

WDR

Fernsehen



Quarks&Co

Westdeutscher Rundfunk Köln

Appellhofplatz 1
50667 Köln

Tel.: (0221) 2 20-36 82
Fax: (0221) 2 20-86 76

E-mail: quarks@wdr.de

Mit Zahlen lügen

www.quarks.de

Script zur wdr-Sendereihe *Quarks & Co*

Quarks&Co



Inhalt Inhalt

- 4 Die Intelligenz der Masse

- 5 Früher war alles billiger! Die gefühlte Statistik

- 7 Das Jobangebot

- 10 Tödliche Zahlen: Medizinstatistiken

- 13 Was macht eigentlich das Statistische Bundesamt?

- 16 Werden die Menschen wirklich immer älter?

- 18 Eine Frau gegen die Statistik – der Fall Sally Clark

- 20 Auf Verbrecherjagd mit Benford

- 22 Der Umfrage-Test

- 24 Die schlechtesten Grafiken der Welt

Text: Axel Bach, Reinhard Brüning, Katrin Krieft, Hilmar Liebsch, Martin Rosenberg;
Kooperation: Der Einleitungstext entstand in Kooperation mit dem Lehrstuhl für
Wissenschaftsjournalismus in Dortmund. Autorin: Jennifer Dacqué; **Redaktion:**
Tilman Wolff; **Copyright:** wdr, Oktober 2006; **Gestaltung:** Designbureau Kremer &
Mahler, Köln

Bildnachweis: alle Bilder Freeze wdr 2006 **außer** S. 6: Nature November 2006

Mit Zahlen Mit Zahlen lügen lügen

■ Bereits vor dem Frühstück geht es los: Das Etikett des Shampoos verspricht 70 Prozent weniger Haarausfall, die Zahnpasta dreimal weißere Zähne als fünf Vergleichsprodukte und im Radio verkündet der Sprecher die neuen Arbeitslosenzahlen. Statistiken begleiten uns von morgens bis abends, denn nahezu alles wird heute in Zahlenreihen erfasst, ausgewertet und verglichen.

■ Glauben wir Zahlen eher als Worten? Eine Aussage gestützt durch eine Statistik klingt schließlich gleich viel seriöser. Doch ist sie das auch immer?

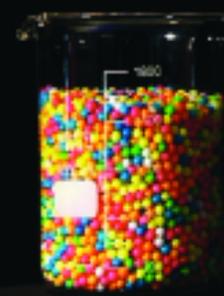
„Ich glaube nur der Statistik, die ich selbst gefälscht habe.“ Diesen Satz hängte die Nazi-propaganda einst dem britischen Premierminister Winston Churchill an, um den Kriegsgegner unglaublich zu machen. Dabei ist der Satz an sich richtig: Mit kleinen Tricks lässt sich fast jede Statistik so frisieren, dass sie praktisch jede Aussage untermauert. Beispiele gibt es genug. Einige haben wir hier für Sie zusammengestellt.

Viel Spaß beim Lesen wünscht das Quarks-Team!

■ Weitere Informationen, Link- und Lesetipps finden Sie unter: www.quarks.de



Auf die Intelligenz der Masse stieß vor über hundert Jahren der englische Gelehrte Francis Galton



Im Internet sollten die Quarks & Co Zuschauer die Anzahl der Liebesperlen in diesem Becherglas schätzen

Die Intelligenz der Massen

Wenn viele Leute gemeinsam auf das richtige Ergebnis kommen, dann nennen die Wissenschaftler das die *Intelligenz der Masse*. Darauf gestoßen ist vor 100 Jahren der englische Gelehrte Francis Galton und zwar als er eigentlich das Gegenteil beweisen wollte – nämlich, dass die Masse dumm ist.

durfte man auf das Gewicht eines Ochsen wetten. Wer am nächsten dran lag, konnte gewinnen.

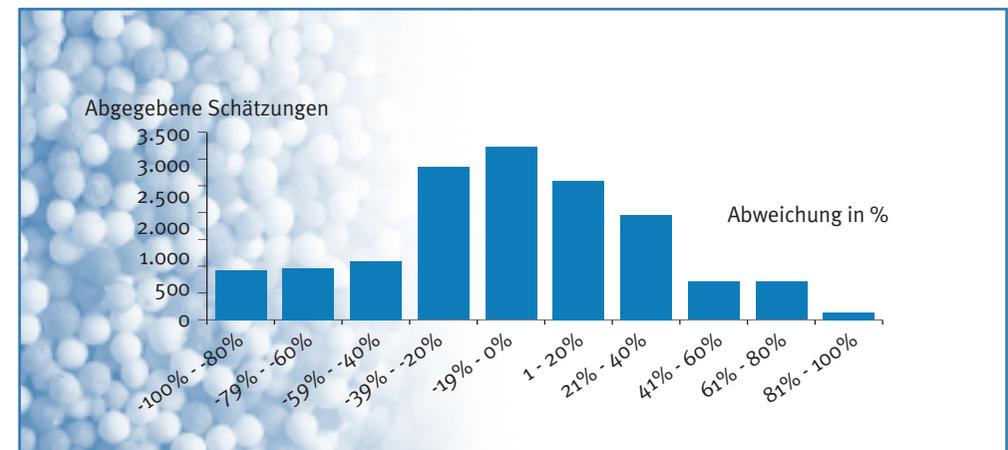
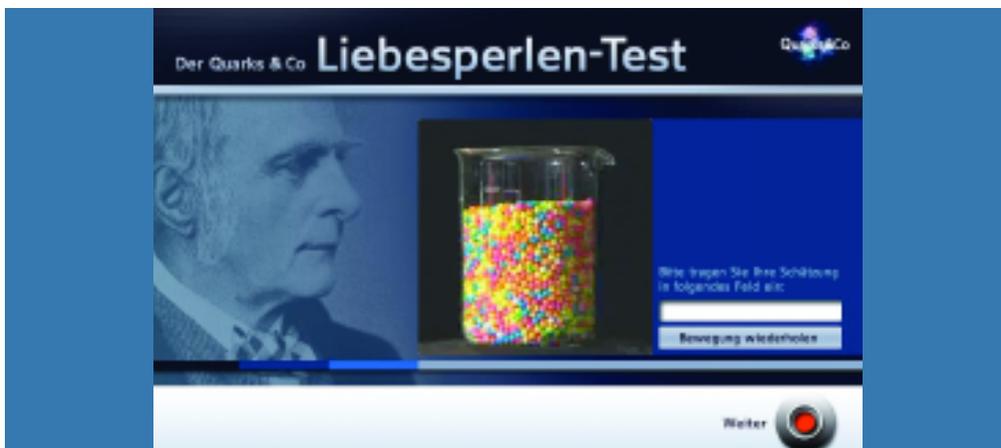
Nachdem Galton die Schätzungen ausgewertet hatte, muss er ziemlich verblüfft gewesen sein: Der Mittelwert der 787 Einzelschätzungen wich nur um ein Pfund vom tatsächlichen Gewicht des Ochsen ab. Und – was noch erstaunlicher war – kein Einzel-Tipp (auch nicht der eines Experten) war genauer als der Mittelwert der großen Masse.

1906 besuchte er die westenglische Nutztiermesse in der Nähe von Plymouth. Dort gab es einen Schätzwettbewerb: Für sechs Pence

Ob das Experiment, das Galton 1907 in der Zeitschrift Nature beschrieben hat, auch mit anderen Schätzaufgaben funktioniert, haben wir mit Ihnen zusammen getestet und waren selbst überrascht über das Ergebnis.

Unter www.quarks.de zeigten wir einen Behälter, in dem genau 5.780 Liebesperlen enthalten waren. Wieviele es waren, sollten Sie schätzen. Hier das Ergebnis Ihrer Schätzungen:

Anzahl Liebesperlen:	5.780
Durchschnittlicher Schätzwert:	5.714
Abweichung:	-1,14 %
Anzahl der Schätzungen:	15.885
Genauere Treffer:	0,47 %





Der Quarks & Co Liebesperlentest war der Wissenschafts-Zeitschrift Nature eine Veröffentlichung wert



Früher war alles billiger!

Die gefühlte Statistik

■ Quarks & Co bei Nature

Dieses Ergebnis hat auch die Herausgeber der renommierten Wissenschafts-Zeitschrift **Nature** so überrascht, daß sie das Quarks-Liebesperlen-Experiment in der Ausgabe vom 2. November 2006 veröffentlichten.

Sweet success shows you can count on the public

Are the masses stupid? A hundred years ago Francis Galton, a half-cousin of Charles Darwin, analysed the results of a guess-the-weight-of-an-ox competition at a country fair. The average of the guesses from 787 participants was virtually spot-on to the real weight of the ox, Galton reported (see *Nature* **75**, 450; 1907). The findings shook his belief in eugenics, a term he himself had coined.



Last month, Ranga Yogeshwar, host of Germany's popular Quarks & Co science television show, repeated the experiment — but with a laboratory beaker filled with sweets. The average guess of the 16,000 viewers was again astonishingly close to the correct number of 5,780; the average was 5,714, and nearly one in 200 hit the right number exactly. "The masses are intelligent," says Yogeshwar. "We can use them for all sorts of science TV-Internet experiments."

■ Preisexplosion, und der Euro ist schuld

Seit 2002 der Euro eingeführt wurde, wird alles teurer – oder? Die Debatte darüber, ob der Euro ein *Teuro* ist, flammte sofort nach der Umstellung auf und ließ sich auch von den Zahlen des Statistischen Bundesamtes nicht aus der Welt schaffen. Dessen abschließende Studie zeigte: in den letzten zweieinhalb Jahren der D-Mark war die Inflation sogar ein Prozent höher als in den ersten beiden Euro-Jahren – die Preissteigerung ging mit dem Euro also wieder etwas zurück. Diese Zahlen sind unanfechtbar. Aber bei den Menschen blieb das *Teuro*-Gefühl bestehen.

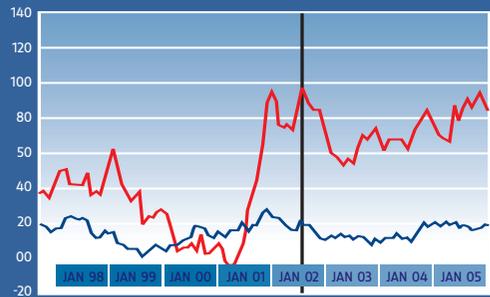
zialist für die Statistik von Preisen, doch er konnte seiner Mutter nicht gleich antworten – aber er nahm ihre Empörung ernst und kam ins Grübeln. Beim Nachdenken über die subjektive Empfindung der Kunden erinnerte er sich an psychologische Entscheidungstheorien. Dort taucht ein Faktor auf, der in Statistiken meist unberücksichtigt bleibt: nämlich die Häufigkeit, mit der etwas gekauft wird. Ein Auto kauft man zum Beispiel eher selten. Deshalb nimmt der Verbraucher hier einen Preisanstieg nur schwach wahr. Lebensmittel oder Kleidung kaufen die Menschen dagegen häufig. Und wenn die teurer werden, empfinden sie das deutlich.

■ Die Mutter gab den Anstoß

Ein Deutscher, der an der Universität von Fribourg in der Schweiz Statistik lehrte, hätte von alledem nichts mitbekommen, wenn er nicht von Zeit zu Zeit in seine alte Heimat nach München gefahren wäre. Dort war seine Mutter empört über die Ergebnisse des Statistischen Bundesamtes: Das sei Schrott, was seine Kollegen da ausrechnen! Sie forderte den Sohn beim Frühstück auf, sich auf dem Tisch umzuschauen: Alles sei teurer geworden! Prof. Hans Wolfgang Brachinger ist zwar Spe-

■ Schnäppchen vergisst man schnell

Hans Wolfgang Brachinger schlug weiter bei den Psychologen nach und entdeckte noch etwas: Jeder Kauf läuft offenbar nach einem starren Schema ab. Der Kunde erinnert sich an einen sogenannten Referenzpreis für das Produkt. Mit dem Preis, den er in Erinnerung hat, vergleicht er den aktuellen Kaufpreis. Ist der neue Preis günstiger, wertet das der Kunde als Gewinn. Ist der erinnerte Referenzpreis dagegen niedriger als der aktuelle Kaufpreis, wird das als Verlust bewertet. Darüber hinaus haben die Psychologen herausgefunden,



Die von Hans Wolfgang Brachinger berechnete gefühlte Inflation (rot) liegt bis heute etwa fünf Prozent über der vom Statistischen Bundesamt berechneten Inflation (blau)

Die gefühlte Statistik

dass hohe Preise doppelt so stark wahrgenommen werden wie die Preise von Schnäppchen – und entsprechend stark und negativ in Erinnerung bleiben.

Die gefühlte Inflation wird gemessen

Mit diesen Erkenntnissen entwickelte Hans Wolfgang Brachinger einen Preisindex für das, was er *gefühlte Inflation* nennt. Dafür arbeitete er eng mit dem Statistischen Bundesamt zusammen. Nach einigen Monaten konnte er seinen neuen, gefühlten Preisindex zum ersten Mal berechnen. Und der Professor war zufrieden, denn die Kurve zeigte tatsächlich einen drastischen Anstieg direkt vor der Währungsumstellung! Kurz bevor der Euro kam, hatten die Preise also stark angezogen – der Unmut der Menschen ging also tatsächlich auf empirische Ursachen zurück. Das schien plausibel. Nur eines war merkwürdig: Die Kurve normalisierte sich wieder, aber das Gefühl der erhöhten Preise, das war ja geblieben! Hans Wolfgang Brachinger war fasziniert von diesem Phänomen – warum halten Menschen so an ihrem schlechten Eindruck fest? Wieder gab die Mutter den Hinweis auf das Problem, denn beim gemeinsamen Einkaufen war Brachinger etwas an seiner Mutter aufgefallen: Sie rechnete alle Preise immer noch auf D-Mark um

und verglich sie mit den alten Mark-Preisen von damals, vor 2002. Sie hatte also einen längst veralteten Referenzpreis, der seit der Währungsumstellung nicht gestiegen ist. Dieser Effekt hat für die gefühlte Inflation dramatische Folgen: Sie steigt mehr und mehr und ist heute bei etwa 16 Prozent angelangt.

Wer nicht umrechnet, ist realistischer

Einige rechnen immer noch um, für die meisten wird der letzte D-Mark-Preis aber nach und nach unwichtiger, sie vergleichen immer mehr mit den Euro-Preisen der letzten Zeit. Das hat der Statistiker auch in seine Zahlen einbezogen. Seine Ergebnisse besagen, dass das Teuerungsgefühl heute bei etwa sieben Prozent im Vergleich zu real vorhandenen zwei Prozent liegt. Also ist die gefühlte Inflation und Teuerung immer noch deutlich höher als die reale. Das Teuro-Gefühl ist durch Hans Wolfgang Brachingers Arbeiten jetzt wissenschaftlich erklärt. Er kann zeigen, dass die Unzufriedenheit der Menschen auf eine empirische Grundlage zurückgeht. Seine Ergebnisse stehen im Einklang zu denen des Statistischen Bundesamtes, denn gefühlte und wirtschaftlich berechnete Inflation sind eben nicht dasselbe.

Das Jobangebot

Das Jobangebot

Blanke Zahlen und Gefühl sind zwei völlig verschiedene Welten und häufig führt uns unser Gefühl aufs Glatteis, selbst wenn es um lebenswichtige Zahlen geht, wie beispielsweise das Gehalt. Schließlich will jeder von uns doch gut bezahlt werden für seine Arbeit. Nehmen Sie an, Ihnen wird ein Jobangebot gemacht: Sie bekommen ein Jahresgehalt von 10.000,- Euro. Außerdem bietet Ihnen der Arbeitgeber eine regelmäßige Gehaltssteigerung an. Sie dürfen wählen ob Sie entweder:

- **Variante A** nach jedem Jahr, das Sie gearbeitet haben, 1000,- Euro mehr bekommen, oder
- **Variante B** nach jedem halben Jahr, das Sie gearbeitet haben, 250,- Euro mehr bekommen.

Also halbjährlich 250,- Euro oder jährlich 1.000,- Euro Gehaltssteigerung.

Das Angebot ist vom Arbeitgeber clever gewählt, denn die meisten Menschen würden sich für die Variante A entscheiden, für die 1.000,- Euro Gehaltssteigerung auf das Jahresgehalt. Das klingt nach mehr – doch ein Geschäft machen bei dieser Variante nicht Sie, sondern das Geschäft macht der Arbeitgeber. Wenn man die beiden Angebote nämlich einmal ausrechnet und vergleicht, ergibt sich, daß Sie mit Variante B, also den 250,- Euro Gehaltssteigerung aufs Halbjahresgehalt deutlich besser abschneiden.

Wenn Sie sich für Variante B entschieden hätten, wäre jedes Jahr mehr in Ihrem Geldbeutel als bei

Jahresgehalt Grundgehalt 10.000,- Euro plus die Gehaltssteigerung:

	1. Jahr	2. Jahr	3. Jahr	4. Jahr	5. Jahr	6. Jahr
Variante A: 1.000,- Euro mehr nach jedem Jahr	€ 10.000,-	€ 11.000,-	€ 12.000,-	€ 13.000,-	€ 14.000,-	€ 15.000,-
Variante B: 250,- Euro mehr nach jedem Halbjahr	€ 10.250,-	€ 11.250,-	€ 12.250,-	€ 13.250,-	€ 14.250,-	€ 15.250,-

Das Jobangebot

Variante A. Wie kommt das? Dafür muß man ein bisschen rechnen und die richtigen Grundlagen dafür wählen:

Wenn Sie mal nicht aufs ganze Jahr rechnen, sondern wie es bei Variante B ja angeboten wird aufs halbe Jahr, dann sieht die Rechnung so aus: Im ersten Jahr gibt es das Jahresgehalt von 10.000,- Euro, also 5.000,- pro Halbjahr:

Variante A:		
1000,- Euro mehr nach jedem Jahr	1. Halbjahr	2. Halbjahr
aufs Halbjahr gerechnet	€ 5.000,-	€ 5.000,-

nach einem Jahr die erste Gehaltssteigerung von € 1.000,-, insgesamt bekommen Sie jetzt € 11.000,-, also 5.500,- pro Halbjahr:

Variante A: 1000,- Euro mehr		
nach jedem Jahr	3. Halbjahr	4. Halbjahr
aufs Halbjahr gerechnet	€ 5.500,-	€ 5.500,-

nach dem zweiten Jahr wieder € 1.000,- mehr, also insgesamt € 12.000,- und aufs Halbjahr gerechnet 6000,- pro Halbjahr:

Variante A: 1000,- Euro mehr		
nach jedem Jahr	5. Halbjahr	6. Halbjahr
aufs Halbjahr gerechnet	€ 6.000,-	€ 6.000,-

Wenn Sie dieselbe Rechnung für Variante B aufmachen, dann sehen Sie schnell die Abweichungen. Hier ist ja eine Gehaltssteigerung bereits nach einem halben Jahr versprochen und das sieht dann so aus:

Das Grundgehalt sind die € 10.000,-, also ebenfalls 5000,- pro Halbjahr. Aber bereits nach einem halben Jahr bekommen Sie bei dieser Variante eine Steigerung von € 250,-

Variante B: 250,- Euro mehr		
nach jedem Halbjahr	1. Halbjahr	2. Halbjahr
aufs Halbjahr gerechnet	€ 5.000,-	€ 5.250,-

So bekommen Sie bereits im ersten Jahr bei dieser Variante € 250,- mehr, also insgesamt € 10.250,-. Im darauffolgenden Halbjahr bekommen Sie jetzt schon wieder eine Steigerung von € 250,- und nach einem weiteren Halbjahr wieder 250,- mehr:

Variante B: 250,- Euro mehr		
nach jedem Jahr	3. Halbjahr	4. Halbjahr
aufs Halbjahr gerechnet	€ 5.500,-	€ 5.750,-

Genauso geht es dann weiter:

Variante B: 250,- Euro mehr		
nach jedem Jahr	5. Halbjahr	6. Halbjahr
aufs Halbjahr gerechnet	€ 6.000,-	€ 6.250,-

Und wenn Sie beide Varianten in dieser detaillierten Rechnung vergleichen, sehen Sie sofort, dass Variante B die bessere ist – für Sie:

Jahresgehalt Grundgehalt 10.000,- Euro plus die Gehaltssteigerung:						
	1. Halbjahr	2. Halbjahr	3. Halbjahr	4. Halbjahr	5. Halbjahr	6. Halbjahr
Variante A: 1000,- Euro mehr nach jedem Jahr aufs Halbjahr gerechnet	€ 5.000,-	€ 5.000,-	€ 5.500,-	€ 5.500,-	€ 6.000,-	€ 6.000,-
Variante B: 250,- Euro mehr nach jedem Halbjahr aufs Halbjahr gerechnet	€ 5.000,-	€ 5.250,-	€ 5.500,-	€ 5.750,-	€ 6.000,-	€ 6.250,-
Ihr Gewinn bei Variante B		+ € 250,-		+ € 250,-		+ € 250,-

Unser Fazit: bei entscheidenden Zahlen, wie dem Gehalt, aber auch bei Krediten und allen anderen Geschäften mit Banken und Versicherungen, beim Preisvergleich und besonders, wenn es ums Geld geht, sollte man sich die Zeit nehmen und mit Zettel, Bleistift und Taschenrechner nachrechnen. Denn: nicht die Zahlen lügen, sondern wir fallen auf uns selbst herein, wenn wir die Zahlen falsch interpretieren...



Links:
In den Universitätsbibliotheken finden sich mehr fehlerhafte Studien als man denken würde

Mitte:
Wenn der Arzt in diesen Röntgenbildern Brustkrebs findet, hilft das nur einer Frau von 1.000

Rechts:
Manche Fehler in Medizinstatistiken gehen regelrecht ans Herz



Tödliche Zahlen: Medizinstatistiken

Tödliche Zahlen:

Medizinstatistiken

■ Mit Zahlen jonglieren – Beispiel Brustkrebs

Jedes Jahr erscheinen rund zwei Millionen Artikel in den rund 10.000 medizinischen Fachzeitschriften. Hier geht es um die Wirksamkeit eines neuen Medikaments, dort um den Heilerfolg einer Behandlungsmethode. Fehler in solchen Studien können verheerende Folgen haben. Und sie sind gar nicht so selten, wie man vielleicht denken würde. So wurde in den 1990er Jahren in diversen Studien untersucht, wie sinnvoll Vorsorgeuntersuchungen zur Früherkennung von Brustkrebs sind. Dabei wird bei allen Frauen ab 50 regelmäßig mit Röntgenstrahlen nach Hinweisen auf Brustkrebs gesucht, das so genannte Mammografie-Screening. Eine schwedische Forschergruppe fand heraus, dass durch eine solche Reihenuntersuchung das Brustkrebsrisiko um gut 30 Prozent sinkt. Das ist zwar richtig – wenn man sich allerdings die absoluten Zahlen ansieht, sind die Ergebnisse nicht ganz so beeindruckend: Drei Frauen von 1.000 sterben an Brustkrebs, wenn keine Reihenuntersuchung gemacht wird. Mit der Reihenuntersuchung sind es nur zwei. Das bedeutet: Nur eine Frau von 1.000 überlebt, weil sie an der Untersuchung teilgenommen hat. Eindrucksvoller ist natürlich die Risikoberechnung mit 30

Prozent geretteten Patientinnen, obwohl auch die zweite Zahl stimmt – nur eine von tausend profitiert von der Voruntersuchung. Daher beziehen sich Aufklärungsbroschüren und Internetseiten, die das Brustkrebs-Screening propagieren, fast immer auf die 30 Prozent. Auf der anderen Seite aber steht die Belastung mit Röntgenstrahlung, die letztlich ebenfalls Krebs hervorrufen kann. Ein Risiko, das allein die Frauen tragen müssen und in dieser Statistik nicht vorkommt.

■ Falsches Studienziel bei Herzrhythmusstörungen

Dass ein neues Medikament durch Nachlässigkeiten tödliche Folgen haben kann, zeigt ein Beispiel aus den 1980er Jahren. Der in den 1970er Jahren entwickelte Wirkstoff Flecainid zur Behandlung von Herzrhythmusstörungen kam nach erfolgreichen Tests auf den Markt. Eine große Studie hatte bewiesen, dass eine bestimmte Art von Rhythmusstörung seltener auftrat, für die das Mittel entwickelt worden war. Ein folgenschwerer Fehler: Es stellte sich nämlich heraus, dass Flecainid in einigen Fällen eine ganz andere Art von Rhythmusstörungen überhaupt erst auslös-

te. Mehrere Patienten starben daran. Von Beginn an hätte man in der Studie eine andere Frage stellen müssen, nämlich die, ob Flecainid bei den Herzpatienten allgemein die Sterblichkeit senken kann. Nur dann wäre aufgefallen, dass das Medikament tödliche Nebenwirkungen hat.

Solche Nachlässigkeiten geschehen in medizinischen Studien häufig. Insbesondere um die Rechenkünste der Mediziner scheint es nicht gut bestellt zu sein. Eine Untersuchung zeigte, dass es in medizinischen Studien von Rechenfehlern nur so wimmelt: Zahlen werden falsch gerundet (aus 0,92 wird so schnell eine 1 statt der 0,90. Oder Ergebnisse werden durch simple Tippfehler um eine ganze Kommastelle verändert (etwa aus 0,0014 wird 0,014).

■ Nur die Hälfte der Zahlen veröffentlicht

Auf einem Auge blind waren die britischen Forscher, die eine neue Bestrahlungsmethode gegen Kopf-Hals-Tumoren testeten. In den ersten 40 Monaten des Studienverlaufs war die neue Methode tatsächlich besser als die alte. Danach verlor sich der Effekt. Doch die Forscher

veröffentlichten nur die Daten, die den Erfolg belegten – kein Einzelfall bei medizinischen Studien. Rund zwei Drittel aller Ergebnisse über Nebenwirkungen werden nur unvollständig veröffentlicht. Und nicht immer scheint Nachlässigkeit der Grund für die Fehler zu sein: 80 Prozent der industriefinanzierten Studien kommen zu positiven Ergebnissen, was die Wirksamkeit neuer Medikamente oder Verfahren angeht. Bei den unabhängigen Studien sind es nur 50 Prozent.

■ Nebenwirkungen verschwiegen

Auch im Fall Vioxx, dem wohl größten Medikamenten-Skandal der letzten Jahre, sind die Forscher nicht über jeden Verdacht erhaben. Vioxx wurde 1999 als *Super-Aspirin* mit einem ungeheuren Werbeaufwand eingeführt. 2005, nur sechs Jahre später, nahm Hersteller Merck das Mittel freiwillig wieder vom Markt. Grund waren die Ergebnisse einer internen Studie, die ein erhöhtes Risiko für Gefäßerkrankungen zeigte. In diesen Tests stieg nach 18 Monaten das Herzinfarkt- und Schlaganfall-Risiko in der Vioxx-Gruppe an, nicht aber bei den Probanden,



Ob eine neue Bestrahlungsmethode wirklich besser ist, kann man nur beurteilen, wenn man die gesamten Studienergebnisse kennt



Lange Flure - zahllose Büros: 2.800 Mitarbeiter bearbeiten jährlich 390 verschiedene Statistiken

Tödliche Zahlen

die ein *Plazebo* geschluckt hatten. Neu schien dieser Effekt der Firma Merck aber nicht zu sein: Bereits in einer Studie aus dem Jahr 2000, also direkt nachdem Vioxx in den Handel kam, war diese gefährliche Nebenwirkung aufgefallen. Drei Herzinfarkte seien nicht in der Publikation der Fachzeitschrift berücksichtigt worden, obwohl sie den Autoren bekannt gewesen seien. Das sagte später Gregory Curfman, Chefredakteur des New England Journal of Medicine, einer anerkannten Medizinzeitschrift, die ebenfalls die Vioxx-Studie veröffentlicht hatte. Außerdem seien noch zwei Tage vor Einreichen der Arbeit wichtige Daten zu den Nebenwirkungen von einer Diskette gelöscht worden. Die Firma Merck sieht sich nun einer Flut von Klagen der geschädigten Patienten ausgesetzt.

▶ Plazebo

Ein *Plazebo* ist ein wirkungsloses Medikament. Es wird zur Überprüfung von Behandlungen eingesetzt, um die spezielle Wirkung eines Medikamentes von unspezifischen Effekten einer Behandlung zu trennen

■ Der Phantasie freien Lauf lassen – der Fall Sudbø

Man kann es allerdings noch bunter treiben. Anfang 2006 ist ein norwegischer Forscher aufgefliegen. Er hatte alle seine 1.000 Studienteilnehmer frei erfunden. Aufgefallen ist er nur, weil 250 seiner fiktiven Probanden dasselbe Geburtsdatum hatten...

■ Zahlen, bitte!

Wenn es darum geht, zu erfahren, wie viele Menschen in Deutschland leben, welche Waren importiert und exportiert werden oder wie sich die Preise entwickelt haben, dann ist eine Behörde immer mit im Spiel: das Statistische Bundesamt. Rund 2.800 Zahlenexperten verfassen hier jedes Jahr rund 390 verschiedene Statistiken zu Wirtschaft, Gesellschaft und Umwelt. Hier werden so wichtige Fragen beantwortet wie:

- ▶ *Auf wie vielen Quadratmetern leben die Deutschen im Durchschnitt?*
- ▶ *Wieso kann sich Deutschland Exportweltmeister nennen?*
- ▶ *Wie viele Erwerbstätige gibt es hierzulande?*
- ▶ *Wohin reisen die Deutschen im Urlaub?*
- ▶ *Wie viele Jahre habe ich zu leben?*

■ Kein Selbstzweck

Das Statistische Bundesamt entscheidet nicht selbst, was gezählt wird, sondern erhält seine Aufträge vom Bundestag oder der EU. Für jede Statistik wird also eigens ein Gesetz erlassen. So soll dafür gesorgt werden, dass die Statistiken immer auch eine gesellschaftlich relevante Frage behandeln. Nur dann treten die amtlichen Zahlenexperten in Aktion.

■ Anfänge nach dem Krieg

Den Anstoß zur Gründung einer Statistikbehörde gaben die Militärregierungen der westlichen Besatzungszonen gleich nach dem Ersten Weltkrieg. 1950 wird daraus das Statistische Bundesamt, 1954 bekommt die junge Behörde in Wiesbaden ein eigenes Gebäude. Nur drei Jahre später kommt die erste Statistik zu Bevölkerung und Erwerbsleben heraus, der Mikrozensus. 1987 geriet das Statistische Bundesamt monatelang in die Schlagzeilen: Die Volkszählung erhitze die Gemüter. Seit 1991 wird eine gesamtdeutsche Statistik geführt und seit 1996 ist das Statistische Bundesamt im Internet vertreten.



Links:
Immer noch kommt ein Teil der Daten auf Papier.
Sie werden dann eingescannt

Mitte:
Seit 1950 geht es in Deutschland stetig aufwärts:
Die Außenhandelsbilanz

Rechts:
Erschreckende Talfahrt

AUSSENHANDELSÜBERSCHUSS



AUS- UND EINFUHRBILANZ CHINA



Was macht eigentlich das statistische Bundesamt?

Im Jahr 2006 hat das Amt immer noch seinen Hauptsitz in Wiesbaden, mit zwei Nebenstellen in Bonn und Berlin. Die Daten werden mittlerweile fast vollständig digital verarbeitet, nur noch ein kleiner Teil kommt auf Papier an und wird dann eingescannt.

■ Vieles ist freiwillig, doch nicht alles

Der Großteil der Zahlen stammt aus den einzelnen Bundesländern, in denen die Statistischen Landesämter ihre eigenen, länderspezifischen Daten erheben. So werden etwa die Verbraucherpreise oder Wohnraumdaten erst auf Länderebene erfasst und dann an das Statistische Bundesamt weitergeleitet. Die Ergebnisse dürfen nicht davon abhängig sein, ob die Befragten freiwillig mitmachen, daher gibt es bei vielen Erhebungen eine Auskunftspflicht. Unternehmen und Gewerbebetriebe müssen zum Beispiel ihre Angestelltenzahlen oder ihren Umsatz offen legen. Und auch Privatmenschen müssen dem Amt gelegentlich Rede und Antwort stehen: so bei der Volkszählung 1987, die viel Widerstand hervorrief. Drücken ging nicht – es war eine Vollerhebung, alle deutschen Haushalte wurden

befragt. In der Regel macht das Statistische Bundesamt aber nur Stichproben. Es ist also eher Zufall, wenn ein Fragebogen des Statistischen Bundesamtes ins Haus segelt. Wenn man dabei zum Beispiel angeben soll, wie lange man schläft und womit man seine Zeit verbringt, dann handelt es sich um die Zeitbudgeterfassung. Die wird etwa alle zehn Jahren durchgeführt und dient z. B. Medizinern als Grundlage, wenn sie z. B. beurteilen wollen, ob jemand besonders kurz schläft, oder ob er im Schnitt liegt. Die Antworten auf die Zeitbudgeterhebung sind allerdings freiwillig.

■ Rein und raus

Die größte Einzelstatistik, die das Statistische Bundesamt führt, ist die Außenhandelsstatistik. Sie gibt an, wie viele Waren ein- oder ausgeführt werden. An ihr erkennt man welche Rolle Deutschland als Exportnation spielt. Alle ein- und ausgeführten Waren ab einem Wert von 1000,- € werden darin erfasst. Der Warenverkehr mit EU-Staaten wird getrennt vom Warenverkehr mit Nicht-EU-Staaten aufgenommen. Etwa 15 Millionen Datensätze kommen so zusammen. Zieht

man die eingeführten Warenwerte von den exportierten Warenwerten ab, so erhält man die Außenhandelsbilanz. Die Grafik oben links zeigt, dass Deutschland eine positive Außenhandelsbilanz hat: Deutschland ist Exportweltmeister. Interessant ist im Vergleich dazu die Bilanz zwischen Deutschland und China in der Abbildung daneben: Es ist klar zu sehen, dass China in den letzten zehn Jahren deutlich mehr Waren nach Deutschland exportiert hat, Tendenz steigend. Die Kurve sieht beeindruckend aus – wenn man jedoch auf die absoluten Warenwerte schaut, dann wird klar, dass der Handel mit China nur ein Zehntel des Gesamtwarenverkehrs ausmacht. Dass also ausgerechnet chinesische Kleider und Stoffe die deutsche Wirtschaft ruinieren, braucht man nicht zu befürchten.

■ Daten für alle

Die Zahlen und Statistiken des Statistischen Bundesamtes verschwinden nicht im Orkus – sie sind für jedermann zugänglich. Statistiken und Veröffentlichungen können zum Beispiel über das Internet abgerufen werden. Es ist aber auch

möglich, Fachleute des Statistischen Bundesamtes im Datenbestand recherchieren zu lassen. Wer also wissen möchte, ob seine Wohnung im Schnitt liegt, seine Freizeitbeschäftigung normal ist oder ob er mehr verdient als der Durchschnittsdeutsche, der kann in Wiesbaden die Vergleichsdaten erfragen, per Mail oder per Telefon. Übrigens sind die Daten anonym. Wer also wissen möchte, was der Nachbar verdient, ist beim Statistischen Bundesamt falsch. Die web-Adresse des Statistischen Bundesamtes lautet <http://www.destatis.de>.



Irgendwann stirbt jeder. Todesursache und Sterbealter werden in der amtlichen Sterbestatistik erfasst



Wie alt wird Nils?

Werden die Menschen wirklich immer älter?

Werden die Menschen wirklich immer älter?

■ Leben nach Zahlen

Jeder weiß es: Die Römer lebten nicht länger als 35 Jahre, dafür steigt heute rasant die Zahl der rüstigen Rentner, und es gibt Inseln wie Kreta oder Landstriche wie den Kaukasus, die geradezu als Hort der Hundertjährigen bekannt sind. Doch all das sind eigentlich Zahlenspielerien mit der Statistik, genauer: mit der Art und Weise, wie man das durchschnittliche Lebensalter, Sterbedaten und Lebenserwartung berechnet.

■ Hier ein kleiner Dschungelführer:

Das Durchschnittsalter eines Landes berechnen Experten, indem sie das Alter aller Bewohner durch die Anzahl aller Bewohner teilen. Gibt es also viele Kinder, ist das Durchschnittsalter niedrig; überwiegen ältere Menschen, so ist das Durchschnittsalter hoch. So hat Monacos Bevölkerung mit 45,4 das höchste; Uganda mit 15 Jahren das niedrigste Durchschnittsalter. Das Durchschnittsalter in Deutschland beträgt übrigens 42,6 Jahre.

Das mittlere Sterbealter einer Bevölkerung bezeichnet das durchschnittliche Alter zum Zeitpunkt des Todes. In der Medizin wird auch das mittlere

Sterbealter bei bestimmten Krankheiten verwendet. So ist das mittlere Sterbealter von Krebskranken höher als das mittlere Sterbealter von Menschen mit Herz- und Kreislauferkrankungen.

Die Lebenserwartung ist definiert als die erwartete Zeitspanne, die einem Lebewesen ab einem bestimmten Zeitpunkt noch bis zum Tode bleibt. Im allgemeinen Sprachgebrauch versteht man unter der Lebenserwartung die durchschnittliche Lebensspanne bei der Geburt. Sie entspricht dann dem mittleren zu erwartendem Sterbealter der Neugeborenen. Während das mittlere Sterbealter auf einen tatsächlich gemessenen Wert beruht – die Todesdaten werden empirisch ausgezählt – ist die Lebenserwartung also eine reine Vorhersage, eine Hochrechnung aus der Statistik.

■ Im Mittelalter nur 35 Jahre im Durchschnitt

Die Werte der mittleren Lebenserwartung sind erst in den vergangenen 160 Jahren deutlich angestiegen. Noch 1840 wurden schwedische Frauen mit 40 Jahren im Schnitt am ältesten. Heute sind es die Japanerinnen, die mit 86 Jahren den Rekord halten. Dazwischen verläuft der Anstieg der Lebenserwartung konstant: Jedes Jahr kommen drei Monate hinzu. Die Ursache für diesen Anstieg liegt

vor allem in der besseren Gesundheitsversorgung, Ernährung und Hygiene. Das war vor Mitte des 19. Jahrhunderts noch ganz anders, weil man über Infektionskrankheiten wenig wusste. In dieser Zeit war die Säuglings- und Kindersterblichkeit besonders hoch. Hatten die Menschen aber die Risiken der Jugendjahre überstanden, konnten sie auch damals ein hohes Alter erreichen. Wenn also davon die Rede ist, dass im Mittelalter die Lebenserwartung nur 35 Jahre betrug, heißt das ganz und gar nicht, dass die Mehrheit mit Mitte Dreißig starb. In der Statistik ziehen die vielen an Krankheiten gestorbenen Kinder, die im Kindbett gestorbenen jungen Frauen sowie die jungen Männer, die in den Kriegen umkamen, den Wert nach unten.

■ Gutes Leben – hohes Alter

Wer sich aber einen geruhsamen Alltag, Ärzte und eine gute Ernährung leisten konnte – das traf vor allem für Gebildete, Adlige und Geistliche zu – der lebte schon damals viel länger als es die mittlere Lebenserwartung vorhersagte: Die antiken Philosophen Aristoteles und Seneca wurden über 60 Jahre alt, Karl der Große ebenso. Die mittelalterliche Äbtissin Hildegard von Bingen starb mit 81, Goethe wurde 83 und Isaac Newton 84 Jahre alt.

Auch heute noch gilt, dass der Lebensstil und Umweltfaktoren zu zwei Dritteln die Lebenserwartung bestimmen. Das haben Vergleiche zwischen Zwillingen ergeben. Wie alt man also wird, ist weniger durch das Erbgut bestimmt, als durch die Umwelt und das Verhalten.

■ Wie alt werden wir?

Um etwas über die zukünftige Lebenserwartung zu sagen, muss man bestimmte Voraussetzungen mit einbeziehen: Bleiben die Umweltbedingungen gleich? Wie entwickeln sich die Gesundheitsversorgung und die Ernährungssituation? Obwohl diese Faktoren natürlich nicht genau festliegen, können Experten Rahmenbedingungen für die Berechnung der Lebenserwartung stecken: Zum Beispiel erwarten sie, dass Gesundheitsversorgung und Medizin Fortschritte machen. Deshalb gehen die meisten Prognosen von einem weiteren Anstieg der Lebenserwartung aus. Damit die Lebenserwartung allerdings in dem Tempo ansteigt wie bisher, müssen auch ähnliche medizinische Revolutionen stattfinden wie in der Vergangenheit.



Anklage wegen zweifachen Mordes: Sally Clark



Wie oft kommt der plötzliche Kindstod in einer Familie wirklich vor?

Eine Frau gegen die Statistik – der Fall Sally Clark

Eine Frau gegen die Statistik – der Fall Sally Clark

■ Zwei tote Babys in einem Jahr

Manipulationen und falsche verstandene Statistiken können einfach nur ärgerlich sein – aber sie können auch gravierende Folgen haben. In England musste eine Frau ins Gefängnis, weil ein Sachverständiger grundlegende Regeln der Wahrscheinlichkeitsrechnung außer Acht ließ und das Gericht ihm glaubte.

Der Fall Sally Clark beginnt mit dem plötzlichen Tod ihres ersten Kindes: Die Rechtsanwältin aus Manchester bekommt 1996 ihren ersten Sohn Christopher, ein gesundes Kind. Doch nach 11 Wochen ist er tot – *plötzlichen Kindstod*, diagnostizieren die Ärzte. Tragisch für die junge Mutter, doch nicht so ungewöhnlich: Das Schicksal trifft jedes Jahr einen von rund 2.000 gesund geborenen Säuglingen, die genauen Ursachen kennt man nicht.

Die Ärzte raten Sally Clark dazu, bald wieder schwanger zu werden. Dann könne sie das schreckliche Erlebnis psychisch besser verarbeiten. Ein Jahr später, am 29.11.1997 kommt der zweite Sohn Harry auf die Welt.

Aber auch Harry stirbt im Alter von acht Wochen – jetzt werden die Behörden misstrauisch. Sally Clark wird angeklagt wegen zweifachen Mordes.

■ Schuldig durch falsche Wahrscheinlichkeit?

Neben undurchsichtigen und widersprüchlichen Gutachten der Pathologen ist es vor allem die Stellungnahme eines Mannes, die das Gericht überzeugt: Der Kinderarzt Professor Roy Meadow hält es für äußerst unwahrscheinlich, dass in einer Familie zweimal hintereinander der plötzliche Kindstod auftritt. Meadow führt schon länger einen Kampf gegen diese Diagnose, denn seiner Meinung nach verbirgt sich dahinter meistens Mord durch die Eltern. Meadows rechnet vor: In einer von 8.500 Familien, deren Situation mit der von Sally Clark vergleichbar ist, stirbt ein Säugling zufällig an plötzlichem Kindstod, also besteht dafür eine Wahrscheinlichkeit von 1 : 8.500. Für das zweite Kind besteht eine genauso große Wahrscheinlichkeit, zu sterben. Meadows multipliziert also die Wahrscheinlichkeiten und kommt auf eine Zahl von 1 : 72.250.000 Soll heißen: Nur in einer von 72 Millionen Familien würde das Schicksal tatsächlich zweimal auf so grausame Weise zuschlagen. In England wäre das alle hundert Jahre einmal der Fall. Meadows interpretiert sein Ergebnis als Wahrscheinlichkeit für Sally Clarks Unschuld – die damit also äußerst gering wäre. Diesem Argument folgt das Gericht, und Sally Clark wird im November 1999 wegen zweifachen Mordes zu zweimal lebenslanglich verurteilt.

■ Widerspruch von Statistik-Experten

Die Königliche Statistische Gesellschaft protestiert gegen den *Missbrauch von Statistik* vor Gericht: Zunächst einmal ist es rein statistisch überhaupt nicht zulässig, die Zahl von 8.500 einfach zu multiplizieren. Das käme nur in Frage, wenn die beiden Todesfälle völlig unabhängig voneinander wären. Man vermutet aber gerade beim plötzlichen Kindstod, dass es noch unbekannte Ursachen in den Genen oder in der Umwelt gibt, die solche Todesfälle zur Folge haben. Die Wahrscheinlichkeit, dass in einer Familie zwei Kinder sterben, ist also deutlich höher, als Meadows berechnet hat.

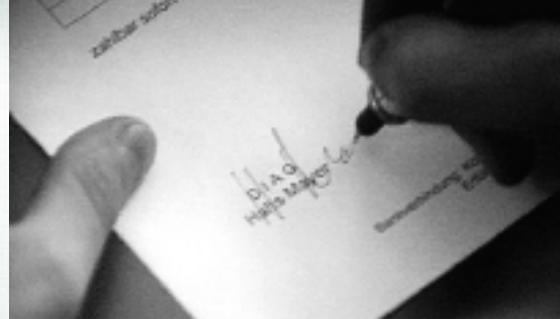
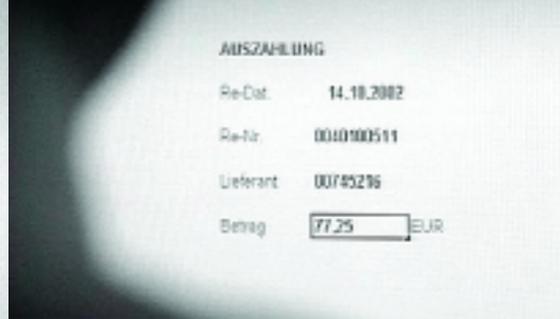
■ Wie klein ist klein genug?

Es gibt aber noch ein weiteres Problem bei Meadows' Argumentation, unter Statistikern bekannt als *prosecutor's fallacy*, was man etwa als *Trugschluss des Anklägers* übersetzen könnte. Angenommen, die Wahrscheinlichkeit für den plötzlichen Kindstod sei mit 1 : 72,25 Millionen korrekt berechnet – die Zahl alleine sagt überhaupt noch nichts aus. Der Staatsanwalt interpretiert sie aber als *Wahrscheinlichkeit für Unschuld* und legt nahe: „Weil die

Zahl klein ist, muss die Wahrscheinlichkeit für Schuld groß sein.“ Korrekterweise müsste er jedoch die *Wahrscheinlichkeit für Schuld* unabhängig davon ermitteln, und dann die beiden Zahlen gegenüberstellen. Man müsste also überschlagen, wie viele Mütter ihre beiden Kinder töten: Macht so etwas eine von 100 Millionen oder eine von einer Milliarde? Erst durch den Vergleich dieser beiden Zahlen könnte man überhaupt eine Aussage treffen, aber man sieht auch sofort, wie absurd so eine Aussage wäre.

■ Freispruch nach drei Jahren

Sally Clark bleibt noch bis Januar 2003 inhaftiert. Ihr drittes Kind wird zu Pflegeeltern gegeben. Ihr Leben ist zerstört, weil ein Medizin-Professor mit Zahlen jongliert hat. Erst nach drei Jahren erreicht ihr Mann eine Wiederaufnahme des Verfahrens. Und diesmal wird sie freigesprochen, allerdings aus anderen Gründen: Die Anklage hatte im ersten Prozess verschwiegen, dass der kleine Harry eine starke Bakterien-Infektion hatte.



Links:
Betrügerische Mitarbeiter können ein Unternehmen im Extremfall in den Ruin treiben

Mitte:
Verdächtig ist es, wenn oft Beträge knapp unterhalb der Bagatellegrenze gebucht werden

Rechts:
Am schwersten zu entdecken: Mitarbeiter steckt mit Lieferant unter einer Decke und zeichnet falsche Rechnungen ab

Auf Verbrecherjagd mit Benford

Auf Verbrecherjagd mit Benford

Kann man mit bloßer Ziffern-Analyse Verbrechern auf die Schliche kommen? Diese Frage scheint auf den ersten Blick absurd. Aber die Antwort lautet: man kann. Der amerikanische Professor des Buchhaltungswesens Mark Nigrini hat sich durch Benfords Gesetz inspirieren lassen: Er schrieb ein einfaches Computer-Programm, mit dem man große Zahlenmengen auf ihre Verteilung nach Benford analysieren kann. Seine Idee war: Wenn Zahlen in der Buchhaltung eines Betriebs von der Benford-Verteilung statistisch signifikant abweichen, könnten dahinter eventuell betrügerische Mitarbeiter stecken.

Quarks & Co wollte genauer wissen, was mit solch einer Ziffern-Analyse herauszufinden ist und bat die Mitarbeiter der Revisionsabteilung des WDR um Hilfe. Aus dem WDR-System luden sie die Rechnungen der letzten zwei Monate herunter: 12.372 Posten. Eine Überprüfung nach Benford ergab: Die Zahlen sind nicht auffällig. Mit ihnen war also alles OK. Nun manipulierten wir die (anonymisierten) Zahlen dreimal auf unterschiedliche Art und Weise und ließen sie von Wirtschaftsprüfern untersuchen. Würden sie herausfinden, was mit den Zahlen passiert ist? Drei Fälle haben wir nachgestellt:

■ Erster Fall: Die Unterschriftengrenze

Ein Mitarbeiter im Einkauf bevorzugt einen bestimmten Lieferanten, obwohl der nicht der günstigste ist. Solange die Bestellungen aber 5.000 Euro nicht überschreiten, merkt das niemand. Seinen Chef muss er nämlich erst bei größeren Anschaffungen informieren. Die Lieferfirma zeigt sich erkenntlich und spendiert hin und wieder einen Kurzurlaub ...

Für diesen ersten Fall veränderten wir die ursprünglichen Rechnungsdaten: Bei 63 Rechnungen eines Lieferanten gingen wir auf Beträge knapp unter der gedachten Unterschriftengrenze von 5.000 Euro. Wir waren gespannt: Würden die 63 geänderten Rechnungen unter den 12.372 anderen auffallen? Hier die Analyse der Experten von Ernst & Young: „Bei der Untersuchung dieser Daten haben wir beim Erste-Ziffer-Test Auffälligkeiten bei der Vier gefunden. Daraufhin sind wir auf einen Lieferanten gekommen, der eine vermehrte Häufigkeit bei dieser Ziffer Vier aufweist. Bei genauerem Hinsehen erkannten wir, dass diese Beträge mit der Vier als 1. Ziffer alle knapp unter 5.000 Euro liegen. Dafür kann es mehrere Gründe geben: zum Beispiel eine Unterschriftenregelung oder Ähnliches im Unternehmen.“

■ Zweiter Fall: Die Bagatellegrenze

Ein schönes Ritual: Jeden Tag überweist ein Mitarbeiter Geld auf sein eigenes Konto. Das fällt nicht auf, weil es in der Firma eine Bagatellegrenze gibt: Kleinbeträge werden nicht überprüft. Auf ein Jahr gerechnet kommt dabei aber einiges zusammen ...

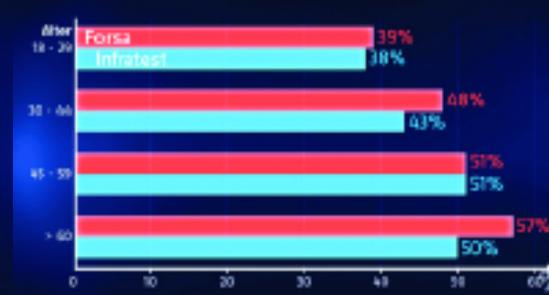
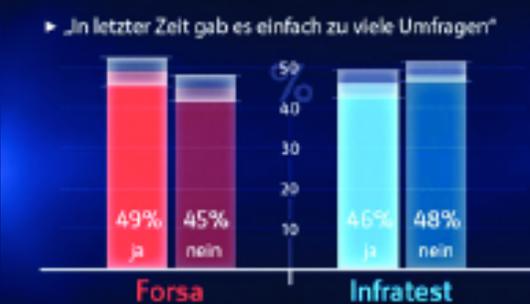
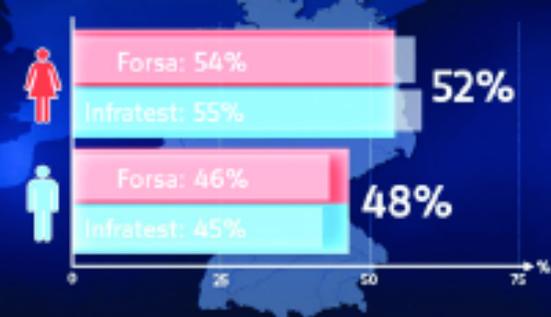
In unserer Datei mit den ungefälschten Daten erfanden wir einen neuen Lieferanten und fügten für jede Woche fünf Überweisungen ein. Alle Beträge lagen zwischen 70 und 80 Euro. Wieder waren die Wirtschaftsprüfer an der Reihe. Dieses Mal waren aber noch weniger Zahlen geändert. Aber trotzdem wurden sie fündig:

„Bei der Ziffer Sieben haben wir eine Auffälligkeit festgestellt. Wir haben uns diese Beträge näher angeschaut und sind auf einen Lieferanten aufmerksam geworden, der bestimmte Rechnungsbeträge zwischen 70 und knapp unter 80 Euro hatte. Es könnte sein, dass es in diesem Unternehmen eine Bagatellegrenze gibt, unter der diese Beträge liegen. Es könnte aber auch ein Systemfehler sein, weil die Beträge immer werktags gebucht wurden.“

■ Dritter Fall: Erfundene Rechnungen

Ein Händler stellt neben den normalen Rechnungen noch weitere aus. In der belieferten Firma zeichnet sie ein Mitarbeiter ab, obwohl dafür gar keine Leistungen erbracht wurden. Am Monatsende machen die beiden halbe-halbe ...

In unserer Originaldatei erfanden wir bei einem Lieferanten für jede echte Rechnung noch eine zusätzliche gefälschte. Dieser Test war besonders schwierig: Insgesamt fügten wir nämlich nur 20 Rechnungsposten ein. Und tatsächlich: In der großen Datenmenge waren diese 20 zufällig ausgedachten Rechnungssummen nicht mehr auffällig. Fazit: Immerhin zwei der drei Betrüger wären schon mit der einfachen Benford-Analyse aufgefliegen.



Links:
Anteil von Frauen und Männern bei der Befragung

Mitte:
Vergleich der Untersuchungsergebnisse von Forsa und Infratest dimap

Rechts:
Vergleich der Untersuchungsergebnisse von Forsa und Infratest dimap in Bezug auf das Alter der Befragten

Der Umfrage-Test

Auf den Zahn gefühlt – wie genau sind Telefon-Umfragen wirklich?

Ob für Zeitung, Radio oder Fernsehen: Was die Menschen in Deutschland wählen, denken oder essen – Meinungsforschungsinstitute versuchen, das herauszufinden. Für jede Umfrage werden mindestens 1.000 Menschen angerufen, die zuvor per Zufallsauswahl festgelegt wurden. Solche Umfragen sollen repräsentativ für die deutsche oder deutschsprachige Bevölkerung sein. Das bedeutet, dass die Ergebnisse der Umfrage unter den 1000 Befragten dem Ergebnis entsprechen sollen, das man bekäme, wenn man nicht 1000, sondern alle deutschen oder deutschsprachigen Menschen befragen würde. Wir wollten wissen: Wie genau sind solche Ergebnisse von Telefonumfragen wirklich? Um das zu testen, haben wir selber eine Umfrage in Auftrag gegeben – bei zwei verschiedenen Instituten: Forsa und Infratest dimap.

Thema der Test-Umfrage: Wir wollten wissen, was Sie von Meinungsumfragen halten. Eine der Fragen lautete: „In letzter Zeit gab es einfach zu viele Meinungsumfragen. Inwieweit stimmen Sie dieser Aussage zu?“ An zwei Tagen im Mai telefonierten die beiden Forschungsinstitute mit 1.000 beziehungsweise 1.009 Menschen und stellten unter anderem unsere Quarks-Fragen. Die Telefon-

nummern wurden bei beiden Instituten zufällig ausgewählt – es wurden also von Forsa andere Menschen befragt als von Infratest dimap. Wie unterscheiden sich die Ergebnisse der beiden Institute? Drei Beispiele:

Erster Test

Wenn tausend Befragte für die gesamte Bevölkerung stehen, dann muss auch der Anteil der angerufenen Frauen und Männern im richtigen Verhältnis sein: 52 Prozent Frauen und 48 Prozent Männer. Forsa liegt um zwei Prozentpunkte daneben, Infratest um drei; solche Abweichungen sind ganz normal und werden bei den Berechnungen der Ergebnisse mit einbezogen: Wurden weniger Männer befragt als es ihrem Anteil an der Bevölkerung entspricht, dann gehen deren Antworten zum Ausgleich etwas stärker in das Endergebnis ein als die der Frauen.

Zweiter Test

Wie stark unterscheiden sich die Antworten auf unsere Testfrage? Wir fragten: „In letzter Zeit gab es einfach zu viele Umfragen.“ Bei Forsa stimmte eine knappe Mehrheit dieser Aussage zu

(49 gegen 45 %). Anders bei Infratest: Hier sieht's genau umgekehrt aus: Eine knappe Mehrheit stimmte mit nein (48 %), (ja meinten 46 %). Ist das ein Widerspruch? Die Forschungsinstitute geben eine Fehlertoleranz an; je nach Institut maximal plus-minus 3,1 Prozentpunkte: Konkret bedeutet das, dass das echte Ergebnis nach der Forsa-Umfrage zwischen 46 und 52 % liegen kann. Für Infratest dimap liegt dieser Bereich zwischen 43 und 49 %. Diese beiden Intervalle überschneiden sich. Und das bedeutet, dass sich die Werte der beiden Institute – zumindest statistisch gesehen – nicht widersprechen. Die Forschungsinstitute schränken ihre Zahlenangaben noch weiter ein: Die ermittelten Werte liegen nur mit 95-prozentiger Wahrscheinlichkeit innerhalb der angegebenen Fehlertoleranz. Das heißt, dass – zumindest theoretisch – eine Umfrage auch völlig daneben liegen kann.

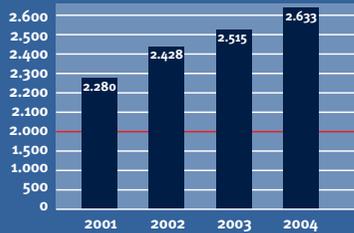
Dritter Test

Wirkt sich das Alter der Befragten auf die Antwort aus? Forsa findet heraus: „Wenn man jetzt etwas genauer in die Daten blickt, kann man zum Beispiel bei den Altersgruppen erkennen, dass die älteren Bundesbürger Umfragen etwas skeptischer gegenüber stehen als jüngere Bundesbürger.“

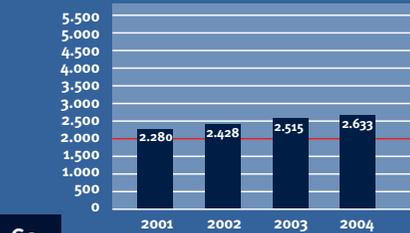
Stimmt das so auch für die Umfrage, die wir bei Infratest dimap in Auftrag gegeben haben? Auf den ersten Blick gibt es große Unterschiede. Bei den Älteren weicht der Wert um sieben Prozentpunkte ab – mehr als die Institute zugestehen. Aber: Der Fehlerbereich (das so genannten Vertrauensintervall) ist jetzt größer: plus-minus sechs bis acht Prozentpunkte. Denn in den einzelnen Altersgruppen sind viel weniger Befragte: nur noch zwei- bis dreihundert. Das *Zentrum für Umfragen, Methoden und Analysen* hat für uns alle Werte getestet. Das Ergebnis: „In keinem der getesteten Fälle lässt sich bei fünf Prozent Irrtumswahrscheinlichkeit die Nullhypothese zurückweisen, dass sich die Altersverteilungen der den Items Zustimmenden nicht unterscheiden.“ Das ist die exakte statistische Formulierung, die schlicht bedeutet: Offenbar sind die Verteilungen beider Untersuchungen gleich, obwohl sich die Ergebnisse um bis zu sieben Prozentpunkte (bei der Altersgruppe der Über-Sechzig-Jährigen) unterscheiden.

Fazit unseres Quarks-Test

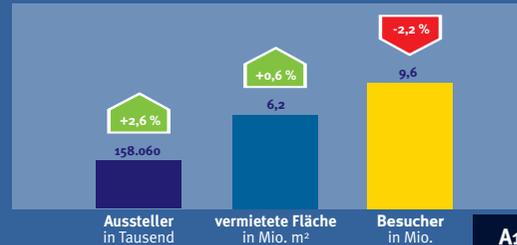
Man kann sich auf die Ergebnisse von Meinungsumfragen verlassen. Mit weitergehenden Interpretationen sollte man aber vorsichtig sein.



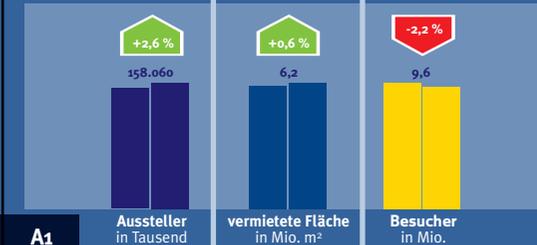
G1



G2



A1



A1

Die schlechtesten Grafiken der Welt

Die schlechtesten

Wenn man statistische Sachverhalte in Diagramme einträgt, erlebt man oft eine Enttäuschung: das Wachstum ist gar nicht so dramatisch, wie man es gerne darstellen möchte oder die Differenzen sind gar nicht so groß. Also muss der Grafiker ein bisschen nachhelfen: er lässt ein paar Teile weg und biegt andere passend zurecht. Das sieht zwar besser aus – aber korrekt ist die Darstellung dann nicht mehr. Die Längen- und Flächenverhältnisse entsprechen dann nicht mehr den wirklichen Größenverhältnissen und die Grafik hat jede Aussagekraft verloren.

Hier sind einige Grafiken dargestellt, wie sie täglich in Zeitungen, Zeitschriften und anderen Medien zu finden sind. Wir haben die Darstellungen jeweils so angepasst, dass sie die echten Größenverhältnisse repräsentieren.

Geschäftsbericht

Einen erfreulichen Geschäftsverlauf wollte die *psd-Bank Rhein-Ruhr* in ihrem Geschäftsbericht 2004 darstellen. Rapide aufwärts muss es mit der Bilanzsumme gegangen sein, wenn man sich die

dazugehörige Grafik ansieht. Als *Ausdruck der positiven Entwicklung der Bank* stehen dort vier Säulen; jede deutlich größer als die links daneben stehende. (Grafik G1)

Doch ein näheres Hinsehen zeigt: Wachstum ja, aber kaum so drastisch, wie die Bank es wohl gerne hätte. Die Grafik ist geschummelt: Horizontale Linien in gleichem Abstand spiegeln eine lineare Skala vor: doppelt so hohe Säulen würden für eine doppelt so hohe Bilanzsumme stehen. Aber weit gefehlt: Die Linien unter der 2.000 haben jeweils einen Abstand von 500 Einheiten, aber darüber geht es nur im Hunderter-Abstand weiter. Auf den ersten Blick fällt das kaum auf. Das Wachstum wirkt dadurch beträchtlich.

Weniger beeindruckend wirken die Zahlen, wenn man sie maßstabsgerecht einträgt. Der Werteabstand muss natürlich bei allen horizontalen Linien gleich sein, nämlich jeweils 500 Einheiten. Und so sieht das Ergebnis in Wirklichkeit aus wie in Grafik G2.

Mit freundlichem Dank an Prof. Dr. Walter Krämer, Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund

Aufwärtstrend

„Die deutsche Messewirtschaft hat sich leicht erholt“ schreibt der Focus zu einer Grafik, auf der drei Säulen nebeneinander gen Himmel streben. Für den flüchtigen Beobachter ist klar: hier geht es aufwärts. Vom dunklen Schwarz über ein neutrales Blau zum Sonnenschein gelb – die dritte Säule ist gut doppelt so hoch wie die erste (Grafik A1).

Doch bei näherem Hinsehen zeigt sich: die drei Zahlenwerte über den Säulen haben überhaupt nichts miteinander zu tun. Die erste Säule zeigt die *Aussteller in Tausend*, die zweite die *vermietete Fläche in Mio. m²*, die dritte die *Besucher in Mio.* Es gibt also keinen Grund dafür, die Säulen unterschiedlich hoch darzustellen. Was sollte auch ein Vergleich aussagen, der zum Inhalt hat *Die Anzahl der Besucher ist ungefähr eineinhalb mal so groß wie die vermietete Fläche?* Der in der Überschrift angesprochene *Aufwärtstrend* zeigt sich nur in den *Sahnehäubchen* über den Säulen: zwei grüne Pfeile, die aufwärts zeigen, und ein roter nach unten. Sie verdeutlichen die Veränderung gegenüber irgendeinem Zeitpunkt in

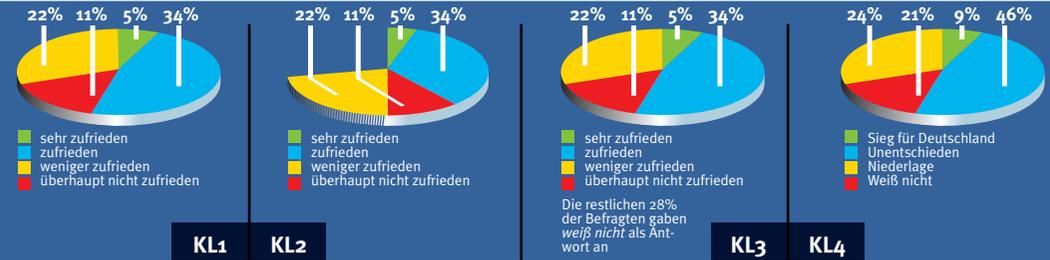
der Vergangenheit. (Welcher das ist, erschließt sich weder aus der Grafik noch aus dem Begleittext – vermutlich ist das Vorjahr gemeint).

Realistischer – aber auch weniger spektakulär – wird die Darstellung, wenn man für alle Zahlen ein gleiches Niveau in der Vergangenheit zugrunde legt und das entsprechende Wachstum relativ zu diesem Niveau in einer zweiten Säule darstellt (Grafik A2). Der grafische Unterschied in der Höhe der Säulen – bei gleichbleibender Breite – stellt dann anschaulich die Veränderung dar. Allerdings sollte man dann doch noch mal über die Überschrift nachdenken: *Aufwärtstrend?*

Quelle: Focus, Nr. 20/2006, S. 187

Kindergeld

In der Broschüre *Antworten zur agenda 2010* wollte das Finanzministerium unter der rot-grünen Regierung darstellen, wie stark zwischen 1998 und 2002 das Kindergeld erhöht wurde. Kinderwagen in unterschiedlicher Größe symbolisieren den scheinbar enormen Anstieg.



Die schlechtesten Grafiken der Welt

Damit der Anstieg richtig zur Geltung kommt, hat der Grafiker den lästigen Teil unter 100 Euro weggelassen, die Skala geht von 100 bis 160. Genau genommen müsste die Überschrift statt *Entwicklung des Kindergeldanteils...* also lauten: *Entwicklung des Kindergeldanteils über 100 Euro...* Wollte man die Entwicklung maßstabgerecht darstellen, müsste man zumindest die Skala ändern. Das sähe zunächst wie in der Grafik K2 aus. Aber hier fehlt noch das schöne Bild von den Kinderwagen. Denn wenn man die vergrößert, vergrößert man sie nicht nur in der Höhe, sondern auch in der Breite. Die Veränderung des Kindergeldes spiegelt sich für den Betrachter in der Veränderung des Flächeninhaltes wider. Wenn man davon ausgeht, dass die Fläche des ersten Kinderwagen für 112,48 Euro steht, ergeben sich daraus im Verhältnis die Flächen für 2000 und 2002 wie in der Grafik K3.

Quelle: Presse- und Informationsamt der Bundesregierung: *Antworten zur agenda 2010*, (November 2003) S. 45

Klinsmann-Torte

Am 12.3.2006, 3 Monate vor der WM ist die Kampagne der Bild-Zeitung gegen den damaligen Bundestrainer Jürgen Klinsmann noch in vollem Gange. Auf der Online-Ausgabe *bild.de* kann man es nachlesen: *Die Sonntags-Frage zur WM*. Und es ist kein Wunder, dass Klinsmann schlechte Noten bekommt: *Nur fünf Prozent sind sehr zufrieden mit Klinsmann*. Ganz abgesehen davon, dass aber nur 11% *überhaupt nicht* zufrieden sind und der ganze große Rest sich also zu 84% auf *zufrieden*, *weniger zufrieden* und *mir egal / weiß nicht* aufteilt, gibt auch die illustrierende Grafik (KL1) Rätsel auf.

Wenn man die abgegebenen Stimmen zusammenzählt, kommt man erstaunlicherweise nur auf 72%. Denn alle *weiß-nicht*-Antworten hat Bild einfach unter den Tisch fallen lassen. Richtig wäre eine Darstellung wie KL2 gewesen:

Nachdem *Bildblog.de* darauf aufmerksam gemacht hat, wird die Grafik nicht etwa korrigiert – die neue Version sieht jetzt einfach so aus (KL3).

Das ist natürlich Blödsinn, denn der Sinn einer Tortengrafik liegt ja darin, eine räumliche Vorstellung von den unterschiedlichen Anteilen zu vermitteln. Und das geht nur, wenn die ganze Torte 100 % entspricht. Oder, wie *Bildblog* schreibt: „Sinn und Zweck einer Tortengrafik ist es, Teilwerte eines Ganzen wie Kuchenstücke aussehen und dadurch anschaulich werden zu lassen. Und das funktioniert natürlich nur, wenn der Zuckerbäcker nicht irgendwelche Teigklumpen unter den Tisch fallen lässt.“

Woher die Grafik ursprünglich stammte, merkte ein aufmerksamer Leser übrigens später. Schon vor dem Länderspiel am 22.3. gegen die USA hatte *Bild.de* gefragt *Gewinnt Deutschland den Test gegen die USA?* (Grafik KL4) Die Überraschung: Bis auf die Zahlen sind die beiden Torten identisch.

Mit freundlichem Dank an Christoph Schultheis und *Bildblog.de*,
Quelle: <http://www.bildblog.de/?p=1225>.



Lesetipps

ZU: TÖDLICHE ZAHLEN: MEDIZINSTATISTIKEN

Statistik verstehen – Eine Gebrauchsanweisung

Autor: Walter Krämer
Verlagsangaben: Campus Verlag 1998,
ISBN: 3-593-36149-3

Gutes Übersichtswerk, das die häufigsten Begriffe der Statistik erklärt und mögliche Quellen für Missverständnisse und Fehlinterpretationen aufzeigt.

Der Hund, der Eier legt

Autoren: Hans-Peter Beck-Bornholdt,
Hans-Hermann Dubben
Verlagsangaben: Rororo 2001
ISBN: 3-499-61154-6

Unterhaltsames Buch, das fragwürdige Gebräuche und Skandale aus der Wissenschaftsstatistik aufdeckt.

Der Schein der Weisen

Autoren: Hans-Peter Beck-Bornholdt,
Hans-Hermann Dubben
Verlagsangaben: Hoffmann und Campe 2002,
ISBN: 3-455-09340-X

Nachfolgeband von Der Hund der Eier legt.

WISSENSCHAFTLICHE LITERATUR ZU:

TÖDLICHE ZAHLEN: MEDIZINSTATISTIKEN

Für diejenigen, die sich die Originalstudien ansehen wollen. Leider gibt es hier keine direkten Links. Zudem sind die Studien ausschließlich in englischer Sprache und sehr fachspezifisch verfasst, also nur für wissenschaftlich vorgebildete Leser geeignet.

Die großen Skandale

Brustkrebs-Reihenuntersuchungen

Nystrom, Rutquist, Wall, Lindgren: *Breast cancer screening with mammography: overview of Swedish randomized trials*, Lancet 1993 Apr.17; 341 (8851): 973-8

Die CAST-Studie zum Medikament Flecainid

Ward, Garratt, Camm: *Cardiac arrhythmia suppression trial and flecainide*, Lancet 1989 Jun 3;1(8649):1267-8

Die Vioxx-Skandal-Studien

Bombardier und andere: *Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis – VIGOR*, NEJM Volume 343 Number 21 (2000): 1520-8

Bresalier und andere: *Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial – Approve*, NEJM 2005 Mar 17;352(11):1092-102. Epub 2005 Feb 15.

Bestrahlung von Kopf-Hals-Tumoren

Saunders und andere: *Experience with CHART. Continuous, hyperfractionated accelerated radiotherapy*, IntJRadiat OncolBiolPhys, 1991 Aug; 21(3): 871-8

Die Original-Veröffentlichung des kreativen norwegischen Forschers

Sudbø und andere: *DANN Content as a prognostic marker in patients with oral leukoplakia*, NEJM Volume 344 Number 17 (2001): 1270-8

Studien über Medizin-Studien

An-Wen Chan: *Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials*, JAMA 2004; 291; 2457-2465
Die Autoren haben 102 Studien, aus denen 122 Artikel publiziert wurden, untersucht. 50 Prozent der Ergebnisse zur Wirksamkeit und 65 Prozent der Nebenwirkungsergebnisse wurden nur inkomplett veröffentlicht.

Gross und andere: *Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review*, Jama 2003;289;454-465
Gross und Kollegen haben über 100 Studien untersucht. 80 Prozent der industrie-finanzierten, aber nur 50 Prozent der unabhängigen Studien kommen zu positiven Ergebnissen.

Garcia-Berthou und andere: *Incongruence between test statistics and p values in medical papers*, BMC Medical Research Methodology 2004;4;13
Die Autoren haben Artikel aus Nature und dem British Medical Journal untersucht und festgestellt, daß in etwa 12 Prozent der Publikationen Rechenfehler waren.

Juhl Jørgensen und andere: *Presentation on websites of possible benefits and harms from screening for breast cancer: cross sectional study*, BMJ 2004;328:148-154
Die Autoren haben verschiedene Internetseiten danach untersucht, wie Vor- und Nachteile des Brustkrebs-Screenings dargestellt werden. Zu: Eine Frau gegen die Statistik – der Fall Sally Clark

ZU: EINE FRAU GEGEN DIE STATISTIK – DER FALL SALLY CLARK

Mit an Wahrscheinlichkeit grenzender Sicherheit. Logisches Denken und Zufall

Autoren: Hans-Hermann Dubben,
Hans-Peter Beck-Bornholdt
Verlagsangaben: Rororo science,
Reinbek bei Hamburg, 2005,
ISBN: 3-499-61902-4
Sonstiges: Preis 8,90 Euro.

Nach Der Hund, der Eier legt und Der Schein der Weisen das dritte Buch der beiden Wissenschaftler, in dem sie Fehlschlüsse und Denkfallen aus der Welt der Wahrscheinlichkeiten aufgreifen. Unter anderem auch mit einer näheren Analyse des Falles Sally Clark.

ZU: DER UMFRAGE-TEST

Statistik für Dummies

Autor: Deborah Rumsey
Titel: Statistik für Dummies
Verlagsangaben: Wiley-VCH-Verlag,
ISBN: 3-527-70108-7
Sonstiges: 356 Seiten, Preis: 19,95 Euro

Wer an anderen Statistik-Büchern gescheitert ist, kann hier einen erneuten Anlauf nehmen. Dabei spart das Buch auch harte Themen wie Konfidenzintervalle und -koeffizienten sowie Hypothesentest nicht aus.