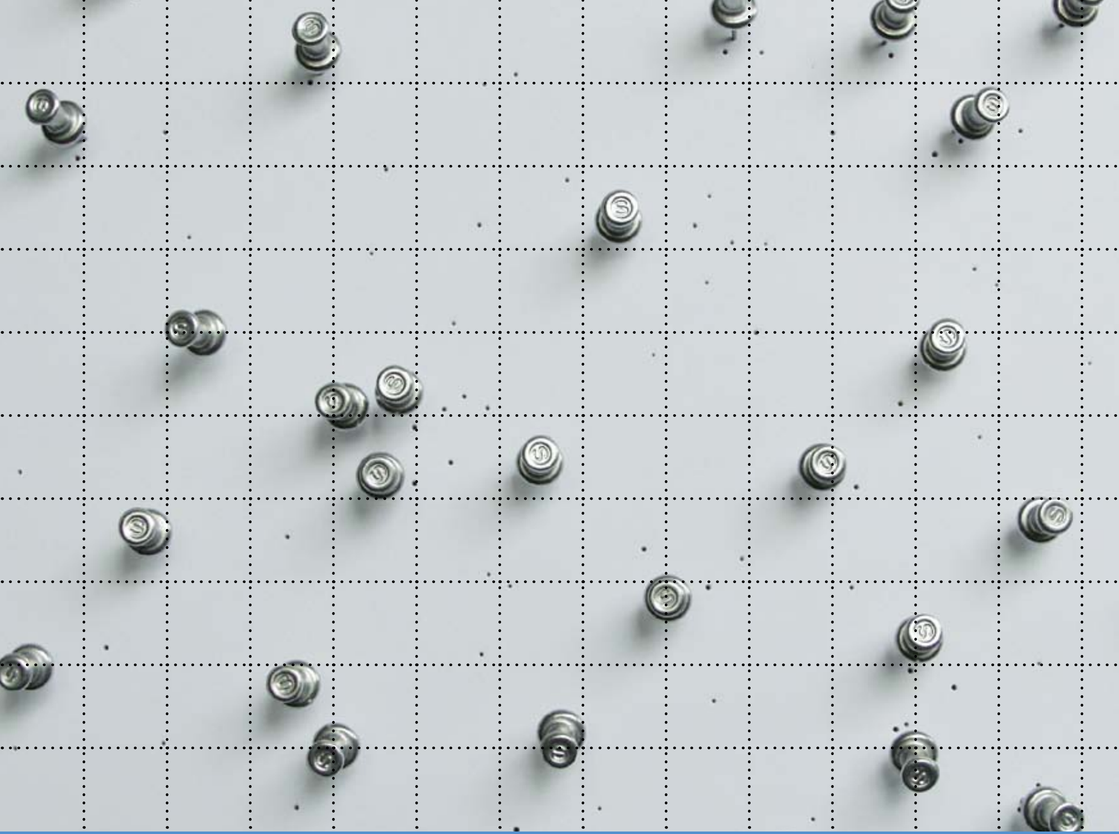


Die Stichprobe:
warum sie funktioniert

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Neuchâtel, 2009

Die Stichprobe: *warum sie funktioniert*

Jürg Zimmermann, Bernhard Morgenthaler, Beat Hulliger

Zu dieser Publikation steht eine Excel-Datei zur Verfügung,
mit der Sie das Prinzip der Stichprobe simulieren können.

Diese Datei kann auf dem BFS-Statistikportal heruntergeladen werden:
http://bit.ly/qZRQKk

Smartphonebesitzer können die Datei laden, indem sie diesen QR-Code mit dem Mobiltelefon und der darauf installierten QR-Lesesoftware fotografieren. Sie werden direkt zu der entsprechenden Seite auf dem BFS-Statistikportal geführt.

Herausgeber: Bundesamt für Statistik (BFS)

Auskunft: Beat Hulliger, BFS, Tel. 032 713 60 14
E-Mail: beat.hulliger@bfs.admin.ch

Autoren: Jürg Zimmermann, Bernhard Morgenthaler
(Sektion Information und Dokumentation, BFS),
Beat Hulliger (Dienst Statistische Methoden, BFS)

Vertrieb: Bundesamt für Statistik, CH-2010 Neuchâtel
Tel. 032 713 60 60 / Fax 032 713 60 61
E-Mail: order@bfs.admin.ch

Bestellnummer: 654-0400 (zweite, unveränderte Auflage)

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: deutsch

Grafik/Layout: 2. stock süd netthoevel & gaberthüel, Biel

ISBN: 3-303-00303-3

Die Stichprobe:
warum sie funktioniert

Rund um Jugendliche, Beziehungen und Familie

Die 4 Grafiken sind eine kleine Auswahl aus vielen verschiedenen Statistiken zum Thema Familie, die das Bundesamt für Statistik erhoben hat. In diesen Bildern steckt einiges: Es sind «Fahrtenschreiber», die anzeigen, in welche Richtung sich unser Leben verändert.

GRAFIK 1

Hotel Papamama dauert länger

Die Jugendlichen leben heute länger mit ihren Eltern zusammen als früher. Sie kommen später als ihre Eltern zu einem eigenen Einkommen und gründen auch den eigenen Haushalt später.

Frauen
Männer

Quelle: Mikrozensus Familie 1994/95

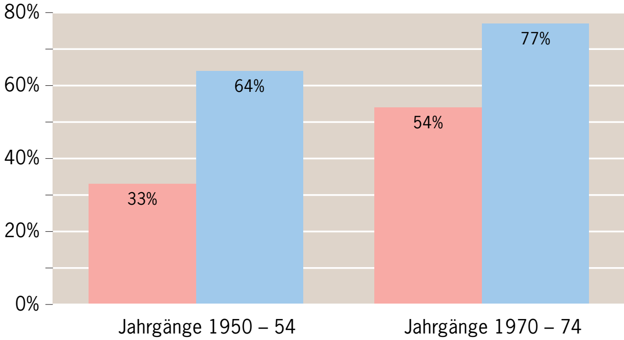
GRAFIK 2

Und dann bald mit Freund und Freundin

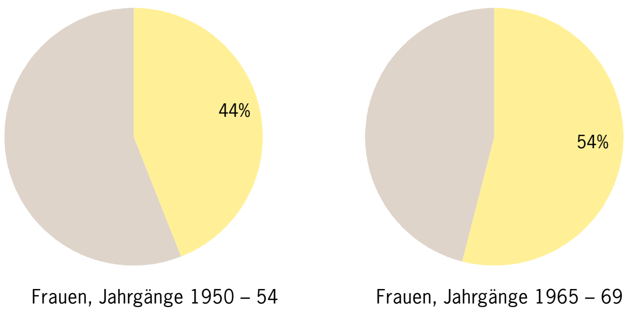
Wenn aber das Elternhaus einmal verlassen ist, wird rascher als früher mit dem Freund oder der Freundin ein gemeinsamer Haushalt gegründet.

Quelle: Mikrozensus Familie 1994/95

Mit 19 wohnen noch bei den Eltern:



Innert 3 Jahren nach dem Auszug aus dem Elternhaus leben mit einem Partner zusammen:



GRAFIK 3

Partnerschaft auf Zeit?

Die Scheidungen nehmen stark zu.

Quelle: Mikrozensus Familie 1994/95

GRAFIK 4

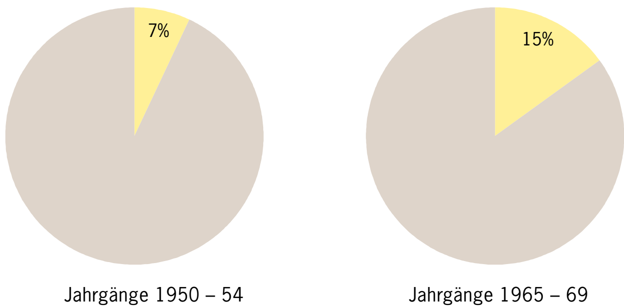
Noch immer der Normalfall: die traditionelle Familie

Die grosse Mehrheit der Eltern mit Kindern lebt noch immer in traditionellen Familien (Mutter, Vater, Kind/er).
Doch leben heute auch viele in Fortsetzungsfamilien (Paar, das gemeinsam mit mindestens 1 Kind aus einer früheren Beziehung wohnt).
Von den Männern und Frauen zwischen 20 und 49, die mit Kind(ern) in einem Haushalt leben, sind es immerhin jeder zwölfte und jede dreizehnte.
Nicht wenige Frauen leben auch in einem Haushalt mit Kind(ern) allein.

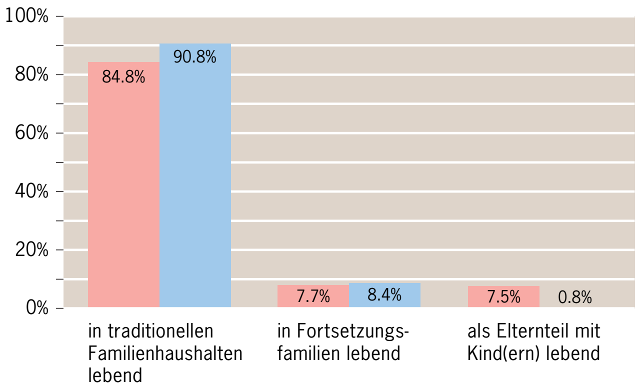
Frauen
Männer

Quelle: Mikrozensus Familie 1994/95

Bis 18 haben die Scheidung der Eltern erlebt:



Familientyp bei Männern und Frauen zwischen 20 und 49



Und woher wissen die das alles?

«Die haben halt Statistiken gemacht.» – «Na schön, aber wie denn? Sie konnten doch nicht sämtliche Leute im Land befragen!» Richtig. Doch interessanterweise ist dies auch nicht immer nötig. Und gerade darin steckt einer der Schlüssel moderner Statistik.

Das Zauberwort heisst **Stichprobe**. Nur – einfach wild drauflos Leute auswählen und fragen, das müsste ja geradezu mit Sicherheit zu falschen Bildern führen! (Was leider bei den populären Laienumfragen nicht selten so ähnlich gemacht wird. Kein Wunder, wenn man hört, mit «Statistik» könne man alles beweisen …)

Wie also soll man die Befragten so auswählen, dass die Auswahl (= die Stichprobe) einen gültigen Schluss auf das Ganze (= die **Grundgesamtheit**) erlaubt?

Hier hat die Wissenschaft «Statistik» eine Antwort. Sie hat gezeigt (und mathematisch bewiesen), dass eine Stichprobe einen gültigen Schluss auf die Grundgesamtheit nur dann erlaubt, wenn sie nach dem Zufallsprinzip gezogen wird.

Bevor wir auf das Thema «Zufallsprinzip» weiter eingehen, müssen wir uns noch etwas klarer darüber werden, was unter «Schluss von der Stichprobe auf die Grundgesamtheit» eigentlich gemeint sein soll.

Statistiker sprechen in diesem Zusammenhang meist von **Schätzung**. Sie machen damit deutlich, dass Aussagen über die Grundgesamtheit, die man auf Grund einer einzelnen Stichprobe machen kann, in der Regel nicht ganz genau sind, also eine gewisse Abweichung vom wahren Wert aufweisen. Die Geschichte mit der Stichprobe hat also auch einen Haken: Wir müssen mit der **Ungenauigkeit** einer Schätzung rechnen! Dass diese Ungenauigkeit aber keineswegs eine verschwommene Sache ist, das ist eine der spannenden Erkenntnisse der wissenschaftlichen Statistik, und davon später (in «Wie genau? Der abschätzbare Fehler»).

Kommen wir nun wieder zurück auf unsere Frage nach der Auswahl einer Stichprobe. Wie wir festgestellt haben, muss dies nach dem Zufallsprinzip geschehen.

Die einfachste Form von Zufallsstichprobe erhalten wir, indem wir zum Beispiel für jedes Element der Grundgesamtheit ein Los in einen Sack (oder in eine Urne) stecken, gut durchmischen und dann blind die gewünschte Anzahl Lose herausholen. Diese Art der Stichprobe heisst **einfache Zufallsstichprobe**. Im Folgenden meinen wir mit «Stichprobe» vorläufig stets die einfache Zufallsstichprobe. Bei der einfachen Zufallsstichprobe hat jedes Element der Grundgesamtheit genau die gleiche Chance, gezogen zu werden. (Später werden wir über kompliziertere Fälle sprechen, die in der Praxis häufig zur Anwendung kommen.)

Wir probieren es aus

Hier ist nicht der Platz, die wissenschaftlichen und mathematischen Grundlagen statistischer Methoden vorzuführen. Doch wollen wir das Prinzip der Stichprobe an zwei einfachen Experimenten veranschaulichen. Wir wollen prüfen, ob man mit einer zufälligen Auswahl von Einzelteilchen aus einem Ganzen den wahren Verhältnissen wirklich nahe kommt. Dabei werden wir merken, dass es gar nicht so einfach ist, die Idee mit den Losen in der Urne anzuwenden. Selbst für eine kleine Grundgesamtheit von 600 Elementen wie im zweiten der folgenden Experimente benötigen wir schon fast einen Computer. Wir beginnen darum zuerst mit einem ganz kleinen Experiment, das wir noch «von Hand» durchführen können: Nehmen wir an, wir möchten die mittlere Schuhgrösse in einer Schulklasse schätzen.

EXPERIMENT I: ***Die mittlere Schuhgrösse der Klasse:***

Nehmen wir mal an, wir stellen die jungen Leute in Sechser-Reihen auf dem Pausenplatz auf. Es macht nichts, wenn in der letzten Reihe Plätze frei bleiben. Wir nehmen zwei Würfel verschiedener Farbe (z. B. blau und rot) und beginnen zu würfeln. Der blaue Würfel zeigt die Reihe an, der rote die Kolonne. Wenn also z.B. blau 2 und rot 3 fällt, dann wird in der 2. Reihe die dritte Schülerin/ der dritte Schüler von links gezogen (Achtung: blau 3, rot 2 führt nicht zum gleichen Platz!). Sie oder er tritt nach vorn, die anderen bleiben an ihrem Platz. Wir würfeln, bis wir 7 Personen gezogen haben; unsere Stichprobengrösse sei also 7. (Würfe, die auf einen leeren Platz treffen, werden wiederholt.)
Wenn wir 7 Personen gezogen haben, notieren wir uns von jeder die Schuhgrösse, summieren sie und dividieren die Summe durch 7. Das ist bereits unser Schätzverfahren: Der Mittelwert der Schuhgrössen in der Stichprobe

schätzt die mittlere Schuhgrösse in der ganzen Klasse. Wir können auch die 7 Schuhnummern der Grösse nach ordnen und die mittlere, also die viertgrösste als Schätzung nehmen. Dieses Schätzverfahren, der so genannte Median, hat auch seine Vorzüge. Zum Beispiel reagiert der Median nicht so stark wie der Mittelwert auf eine extreme Schuhgrösse wie etwa 53. Der Median ist aber meistens ungenauer als der Mittelwert.
Wir können nun auch noch beim Rest der Klasse die Schuhgrösse messen und dann den wahren Mittelwert (oder auch den Median) der Klasse mit unserer Schätzung vergleichen. Wenn wir Spass daran haben, können wir auch noch ganz andere interessante Dinge aus der Stichprobe schätzen, z.B. den Anteil der Einzelkinder in der Klasse oder den Anteil derjenigen, die gerne Ketchup zu Pommes-Frites nehmen. Und wir können natürlich das Experiment wiederholen, um zu sehen, wie der Mittelwert (oder der Median) je nach Stichprobe schwankt.

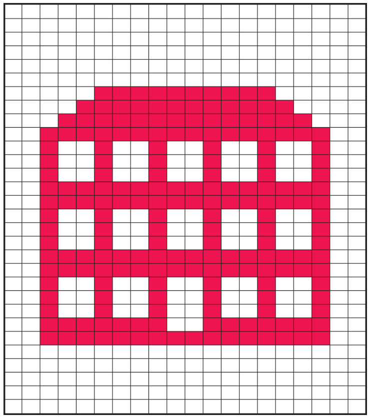
EXPERIMENT II: ***Die Abstimmungs-Vorhersage:*** Jetzt wagen wir uns an ein grösseres Experiment. Den Versuch kann man selbst durchführen mit Häuschenpapier und Klarsichtfolien oder mit einem Computer. Wir stellen uns vor, in einer Gemeinde mit 600 Einwohnern soll in 4 Wochen über die Renovation des Schulhauses abgestimmt werden. Die Schüler sind neugierig und wollen bereits jetzt mit einer Umfrage herausfinden, was die Leute denken.

Wie geht das?
Zuerst brauchen wir eine Grundgesamtheit: die Menge, die wir untersuchen wollen. Wir fabrizieren dazu eine Versuchsfläche mit 600 Elementen. Ein bestimmter Teil davon ist rot (in unserer Umfrage «Schulhaus-Renovation Ja oder Nein?» wären dies z.B. alle Ja-Stimmen). Für unseren Versuch müssen wir die wirkliche Zahl der roten Felder kennen. Wir bestimmen sie hier als 200 (200 rote Felder von 600, d.h. Anteil = 1/3 = 33,3%). Für unser Experiment kennen wir also den wahren Wert, den wir sonst erst mit Hilfe der Stichprobe schätzen können. Als Nächstes ziehen wir eine einfache Zufallsstichprobe der Grösse 150. Und dann berechnen wir unsere Schätzung für den Anteil der roten Felder in der Grundgesamtheit, indem wir einfach die roten Felder in der Stichprobe zählen und durch 150 dividieren.

Schritt für Schritt:

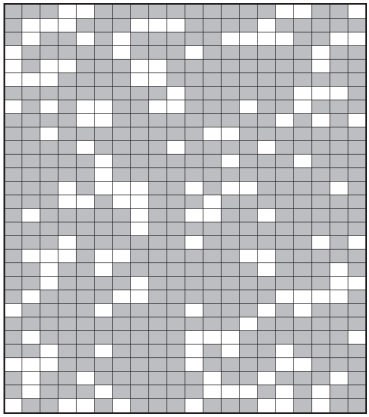
Vorbereitung – die Grundgesamtheit:

Wir stecken auf drei Klarsichtfolien drei rechteckige Versuchsflächen mit je 30 x 20 = 600 gleichen Feldern (Häuschen) ab. In der ersten Versuchsfläche füllen wir beliebige 200 von den 600 Feldern mit einer Farbe aus. (Wir haben hier mit den 200 ein Schulhaus geformt, aber das ist eine völlig unwichtige Spielerei!).

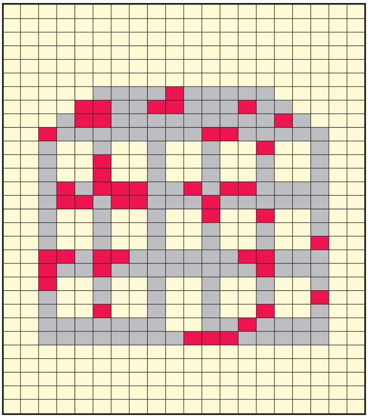


Erste Stichprobe (25% = 150 Felder):

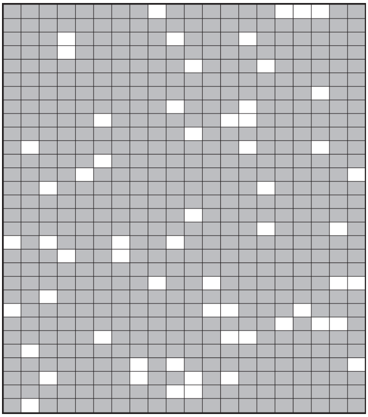
Auf der nächsten Versuchsfläche präparieren wir die erste Stichprobe: nur ein Viertel der Grundgesamtheit soll in die Stichprobe fallen. Dazu müssen wir 150 Felder zufällig auswählen; diese sollen durchsichtig bleiben, die übrigen färben wir mit einer Deckfarbe aus (d.h. die Stichprobe «sieht» nur 150 Elemente). Wir nummerieren die Felder von 1 bis 600, z.B. zeilenweise. Man kann jetzt natürlich Zettelchen mit den Zahlen 1 bis 600 in eine Urne stecken, schütteln und 150 blind herausziehen. Aber das ist aufwändig. Wir haben für die Ziehung einer einfachen Zufallsstichprobe das Computerprogramm Excel eingesetzt.



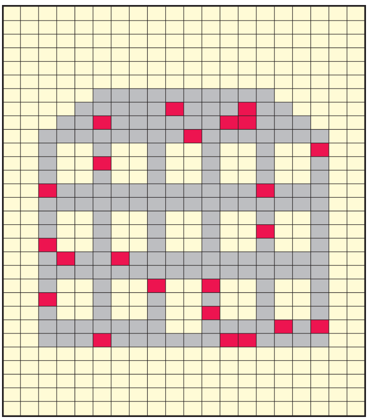
Für diejenigen, die es auf dem Computer selbst ausprobieren möchten: Wir schreiben die Nummern der Felder 1 bis 600 in die erste Kolonne einer Tabelle und in die zweite Kolonne setzen wir eine Zufallszahl zwischen 0 und 1 (Excel-Formel «=ZUFALLSZAHL()»). Dann ordnen wir die beiden Kolonnen nach der zweiten Kolonne, den Zufallszahlen (Es macht nichts, dass diese dabei auch wieder ändern). Die Felder, die jetzt den obersten 150 Nummern in der ersten Kolonne entsprechen, lassen wir dann durchsichtig, d.h. sie bilden unsere einfache Zufallsstichprobe.



Erste Schätzung: Wir legen nun die Stichprobenfolie über unsere Grundgesamtheit. Wir zählen die roten Felder, die wir noch sehen. Bei unserem ersten Versuch waren es 47. Wir dividieren dann die Anzahl rote Felder durch 150. Bei uns gab das 47/150 = 31,3%. Wir sind also recht nahe bei den wahren 33,3%.



Zweite Stichprobe (10% = 60 Felder):
Sie entsteht wie die erste, nur lassen wir hier nur noch 1/10 der Felder durchsichtig, also 60. Wieder: einfache Zufallsstichprobe! In Excel ordnen wir neu nach den Zufallszahlen in der zweiten Kolonne und nehmen nun die Felder, die den obersten 60 Nummern in der ersten Kolonne entsprechen. Wir lassen die Felder mit diesen Nummern auf der Folie durchsichtig.



Zweite Schätzung: Diesmal haben wir 23 rote Felder gezählt. Unsere Schätzung ist also 23/60=38,3%. Die Abweichung vom wahren Wert ist zwar ein bisschen grösser als vorher, aber die Schätzung kann immer noch brauchbar sein.

Es ist tatsächlich so, dass wir bei der kleineren Stichprobe (60 Felder) mit grösseren Abweichungen rechnen müssen. Bei jeder Wiederholung des Experiments werden wir zwar eine leicht andere Schätzung und damit eine etwas andere Abweichung erhalten. Im Durchschnitt wird aber die Abweichung bei der Stichprobe mit 60 Feldern grösser sein als bei der Stichprobe mit 150 Feldern.

Wir haben 60 Stichproben durchgeführt, je 30 kleinere (60 Felder) und grössere (150 Felder).

Unsere Ergebnisse:

Anzahl rote Felder in den 30 kleineren Stichproben (60 Felder):

Die einzelnen Ergebnisse:

21, 21, 17, 22, 18, 22, 17,
23, 17, 17, 21, 18, 25, 23,
18, 14, 25, 17, 23, 18, 20,
21, 16, 21, 24, 18, 27, 21,
16, 22

Ergebnisse zwischen		
14 – 16	III	3
17 – 18	IIIIIIIIII	10
19 – 21	IIIIIII	7
22 – 23	IIIIII	6
24 – 27	IIII	4

Zu erwartender Wert: 20
Durchschnittliche Abweichung absolut: 2,7
Durchschnittliche Abweichung in %: 13,5

Anzahl rote Felder in den 30 grösseren Stichproben (150 Felder):

Die einzelnen Ergebnisse:

53, 46, 46, 47, 51, 50, 50,
47, 44, 51, 53, 38, 64, 55,
48, 56, 46, 58, 55, 43, 59,
57, 46, 54, 54, 54, 51, 48,
47, 40

Ergebnisse zwischen		
38 – 43	III	3
44 – 46	IIIII	5
47 – 53	IIIIIIIIIIII	12
54 – 56	IIIIII	6
57 – 64	IIII	4

Zu erwartender Wert: 50
Durchschnittliche Abweichung absolut: 4,6
Durchschnittliche Abweichung in %: 9,3

Die Schätzungen 1 – 30 schwanken zwischen 14 und 27 in den kleineren bzw. zwischen 38 und 64 in den grösseren Stichproben. (Übrigens kommt ihr Durchschnitt von 20,1 bzw. 50,4 dem wahren Wert erstaunlich nahe!) Doch ein grosser Teil von ihnen schart sich enger um den wahren Wert:

Zum Beispiel weichen bei den kleineren Stichproben

15 Schätzungen	weniger als 3 Felder vom wahren Wert ab
26 Schätzungen	weniger als 5 Felder
… und 28 der 30 Schätzungen immerhin noch	weniger als 6 Felder

Und bei den grösseren Stichproben weichen

19 Schätzungen	weniger als 5 Felder vom wahren Wert ab
25 Schätzungen	weniger als 8 Felder
… und 27 der 30 Schätzungen immerhin noch	weniger als 10 Felder

Und wenn wir jetzt eine Stichprobe von 150 Einwohnern nicht in einer Gemeinde mit 600 Einwohnern, sondern in Tokyo machen? Wird dann die Schätzung nicht zu ungenau, um brauchbar zu sein? Das Erstaunliche ist, dass die Grösse der Grundgesamtheit gar nicht so wichtig ist: Die Stichprobe von 150 Einwohnern in der grössten Stadt der Welt ist fast so genau wie in der kleinen Gemeinde! Etwas mehr darüber steht im Anhang.

Wie genau? *Der abschätzbare Fehler*

Unser kleines Experiment hat uns gezeigt, dass Zufallsstichproben gute Chancen liefern, dem wahren Wert einer Grundgesamtheit erstaunlich nahe zu kommen. Zieht man eine Vielzahl von Stichproben, werden viele der entsprechenden Schätzungen dem wahren Wert sehr nahe kommen, einige werden etwas mehr davon abweichen und sehr wenige werden weit daneben liegen. Wenn wir eine einzige Stichprobe ziehen, können wir daher grundsätzlich voraussagen: Die Wahrscheinlichkeit, dass die entsprechende Schätzung dem wahren Wert nahe kommen wird, ist sehr hoch, die Chance, dass sie weit daneben liegen wird, ist gering. Dieses Wissen können Statistiker ausnützen, um aus einer einzigen Stichprobe eine Aussage abzuleiten über die typische Abweichung von solchen Schätzungen vom wahren Wert. Sie stützen sich dabei auf die so genannte **Standardabweichung**. Den mathematischen Weg zu ihrer Berechnung müssen wir hier nicht darlegen, aber wir wollen noch einmal betonen, dass die Stichprobe das Prinzip der Zufallsauswahl erfüllen muss.

Die Standardabweichung ist also ein Mass für den Fehler, mit dem wir rechnen müssen. Sie hilft uns, die Ungenauigkeit der Schätzung und damit unsere Unsicherheit genauer zu beschreiben. Oft drückt man diese Unsicherheit mit Hilfe eines Zahlenbereichs aus, der den wahren Wert mit grosser Sicherheit (z.B. mit einer Wahrscheinlichkeit von 95%) umfasst. So ein Bereich heisst **Vertrauensbereich**, oder Vertrauensintervall. Er ist umso grösser, je ungenauer eine Stichprobe ist.

In unserer ersten Stichprobe (mit 150 Feldern) für die «Abstimmungs-Vorhersage» haben wir den Ja-Stimmen-Anteil auf 31,3% geschätzt. Der geschätzte Wert der Standardabweichung ist 3,3%. Wir können sagen: Der wahre Wert liegt mit einer Wahrscheinlichkeit von 95% bei 31,3% plus oder minus das Zweifache der Standardabweichung, also im Bereich zwischen 24,7% und 37,9%. Für die kleinere Stichprobe (mit 60 Feldern) resultiert eine grössere Standardabweichung (5,5%) und damit auch ein grösserer Vertrauensbereich (38,3% ± 11,0%). (Ausführlicheres zu Standardabweichung und Vertrauensbereich steht im Anhang.)

Statistiker liefern die Ergebnisse von Stichprobenerhebungen oft mit solchen Vertrauensbereichen. Zu unserer Grafik «**Noch immer der Normalfall: die traditionelle Familie**» (S. 6/7) lautet das genaue Resultat aus der Erhebung so:

Der 95%-Vertrauensbereich für den Anteil der Frauen in Eineltern-Haushalten unter allen Frauen im Alter zwischen 20 und 49 beträgt 7,45% ± 1,22%, reicht also von 6,23% bis 8,67%.

Je nach Fragestellung, nach Tragweite, vorhandenem Geld usw. verlangen wir hohe Genauigkeit – oder können uns schon mit einer geringeren zufrieden geben. Von gewissen Zahlen können millionenschwere Entscheidungen abhängen!

Die Sache mit dem Zufall

So viel steht also fest: Stichproben sind nur sinnvoll, wenn sie nach Zufallsprinzip gezogen werden. Aber da heisst es aufpassen. «Blind» heisst nicht immer «Zufall»!

Angenommen, wir möchten herausfinden, welcher Prozentsatz der Leute in unserer Umgebung für den EU-Beitritt der Schweiz ist. Also: wir machen eine Umfrage! Wir stehen zwei Vormittage vor drei wichtige Einkaufszentren, fragen, zählen alle die Ja- und die Nein-Antworten sauber zusammen …

Das Resultat kann nur falsch sein. Wer besucht Dienstag und Mittwoch vormittags ein solches Einkaufszentrum?

Das ist natürlich ein krasses Beispiel. Aber wir kennen die populären Leser- oder Hörerumfragen. Welche Auswahl haben wir getroffen, wenn wir z.B. nur Leute zählen, die eine bestimmte Sendung einschalten?

Manchmal merkt man auch nicht, dass ein Teil der gemeinten Grundgesamtheit überhaupt keine Chance hatte, in die Stichprobe zu kommen – dass man sozusagen mit dem falschen Netz gefischt hat.

Einen peinlichen Fall gibt es dazu aus der Geschichte der USA. Für die Präsidentenwahlen im Jahr 1948 sagte ein Meinungsforscher einen Erdrutschsieg des republikanischen Kandidaten Dewey über den Demokraten Truman voraus – und traf haushoch daneben. Warum? Er hatte doch eine umfangreiche telefonische Stichprobenumfrage gemacht.

Nur – etwas war ihm dabei entgangen: 1948 hatten vor allem die reicheren Bevölkerungsschichten Telefonanschlüsse – und Truman war der Favorit der unteren Schichten.

In unserem Experiment II mit der Abstimmungs-Vorhersage wäre das etwa so, als würde in einem dunklen Raum ein Lichtkegel nur auf den oberen Teil der gesamten Folie fallen, so dass man vom Rest gar nichts bemerken würde.

Wir können also nicht einfach ins Blaue hinaus «stichpröbeln»! Wir brauchen Auswahl-. und Kontrollverfahren, um zu gewährleisten, dass jedes Element der Grundgesamtheit eine kontrollierte Chance hat, in die Stichprobe zu gelangen.

Die Welt ist nicht einfach: *Köpfchen gefragt!*

In unseren kleinen Experimenten hatte jedes Element die gleiche Chance, in die Stichprobe zu kommen. In der Realität muss man aber oft Stichproben benützen, bei denen alle Elemente zwar eine genau kontrollierte, aber nicht die gleiche Ziehungs-Chance haben. In Fachausdrücken gesagt: Sehr häufig müssen die Fachleute komplexe Stichprobenpläne und spezielle Schätzverfahren entwickeln.

Nehmen wir als Beispiel eine Statistik über ein bestimmtes Hautleiden nach Altersklassen, und zwar in Altersstufen von jeweils 10 Jahren: 0 – 9-Jährige, 10 – 19-Jährige, …, 90 – 100-Jährige. Sollte eine einfache Zufallsstichprobe auch für die Altersklasse der 90 – 100-Jährigen noch genügend viele Daten enthalten, dass eine aussagekräftige Schätzung über diese Altersklasse möglich wird, so müsste die gesamte Stichprobe untragbar gross sein.

In einem solchen Fall drängt sich die Methode der **geschichteten Stichprobe** auf. In unserem Beispiel wird dabei für die Altersklasse der 90 – 100-Jährigen eine gleich grosse einfache Zufallsstichprobe gezogen wie für diejenige der 60 – 70-Jährigen, obschon ihre Zahl in der Grundgesamtheit um ein Vielfaches kleiner ist. Damit ist nun die Chance, in die Stichprobe zu gelangen, nicht mehr in jeder der Schichten gleich. Anders als in unseren kleinen Experimenten kann uns der einfache Durchschnitt in der Stichprobe (das «Stichprobenmittel») keine brauchbare Schätzung für die Verbreitung des Hautleidens in der gesamten Population liefern. Wir müssen nun eben mit einer Gewichtung berücksichtigen, dass die 90 – 100-jährigen in der Stichprobe übervertreten sind. Doch haben wir auch einen Vorteil gewonnen: Ein solches Vorgehen führt dazu, dass wir auch für kleine Untergruppen (hier die 90 – 100-jährigen) zu relativ genauen Schätzungen kommen, ohne enorme Kosten zu verursachen.
Voraussetzung für dieses Vorgehen ist allerdings, dass wir bereits vor der

Ziehung von jedem Element der Grundgesamtheit wissen, in welche Altersklasse es fällt.

Es machen selten alle mit: *Antwortausfälle*

Ein weiterer Grund, weshalb Schätzverfahren oft etwas komplizierter sind als in unseren kleinen Experimenten, sind die so genannten Antwortausfälle: also die Tatsache, dass praktisch nie alle Befragten eine Antwort geben. So könnte es z.B. gut sein, dass bei Befragungen über das Einkommen viele die Antwort verweigern. Vielleicht sind es gerade diejenigen, deren Einkommen besonders hoch ist (oder auch besonders tief?). Mit dem einfachen Durchschnitt aus den Antworten käme dann ein schiefes Bild heraus, auch wenn wir ursprünglich eine einfache Zufallsstichprobe gezogen haben.

Wie zahlreich die Ausfälle bei einer Erhebung sein können, zeigt das folgende Beispiel: Es stammt aus dem Landesindex der Konsumentenpreise, dem bekannten «Fiebermesser der Teuerung». Er wird nach einem so genannten «Warenkorb» berechnet. Dieser Warenkorb repräsentiert die durchschnittlichen Ausgaben eines Schweizer Haushaltes für Nahrungsmittel, Miete, Kleider, Versicherungen usw. Er wird ermittelt durch die Befragung einer Stichprobe von Haushalten. Der Anteil der Haushalte, die bei der Befragung von 2000 effektiv mitmachten, liegt bei rund einem Drittel der ursprünglichen Stichprobe. Zwei Drittel der ausgewählten Haushalte wurden nicht erreicht oder waren nicht teilnahmefähig (Sprach-, Alters- oder Gesundheitsprobleme), verweigerten die Teilnahme oder fielen während der Erhebung aus. Es liegt nahe, dass bestimmte Haushalt-Typen unter den Ausfällen besonders häufig vertreten waren – etwa Haushalte mit älteren Menschen oder mit jungen Alleinlebenden. Diese Haushalt-Typen wären dann nicht richtig im Landesindex vertreten, und dieser ergäbe so vielleicht ein schiefes Bild der Teuerung.

Es gibt verschiedene Verfahren, wie man mit solchen Ausfällen umgehen kann. Man kann z.B. die Schätzungen vergleichen mit Informationen über die Grundgesamtheit aus anderen Quellen wie Gemeinderegistern oder anderen Erhebungen. Allerdings muss man da immer sehr sorgfältig ans Werk gehen und Acht geben, dass zwischen diesen Quellen nicht unvereinbare Unterschiede bestehen – etwa in der Definition der Grundgesamtheit oder in den Erhebungsmethoden.

Eine besondere Form von Quellen, aus der sich oft Zusatzinformationen über die Grundgesamtheit ziehen lassen, sind Erhebungen, welche die volle Grundgesamtheit untersucht haben, wenn auch in einem anderen Zusammenhang. Eine solche *Vollerhebung* ist die Eidgenössische **Volkszählung**.

Die Volkszählung liefert uns für unser Beispiel zum Konsum der privaten Haushalte die Angaben darüber, wie viele von welchen Haushalttypen es bei uns gibt (mit 1, mit 2 oder mehr Personen, mit oder ohne Kinder etc.). Wenn wir nun unter den Antwortenden zu viele vom einen Typ und zu wenige vom andern haben, können wir dies bei der Schätzung korrigieren, indem wir die Stärke der verschiedenen Gruppen so **gewichten**, dass sie der Grundgesamtheit wieder entsprechen.

Fazit *oder: wann ist eine Erhebung «repräsentativ»?*

Wir halten fest: Damit die Ergebnisse einer Erhebung brauchbar sind, müssen folgende Bedingungen erfüllt sein:

- Es braucht eine Stichprobe, die nach dem **Zufallsprinzip** gezogen worden ist. In der Regel wird es sich dabei nicht um eine einfache Zufallsstichprobe handeln, sondern um eine komplexere Art von Stichprobe. Um einen guten Stichprobenplan zu erstellen, müssen wir oft bereits einiges über die Grundgesamtheit wissen.

- Es ist ein **Schätzverfahren** notwendig, um von der Stichprobe auf die Grundgesamtheit zu schliessen. Dieses Schätzverfahren muss berücksichtigen, wie die Stichprobe gezogen worden ist und welche Antworten bei der Erhebung ausgefallen sind. Es berücksichtigt häufig auch Informationen über die Grundgesamtheit, die aus anderen Quellen als der Stichprobe stammen.

- Ob die Ergebnisse in einem konkreten Fall brauchbar sind, hängt natürlich auch von ihrer **Genauigkeit** ab. Sie wird, wie wir gesehen haben, unter anderem durch die Grösse der Stichprobe beeinflusst. Ob die Genauigkeit genügt, wird schliesslich auch vom **Zweck der Erhebung** bestimmt: je nach Zweck kann sie geringer oder muss sie höher sein.

Eine Erhebung, die diese Bedingungen erfüllt, vermittelt ein ausreichendes Bild der Grundgesamtheit. Wir bezeichnen sie in diesem Fall oft als repräsentativ.

Anhang: *Mathematisches zum Thema «Stichprobe»*

Berechnung der Standardabweichung und des Vertrauensbereichs

Die Statistiker haben eine Formel entwickelt, mit der sich die **Standardabweichung** aus der Stichprobe berechnen lässt. Wir führen diese Berechnung für unser Experiment mit der «Abstimmungs-Vorhersage» (Kapitel «Wir probieren es aus») durch, ohne den Rechnungsweg näher herzuleiten.

Vorgaben:

N = 600 Grundgesamtheit

n = 150 Stichprobenumfang

m = 47 Anzahl rote Felder in unserer Stichprobe

$p_m = m : n$ = 47 : 150 = 0,313 = 31,3 % Anteil der roten Felder in der Stichprobe

Die «Standardabweichung» des Stichprobenanteils p_m bei einer einfachen Zufallsstichprobe wird so geschätzt:

$$\hat{\sigma} = \sqrt{1-\frac{n}{N}}\sqrt{\frac{p_m\,(1-p_m)}{n-1}} = \sqrt{1-\frac{150}{600}}\sqrt{\frac{0{,}313\times(1-0{,}313)}{149}} = 0{,}866\times 0{,}038 = 0{,}033$$

Der **Vertrauensbereich** ist im einfachsten Fall ein Intervall, also ein durchgehender Abschnitt zwischen zwei Zahlenwerten; wir sprechen dann von einem Vertrauensintervall (auch «Konfidenzintervall» genannt). Ein Vertrauensintervall wird besonders oft für eine Sicherheit von 95% berechnet und so angegeben, dass man zur Schätzung die doppelte Standardabweichung hinzuzählt und abzieht. Für unser Experiment ist das Vertrauensintervall somit ungefähr gleich

$$p_m \pm 2\hat{\sigma} = 0{,}313 \pm 2\times 0{,}033 = 31{,}3\,\% \pm 6{,}6\,\%$$

Wir können nun sagen: Der wahre Anteil roter Felder liegt mit 95 % Sicherheit zwischen 31,3 % ± 6,6 %, also zwischen 24,7 % und 37,9 % roter Felder.

Der Ausdruck «mit 95 % Sicherheit» ist natürlich eine vereinfachte Sprechweise. Der wahre Wert liegt nämlich entweder innerhalb 31,3 % ± 6,6 % oder eben ausserhalb. Wenn wir aber sehr viele Stichproben ziehen würden, dann würde der wahre Wert in ungefähr 95 % der Stichproben tatsächlich im

Vertrauensintervall liegen. Ganz korrekt würde es heissen: Das genäherte 95%-Vertrauensintervall für den wahren Anteil der roten Felder ist 31,3% ±6,6%.

«Wie gross müssen Stichproben sein?» *oder: Das Verhältnis zur Grundgesamtheit*

Man hört oft, dass es bei Umfragen eine optimale Stichprobengrösse gibt, welche unabhängig von der Grösse der Grundgesamtheit ist. Anders gesagt: Führen wir eine Umfrage in der Bevölkerung der Schweiz und eine gleiche in den USA durch, und befragen wir dabei beide Male 2000 Personen, so fällt das Resultat für die Schweiz interessanterweise nicht wesentlich genauer aus als für die USA, obschon doch die Bevölkerung der USA rund 35 mal grösser ist als die der Schweiz. Wie kommt das?

Das kann man der obigen Formel für die Standardabweichung $\hat{\sigma}$ des Anteils in der Stichprobe ansehen: Das N, die Grösse der Grundgesamtheit, kommt nur im ersten Term vor:

$$\sqrt{1-\frac{n}{N}}$$

Wenn wir diesen Term ausrechnen für die Schweiz, ergibt sich:

$$\sqrt{1-\frac{2000}{7'000'000}} = 0{,}999857$$

Für die USA wird der Term nur unwesentlich grösser:

$$\sqrt{1-\frac{2000}{250'000'000}} = 0{,}999996$$

Die Standardabweichung wird also bei der gleichen Stichprobengrösse in der Schweiz nur unwesentlich kleiner sein als in den USA.

Die Statistiker können die Formel für die Standardabweichung auch umkehren und damit die Grösse der Stichprobe berechnen, welche benötigt wird, um eine gewünschte Genauigkeit (Standardabweichung) zu erreichen. Allerdings brauchen sie dafür mindestens eine grobe Ahnung, wie gross der Anteil p_m in Wahrheit ist.

Bundesamt für Statistik
Espace de l’Europe 10
2010 Neuchâtel

www.statistik.admin.ch