

Deutsche
Forschungsgemeinschaft

**Qualitätskriterien
der Umfrageforschung**

Quality Criteria
for Survey Research



Akademie Verlag

Deutsche
Forschungsgemeinschaft

**Qualitätskriterien
der Umfrageforschung**

**Quality Criteria
for Survey Research**

Denkschrift
Memorandum

Herausgegeben von / Edited by
Max Kaase



Akademie Verlag

DFG

Deutsche Forschungsgemeinschaft
Geschäftsstelle: Kennedyallee 40, D-53175 Bonn
Postanschrift: D-53170 Bonn
Telefon: ++49/2 28/8 85-1
Telefax: ++49/2 28/8 85-27 77
E-Mail: (X.400): S = postmaster, P = dfg, A = d400, C = de
E-Mail (Internet RFC 822): postmaster@dfg.de

Das vorliegende Werk wurde sorgfältig erarbeitet. Dennoch übernehmen Herausgeber und Verlag für die Richtigkeit von Angaben, Hinweisen und Ratschlägen sowie für eventuelle Druckfehler keine Haftung.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Qualitätskriterien der Umfrageforschung : Denkschrift = Quality criteria
for survey research / Deutsche Forschungsgemeinschaft.

Hrsg. von Max Kaase. – Berlin: Akad. Verlag, 1999
ISBN 3-05-003455-6

© Akademie Verlag GmbH, D-10243 Berlin (Federal Republic of Germany), 1999

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.

Alle Rechte, insbesondere die der Übersetzung in andere Sprachen, vorbehalten. Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Photokopie, Mikroverfilmung oder irgendein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsmaschinen, verwendbare Sprache übertragen oder übersetzt werden.

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers.

Umschlaggestaltung und Typographie: Dieter Hüsken
Datenkonvertierung/Satz: Werksatz Schmidt & Schulz, D-06773 Gräfenhainichen.
Druck und Bindung: Druckhaus „Thomas Müntzer“, Bad Langensalza.

Printed in the Federal Republic of Germany.

Inhalt

Vorwort	1
1 Einführung	3
Teil 1 Methoden, Qualitätskriterien und Perspektiven der Umfrageforschung	9
2 Bestimmung des Gegenstandsbereichs <i>Umfrageforschung</i>	11
2.1 Bestimmung	11
2.2 Welche Informationen werden erhoben?	13
2.3 Was heißt repräsentative Befragung?	13
2.4 Welche Befragungsarten unterscheidet man?	14
2.5 Wie ist der Fragebogen zu gestalten?	15
2.6 Welche Untersuchungspläne werden verwendet?	15
3 Stichproben in Theorie und Praxis	16
3.1 Grundgesamtheit	16
3.2 Stichproben	16
3.3 Vor- und Nachteile der unterschiedlichen Befragungsarten aus statistischer Sicht	18
3.3.1 Persönlich-mündliche Befragungen	18
3.3.2 Telefonische Umfragen	19
3.3.3 Schriftliche Befragungen	19
4 Fragebogen und Fragebogenmodelle	20
4.1 Vorbemerkung	20
4.2 Zur Problematik der Erfassung der öffentlichen Meinung	21
4.3 Konkrete Probleme	21
4.3.1 Unangemessene Komplexitätsreduktion	21
4.3.2 Einflüsse des Fragenumfeldes	22
4.3.3 Wirkungen unterschiedlicher Fragemodelle	23
4.3.4 Einflüsse des Befragungsmodus	23
4.4 Qualitätskriterien für Fragebogen und Fragebogenmodelle	24
4.5 Effekte technischer Entwicklungen der Umfrageforschung im Bereich Fragebogen und Fragebogenmodelle	25

5	Feldarbeit	27
5.1	Vorbemerkung	27
5.2	Rahmenbedingungen	28
5.3	Feldgeschehen und Schwachstellen der Feldarbeit	29
5.4	Auswirkungen auf die Qualität der Umfrageergebnisse	31
5.5	Sicherung und Verbesserung der Qualität in der Feldarbeit – <i>best practices</i> und <i>good practices</i>	32
5.5.1	Öffentlichkeitsarbeit	33
5.5.2	Kodifizierung von <i>good practices</i>	33
5.5.3	Definition von Qualitätskriterien	34
5.5.3.1	Stichprobe	35
5.5.3.2	Offenlegung der Stichprobenrealisierung	35
5.5.3.3	Kontaktierung	36
5.5.3.4	Kontrolle der Kontaktierung	36
5.5.3.5	Legitimation der Interviewer	37
5.5.3.6	Interviewerkontrollen	37
5.5.3.7	Anzahl der Interviews pro Interviewer	38
5.5.3.8	Nutzen und Anreize	38
5.6	Vorläufiges Fazit	39
5.7	Telefoninterviews	40
5.8	<i>Access Panels</i>	41
6	Standards der Umfrageforschung, Prozeßkontrolle und internationale Situation	43
6.1	Vorbemerkung	43
6.2	Vorbildliche Umfrageforschung	44
6.2.1	Einführung	44
6.2.2	AAPOR-Standards	44
6.2.2.1	Kosten-Nutzen-Gesichtspunkte	45
6.2.2.2	Stichproben	47
6.2.2.3	Pretest: Entwicklung und Erprobung von Frageinstrumenten und -techniken	49
6.2.2.4	Interviewerauswahl und -training	49
6.2.2.5	Statistische Aufbereitung und Auswertung	50
6.3	Prozeßorientierte Qualitätskontrollen	51
6.3.1	Zum internationalen Diskussionsstand	51
6.3.2	Zertifizierung nach ISO 9000	51
6.3.3	Nationale Standards	53
6.3.4	Internationale, länderübergreifende prozeßorientierte Qualitätsstandards	54
6.3.5	Zur Situation in Deutschland	55
6.3.6	Zur Qualitätsdiskussion im System der europäischen amtlichen Statistik	56
Teil 2	International vergleichende Umfrageforschung und Archivierung von Umfragen	57

Inhalt

7	International vergleichende Umfrageforschung	59
7.1	Vorbemerkung	59
7.2	Effekte der Internationalisierung	60
7.3	Perspektiven der international vergleichenden Sozialforschung	60
7.4	Harmonisierungsbestrebungen in der europäischen Statistik	61
7.5	Entwicklungsprioritäten	62
8	Archivierung von Umfragedaten	64
8.1	Vorbemerkung	64
8.2	Eingangskontrollen der Umfragen	65
8.3	Datenaufbereitung und Dokumentation	65
8.4	Langzeitspeicherung	66
8.5	Bereitstellung für Sekundäranalysen	66
Teil 3	Ausbildung und Massenmedien	69
9	Die universitäre und praktische Ausbildung im Bereich der Empirischen Sozialforschung	71
9.1	Diagnose	71
9.2	Optionen für eine Verbesserung der Situation	72
10	Umfrageforschung und Öffentlichkeit	73
10.1	Vorbemerkung	73
10.2	Veröffentlichung von Umfrageergebnissen	74
10.3	Aus- und Weiterbildung von Journalisten	76
Teil 4	Forschungsfreiheit und Datenschutz	79
11	Forschungsfreiheit und Datenschutz: Bestandsaufnahme und Ausblick	81
11.1	Vorbemerkung	81
11.2	Datenschutz und Empirische Sozialforschung	81
11.3	Zur Aushöhlung des Forschungsbegriffs	84
Teil 5	Entwicklungsdesiderate	87
12	Zur Zusammenarbeit privatwirtschaftlich und akademisch verfaßter Sozialforschung	89
12.1	Ausgangslage	89
12.2	Zur heutigen Situation	91
12.3	Folgerungen	92
13	Wandel und Perspektiven der Umfrageforschung	93
13.1	Vorbemerkung: Qualität und Wandel	93
13.2	Qualitätsstandards und ihre Voraussetzungen	94
13.2.1	Qualitätskriterien: Richtige Ergebnisse und korrekte Verfahren	94
13.2.2	Historische Bedingungen: Die Infrastruktur der Umfrageforschung	95

Inhalt

13.2.3	Verfahrensmodelle	96
13.3	Entwicklungstrends der Umfrageforschung	97
13.3.1	Neue Informations- und Kommunikationstechniken	97
13.3.2	Der Markt für Umfragen: Angebot und Nachfrage	98
13.3.3	Pluralisierung der Verfahrensmodelle	99
13.4	Kontroversen und Perspektiven	99
13.4.1	Stichprobenqualität	100
13.4.1.1	Ein Kriterienraster	100
13.4.1.2	Ausgewählte Qualitätsaspekte von Bevölkerungsumfragen	101
13.4.2	Interviewmethoden und neue Techniken	105
13.4.3	Aufgaben der sozialwissenschaftlichen Qualitätsdebatte	107
Teil 6	Empfehlungen	109
14	Zusammenfassende Analyse und Empfehlungen	111
15	Literaturverzeichnis	117
Teil 7	Anhang/Dokumentation	127
16	Codes of Ethics der American Association for Public Opinion Research (AAPOR)	129
16.1	Vorbildliche Praktiken der Umfrageforschung	129
16.1.1	Einleitung	129
16.1.2	Vorbildliche Praktiken der Umfrageforschung	130
16.1.3	Umfragepraktiken, die AAPOR mißbilligt	138
16.1.4	Quellen	139
16.2	Code of Professional Ethics and Practices	140
16.3	Research Industry Coalition (RIC): Professional and Ethical Standards For Marketing, Opinion, and Media Research	142
16.3.1	Research Industry Coalition (RIC)	142
16.3.2	RIC Statement of Professional and Ethical Standards For Marketing, Opinion, Media, and Related Research	143
17	DFG-geförderte Projekte in den Sozialwissenschaften 1987 bis 1996	145
	Verzeichnis der Abkürzungen	147
	Mitglieder der Arbeitsgruppe	149

Contents

Foreword	153
1 Introduction	155
Part 1 Methods, Quality Criteria and Perspectives in Survey Research	159
2 Defining <i>Survey Research</i>	161
2.1 Definition	161
2.2 What Information is Collected?	162
2.3 What is a Representative Sample Survey?	163
2.4 What Types of Interview Are There?	164
2.5 How Should the Questionnaire be Designed?	164
2.6 What Research Designs are Used?	165
3 Sampling in Theory and Practice	166
3.1 Population	166
3.2 Samples	166
3.3 Advantages and Disadvantages of Different Interview Techniques from a Statistical Point of View	167
3.3.1 Face-to-face interviewing	168
3.3.2 Telephone interviewing	168
3.3.3 Mail interviewing	169
4 Questionnaires and Questionnaire Models	170
4.1 Introductory Remarks	170
4.2 Problems in Recording Public Opinion	171
4.3 Specific Problems	171
4.3.1 Inappropriate reduction of complexity	171
4.3.2 Influence of the question context	172
4.3.3 Influence of different question models	172
4.3.4 Influence of the interviewing mode	173
4.4 Quality Criteria for Questionnaires and Questionnaire Models	173
4.5 Effects of Technological Developments in Survey Research on Questionnaires and Questionnaire Models	174

Contents

5	Field Work	176
5.1	Introductory Remarks	176
5.2	General Setting	177
5.3	Working in the Field and Where the Weak Spots Are	178
5.4	Impact on the Quality of Survey Results	179
5.5	Assuring and Improving Field Work Quality: <i>Best Practices</i> and <i>Good Practices</i>	180
5.5.1	Public relations	180
5.5.2	Codifying good practices	181
5.5.3	Defining quality criteria	182
5.5.3.1	Sampling	182
5.5.3.2	Information about sample implementation	183
5.5.3.3	Making contact	183
5.5.3.4	Monitoring contacts	184
5.5.3.5	Interviewer legitimisation	184
5.5.3.6	Interviewer supervision	185
5.5.3.7	Number of interviews per interviewer	185
5.5.3.8	Benefits and incentives	186
5.6	Interim Conclusion	186
5.7	Telephone Interviews	187
5.8	Access Panels	188
6	Standards of Survey Research, Process Control and the International Situation	190
6.1	Introductory Remarks	190
6.2	Excellence in Survey Research	191
6.2.1	Introduction	191
6.2.2	AAPOR standards	191
6.2.2.1	Cost–benefit considerations	192
6.2.2.2	Samples	194
6.2.2.3	Pretesting: developing and testing interview tools and techniques	195
6.2.2.4	Interviewer recruitment and training	195
6.2.2.5	Statistical analysis and interpretation	196
6.3	Process-Related Quality Management	197
6.3.1	The international discussion	197
6.3.2	ISO 9000 certification	197
6.3.3	National standards	198
6.3.4	International, crossborder process-related quality standards	200
6.3.5	The situation in Germany	200
6.3.6	The quality debate in the system of European official statistics	201
Part 2	International Comparative Survey Research and the Archiving of Surveys	203
7	International Comparative Survey Research	205
7.1	Introductory Remarks	205

Contents

7.2	Effects of Internationalisation	206
7.3	Prospects for International Comparative Social Research	206
7.4	Harmonisation of European Statistics	207
7.5	Development Priorities	208
8	Archival Storage of Survey Data	210
8.1	Introductory Remarks	210
8.2	Incoming Inspection of Survey Material	211
8.3	Data Editing and Documentation	211
8.4	Long-Term Storage	212
8.5	Availability for Secondary Analysis	212
Part 3	Training and the Mass Media	215
9	University and Practical Training in Empirical Social Science	217
9.1	Diagnosis	217
9.2	Options for Improving the Situation	218
10	Survey Research and the Public	219
10.1	Introductory Remarks	219
10.2	Publishing Survey Findings	220
10.3	Initial and Further Training for Journalists	222
Part 4	Freedom of Research and Data Protection	225
11	Freedom of Research and Data Protection: Status and Prospects	227
11.1	Introductory Remarks	227
11.2	Data Protection and Empirical Social Research	227
11.3	Erosion of the Research Concept	229
Part 5	Development Desiderata	231
12	Collaboration between Private and Academic Social Research	233
12.1	The Initial Situation	233
12.2	The Current Situation	235
12.3	Conclusions	236
13	Change and Prospects in Survey Research	237
13.1	Introductory Remarks: Quality and Change	237
13.2	Quality Standards and the Prerequisites	238
13.2.1	Quality criteria: accurate results and appropriate methods	238
13.2.2	Historical conditions: the infrastructure of survey research	239
13.2.3	Procedural models	240
13.3	Trends in Survey Research	241
13.3.1	New information and communication technologies	241
13.3.2	The market for surveys: supply and demand	241

Contents

13.3.3	Pluralising procedural models	242
13.4	Controversies and Perspectives	242
13.4.1	Sample quality	243
13.4.1.1	A quality checklist	243
13.4.1.2	Selected quality aspects of population surveys	244
13.4.2	Interview methods and new technologies	247
13.4.3	The task of the social science debate on quality	249
Part 6	Recommendations	251
14	Summary Analysis and Recommendations	253
15	Bibliography	258
Part 7	Appendix/Documentation	267
16	Codes of Ethics of the American Association for Public Opinion Research (AAPOR)	269
16.1	Code of Best Practices for Survey and Public Opinion Research	269
16.1.1	Introduction	269
16.1.2	Best Practices for Survey and Public Opinion Research	270
16.1.3	Survey Practices That AAPOR Condemns	276
16.1.4	Sources	277
16.2	Code of Professional Ethics and Practices	278
16.3	Research Industry Coalition (RIC): Professional and Ethical Standards For Marketing, Opinion, and Media Research	280
16.3.1	Research Industry Coalition (RIC)	280
16.3.2	RIC Statement of Professional and Ethical Standards For Marketing, Opinion, Media, and Related Research	281
17	Social Science Projects Sponsored by the DFG 1987–1996	284
	Abbreviations	286
	Members of the Expert Group	288

Vorwort

Die Deutsche Forschungsgemeinschaft dient – gemäß § 1 ihrer Satzung – der Wissenschaft. Die Satzung präzisiert mehrere Felder dieser Dienstleistung: die finanzielle Unterstützung von Forschungsaufgaben, die Förderung der Zusammenarbeit unter den Forschern mit Partnern im In- und Ausland, die Ausbildung und Förderung des wissenschaftlichen Nachwuchses und die Beratung von Parlamenten und Behörden in wissenschaftlichen Fragen.

Eine die Erfüllung aller dieser Aufgaben sichernde Grundlage erwähnt die Satzung nicht ausdrücklich; sie wird aber, je länger die Deutsche Forschungsgemeinschaft besteht, immer wichtiger: die Qualitätssicherung der Forschung. „Lege artis zu arbeiten“, definiert eine 1998 erschienene Denkschrift der Deutschen Forschungsgemeinschaft als eine Grundvoraussetzung für gute wissenschaftliche Praxis. Was ist die „lex artis“ auf einem Gebiet wie der Empirischen Sozialforschung, die sich mit Wechselbeziehungen zwischen Menschen, zum Beispiel in Familien, Institutionen, Gemeinschaften und Staaten, und ihren Lebensverhältnissen, Gewohnheiten, Anschauungen und Werthaltungen beschäftigt und deren Methoden sich dank der Ausweitung und den revolutionären Veränderungen in den Möglichkeiten der Kommunikation mit und zwischen ihren Forschungsgegenständen rapide fortentwickeln?

Sozialwissenschaftliche Forschung und – in einem engeren Kontext, um den es in dieser Denkschrift geht – Befragungen unter Verwendung sozialwissenschaftlicher Methoden bilden heute eine wichtige Grundlage von Entscheidungen auf so verschiedenen Feldern wie der Gesetzgebung, der Verkehrswegeplanung, der Formulierung und der öffentlichen Darstellung politischer Ziele durch Regierungen und Parteien oder auch der Einführung neuer Produkte in Märkte. Für alle diese und viele weitere alltägliche Entscheidungen werden Zustandsbeschreibungen, Analysen und Erklärungen benötigt, die sich der Methodik der Umfrageforschung bedienen. Umfragen waren und bleiben ein wichtiges Instrument sozialwissenschaftlicher Forschung zum Beispiel zu den Voraussetzungen und den Prozessen der deutschen Einigung und des langsamen Zusammenwachsens von Europa, wenn es um den Blickwinkel des Bürgers, des demokratischen Souveräns, geht.

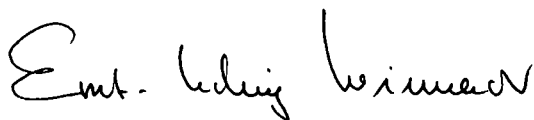
Für Umfragen gibt es einen Markt. Neue Methoden wie zum Beispiel computergestützte Telefoninterviews verändern die Verhältnisse auf diesem Markt. Sie können innovative Möglichkeiten des Qualitätsnachweises schaffen helfen, können aber auch die Markttransparenz verringern, wenn einzelne Anbieter mit ungewohnten

Preis-Leistungs-Relationen aufwarten, ohne daß dem Nachfrager ein informierter Vergleich möglich wäre.

Die vorliegende Denkschrift der Deutschen Forschungsgemeinschaft beschreibt den aktuellen „Stand der Kunst“ in der Umfrageforschung und setzt damit Maßstäbe für das, was das Prädikat „wissenschaftlich“ verdient. Sie analysiert Perspektiven der Methodenentwicklung und zeigt offene Fragen auf, die durch weitere Forschung geklärt werden müssen. Sie nutzt dafür Erkenntnisse aus anderen Ländern mit einer langen Tradition in Theorie und Praxis der Umfrageforschung, vor allem aus den Vereinigten Staaten von Amerika.

Die Denkschrift leistet aber noch mehr, denn sie formuliert Qualitätsmaßstäbe der Umfrageforschung in einem breiten Konsens von Sozialwissenschaftlern in Hochschulen und außeruniversitären wissenschaftlichen Forschungsinstituten, Nutzern und Anwendern von Umfrageergebnissen und führenden Vertretern von Anbietern auf dem Umfrage-Markt, also der für Auftraggeber arbeitenden, privatwirtschaftlich verfaßten Markt- und Meinungsforschungsinstitute. Sie behandelt Qualität in unterschiedlichen Dimensionen von der handwerklichen Korrektheit über forschungsethische Grundsätze des Umgangs mit Befragten und ihren Aussagen bis hin zu Fragen der Dokumentation und Offenlegung der aus Umfragen hervorgehenden Datensätze. Schließlich und fast nebenbei gibt sie dem allgemein interessierten Leser aufschlußreiche Einblicke in die thematische Vielfalt, die soziale Relevanz und die methodische Strenge quantitativer Sozialforschung. Sie wendet sich damit an die „Profession“ ebenso wie an Politik, Unternehmen, Öffentlichkeit und hier insbesondere die Massenmedien, die an qualitativ hochwertigen, gut abgesicherten und als Grundlage für Entscheidungen tauglichen Umfrageergebnissen interessiert sind und interessiert sein müssen.

Die Denkschrift ist in mehrjähriger, intensiver, ehrenamtlicher Tätigkeit einer vom Präsidium der Deutschen Forschungsgemeinschaft berufenen Arbeitsgruppe entstanden. Das Präsidium hat der Veröffentlichung des Ergebnisses ihrer Beratungen dankbar zugestimmt. Max Kaase als ihrem Vorsitzenden und allen ihren Mitgliedern und Gästen sage ich meinen herzlichen Dank und wünsche der Denkschrift viele Leser und eine nachhaltige Wirkung zugunsten der Forschung und der Anwendung ihrer Erkenntnisse in Politik, Wirtschaft und Gesellschaft.



Prof. Dr. Ernst-Ludwig Winnacker
Präsident der Deutschen Forschungsgemeinschaft

1 Einführung

Empirische Sozialforschung, das heißt die systematische, wissenschaftlich fundierte Erfassung sozialer, politischer und kultureller Verhältnisse der in territorialen Einheiten und gesellschaftlichen Verbänden lebenden Menschen, hat eine lange Tradition. Sie reicht von den bereits im Neuen Testament erwähnten Volkszählungen bis hin zu den Sozialenketten des 18. Jahrhunderts in Großbritannien (Scheuch 1965; Maus 1973; Kaase u. a. 1983; Oberschall 1997). Ihre eigentliche Bedeutung als *Umfrageforschung* – der Gegenstand dieser Denkschrift – hat sie jedoch erst über zwei wissenschaftliche Entwicklungen im 20. Jahrhundert erlangt: zum einen über die wahrscheinlichkeits-theoretisch fundierte Konzeption von Zufallsstichproben, die auf der Grundlage einer geringen Zahl von Einheiten zuverlässige Schlußfolgerungen über beliebig große Grundgesamtheiten zulassen, und zum anderen über die Methodik der mündlichen Befragung der Angehörigen solcher Stichproben mit standardisierten, das heißt für alle Erhebungseinheiten gleich gestalteten Fragebogen (Kaase 1997).

Trotz der europäischen Wurzeln dieser Verfahren und Vorgehensweisen erfolgte ihre wissenschaftliche Entwicklung vornehmlich in den Vereinigten Staaten von Amerika seit den zwanziger und danach in den vierziger und fünfziger Jahren. Dabei wurden wesentliche intellektuelle Voraussetzungen nicht zuletzt durch europäische Wissenschaftler geschaffen, die wegen der Verfolgung durch den Nationalsozialismus gezwungen worden waren, in die USA zu emigrieren (Lepsius 1981).

Besonders bemerkenswert ist, daß die Techniken der Empirischen Sozialforschung, und hierbei vor allem die Entwicklung von Forschungskonzepten zur Befragung repräsentativer Bevölkerungsstichproben, in den USA von Anfang an auch eine demokratietheoretische Einbettung erfahren haben. Gelang es doch auf diese Weise erstmals, die politischen Meinungen, Einstellungen und Verhaltensweisen der Bevölkerung zu beschreiben und sie in ihren sozialen und politischen Entstehungszusammenhängen zu analysieren.

Der europäische Demokratisierungsprozeß des ausgehenden 19. und frühen 20. Jahrhunderts verlief in erster Linie über die Einführung des gleichen und allgemeinen Wahlrechts. Dabei lag es in der Natur dieses Entscheidungsverfahrens, daß es zur Wahl und Legitimierung von Regierungen nur in regelmäßigen, zeitlich größeren Abständen (in der Regel vier Jahre) verwendet werden konnte. Mit der Methodik der repräsentativen Umfrageforschung wurde es möglich, die politischen Meinungen

und Einstellungen der Bevölkerung auch zwischen Wahlen jederzeit zuverlässig zu erfassen und damit den demokratischen Souverän – die wahlberechtigte Bevölkerung – in vielfältiger Weise in den kontinuierlichen politischen Willensbildungs- und Entscheidungsprozeß in Demokratien einzubeziehen.

In demokratiethoretischer Perspektive ist diese Dimension der Umfrageforschung stets ambivalent bewertet worden. Einerseits wurde und wird heute noch immer argumentiert, daß die Qualität des demokratischen Prozesses durch die über Massenmedien vermittelte Einbringung von mehr oder weniger fluiden Bevölkerungsmeinungen entscheidend geschmälert würde, weil die Politiker zunehmend auf letztlich nicht sonderlich belastbare Informationen reagierten und damit die politischen Deliberationen durch ein problematisches, plebiszitäres Element verfälscht würden (Hennis 1957). Andererseits – und hierfür stehen vor allen Dingen Sozialwissenschaftler in den Vereinigten Staaten von Amerika – wurde der Gesichtspunkt hervorgehoben, daß die repräsentative Umfrageforschung zu einer entscheidenden Verbesserung der Qualität dieser Deliberationen führen würde. Diese Position ist vielfältig belegbar auch als Grundlage für die Bemühungen der amerikanischen Besatzungsmacht nach Ende des Zweiten Weltkrieges, in Deutschland eine wissenschaftlich fundierte Empirische Sozialforschung aufzubauen (Crespi 1952; Braun und Articus 1984; Reimann 1989; Kutsch 1995; Merritt 1995).

Die wissenschaftliche Grundlegung der Methodik der repräsentativen Umfrageforschung mit den Schwerpunkten der Entwicklung von tragfähigen Stichprobenplänen und eines theoretisch abgesicherten Instrumentariums der Informationsgewinnung durch Befragungen erfuhr einen ersten Höhepunkt in den fünfziger Jahren (siehe zum Beispiel Lazarsfeld und Rosenberg 1955; Hyman 1957). Danach galten diese Verfahren lange Zeit als wissenschaftlich so abgesichert, daß in diesem Bereich grundlegende Neuerungen durch Methodenforschung nicht mehr für notwendig gehalten wurden. Erst mit Weiterentwicklungen insbesondere in der Psychologie sowie mit dem zunehmenden Einsatz von Computern und der Entwicklung von Standardverfahren der statistischen Datenauswertung über Programmsysteme wie zum Beispiel das Statistical Package for the Social Sciences (SPSS) hat die Empirische Sozialforschung noch einmal einen neuen, großen Anstoß erhalten.

Dabei gab es – und das schließt sowohl die akademisch und die privatwirtschaftlich verfaßte Umfrageforschung als auch die amtliche Statistik ein – stets eine weitgehende Übereinstimmung über das, was auf diesem Gebiet zu gegebener Zeit als wissenschaftlich abgesicherte Verfahrensweisen zu gelten hatte. Diese Situation hat sich jedoch in den letzten Jahren durch die Entwicklung neuer Erhebungsverfahren vor allem in Form telefonischer Befragungen wesentlich verändert. Daraus resultierende Unsicherheiten sind auch ein wichtiger Anlaß für die Einsetzung der Arbeitsgruppe gewesen, die diese Denkschrift verfaßt hat.

Bereits ein Blick auf die Dynamik der Verbreitung der elektronischen Datenverarbeitung und die Entwicklung von darauf zugeschnittenen Programmen und Programmpaketen zur Analyse sozialwissenschaftlicher Daten zeigt, in welchem Umfang die Empirische Sozialforschung durch solche Innovationen Entwicklungsschübe erfahren hat. Inzwischen ist hier, nicht zuletzt durch die flächendeckende Einführung von Personalcomputern in der akademisch wie privatwirtschaftlich verfaßten Sozialforschung, eine Routinisierung dieser Aspekte der Forschung eingetreten. Wie in der

Denkschrift gezeigt werden wird, könnte die inzwischen in fast allen Ländern Europas umfassende Ausstattung der Haushalte mit Telefon und die zunehmende Verfügbarkeit von Anschlüssen an elektronische Netze (Lass 1997) der Empirischen Sozialforschung über entsprechende methodische Innovationen noch einmal einen Anstoß verleihen, und zwar in einer Reihe von sehr unterschiedlich zu bewertenden Facetten.

Insbesondere der starke Anstieg des Anteils telefonisch durchgeführter Interviews in der privatwirtschaftlich verfaßten Markt- und Meinungsforschung hat neue Akzente in der Sozialforschung gesetzt. Allein schon die Mitgliedsinstitute des Arbeitskreises deutscher Markt- und Sozialforschungsinstitute (ADM) haben 1996 rund 6,8 Millionen und 1997 rund neun Millionen standardisierte Interviews durchgeführt. Die Zahl der telefonischen Interviews betrug hierbei 1996 rund drei Millionen und 1997 rund 3,6 Millionen, also ein erheblicher absoluter Zuwachs (bei prozentual allerdings leichtem Rückgang von 44 Prozent 1996 auf 40 Prozent 1997; siehe dazu Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. 1997a: 5; 1998a: 5). Mit der Methode der telefonischen Befragung ist es nun zum Beispiel möglich, über zentral platzierte Telefonstudios kurzfristig zuverlässige, repräsentative Informationen über Einstellungen der Bevölkerung zu bestimmten Ereignissen zu erhalten und damit Prozesse der Bildung der öffentlichen Meinung zu analysieren. In der Wahlforschung sind zum Beispiel Telefoninterviews bis kurz vor dem Termin einer spezifischen Wahl eine Selbstverständlichkeit geworden; auf diese Weise gelingt es, die Dynamik der politischen Entscheidung der Wähler praktisch bis zum Wahltag abzubilden.

Es ist kein Zufall, daß die Methodik der Telefoninterviews sehr viel schneller in der Markt- und Meinungsforschung als in der akademisch verfaßten Sozialforschung (Gräf und Rohlinger 1997: XVII) und in der deutschen amtlichen Statistik Eingang gefunden hat. Beinhaltet diese Form der Befragung doch zwei Einschränkungen, mit denen an Universitäten tätige Sozialwissenschaftler nicht leicht leben können: die nach bisheriger Erfahrung wegen des Widerstands der Befragten nur in Ausnahmefällen überwindbare Beschränkung der Interviewdauer auf etwa 30 Minuten (diese Beschränkung gilt zum Beispiel nicht notwendigerweise für an einem Befragungsthema hoch interessierte Personen) und der Zwang, auf bewährte und mit großer Mühe entwickelte, zum Teil visualisierende komplexe Erhebungsinstrumente zu verzichten, weil diese über das Telefon nicht an die Befragten zu vermitteln sind. Aus solchen und anderen Gründen wird gelegentlich bereits die Ansicht vertreten, auch telefonisch durchgeführte Interviews bei repräsentativen Stichproben der Bevölkerung oder besonderen Teilgruppen stellen nur einen Zwischenschritt auf dem Wege zu Stichproben von Haushalten und Personen dar, die über den Anschluß an elektronische Netze in Quasi-Vertragsverhältnissen kontinuierlich zu unterschiedlichen Themen befragt werden können.

Die kurz beschriebene Entwicklung zu anderen Befragungsformen als dem persönlichen Interview in der Wohnung des Befragten zeichnet sich zum einen durch die große Geschwindigkeit aus, in der sie abläuft. Zum anderen wirft sie neue methodische Fragen auf wie zum Beispiel die nach der Vergleichbarkeit von einerseits durch persönliche Interviews vor Ort und andererseits durch Telefoninterviews gewonnenen Informationen. Gerade die Beantwortung solcher Fragen ist von entscheidender Bedeutung für die weitere Historisierung der Empirischen Sozial-

forschung in dem Sinne, daß über Datenarchive inzwischen differenzierte Informationen über Einstellungen, Mentalitäten und Verhaltensweisen der Bevölkerung im zeitlichen Verlauf zur Verfügung stehen, welche die Optionen für die geschichtswissenschaftliche Analyse gesellschaftlicher und politischer Entwicklungen erweitert und bereichert haben. Das setzt jedoch voraus, daß man sich der Gefahr inhaltlicher Fehldeutungen durch Methodenartefakte bewußt ist und sie methodisch unter Kontrolle bringt.

Vor allem durch die wachsende Zahl der Telefoninterviews ist also ein neues Feld für systematische Methodenforschung entstanden. In diesem Zusammenhang werden auch weitere Fragen nach der Qualität, Reichweite und Belastbarkeit der über Umfragen gesammelten sozialwissenschaftlichen Informationen an Bedeutung gewinnen. Die Janusköpfigkeit dieser Entwicklungen läßt sich an folgendem Beispiel illustrieren: Einerseits werden diese Informationen über telefonische Befragungen kontrollierter und vor allem schneller als in der Vergangenheit zugänglich werden. Andererseits schaffen die inzwischen relativ kostengünstige Verfügbarkeit von Hardware für telefonische Befragungen (Telefonstudios) und die Software für die Steuerung von Befragungen (zum Beispiel computerunterstützte Telefonbefragungen – CATI) günstige Bedingungen für den Eintritt von neuen Anbietern auf dem Markt der Sozialforschung und damit eine erhebliche Ausdifferenzierung der Anbieterstruktur. Es erhebt sich jedoch die Frage, wie sichergestellt werden kann, daß seitens dieser Anbieter die Informationsgewinnung mit der notwendigen Expertise erfolgt, und das bedeutet nicht zuletzt die Notwendigkeit des Rückgriffs auf qualifiziertes Personal und damit auf eine bestimmte Mindestbetriebsgröße.

Die Facetten, Bedingungen und möglichen Konsequenzen dieser Entwicklung werden im folgenden in dieser Denkschrift dargestellt und bewertet. Dabei kann es aber nicht nur um die Folgen gehen, die technisch bedingte Umwälzungen für die Empirische Sozialforschung besitzen. Vielmehr müssen auch die möglichen Wirkungen auf den Prozeß der Bildung der öffentlichen Meinung in demokratisch verfaßten Gesellschaften in den Blick kommen. Wie später gezeigt werden wird, handelt es sich bei der Umfrageforschung vor allem in ihrer privatwirtschaftlichen Ausgestaltung um einen Wachstumsmarkt. Hierzu hat entscheidend beigetragen, daß die Attraktivität der so gewonnenen Informationen durch die Massenmedien und hierbei insbesondere durch das Fernsehen entdeckt worden ist (Brettschneider 1996; Foundation for Information 1997). Ergebnisse der Empirischen Sozialforschung wirken zunehmend auf den Prozeß der öffentlichen Meinungsbildung ein. Damit ist die Frage nach der Qualität der Gewinnung und Darstellung von Ergebnissen der Umfrageforschung nicht mehr nur eine Angelegenheit des Wissenschaftssystems, sondern der gesamten Gesellschaft geworden.

In vielen Fällen wird diese Art von Information schon auf kurze – und sicherlich in den meisten Fällen auf mittlere – Sicht letztlich ohne Bedeutung für den gesellschaftlichen Diskurs bleiben. Da jedoch für die Öffentlichkeit im Prinzip nicht unterscheidbar ist, welche Ergebnisse der Umfrageforschung nach den Regeln der Kunst gewonnen worden sind und welche nicht, entstehen um so höhere Anforderungen an die wissenschaftliche Kontrolle der Qualität der Befunde und an die hauptsächlichen Vermittler der so gewonnenen Information an die Öffentlichkeit, nämlich die Massenmedien. Nicht zuletzt in diesem Zusammenhang muß die Forderung gesehen wer-

den, in den Massenmedien Standards für die Berichterstattung über Ergebnisse der Umfrageforschung zu etablieren und dabei auch an Instrumente und Verfahren zu denken, über die ein verantwortungsloser oder sogar bewußt verfälschender Umgang mit dieser Art von Informationen öffentlich gemacht und sanktioniert werden kann.

Vor diesem Hintergrund werden im folgenden Methoden, Erkenntnisse und Entwicklungsperspektiven der Umfrageforschung aufgezeigt und aus ihnen Schlußfolgerungen sowohl für die Forschung als auch für die Verwendung der Ergebnisse in Wissenschaft und Öffentlichkeit gezogen. Dieser Aufgabe stellt sich die Denkschrift mit dem Versuch, den Kenntnis- und Diskussionsstand über die Umfrageforschung am Ende des 20. Jahrhunderts in Deutschland zumindest in den wichtigsten Konturen zusammenfassend zu beschreiben und dabei auch auf Konsequenzen einzugehen, die sich aus ihrer inzwischen weltweiten Etablierung ergeben. Insofern liegt die Schwerpunktsetzung zwar auf der Situation in der Bundesrepublik Deutschland; diese Begrenzung spiegelt jedoch lediglich einen praktischen, nicht aber einen systematischen Gesichtspunkt wider.

Die Umfrageforschung ist in Deutschland seit längerem durch eine konstruktive Kooperation und einen kontinuierlichen Dialog zwischen akademisch und privatwirtschaftlich verorteten Forschern gekennzeichnet. Dieser Philosophie und Erfahrung sieht sich diese Denkschrift ebenfalls verpflichtet. Daraus ergibt sich aber auch, daß die Zielgruppe, für welche die Denkschrift bestimmt ist, nicht nur im Bereich der Hochschulen und vergleichbaren Einrichtungen liegen kann. Sie soll mindestens mit gleicher Intensität und Reichweite die in der privatwirtschaftlich verfaßten Umfrageforschung arbeitenden Wissenschaftler, sei es in den Instituten der Markt- und Meinungsforschung oder sei es bei den Unternehmen, Interessengruppen usw., erreichen. Angesichts der Öffentlichkeitswirkung von Ergebnissen der Empirischen Sozialforschung liegt ein dritter gewichtiger Adressatenkreis bei den Massenmedien, ebenso ist die Politik als Auftraggeber, Abnehmer und Verwerter von Umfrageergebnissen zu berücksichtigen.

Die Autoren der Denkschrift sind sich angesichts des rapiden Wandels in den Methoden der Umfrageforschung klar darüber, daß sie lediglich eine Bestandsaufnahme der Verhältnisse am Ende dieses Jahrhunderts vorlegen können. Allerdings sind schon jetzt Zäsuren und Entwicklungsperspektiven erkennbar, die es als besonders sinnvoll erscheinen lassen, diese Analyse gerade jetzt vorzunehmen (dieser Einschätzung entspricht, daß sich auch die Markt- und Meinungsforschung der Qualitätsproblematik annimmt; siehe dazu planung & analyse 1998). Sie soll und kann dazu beitragen, neue Akzente für die Forschung zu setzen, die Zusammenarbeit zwischen universitärer und privatwirtschaftlich verfaßter Sozialforschung sowie der amtlichen Statistik weiter zu intensivieren und dafür Sorge zu tragen, daß die wissenschaftlichen Grundlagen der Methoden der Empirischen Sozialforschung und ihre Weiterentwicklung wie auch der verantwortungsvolle, sachkundige Umgang mit ihren Ergebnissen in der Öffentlichkeit zu keiner Zeit aus dem Blick geraten.

Teil 1

Methoden, Qualitätskriterien
und Perspektiven
der Umfrageforschung

2 Bestimmung des Gegenstandsbereichs *Umfrageforschung*

2.1 Bestimmung

Mit dem Begriff *Umfrageforschung* als einem wichtigen Bereich der Empirischen Sozialforschung wird im Deutschen die Erhebungstechnik der *Befragung* als wesentliches Merkmal einer Untersuchungsart herausgehoben, für die im Englischen der Ausdruck *survey research* oder, konkreter, *representative sample surveys* üblich ist. Im Deutschen ist auch der Begriff *Demoskopie* gebräuchlich, von dem sich eher ein Bezug zum *survey research* herstellen läßt, leitet sich doch der Ausdruck *survey* vom lateinischen *supervidere* und die „Volksschau“ vom griechischen *σκοπειν* ab. Allgemein kann man Umfrageforschung als *Erhebung von Informationen über Bevölkerungen auf der Basis von Stichproben mit Hilfe von standardisierten Befragungsinstrumenten (Fragebogen)* definieren. Diese Definition enthält drei Kernelemente:

- die Befragung als Methode der Datenerhebung,
- die Stichprobe als Element der Auswahl der Befragten und
- eine Bevölkerung als Beschreibung der *Grundgesamtheit/Population*, zum Beispiel alle bei Bundestagswahlen wahlberechtigten Bürger.

Bei jedem dieser Kernelemente sind Einschränkungen oder Erweiterungen denkbar, die dann aber nicht mehr die Bedeutung der Umfrageforschung als Schnittmenge der Kernelemente abdecken. Die oben formulierte Definition grenzt die Umfrageforschung im engeren Sinn von derartigen Einschränkungen und Erweiterungen ab. Man könnte alle mit dem Mittel der Befragung gewonnenen Daten als Umfrageforschung bezeichnen. Darunter fiele dann auch die *Volkszählung*. Der Erhebungsbogen für die Volkszählung (*Makrozensus*) dient der Feststellung von Merkmalen, die mehr oder weniger als objektiv gelten, und deswegen wurde diese Erhebungsart ursprünglich als Zählung und nicht als Umfrage aufgefaßt. Hinzu kommt bei diesem Beispiel, daß es sich bei Makrozensen – im Unterschied zu den gebräuchlichen Stichprobenverfahren, die man einsetzt, um von einer kleinen Zahl ausgewählter Einheiten auf die Grundgesamtheit zu schließen – um Vollerhebungen handelt. Die mit *Stichproben* und *standardisierten Fragebogen* arbeitende Umfrageforschung weicht von der oben gegebenen Definition weiterhin oft insofern ab, als nicht Informationen über Bevölkerungen gesammelt werden, sondern Informationen über Organisa-

tionen, Betriebe oder juristische Personen, also Populationselemente im weiteren Sinn. Läßt man diese Erweiterung des Begriffs zu, gewinnen die in der Umfrageforschung im engeren Sinn verbreiteten Fragen zur Erhebung subjektiver Tatbestände eine andere Bedeutung. Tatsächlich können zwar immer nur natürliche Personen befragt werden, es macht aber einen Unterschied, ob sie über sich und ihre Einstellungen Auskunft geben oder ob sie als Datenquelle für Organisationen dienen. Die zu definierenden Qualitätskriterien der Umfrageforschung gelten für die oben gegebene Definition der Umfrageforschung im engeren Sinn.

Repräsentative Umfragen haben sich seit dem Zweiten Weltkrieg zum unentbehrlichen Mittel der Datenerhebung über größere Bevölkerungen entwickelt: in der Wahlforschung über Präferenzen und Meinungen der Wähler, in der Konsumforschung über Kaufverhalten und Konsumwünsche und in der Gesundheitsforschung über die Gesundheitszustände und Inanspruchnahme medizinischer Leistungen usw. In bestimmten Bereichen der Empirischen Sozialforschung wie zum Beispiel der Bildungsforschung werden oft Populationen innerhalb bestimmter Einrichtungen wie Schulen untersucht. Dagegen kann man die *allgemeinen Bevölkerungsumfragen* als Informationssammlung mit Hilfe standardisierter Befragungen über die Wohnbevölkerung in räumlich abgegrenzten Gebieten verstehen, wobei wegen der Identität von *Erhebungs-* und *Analyseeinheit* (in beiden Fällen: der einzelne Befragte) eine Einschränkung auf befragungsfähige Alters- und Sprachgruppen erfolgen kann. Eine typische allgemeine Bevölkerungsumfrage ist die Befragung der Wohnbevölkerung ab 14 Jahren in Privathaushalten in der Bundesrepublik Deutschland; dabei werden Personen mit nichtdeutscher Staatsangehörigkeit manchmal einbezogen und manchmal – wie in der Wahlforschung – ausgeschlossen. Diese *Zielpopulation* wird als repräsentative Stichprobe in der Regel an ihrem Wohnsitz und in ihrem Privathaushalt mündlich, telefonisch oder schriftlich befragt.

Allgemeine Bevölkerungsumfragen haben über die erwähnten Inhalte hinaus in modernen Demokratien auch deshalb eine besondere Bedeutung, weil mit ihnen die *öffentliche Meinung* erhoben werden kann. Man kann öffentliche Meinung als *die mit den Mitteln der repräsentativen Befragung festgestellten Meinungen und Einstellungen der Bevölkerung eines Landes zu Themen von öffentlichem Interesse* verstehen.¹ Politik und Massenmedien sind in erster Linie die Abnehmer dieser Ergebnisse der Umfrageforschung. Für ausgewählte Themen stehen auch Zeitreihen zur Verfügung, die den Prozeß der Meinungsbildung im *Aggregat* abbilden. Gegenüber dem alten Vorurteil, daß die öffentliche Meinung unerklärliche Schwankungen aufweise, hat die neuere Forschung feststellen können, daß die Meinung der Bevölkerung zu bestimmten zentralen Themen der Politik im Aggregat relativ stabil ist und daß Meinungsschwankungen sich in der Regel auf spezifische Ereignisse zurückführen lassen, auf welche die öffentliche Meinung dann reagiert (Page und Shapiro 1991).

1 Nach Converse (1987) hat sich diese Definition von öffentlicher Meinung größtenteils durchgesetzt (siehe dazu ferner Page und Shapiro 1991). Hennis (1957) hat gegen diese Konzeption von öffentlicher Meinung demokratietheoretische, Noelle-Neumann (1980) sozialpsychologische Einwände ins Feld geführt.

2.2 Welche Informationen werden erhoben?

Das Spektrum der zu erhebenden Informationen ist groß. Es reicht von der Feststellung oder „Zählung“ objektiver Sachverhalte wie Haushaltsgröße, Staatsangehörigkeit usw. über die Erfragung von Verhaltensabsichten und vergangenen Verhaltensakten („Welche Partei würden Sie wählen?“; „Welche Partei haben Sie gewählt?“) bis zur Messung subjektiver Phänomene wie Einstellungen, Meinungen, Wahrnehmungen und Präferenzen. Gerade die Erhebung subjektiver Phänomene macht die Umfrageforschung für bestimmte Forschungszwecke unentbehrlich (Turner und Martin 1984). In diesem Sinne konnte mit den ersten systematischen Umfragen bereits in der Zwischenkriegszeit (Crossley 1937; Noelle 1940; Oberschall 1997) die Informationsbasis über größere Bevölkerungen entscheidend erweitert werden.

2.3 Was heißt repräsentative Befragung?

Mit dem Ausdruck *Repräsentativerhebung* oder *-befragung* ist gemeint, daß auf der Basis einer dem statistischen Zufallsprinzip folgenden Auswahl der zu Befragenden aus der Gesamtpopulation ein *Schluß von der Stichprobe auf die Gesamtheit möglich ist* (zu den in diesem Abschnitt angesprochenen Aspekten siehe im einzelnen Kapitel 3). Dabei kommt es entscheidend darauf an, daß jedes Element der Zielpopulation eine *berechenbare Chance* hat, in die Stichprobe aufgenommen zu werden. Bei der *einfachen Zufallsauswahl* sind die Chancen der Angehörigen der Zielpopulation, in die Stichprobe zu gelangen, gleich; bei einer „*geschichteten*“ *Zufallsauswahl* sind die Auswahlchancen für die Mitglieder unterschiedlicher Teilpopulationen zwar unterschiedlich, aber bekannt beziehungsweise berechenbar (und durch *Gewichtung* kompensierbar; zur Gewichtung siehe die Abschnitte 3.2, 6.2.2.5 und 13.4.1.2). Bei Haushaltsstichproben wird außerdem noch die zu befragende Person im Haushalt nach einem Zufallsverfahren bestimmt.

Neben der Zufallsauswahl sind auch *Quotenauswahlen* gebräuchlich, bei denen nicht bestimmte Personen, sondern vom Interviewer zu bestimmende Träger von vorgegebenen Merkmalskombinationen befragt werden. Dabei wird durch die Quotenvorgaben je nach dem durch die Volkszählung (beziehungsweise ihre teilweise Fortschreibung durch den *Mikrozensus*²) ermittelten Anteil der Bevölkerungsgruppe an der Gesamtbevölkerung die Zusammensetzung der Stichprobe gesteuert. Unzulässig ist jedoch eine Auswahl nach Gutdünken zum Beispiel von Personen, die sich freiwillig für solche Aufgaben zur Verfügung stellen, denn damit wird die Verlässlichkeit des Schlusses von der Stichprobe auf die Gesamtheit in nicht berechenbarer Weise ver-

2 Der Mikrozensus ist die jährliche Bevölkerungsumfrage der amtlichen Statistik in Deutschland, durchgeführt mit einer Ein-Prozent-Stichprobe der Haushalte. Für das Kernprogramm des Fragebogens besteht eine Auskunftspflicht. Zum Mikrozensus siehe auch Esser u. a. 1989.

zerrt. Unabhängig davon folgt aus der Logik der Stichprobentheorie (siehe dazu Kapitel 3), daß jede aus einer Stichprobe abgeleitete Schätzung des „wahren“ Wertes in der Grundgesamtheit mit einem statistischen Zufallsfehler – dem *Stichprobenfehler* – behaftet ist, den man nach vorgegebenen Regeln berechnen kann. Die Größe dieses Fehlers (Fehlertoleranz) ist auch von der Größe der Stichprobe abhängig. Dabei ist es fast unerheblich, wie groß der Anteil der Stichprobe an der Grundgesamtheit ist; es kommt allein auf die Größe der Stichprobe an: auch für eine sehr große Bevölkerung von mehreren Millionen Personen führen Stichproben von tausend Befragten zu verlässlichen Ergebnissen.

Die Aussagefähigkeit von Umfragedaten wird durch den Stichprobenfehler und durch Probleme bei der Realisierung des Stichprobenplans, zum Beispiel infolge mangelnder Beteiligung der nach dem Stichprobenplan zu befragenden Personen (*Ausschöpfung*), beeinträchtigt. Die schwer erreichbaren Personen und diejenigen, die eine Befragung verweigern, können sich in vielen Merkmalen, die mit der Stichprobe erhoben werden sollen, systematisch von der Gesamtpopulation unterscheiden, so daß die Ausfälle in der Regel nicht als neutral betrachtet werden können.

2.4 Welche Befragungsarten unterscheidet man?

Die Befragung kann mündlich in einem persönlichen Interview, telefonisch oder schriftlich erfolgen. Die Wahl der Befragungsart hängt wesentlich vom Inhalt der zu erhebenden Information ab. So gelten zum Beispiel schriftliche Befragungen bei der Erhebung subjektiver Phänomene als problematisch, weil sich die Situation, in der die Zielperson den Fragebogen ausfüllt, nicht kontrollieren läßt. Innerhalb einer allgemeinen Bevölkerungsumfrage sind aber durchaus verschiedene Methoden der Datenerhebung kombinierbar. So können Befragte zum Beispiel während des Interviews gebeten werden, selber einen Einstellungsfragebogen auszufüllen. Im allgemeinen gilt, daß nach wie vor in mündlichen Befragungen sehr viel komplexere Informationen erhoben werden können als durch eine schriftliche Befragung. Das Hauptproblem von Telefoninterviews ist die Zeitdauer der Befragung, die im Durchschnitt wesentlich kürzer ist als die bei persönlichen Interviews, und der Verzicht auf visuelle Hilfsmittel bei der Befragung. Als Vorteil ist vor allem die bessere Kontrolle des Stichprobenplans und der Interviewer zu erwähnen.

Immer häufiger kommt inzwischen die computerunterstützte Befragung vor, bei der die Datenaufnahme sofort maschinenlesbar erfolgt, also nicht unter Verwendung von Bleistift und Papier und einer anschließenden Verkodung der Antworten. Beim Telefoninterview spricht man von *computer assisted telephone interviews (CATI)* und bei mündlichen Befragungen von *computer assisted personal interviews (CAPI)*.

2.5 Wie ist der Fragebogen zu gestalten?

Da die Umfrageforschung mit zumindest teilweise standardisierten Fragebogen arbeitet, hängt die Qualität der Ergebnisse wesentlich von der Formulierung der Fragen und ihrer Zusammenstellung in einem Fragebogen ab (siehe dazu Kapitel 4). Dies gilt insbesondere für die Erfassung subjektiver Phänomene. Die Frage wird dabei als verbaler Reiz aufgefaßt, auf den die Befragten je nach ihrer Meinung, Einstellung usw. unterschiedlich reagieren (Schnell u. a. 1992). Bei der *geschlossenen Frage* sind die Antwortkategorien vorgegeben, zwischen denen sich der Befragte entscheiden soll; bei *offenen Fragen* protokolliert der Interviewer die frei formulierten Antworten des Befragten. Die Antworten können sowohl von der Frageformulierung als auch vom Fragenkontext, zum Beispiel von den unmittelbar vorher gestellten Fragen, beeinflusst werden. Dieses Problem ist bei der Platzierung der Frage im Fragebogen zu berücksichtigen.

Abweichungen von der voll standardisierten Befragungsform führen zu verschiedenen Erweiterungen in Richtung qualitativer Verfahren wie *Leitfadeninterviews* (Vorgabe von anzusprechenden Themen statt von festen Fragen) oder *narrativen Interviews*. Bei letzteren wird lediglich ein „Großthema“ vorgegeben, zu dem der Befragte seine Sichtweise als „Erzählung eigenerlebter Geschichten“ in seinen Worten und Akzentuierungen mitteilt (Schnell u. a. 1992: 354–355). Je offener diese Gesprächsformen sind, desto schwieriger sind sie als Bestandteil einer größeren Repräsentativbefragung mit einer Vielzahl von Interviewern einsetzbar.

2.6 Welche Untersuchungspläne werden verwendet?

Allgemeine Bevölkerungsbefragungen werden in der Regel als *Querschnittsstudien* durchgeführt. Die ausgewählten Zielpersonen werden innerhalb eines bestimmten Zeitraums einmal befragt. Von Querschnittsbefragungen sind *Wiederholungs-* oder *Panelbefragungen* zu unterscheiden, bei denen dieselben Zielpersonen nach einem bestimmten Zeitabstand erneut befragt werden. Diese Befragungsform ist von wiederholten Querschnittsbefragungen mit identischen Fragen zu unterscheiden, die auf verschiedenen Stichproben derselben Population nach dem gleichen allgemeinen Stichprobenplan beruhen und deren Ergebnisse in aggregierter Form für die einzelnen Zeitpunkte der jeweiligen Querschnittsbefragungen zu *Zeitreihen* zusammengestellt werden (*Trendstudien*). So lassen sich zum Beispiel die Autokaufwünsche der Konsumenten über einen Zeitraum von 20 Jahren bei halbjährlicher Erhebung mit den tatsächlichen Absatzzahlen vergleichen. Eine andere Anwendungsform sind die „Populäritätsfunktionen“ der Regierung, das heißt die Sympathiewerte, die die Regierung eines Landes bei der wahlberechtigten Bevölkerung zu verschiedenen Zeitpunkten erzielt, oder auch die regelmäßige Erhebung von Parteisympathien beziehungsweise Wahlabsichten.

3 Stichproben in Theorie und Praxis

3.1 Grundgesamtheit

Bei der Definition der Grundgesamtheit ist zu unterscheiden zwischen der *Untersuchungsgesamtheit* (auch Zielpopulation genannt), die alle Personen enthält, über die man letztlich Aussagen machen möchte (zum Beispiel Wahlbevölkerung), und der *Erhebungsgesamtheit* (auch *Stichprobenrahmen* genannt). Die Erhebungsgesamtheit besteht aus den Einheiten, über die man sich Zugang zu den Daten der Untersuchungseinheiten verschafft (zum Beispiel Wählerverzeichnisse). In vielen Fällen sind in der Umfrageforschung Erhebungs- und Untersuchungsgesamtheit identisch. Eine möglichst exakte Definition der Grundgesamtheit ist für die Entwicklung von guten Stichprobenplänen unabdingbar.

3.2 Stichproben

Da in der Umfrageforschung die Einbeziehung aller der Grundgesamtheit zuzurechnenden Einheiten in eine Untersuchung in der Regel viel zu kosten- und zeitaufwendig wäre, beschränkt man sich auf eine gewisse Anzahl ihrer Vertreter (*Stichprobe*), um Aussagen über Erscheinungen in der Grundgesamtheit machen zu können. Der erste, der für eine solche repräsentative Methode in Form von Stichproben eintrat, war 1895 der Norweger Kiaer. Der Zweig der diese Verfahren begründenden Stichprobentheorie entstand innerhalb der Statistik und erhielt von Godambe (1955) ein einheitliches theoretisches Fundament.

Eine Stichprobe ist ein Teil einer Gesamtheit. Wünschenswert sind Stichproben, deren Struktur in bezug auf Merkmale nicht von der Struktur der Merkmale und Merkmalskombinationen in der Gesamtheit abweicht, Stichproben also, die ein *Miniaturbild der Gesamtheit* sind. In der Praxis existieren Stichproben, die diesem hohen Anspruch für alle Merkmale exakt genügen, nur in mehr oder weniger guter Annäherung. Fehler können innerhalb berechenbarer Toleranzbereiche zufallsbedingt sein oder in systematischer Weise auftreten. Der in diesem Zusammenhang

einschlägige Begriff der Repräsentativität wird zwar gerne bemüht, jedoch in vielfältiger Weise mißbraucht (Kruskal und Mosteller 1979a, 1979b, 1979c, 1980).

Die Entscheidung für eine bestimmte Art der Stichprobenerhebung kann nach ganz unterschiedlichen Kriterien erfolgen. Eine besondere Bedeutung wird Stichproben zugeschrieben, die nach dem Zufallsprinzip ausgewählt werden und so in berechenbaren Fehlertoleranzen statistische Schlüsse auf die zugrundeliegende Grundgesamtheit zulassen. Entsprechend ist die Entwicklung von *Schätzfunktionen* mit günstigen Eigenschaften wie Erwartungstreue bezüglich des *Stichprobendesigns* oder Optimalitätseigenschaften von besonderem Interesse. Dabei zeigt sich, daß im Gegensatz zur klassischen Statistik eine auf allen Dimensionen „beste“ Schätzfunktion nicht existiert.

Die Bausteine der Stichprobentheorie sind eine Wahrscheinlichkeitsverteilung über die Menge aller möglichen Stichproben und eine Vorschrift, wie aus den ermittelten Stichprobendaten auf die interessierenden Kennzahlen der Gesamtheit geschlossen werden soll. Betrachtet man beide Bausteine gemeinsam, so spricht man von einer *Stichprobenstrategie*. Der erste Teil der Strategie, wie man zu den ausgewählten Einheiten der Grundgesamtheit gelangt – der sogenannte *Stichprobenplan* –, wird bei der Analyse der Daten in der Praxis oft vernachlässigt. Der Grund hierfür liegt zum einen an der Komplexität der in der Umfrageforschung verwendeten Stichprobenpläne, die dem Nutzer der Daten bei *Sekundäranalysen* häufig noch nicht einmal genau bekannt sind. Zum anderen stehen bei Datenanalysen Verteilungsmodelle der Merkmale im Vordergrund, in denen der Prozeß der Datenerhebung keine Rolle spielt. Die Qualität der Ergebnisse steht und fällt dann mit der Gültigkeit des Modells, das in der Regel unabhängig identisch verteilte *Zufallsvariablen* unterstellt. Solche Variablen (Merkmale) werden exakt jedoch nur im Falle einer einfachen Zufallsauswahl mit Wiederholung erzeugt.

Berücksichtigt man bei der Hochrechnung von der Zufallsstichprobe auf Kennzahlen der Gesamtheit das zugrundeliegende Design in angemessener Weise, lassen sich Vertrauensgrenzen (*Konfidenzintervalle*, *Fehlertoleranzen*) für die interessierenden Kennzahlen berechnen. In der Regel sind die Schätzergebnisse um so genauer, je größer der Stichprobenumfang ist. Ergibt sich etwa bei einer uneingeschränkten Zufallsauswahl vom Umfang $n = 1000$ ein Anteil an Befragten mit Abitur von zehn Prozent, wird man mit einer Wahrscheinlichkeit von 0,95 den entsprechenden Anteilswert in der Grundgesamtheit zwischen 8,14 Prozent und 11,86 Prozent erwarten dürfen. Bei einem Wert von 50 Prozent in der Stichprobe lägen die Anteilswerte in der Grundgesamtheit zwischen 46,9 Prozent und 53,1 Prozent.

In der Praxis berührt der Umstand, daß sich nicht alle Mitglieder einer Zufallsstichprobe an einer Untersuchung beteiligen, den Stichprobenplan mehr oder weniger; die Gefahr der *Verzerrung* einer Stichprobe entsteht. In der Umfrageforschung wird diesem Problem so Rechnung getragen, daß durch ein *Anpassungsgewicht* bei den Befragten die Verteilung der Merkmalsausprägungen derart korrigiert wird, daß nach der Gewichtung die Struktur von für wichtig gehaltenen Merkmalen in der Stichprobe mit der Struktur in der Gesamtheit übereinstimmt. Inwieweit sich diese Art der Gewichtung auf sonstige in der Befragung erhobene Merkmale auswirkt, kann von Fall zu Fall sehr unterschiedlich sein und ist Gegenstand intensiver Fachdiskussionen (zum Beispiel Schnell 1997: 245–247).

3.3 Vor- und Nachteile der unterschiedlichen Befragungsarten aus statistischer Sicht

Prinzipiell sind repräsentative Bevölkerungsbefragungen derzeit als persönlich-mündliche (*face to face*), telefonische und schriftliche Befragungen durchführbar.³ Hinsichtlich der Stichprobenanlage und der praktischen Umsetzung der Designs in der Feldphase muß jedoch jeweils auf einige Probleme verwiesen werden, welche die Qualität der Stichproben beeinflussen können.

3.3.1 Persönlich-mündliche Befragungen

Gegenwärtig existieren in Deutschland für persönlich-mündliche Bevölkerungsbefragungen zwei hauptsächliche Stichprobendesigns auf der Grundlage von Zufallsauswahlen. Zu nennen ist als erstes das *random-route*- beziehungsweise das *Adreßvorlaufverfahren* (der Haushalt wird vom Interviewer entweder nach Regeln zufällig ausgewählt oder der Interviewer erhält eine genaue Adressenvorgabe eines zuvor nach dem Zufallsprinzip ausgewählten Haushalts). Die Auswahl erfolgt allgemein auf Grundlage der Standardstichproben des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM). Dabei werden, wie gerade beschrieben, zunächst Haushalte und dann in einem weiteren Schritt eine Person im Haushalt nach Regeln zufällig ausgewählt (Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt 1994). Das zweite Verfahren ist die Auswahl über eine *Einwohnermeldeamtsstichprobe*, das heißt eine Zufallsauswahl namentlich genau bestimmter Personen aus den Meldeeregistern der in der Flächenstichprobe vertretenen Gemeinden.

Die Umsetzung dieser Designs wird in der Praxis durch die Interviewer und die Interviewten (zum Beispiel Einhaltung der Begehungsvorschriften für die Zielhaushalte sowie der Vorschrift zur Ermittlung der Zielperson beim ADM-Design, ferner Verweigerungen und Nichterreichbarkeit) beeinflusst. Die dabei ablaufenden Prozesse können zwar als Randomisierungsprozesse aufgefaßt werden, die das ursprünglich im Stichprobenplan vorgesehene Zufallsverfahren überlagern (Lessler und Kalsbeek 1992), lassen sich bisher jedoch stichprobentheoretisch nicht befriedigend beschreiben. Daher sind Schlüsse von der Stichprobe auf die ihr zugrundeliegende Grundgesamtheit mitunter nicht zuverlässig möglich (Stenger 1994: 44; Bradburn 1992: 394–397).

Ein Hauptproblem bei persönlich-mündlichen Befragungen liegt darin, daß der Entscheidungsspielraum für die Interviewer im Rahmen der Implementation der genannten Stichprobenpläne – zum Beispiel im Vergleich zu Telefonumfragen – relativ groß ist (siehe dazu ausführlich Kapitel 5). Aber selbst im Vergleich der Designs untereinander gibt es Unterschiede: Insbesondere bei *random-route*-Stichproben, aber auch beim Adreßvorlaufverfahren ist der Entscheidungsspielraum der Interviewer deutlich höher als bei Einwohnermeldeamtsstichproben. Da sich daraus aller Erfahrung nach Konsequenzen auch für die korrekte Umsetzung des Designs erge-

3 Umfragen via Internet beziehungsweise per e-mail sind aufgrund der zu geringen Ausstattung der Bevölkerung mit diesen Medien gegenwärtig nicht sinnvoll. Sie sind bisher lediglich für Expertenbefragungen oder Umfragen in Intranets geeignet.