

# Vom Kohlenstoff zum Diamanten – oder vom Faktum zur Information



**Dr. Carmina Brenner**

**Auszüge eines Vortrags über einige Merkwürdigkeiten der erhebenden, analysierenden und darstellenden Statistik von Dr. Carmina Brenner, Präsidentin des Statistischen Landesamtes Baden-Württemberg bei der Heinrich-Hertz-Gesellschaft am 04. Juni 2008 in Karlsruhe.**

„Da ich noch als ‚ordentliches‘ Mädchen in meiner Schulzeit sozialisiert wurde, entsteht in meinem Magen beim Wort Statistik automatisch ein Kloß. Von Zahlen durften nur junge und hässliche Brillenschlangen etwas verstehen. Ansonsten war es charmant und förderlich für den Umgang mit Männern, bei Zahlen möglichst ratlos zu gucken und eine Art soziale Debität zu kultivieren. Ich leide zwar noch an den Folgeerscheinungen. Aber erfreulicherweise gelang es mir über den Kopf, dann doch noch einen späten Zugang zu gewinnen. Heute lese ich die neuen statistischen Jahrbücher in manchen Absätzen wie Krimis. Vor allem wenn sie verständlich und übersichtlich gemacht sind.“ Diese Worte stammen nicht von mir, obwohl es mir nicht anders erging. Sie stammen von der WDR-Redakteurin *Carmen Thomas*.

Dichter und Philosophen, nicht nur Mathematiker haben sich mit Zahlen befasst. Für *Pythagoras* war die Zahl etwas Umfassendes, Wesenhaftes und keine Ansammlung von Ziffern und schon gar nichts zum Abzählen der Wirklichkeiten. Für *Talleyrand* waren Ziffern genau das Gegenteil, sie waren ein politisches Instrument, dessen Gültigkeiten zu prüfen waren. Insofern sehr modern. Und der Mathematiker *Gauß* stand über all dem. „Der Mangel an mathematischer (statistischer) Bildung gibt sich durch nichts so auffallend zu erkennen, wie durch maßlose Schärfe im Zahlenrechnen.“

Was sagt uns die Zahl 35 752? Wahrscheinlich nichts – oder wir müssten raten. Es ist die Fläche unseres Landes in km<sup>2</sup>. Und nun sprudeln bei uns die Assoziationen: Heidelberg, Bodensee, Schwäbische Alb, Schwarzwald, Trollinger, Kehrwoche, Stuttgarter Oper, KSC, VfB und und und. Bei jedem etwas anderes und doch viel Ähnliches. Bei anderen sprudelt beim Namen „Baden-Württemberg“ und der Zahl 35 752 wenig bis nichts. Eine Zahl ohne be-

schreibende Metadaten ist eine Ansammlung von Ziffern. Mit Metadaten wird sie definierbar und mit Erfahrung manchmal sogar wertvoll.

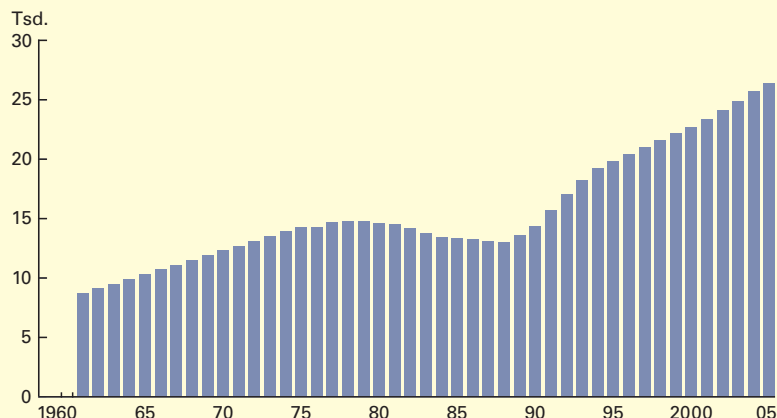
Beim Umgang mit Daten offenbaren wir manchmal manische Züge zum Beispiel beim Thema Aktualität: Vor Jahren erbat eine oberste Landesbehörde die aktuellsten Daten zu „Landesbevölkerung und Landesfläche“. Die Feststellung einer Mitarbeiterin, dass wir seit 1945 keine kriegerischen Auseinandersetzungen in und mit Baden-Württemberg hatten und deshalb keine territorialen Verschiebungen stattfanden, wurde von jenem Ministerialbeamten nur mit Unmut akzeptiert. Sehen wir von Tsunami-Katastrophen, Erdbeben, Pest, Cholera, Börsencrashes oder Maueröffnungen ab, ändern sich Bestände und Strukturen in der Wirklichkeit nur stetig oder regelhaft – auch wenn wir die Regeln häufig nicht kennen. Entscheidungen deshalb zu verschieben, weil die aktuellsten Daten noch nicht vorliegen, mutet oft als Entscheidungsschwäche an.

Oder die Manie des Wachstums: Wachstum ist in unserer Gesellschaft immer gut, macht man uns glauben. Wenn Umsätze oder Gewinne steigen oder die Zahl der Arbeitsplätze zunimmt, wird das als Erfolg betrachtet. Würde *Schaubild 1* die Entwicklung der Beschäftigtenzahl



Dr. rer. pol. Carmina Brenner ist Präsidentin des Statistischen Landesamtes Baden-Württemberg.

S1



Statistisches Landesamt Baden-Württemberg

799 08

einer GmbH seit 1961 darstellen, ist dies eindeutig und somit positiv zu deuten. Was aber, wenn die Grafik die Bevölkerungsentwicklung in Afghanistan darstellen würde? Dann würden wir sehr nachdenklich werden.

Bei der Bewertung von Wachstum oder negativem Wachstum (auch eine akademische Stillblüte) stellen sich immer die Fragen: Für wen ist Wachsen oder Schrumpfen gut und für wen schlecht? Wo ist Wachsen oder Schrumpfen gut und wo schlecht? Warum ist Wachsen immer gut oder Schrumpfen immer schlecht? Nehmen wir als Beispiel die Demografie. Eine Million Einwohner weniger in Baden-Württemberg bedeutet auch weniger Umweltbelastung, weniger Aufwendungen für Humaninfrastruktur, weniger „Flächenverbrauch“, statt öffentlicher Neuinvestitionen nur noch Sanierung usw. Eine Million mehr bedeutet mehr Umweltbelastung, mehr Aufwendungen für Humaninfrastruktur, mehr „Flächenverbrauch“, neue Kindergartengruppen, neue Klassenzimmer. Trotzdem wäre weiteres Bevölkerungswachstum bei uns vorerst kein Problem, in Äthiopien und dem Gazastreifen dagegen schon.

Oder die Sucht nach Genauigkeit. Wie viele Menschen leben auf der Erde? Genau wissen wir das nicht, obwohl bis auf den letzten Einwohner ausgezählte Angaben durch die Welt gehen. Erstsemester könnten die „Weltbevölkerungsuhr“ zitieren: Am 1. Januar 2008 lebten 6 677 319 561 Menschen auf der Erde. Satiriker und Kabarettisten würden die 3 Astronauten des 1. Januar 2008 addieren. Statistiker würden sagen, dass der 1. Wohnsitz der Astronauten – amtlich gesehen – nicht im irdischen All war, sondern dass sie auf der Erde mitgezählt wurden. Die Ermittlung der tatsächlichen Zahl würde Milliarden Euro kosten. Und falsch wäre die Zahl nicht nur im Einser-, sondern im Millionenbereich, weil die statistischen Instrumente nicht weltweit einsetzbar sind. Eine solche, zwangsläufig mit Schätzungen behaftete Zahl und deren Genauigkeit wäre ihr Geld nicht Wert.

Oder die Rankomanie: Statistik sollte Massenerscheinungen erkennen und beschreiben. Zu Massenerscheinungen gehört auch das Normale, das Durchschnittliche. Tatsächlich stürzen wir uns „rankomanisch“ auf kleinste Abweichungen – wie bei den Pisa-Studien. Die Masse in der Mitte wird gerne ignoriert. Normales ist nicht berichtenswert, und „bad news are good news“. Eine minimale Änderung der Punktzahl kann bei Pisa bemerkenswerte Rangverschiebungen erbringen – also beruhigen oder aufregen. Über eines müssen sich die Betroffenen bei derartigen Rankings im Klaren sein: sind die Punktedifferenzen gering,

kann man sehr schnell nach oben klettern oder nach unten wegbrechen – wie diesmal Österreich, das um 5 Ränge hinter Deutschland abfiel, was zu bemerkenswerter Aufregung unserer Nachbarn führte. Die Punktedifferenz zwischen Deutschland und Österreich betrug gerade einmal 3 %. „Gute“ Statistiker würden das unter „Fehlerrange“ subsumieren. Es wäre allerdings auch vermessen, eine analytische 1 %-Genauigkeit in einer international unvergleichbaren Bildungswelt zu fordern.

Oder der taktische Umgang mit Daten: Eine Eigenschaft der Diamanten erreichen Daten nur selten. Diamanten sind optisch isotrop, das heißt sie haben keine Doppelbrechung des Lichtes. Statistische Daten haben die gegenteilige Eigenschaft. „Mit ihnen kann man alles beweisen – auch das Gegenteil“ meinte einst die Wirtschaftsjournalistin *Fides Krause-Brewer*. Literaturkritiker *Marcel Reich-Ranicki* ergänzte: „Mit Statistik kann man alles beweisen, auch die Wahrheit.“ Als Beispiel nehmen wir einen typischen Verlauf eines Start-up-Unternehmens (*Schaubild 2*). Anlaufschwierigkeiten, dynamische Phase, Konsolidierung oder Marktsättigung. Nicht ungewöhnlich, nicht dramatisch. Doch nun wird aus einem erfolgreichen Unternehmen ein scheinbarer „Verlierer“ (*Schaubild 3*).

Mit der 2. Ableitung einer Funktion und einer zwar korrekten, aber nicht sofort erfassbaren Beschreibung eines Sachverhaltes oder einer Grafik lässt sich – ohne eigentlich zu lügen – scheinbar das Gegenteil „beweisen“. Welcher Zeitungsleser oder „Fernseher“ wird sofort erkennen, dass das Wort „Wachstumsdynamik“ die „Veränderung der Veränderung“ und damit die 2. Ableitung ist? Die Basisdaten sind dieselben und sie sind richtig.

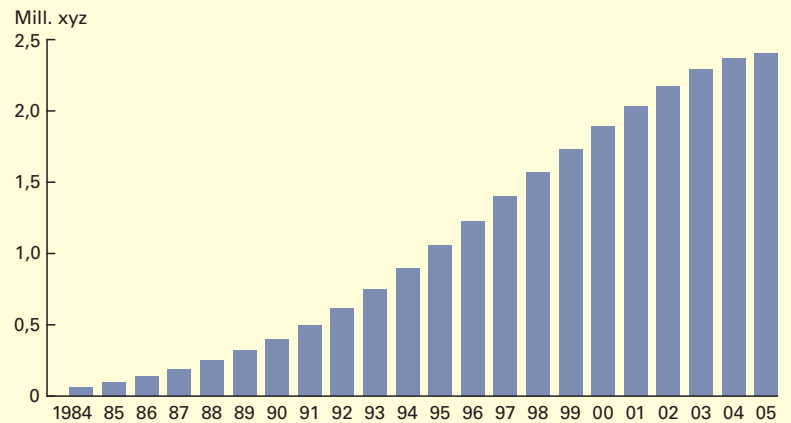
Oder der Wunsch, eine Zählung vollständig auswerten zu wollen: Nehmen wir ein stark vereinfachtes Beispiel aus der Volkszählung 1987. Damals wurde in über 8 500 Gemeinden gezählt. Es gab 20 Merkmale mit über 5 000 Merkmalsausprägungen wie Geschlecht mit männlich, weiblich, Familienstand mit ledig, verheiratet, verwitwet und geschieden, Geburtsjahr 1870, 1871, ..., 1987, jeweils vor dem 27. Mai oder danach, Berufe, Wirtschaftszweige. Wie viele Merkmalsausprägungen damals tatsächlich entstanden, konnte niemand genau sagen. Die Angaben streuen gewaltig. Letztlich spielt das hier aber keine Rolle. Es stellt sich die trivial anmutende Frage wie viele Tabellen beliebiger Größe aus 5 000 Merkmalsausprägungen und aus 8 500 Gemeinden erzeugt werden können. Sie werden überrascht sein. Es sind etwa  $10^{4000}$ . Unser Sonnensystem könnte gerade einmal 10 hoch 43 Blatt Papier verstauen – vielleicht

ein paar mehr oder weniger. Das heißt, jede Erhebung beinhaltet Informationspotenziale, die nie ausgeschöpft werden (können). Es ist die Kunst der Statistiker zu ahnen, welche Informationen von Interesse sein werden. Dabei müssen die Daten vorgehalten werden, mit denen sich schnell Fragen beantworten lassen. Und zwar auch solche, an die die Statistiker zunächst nicht gedacht hatten.

Lassen Sie mich nun in die nahe Zukunft blicken, und zwar auf die kommende Volkszählung – den Zensus im Jahr 2011. Es werden nur diejenigen Fragebögen erhalten, die Eigentümer von Wohnraum sind oder in eine größere Stichprobe fallen. Wir werden etwa 2 Milliarden Daten erfassen. Wozu diese Datenmenge, werden Sie sich fragen? Nicht weil die EU es so vorschreibt, was sie zwar tut, sondern weil grundlegend wichtige Statistiken eine aktuelle Datengrundlage benötigen – insbesondere Statistiken, die als Fortschreibung des nunmehr über 20 Jahre alten Volkszählungsergebnisses laufen, wie die Fortschreibung des Bevölkerungsstands. Menschen melden sich zwar an, wenn sie in einen neuen Wohnort umziehen, aber nicht unbedingt von ihrem bisherigen Wohnort ab. Dieses mangelhafte Meldeverhalten mag in der amtlichen Statistik nicht immer zu Fehlern führen, wenn die Umzüge innerhalb Deutschlands stattfinden, aber häufig doch, wenn Menschen – Deutsche und Ausländer – ins Ausland fortziehen. Wir schätzen derzeit, dass in Deutschland etwa 1 bis 2 Millionen Menschen weniger leben, als die amtliche Statistik meldet. Und etwa 2 bis 3 Millionen haben ihren Lebensmittelpunkt nicht dort, wo sie ihn einst angegeben haben. Dies hat zu sich kumulierenden Fehlern in der Bevölkerungsfortschreibung geführt. Welche Auswirkungen die zu erwartenden Ergebnisse auf die Rentenberechnungsmodelle, die Abgrenzung der Wahlkreise, die Ausweisungen von Bauland, die Raumplanung, die Länderfinanzausgleiche und, und, und ... haben werden, müssen wir abwarten. Es wird spannend, dessen bin ich gewiss, zumal sogar ein Paradigmenwechsel in der Erhebungsmethodik „von Zählungsdaten zu Registerdaten“ erfolgen wird.

Unter Paradigmen verstehen wir gedankliche Konzepte, die in einer intellektuellen Welt von allen oder den meisten als gültig akzeptiert werden. Bei einem Paradigmenwechsel wird in der idealtypischen Annahme das eine Konzept durch ein anderes disjunktes – Konzept ausgetauscht: „die Erde ist eine Scheibe“ > Wechsel > „die Erde ist eine Kugel“. Wie ist das aber in den Sozialwissenschaften und speziell in der amtlichen Statistik? Es wird auch dort

**S2** Umsatz der EXAMPLE Ltd. seit 1984

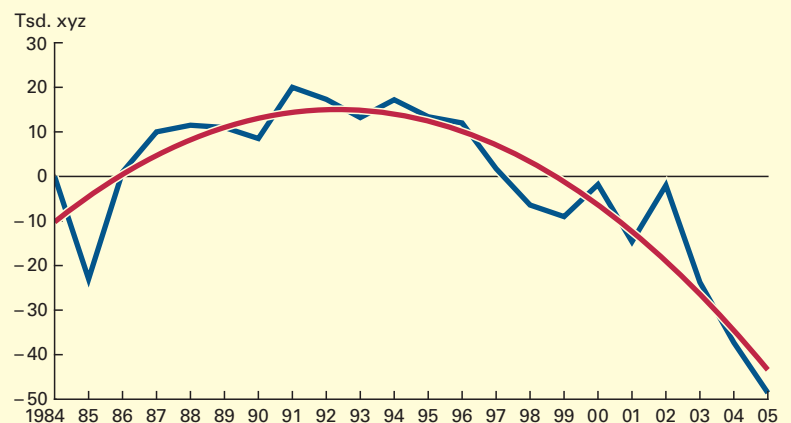


Statistisches Landesamt Baden-Württemberg

800 08

viel von Paradigmenwechseln gesprochen. „Weg von Primärerhebungen hin zur Sekundärnutzung von Registern.“ Beide Formen der Informationsbeschaffung haben aber immer nebeneinander existiert. Es wurde gezählt (vgl. Lukas-Evangelium) und Listen oder Register haben alle hoch entwickelten und verwaltungsorientierten Staaten angelegt. Einen stetigen aber dennoch bemerkbaren Wandel erfahren wir derzeit dennoch. „Der gläserne Mensch“ war in Zeiten der Volkszählungsdiskussion in den 80er-Jahren ein absolutes Meteteckel. Heute wird er sogar individuell und freiwillig selbst produziert, wie die vielen Blogs und Foren im Internet belegen. Es würde mich sehr wundern, wenn beim Zensus 2011, bei dem die Nutzung von Registern im Vordergrund steht, ein ähnlicher Aufruhr entstünde wie 1982 bis 1987.

**S3** Wachstumsdynamik bei Umsätzen der EXAMPLE Ltd. im Keller



Statistisches Landesamt Baden-Württemberg

801 08

Dieser scheinbare Paradigmenwechsel führt uns in die statistischen Geheimhaltungswelten. Lassen Sie mich diese etwas ausleuchten. In einigen Gebieten Ostafrikas dürfen Menschen nicht abgezählt werden, da böse Geister den Abgezählten bedrohen können. In einigen Schweizer Kantonen sind die Steuererklärungen einsehbar. In Deutschland werden die Daten dann geheim gehalten, wenn einzelne Personen oder Betriebe reidentifizierbar wären. An die USA dürfen Individualdaten von Fluggästen gemeldet werden. Die Tatsache, dass es in einem Ort nur einen Menschen mit Hochschulabschluss gibt (und sei es der Pfarrer), ist in Deutschland geheim, wenn diese Person leicht zu identifizieren wäre. Die Geheimhaltung statistischer Einzeldaten scheint kein definierbares und weltweites Naturrecht zu sein.

Aber es gibt auf der anderen Seite gewichtige Gründe für unsere amtliche Statistik, in den gebotenen Fällen die Geheimhaltung strikt zu praktizieren. Die allermeisten Erhebungen sind mit einer Auskunftspflicht der Befragten belegt. Betriebe und Unternehmen, Privatpersonen oder andere Befragte müssen Auskunft an die amtliche Statistik geben. Schon allein dies verpflichtet uns, darauf zu achten, dass durch unsere veröffentlichten Informationen niemand – zumindest mit leichtem Aufwand – als Person oder Betrieb wieder erkennbar wird.

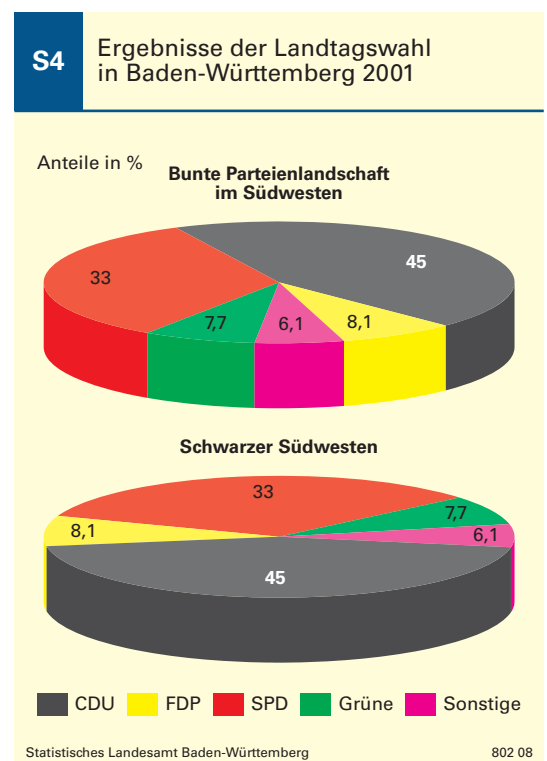
Zu welchen Blüten und Befürchtungen allerdings die Geheimhaltungsdiskussion auch führen kann, möchte ich Ihnen hier an einem – zugegebenermaßen abstrusen – Beispiel zeigen: Einer hiesigen Tageszeitung fiel auf dem damaligen Volkszählungsbogen ein kleines etwa 6x6 mm großes Signierfeld auf, das nur mit „v/n“ beschriftet war. Es war den Journalisten klar, trotz aller Beteuerungen der Statistiker sollten in dieses Feld der v(ornname) und der n(achname) eines jeden Bürgers eingetragen werden. Der Weg zu Bspitzelung der Nation war damit offen. Um den Namen „Jeremias Gotthilf Hintertalhofer“ in einem 6x6 mm-Feld zu signieren, hätte es der Nanotechnologie bedurft. Was hatte es mit dem Feld nun auf sich? V war zu signieren, wenn man vor dem VZ-Stichtag Geburtstag hatte, N war zu signieren, wenn man nach dem VZ-Stichtag Geburtstag hatte. Diese Angaben waren für die korrekte Berechnung des Alters erforderlich. Heute gibt man freiwillig ohne Bedenken – beispielsweise durch Nutzung einer Payback-Karte – seine Essgewohnheiten preis, welche Farbe der Unterwäsche man bevorzugt, ob man ein Schnäppchenjäger-Talent besitzt und vieles, vieles mehr.

Die amtliche Statistik hat kaum Einblick in die Unternehmen, da sie von Interna, wie Auftragslage, Investitionen, Warenbeständen erst so spät erfährt, dass die Daten längst überholt sind, wenn sie veröffentlicht werden. Trotzdem müssen Zahlen in detaillierten Tabellen so lange zusammengefasst oder sogar herausgenommen werden, bis Einzelunternehmen, die durch ihre schiere Größe die Tabelleninhalte (etwa Umsätze, Auftragsbestände und -einzüge) bestimmen, dort nicht im Einzelnen erkennbar werden. Beispiele aus der Wirklichkeit:

- „Volkswagen“ in Wolfsburg
- „SAP“ in Walldorf (Rhein-Neckar-Kreis)
- Ehem. „BASF“ in Ludwigshafen
- „Würth“ in Künzelsau (Hohenlohekreis)
- „Schlecker“ in Ehingen (Alb-Donau-Kreis)

In kleinen Branchen, bei denen sich womöglich pro Landkreis nur 1 oder 2 Unternehmen finden, dürfen wir nicht veröffentlichen, sondern müssen so lange aggregieren, bis kein Rückschluss mehr möglich ist. Also funktioniert die Geheimhaltung ähnlich wie Sudoku, nur andersherum.

Lassen Sie zum Abschluss noch den Spannungsbogen zwischen Veredelung und Verelendung aufzeichnen. Beide schräg gestellten Tortendiagramme in *Schaubild 4* spiegeln die Ergebnisse der Landtagswahl 2001 in Baden-Württemberg wider, das obere Diagramm die bunte Parteienlandschaft im Südwesten, das untere den schwarzen Südwesten.



... wer würde auf den ersten Blick erkennen, dass er bei dieser Parteiendarstellung über den Tisch gezogen wird. Die moderne Grafiksoftware macht es uns so einfach, dass wir manchmal gar nicht merken, was wir tun. Und wenn es Absicht sein sollte, können wir immer noch *Churchill* zitieren: You must look at the facts because they look at you. „Ich glaube nur der Statistik, die ich selbst gefälscht habe.“ stammt übrigens nicht von *Churchill*, ist in Großbritannien völlig unbekannt und muss dem Reichspropagandaministerium zugeordnet werden. Deshalb nut-

zen Sie am Besten [www.statistik-bw.de](http://www.statistik-bw.de), da sind Sie auf der sicheren Seite. Wenn Sie zum Beispiel wissen möchten, wo überall in Baden-Württemberg Spargel angebaut wird. Das wissen wir in Flächenerträgen. Allerdings wissen wir nicht, wie viele Professoren an wie vielen Tagen im Jahr welche Grammzahl Spargel essen. Wenn man es genau bedenkt, braucht man das auch nicht zu wissen. Hauptsache die Professoren, die Spargel lieben, sind rechtzeitig dort, wo es ihn fertig gekocht gibt und noch wichtiger: der Spargel reicht heute für uns alle. ■

## kurz notiert ...

### Der Mikrozensus, die größte amtliche Haushaltsbefragung in Deutschland, startet wieder im Januar 2009

Am 5. Januar 2009 startet in Baden-Württemberg, wie auch in ganz Deutschland, die Befragung zum Mikrozensus 2009. Der Mikrozensus ist eine gesetzlich angeordnete Befragung über die Bevölkerung und den Arbeitsmarkt, die seit 1957 jedes Jahr bei 1 % aller Haushalte in Deutschland durchgeführt wird. In Baden-Württemberg werden jährlich rund 45 000 Haushalte durch das Statistische Landesamt befragt. Zusammen mit dem Mikrozensus wird in allen auskunftspflichtigen Haushalten auch die EU-Arbeitskräftestichprobe durchgeführt.

Die Stichprobenauswahl des Mikrozensus ist im Mikrozensusgesetz vorgeschrieben. Danach werden bei der Stichprobenziehung Gebäude ausgewählt. Die Haushalte, die in diesen durch ein mathematisches Zufallsverfahren ausgewählten Gebäuden wohnen, sind auskunftspflichtig. Die vom Gesetzgeber angeordnete Auskunftspflicht dient dazu, dass mit dem Mikrozensus zuverlässige und aktuelle statistische Informationen bereitgestellt werden können.

Der Stichprobenumfang von etwa 45 000 Haushalten wird gleichmäßig auf alle Monate und Wochen des Jahres verteilt. Somit werden in Baden-Württemberg pro Woche rund 865 Haushalte von den Interviewern des Statistischen Landesamtes befragt. Die Angaben beziehen sich dann jeweils auf die Woche vor dem Interview. Die Vorteile dieses Konzeptes liegen in der Aktualität und Qualität der Ergebnisse, die als Quartals- und als Jahresdurchschnittsergebnis vorliegen werden und sowohl saisonale Spitzen als auch flexible Arbeitsverhältnisse abbilden können.

Datenschutz und Geheimhaltung sind, wie bei allen Erhebungen der amtlichen Statistik, um-

fassend gewährleistet. Die Interviewerinnen und Interviewer, die die Mikrozensusbefragung durchführen, sind zur strikten Verschwiegenheit verpflichtet. Sie kündigen sich einige Tage vor ihrem Besuch *schriftlich* bei den Haushalten an und übergeben mit dieser Ankündigung zudem auch Informationsmaterial über die Erhebung. Die Interviewer weisen sich mit einem Interviewerausweis des Statistischen Landesamtes aus. Die Befragung wird mit einem Laptop durchgeführt.

Die Präsidentin des Statistischen Landesamtes, Dr. Carmina Brenner, bittet alle auskunftspflichtigen Haushalte um Unterstützung: „Um repräsentative Ergebnisse zu gewinnen, ist es notwendig, dass alle in die Erhebung einbezogenen Haushalte die Fragen des Mikrozensus beantworten. Die Auskünfte von älteren Personen oder Rentnern sind genauso wichtig wie die Angaben von Angestellten, Selbstständigen, Studenten oder Erwerbslosen.“ Um qualitativ zuverlässige Ergebnisse zu erhalten, hat der Gesetzgeber daher die meisten Fragen mit einer Auskunftspflicht belegt. Das Statistische Landesamt bittet jedoch, auch die freiwilligen Fragen zu beantworten.

Die Daten des Mikrozensus bilden für Politik, Wirtschaft, Verwaltung, Wissenschaft, Presse und nicht zuletzt für interessierte Bürgerinnen und Bürger eine unverzichtbare und aktuelle Informationsquelle über die wirtschaftliche und soziale Lage der Bevölkerung, der Familien und der Haushalte, den Arbeitsmarkt, die Berufsstruktur und die Ausbildung. Meldungen wie zum Beispiel „Erwerbsbeteiligung in Baden-Württemberg im europäischen Vergleich überdurchschnittlich hoch“, „Jede dritte Familie in Baden-Württemberg ist eine Migrantenfamilie“, „Immer mehr Baden-Württembeger leben in den eigenen vier Wänden“ oder „Berufliche Qualifikation: Frauen holen auf“ basieren auf Ergebnissen des Mikrozensus. ■