darüber, wie groß die Chance ist, daß die durch die Auslosung verursachten Abweichungen innerhalb eines vom Demoskopien frei wählbaren Spielraums liegen (Dieser wird – allerdings unzutreffend – auch als Richtigkeitskorridor oder Fehlerbereich bezeichnet) Soll zum Beispiel die Voraussage der Parteistärken auf ein Prozent genau erfolgen, so wählt der Demoskop ein Spielraum von einem Prozent - das heißt er gibt das Resultat in der Form CDU 46%-47%, SPD 35%-36%, FDP 6%-7%, Grüne 9%-10% an - mit der Konsequenz, daß die Chance der „Repräsentativumfrage“, richtig zu sein, praktisch Null ist. Wählt der Demoskop umgekehrt einen großen Spielraum –beispielsweise 8% für die großen Parteien und 5% für die kleine; was etwa zu einer Prognose-Form CDU 42%-50%, SPD 35%-43%, FDP und Grüne je 4,5%-9,5% führt – dann ist die Chance der „Repräsentativumfrage“, richtig zu sein, etwa 95% (bei 1000 Interviews). Diese Beispiele machen klar:

Eine Prognose ist nur dann sinnvoll, wenn sie sowohl eine akzeptable Chance hat, richtig zu sein, als auch ihr Spielraum genügend klein ist. 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. Aus optischen Gründen werden bei Meinungsumfragen und Wahlprognosen oft genaue Prozentzahlen (manchmal mit Nachkommastelle) angegeben, das heißt es werden die in der „Repräsentativumfrage“ ermittelten Prozentzahlen auf die ganze Bevölkerung übertragen. Diese Prozentzahlen treffen mit Sicherheit nicht zu, sie sind pseudoexakt. Im demoskopischen Alltag wird die Problematik Spielraum-Chance dadurch „gelöst“, daß die Chance der Prognosen überhaupt nicht bzw. inkorrekt angegeben wird –Ronge praktiziert je nach Umfeld beides:

a) In seiner Replik führt er hierzu aus:

„...solche (pseudoexakte) Zahlen stammen in den seltensten Fällen von Meinungsforschern. Diese wissen um den Richtigkeitskorridor ihrer Zahlen und sagen dies in ihren Analyseberichten auch ihren Auftraggebern. Die freilich publizieren nicht

selten, aus unterschiedlichen Motiven pseudoexakte Zahlen. Herr Ulmer müßte sich doch denken können, daß sich die Parteien-Auftraggeber nicht mit pseudo-exakten Zahlen „an der Nase herumführen“ lassen, sondern für ihre ...sechsstelligen Summen seriöse Informationen – daß heißt inklusive ihrer Sicherheitsgrade – abfordern... Jeder seriöse Wahlforscher kapituliert vor der Forderung nach Prognosen, bei denen es auf Nachkommastellen-Genauigkeit ankommt. In den Wahlprognosen insbesondere bedeutsam bei der 5%-Klausel. Ansonsten bieten Korridorprognosen (zum Beispiel 44-47% für die CDU) mit relativ großem Sicherheitsgrad eine durchaus hilfreiche Information.“

Für die großen Parteien wählt Ronge also einen Richtigkeitskorridor (=lotteriebedingter Spielraum) von 3%. Da für die kleinen nur die Stelle nach dem Komma nicht mehr zuverlässig sein soll, läuft dies auf einen Spielraum von einem Prozent hinaus. Dies führt zum Beispiel zu einer Prognose-Form CDU 47%-50%, SPD 37%-40%, FDP und Grüne je 5%-6%. Auf Grund seiner Beteuerungen hätte man erwartet, daß er den Sicherheitsgrad (=Chance) konkret angibt, anstatt mit unverbindlichen Ausdrücken wie „relativ groß“ und „durchaus hilfreicher Information“ um den Kern der Sache herumzureden. Hätte Ronge gerechnet, dann hätte er festgestellt, daß eine Vier-Parteien-Prognose bei 1000 Interviews wegen seines viel zu kleinen Richtigkeitskorridors eine Chance von etwa 85% hat, falsch zu sein! (und fast 70% bei 2000 Interviews). Selbst wenn er den Richtigkeitskorridor – den er ja wählen kann – für die CDU und SPD auf 3,5% und für die FDP bzw. die Grünen auf 2% vergrößert, dann beträgt die Chance, daß eine auf 1 000 Interviews basierende Vier-Parteien-Prognose falsch ist, noch immer fast 60% (bei 2000 Interviews über 30%). Würde ein Auftraggeber - der brauchbare Information sucht (nicht zu verwechseln mit verkaufbarer) – einen sechsstelligen Betrag für ein solches statistisches Nullprodukt ausgeben, wenn ihm der tatsächliche Unsicherheitsgrad der Daten offengelegt wird? Die Schuld an der Irreführung der Öffentlichkeit durch die Produktion und Vermarktung von pseudoexakten Zahlen versucht Ronge auf die Medien und Politiker abzuschieben.

Nur: Die meisten Institute gelangen auch direkt an die Öffentlichkeit - recht publikumswirksam wie zum Beispiel Allensbach in „Die Welt“ und die Forschungsgruppe Wahlen im Politbarometer im ZDF usw. Der große Spielraum, der diesen - auf 1000 Interviews basierenden - Prognosezahlen anlastet, wird immer verschwiegen.

b) Bei der eigenen Arbeit mit Daten ist Ronge in puncto pseudoexakte Zahlen und „repräsentativer Querschnitt“ noch flexibler: In seinem Buch „Von drüben nach hüben - DDR-Bürger im Westen“ (Verlag 84 Hartmann und Petit 1985, S.14 ff.) gibt er bei Stichprobenumfängen bis unter hundert eine Nachkommastelle an. Über den Richtigkeitskorridor, auf den er in seiner Replik so viel Wert legt und der bei diesen kleinen

Aus: Bergische Blätter Nr. 20/1987

Stichprobenumfängen gigantisch wird – was seine harte Daten butterweich macht – sagt er nichts. Über die offensichtlich nicht gegebene Chancengleichheit seines Auswahlverfahrens und einer Ausfallrate von 75% geht er mit der lapidaren Behauptung hinweg: „Hinsichtlich der Fragebogenempfänger kann Repräsentativität insgesamt unterstellt werden. Systematische Abweichungen zwischen den Fragebogenempfängern und den Respondenten sind zwar nicht auszuschließen, sie sind aber nicht ersichtlich.“ Mit solchen Daten und etwas sozialem empirischem Draufgängertum läßt sich vieles „beweisen“, manchmal auch das Gegenteil.

Wie kann das Dilemma zwischen Spielraum und Chance gelöst werden? Theoretisch könnte man dies durch Vergrößerung des „repräsentativen Querschnittes“ erreichen. Zum Beispiel könnte man bei 14000 Interviews den Spielraum für CDU und SPD auf 2% und für FDP und Grüne auf 1.2% reduzieren und dies bei einer Chance von gut 95%. Praktisch sind diesem Vorgehen Grenzen gesetzt, denn bei Prognosen treten neben dem Lotteriereffekt weitere unvermeidbare Fehler auf. Die werden im sogenannten Interviewfehler zusammengefaßt. Sie entstehen u.a. dadurch, daß die im Interview gegebene Antwort nicht mit dem übereinstimmt, was der Befragte dann auf dem Wahlzettel in die Urne legt Außerdem gibt fast ein Drittel der Ausgelosten keine Antwort oder sie können vom Interviewer nicht erreicht werden. Aus Kostengründen werden bei der Erstellung eines „repräsentativen Querschnittes“ in puncto Chancengleichheit erhebliche Abstriche gemacht, was unkontrollierbare Auswirkungen zur Folge hat. Aus diesem Grund hat es keinen Sinn, den Umfang eines „repräsentativen Querschnittes“ über ein gewisses Maß zu erhöhen. Damit würde man nur den lotteriebedingten Spielraum verkleinern, der Interviewfehler aber würde als Balken im Auge zurückbleiben.

Der Interviewfehler ist schwer quantifizierbar, er kann - wie aus den eklatanten Fehlprognosen hervorgeht (zum Beispiel Saarwahl 1985: CDU 37%, Prognose 47%) – weit größer sein als der lotteriebedingte Spielraum und 5% übersteigen. Wie groß der Interviewfehler wirklich ist, spielt für den Demoskopie-Konsumenten keine große Rolle. Denn Wahlprognosen sind allein schon wegen des lotteriebedingten Spielraums nichts aussagend und wertlos. Jeder weiß, daß die CDU/CSU zwischen 42,3% und 49,2%, die SPD zwischen 32,9% und 41,1% die FDP und die Grünen zwischen 5,8% und 10,2% liegen werden. (Dies war die Prognose mit einem angeblichen Sicherheitsgrad von 90% für die Bundestagswahl 1987, die das bekannte Meinungsforschungsinstitut Emnid versiegelt bei einem Notar deponierte; im „Spiegel“ konnte man die nichtversiegelte Version in Form von pseudoexakten Zahlen lesen.) Über diesen Punkt – nämlich die Nutzlosigkeit solcher Prognosen, was meine Kernaussage darstellt – geht Ronge mit Rhetorik hinweg:

Meine Ausführungen bezeichnet er als „Banalitäten schlechter Mixtur, ... die der Mathematikprofessor verbreitet“, denn – so lautet die Quintessenz seiner „Richtigstellung“ - der Interviewfehler sei doch weit wichtiger als der lotteriebedingte Spielraum. Als würde der Interviewfehler die Prognosen zuverlässiger machen... Durch diesen wird die Situation offensichtlich noch schlimmer, als ich sie beschrieben habe- Zur „Untermauerung“ stellt er die – falsche – Behauptung auf, ich sei auf die „tatsächlichen“ Probleme der Demoskopie – Auswahlverfahren (random route, Quote), mangelhafte Ausschöpfung der Stichprobe, Gewichtung – nicht eingegangen. Zu seiner Unterstellung, ich hätte die eigentlichen Hauptaussagen der Demoskopie – nämlich zeitliche Trends - nicht einmal erwähnt: Die Adabsurdumführung der mit „repräsentativen Querschnitten“ ermittelten Trends war der zentrale Punkt meines Artikels „Der Orakelspruch mit dem repräsentativen Querschnitt“ in „Bild der Wissenschaft“.

Im übrigen zeugt seine Behauptung – mit Hilfe des „zeitlichen Trends“ könnten die Probleme, die aus dem lotteriebedingten Spielraum resultieren, erheblich relativiert werden – nicht gerade von statistischem

Sachverstand: Denn bei der Differenzbildung unabhängiger Größen nimmt die Streuung und somit auch der lotteriebedingte Spielraum um über 40% zu. Folglich sind die von den Instituten gelieferten Daten über das zeitliche Auf und Ab der Parteienstärken - zum Beispiel CDU -1.8%, SPD+2.2%, FDP +0.9%, Grüne -1.2% - mit einer noch größeren Unsicherheit belastet als die Angaben über die Parteienstärken. Solche Trendaussagen haben reinen Horoskopcharakter und täuschen Veränderungen vor, die qualitativ mit an Sicherheit grenzender

Wahrscheinlichkeit falsch sind und die häufig nicht einmal qualitativ richtig sind. Die Vermarktung solcher Trenddaten sollte verboten werden, sie erfüllt m.E. den Tatbestand des unlauteren Wettbewerbs. Selbst große Veränderungen bei Meinungsumfragen beweisen wenig: Sind nämlich die Parteienstärken konstant geblieben, dann täuschen zwei aufeinanderfolgende Meinungsfragen von je 1000 Interviews mit einer Wahrscheinlichkeit von fast 50% vor, daß für mindestens eine Partei eine markante Veränderung eingetreten ist; das heißt laut Meinungsumfragen ist eine der großen Parteien um mindestens 3% oder eine der kleinen um mindestens 1,7% gestiegen/gefallen, während sich in Wirklichkeit nichts geändert hat!

Wie kommt es, daß Ronge – als Geschäftsführer a.D. von Infratec, dem Umsatzriesen in der Prognoseindustrie – die Aussagefähigkeit seiner Prognosen dermaßen überschätzt? In den demoskopischen Standardbüchern wird bei der Berechnung der Sicherheit (=Chance) von Prognosen durch die Verwendung der sogenannten Binomialverteilung stillschweigend davon ausgegangen, daß aus einem repräsentativen

Aus: Bergische Blätter Nr. 20/1987

Querschnitt die Stärke von nur einer Partei ermittelt wird, und – man staune – daß der Interviewfehler Null ist. Aber welcher Demoskop erstellt einen Fragebogen mit einer einzigen Frage? Welches Institut macht eine Wahlprognose, in welcher nur eine Parteistärke angegeben wird? Selbst unter diesen abwegigen Annahmen ergibt die Berechnung eine Sicherheit von nicht einmal 70% für den von Ronge gewählten Spielraum von 3% für CDU oder SPD (Basis 1000 Interviews) - bzw. eine Sicherheit von unter 85% bei 2000 Interviews - was dem üblichen Standard von 95% bei weitem nicht entspricht – um die gleiche Sicherheit für die FDP oder die Grünen zu gewährleisten, ist der Spielraum auf etwa 1.7% zu reduzieren. Auf Grund der von Ronge vorgebrachten Argumente - der Dominanz des Interviewfehlers und der Vielzahl der Merkmale, die gleichzeitig in einem repräsentativen Querschnitt ermittelt werden (Stärke aller Parteien usw.), wird klar, daß seine Vorstellungen vom „relativ großen Sicherheitsgrad“ und „durchaus hilfreicher Information“ auf einer binomialen Milchmädchenrechnung beruhen. Wegen der Vielzahl der Merkmale ist die Multinomialverteilung anzuwenden, was die aufgeblasene Sicherheit wie einen Ballon zum Platzen bringt. Bei 1000 Interviews schrumpft sie für eine Vier-Parteienprognose von 70% auf unter 30%, und bei 2.000 Interviews von 85% auf unter 55% (Spielraum für CDU & SPD 3%, für FDP & Grüne 1,7%). Die Konsequenz hiervon ist, daß trotz Ronges großem Spielraum von 100 Vier-Parteien-Prognosen 45 bis 70 falsch sind (je nach Anzahl der Interviews). Dabei ist Ronges Restrisiko - der Interviewfehler - nicht berücksichtigt. Restrisiko Null bedeutet in diesem Fall, daß „der repräsentative Querschnitt“ – nachdem er für teures Geld per Lotterie erstellt worden ist – nicht mehr interviewt wird, sondern den Stimmzettel in den Briefkasten des nächstgelegenen Meinungsforschungsinstitutes wirft, anstatt damit zur Urne zu gehen.

Demoskopie und ihr Instrumentarium - Meinungsumfragen - können in gewissen Situationen brauchbare Angaben liefern, zum Beispiel dann, wenn man für einige unbekannte Größen mit einer relativ groben Abschätzung zufrieden ist, wo es auf ein paar Prozente mehr oder weniger nicht ankommt und der relative Fehler 10 - 30% betragen kann. Aber die Demoskopie vermag bei weitem nicht das zu leisten, was die Prognoseindustrie mit der Produktion und Vermarktung von pseudoexakten Zahlen seit Jahrzehnten suggeriert. Weil bei der gegenwärtigen politischen Situation in der BRD der Wahlausgang von geringen Änderungen der Parteienstärken abhängig ist, sind wissenschaftliche haltbare Wahlprognosen mit Meinungsumfragen nicht möglich. Die hierfür erforderliche Genauigkeit ist und bleibt wegen der Größe des lotteriebedingten Spielraums und des Interviewfehlers eine Utopie.