Lügen und Statistik

FRANK DUCKWORTH, ENGLAND

***Zusammenfassung**: Dies ist eine gekürzte und bearbeitete Fassung der Royal Statistical Society's Lecture des Jahres 2004.*

1 Statistik und Statistik

Es wird mein Ziel sein, alle Mythen über den Zusammenhang zwischen Statistik und Lügen zu prüfen und zu zerstreuen.

In den 1950-er Jahren berichtete *The (London) Times*, dass Sir Winston Churchill Mark Twain zitiert habe, der seinerseits dem britischen Premierminister Disraeli folgendes Zitat zugesprochen haben soll:

> Es gibt drei Arten von Lügen: Lügen, verdammte Lügen und Statistik.

Seitdem müssen wir armen Statistiker mit diesem Ausspruch leben. (Mittlerweile gibt es erhebliche Zweifel an der wahren Quelle dieses Zitats. Vermutlich stammt es von Leonard Henry Courtney (1832 – 1918).)

Natürlich ist es sehr ungerecht, Statistik mit Lügen gleichzusetzen (als Statistiker muss ich so etwas natürlich sagen). Ich werde im Folgenden also die Statistik und die Statistiker gegen den Lügen-Vorwurf verteidigen, und das mache ich, indem ich Beispiele anführe, bei denen Statistik und Lügen fälschlicherweise in Beziehung gesetzt wurden.

In Wirklichkeit gibt es zwei verschiedene Bedeutungen von „Statistik". Diese kann einerseits aus Daten bestehen (und zwar ausschließlich aus Daten ohne jegliche Schlussfolgerungen), andererseits ist Statistik aber auch eine ausübbare Wissenschaft wie Mathematik oder Physik.

Wenn man Statistik betreibt, verwendet man mathematische Techniken, die gelernt und richtig angewendet sein wollen und die man nur richtig versteht, wenn man sich mit Daten befasst hat.

Statistische Techniken wurden während der 1920-er und 1930-er Jahre prominent, und die Pioniere waren Briten wie etwa Sir Ronald Fisher, der an der landwirtschaftlichen Versuchsanstalt in Rothamsted arbeitete. Dies geschah lange nach dem Tod von Disraeli (und dem von Courtney), so dass schon deswegen deren Zitat auf die modernen statistischen Techniken gar nicht zutreffen kann.

2 Irreführende Statistik

Zwar kenne ich nicht den Kontext des erwähnten Zitats, nehme aber an, dass deutlich gemacht werden sollte, dass Statistik irreführend sein kann. Aber das liegt nicht am Statistiker, sondern ist die Schuld unqualifizierter Laien, die statistische Aussagen falsch verstehen oder falsche Schlüsse daraus ziehen. Einige Beispiele:

Das erste Beispiel bezieht sich auf Schul-Rankings, wobei die Rankings durch Abiturnoten definiert werden. Eine Schule, die beim Ranking an der Spitze steht, muss aber nicht die beste Schule sein. Um Sieger beim Ranking zu werden, braucht man zwei Voraussetzungen: Guten Unterricht und gute Schüler. Wenn eine Schule Eingangstests erheben kann, hat sie ohnehin beim Ranking Vorteile. Wenn eine Schule viele Fördermaßnahmen für schwächere Schüler anbieten muss, wird sie vermutlich nicht beim Ranking gut abschneiden, obwohl sie eine gute Schule ist.

Daher sind Rankings kein Maß für die Unterrichtsqualität einer Schule, so dass sie keine Entscheidungsgrundlage für Eltern sein sollten, wenn sie sich überlegen, wohin sie ihre Kinder schicken sollen. Ein besseres Qualitätsmaß würde den Zugewinn bei den Schülern messen und nicht nur das am Ende erreichte Niveau.

Das zweite Beispiel bezieht sich auf die Besucherzahlen bei Fußballspielen in England. In den Jahren kurz nach dem zweiten Weltkrieg waren die Besucherzahlen höher als jemals zuvor. Während der 1950-er und in den frühen 1960-er Jahren gab es dann einen ständigen Niedergang, da der Besuch eines Fußballspiels am Samstagnachmittag mit anderen aufkommenden Interessen wie Fernsehen oder Autofahren konkurrieren musste. Nach dem Sieg von England bei der Weltmeisterschaft 1966 gingen die Zuschauerzahlen wieder nach oben und blieben auch für einige Jahre hoch. Als 1970 England vorzeitig ausscheiden musste, sanken sie wiederum.

Während der frühen 1970-er Jahre nahm sich die Presse sehr stark der rückgängigen Zuschauerzahlen an und veröffentlichte jede Woche die aktuellen Zahlen, die immer niedriger als in der Saison davor zu sein schienen. Ich fing an, mich dafür zu interessieren, da ich den Verdacht hatte, dass es keine ordentliche und saubere statistische Analyse dieser Daten vor der Veröffentlichung gegeben hatte. In der Tat fand ich einige Ungereimtheiten. Hier ist ein Bericht des *Daily Telegraph* mit den entsprechenden Zahlen; man beachte die Bemerkung „Verglichen mit dem letzten Jahr, als es drei Spiele mehr gab“:

> Die Besucherzahl letzten Samstag war um 48.514 Zuschauer niedriger als am entsprechenden Samstag des letzten Jahres, als es drei Spiele mehr gab. Hier sind die Zahlen (mit den entsprechenden Vorjahreszahlen in Klammern):

Bereich 1	175.902	(234.523)
Bereich 2	143.222	(136.008)
Bereich 3	82.155	(78.991)
Bereich 4	53.091	(53.362)
Summe	454.370	(502.884)

Ich fand heraus, dass die drei zusätzlichen Spiele des Vorjahres alle im Bereich 1 stattgefunden hatten. Wenn man das weiß, sieht man, dass in Wahrheit der Trend in den Zuschauerzahlen aufwärts ging. Die veröffentlichte Tabelle erzeugt die völlig falsche Botschaft.

3 Was kann Statistik, und was kann sie nicht?

Die Statistik hat die schwierige Aufgabe, Daten zu untersuchen und aus ihnen Schlussfolgerungen zu ziehen. Das Hauptproblem dabei ist, dass ein Datensatz auf ganz unterschiedliche Weisen zustande gekommen sein kann. Wenn man die datenerzeugende Situation kennt, kann man sich überlegen, wie der resultierende Datensatz aussehen könnte. Aber es ist eine ganz andere Aufgabe, allein aus dem Datensatz auf die datenerzeugende Situation zu schließen. Da ein solcher Schluss nie sicher ist, kann es vorkommen, dass die datenerzeugende Situation in Wirklichkeit ganz anders war. Diese Tatsache bringt womöglich manche Leute dazu, Statistik mit Lügen in Verbindung zu bringen, da sie die unsichere Natur statistischer Schlüsse nicht verstehen.

Nehmen wir zum Beispiel eine Stichprobe mit 20 Walisern und eine andere mit 20 Engländern. Nehmen wir an, dass die Durchschnittsgröße in der walisischen Stichprobe 1,80 m beträgt und die bei den Engländern 1,73 m. Kann man hieraus den Schluss ziehen, dass die Waliser im Durchschnitt größer sind als die Engländer?

Natürlich nicht. Es gibt sehr viel Variation in den Körpergrößen, und es mag sein, dass die walisische Stichprobe rein zufällig viele große Leute umfasste. Aber wenn die Stichprobengröße jeweils 200 betragen würde und wenn die walisische Stichprobe einen Durchschnittswert von 1,78 m und die englische einen von 1,73 m, sähe die Sache etwas anders aus. Obwohl die Differenz nun viel kleiner ist, könnte man eher zur Ansicht neigen, die Waliser seien größer, da ja die Stichproben nun viel umfangreicher waren.

Aber in keinem Fall kann man aufgrund der Stichproben *beweisen*, dass die Waliser größer sind als die Engländer. Das einzige, was die Statistik leisten kann, ist zu sagen, mit welcher Wahrscheinlichkeit die vorliegenden Daten mit einer gewissen Hypothese vereinbar sind oder nicht.

Oft ist es viel zu aufwändig, die gesamte Population zu messen. Daher zieht man Stichproben und versucht mit Hilfe statistischer Analysen auf die gesamte Population zu schließen.

Aber hier ist das Problem. Mit Stichroben kann man niemals etwas über die Gesamtpopulation *beweisen*. Alles, was man tun kann, ist, eine Hypothese zu formulieren, auf die Daten zu schauen und dann zu sagen, ob sie mit der Hypothese konsistent sind oder nicht. In unserem Fall heißt die Hypothese

> Es gibt keinen Unterschied in der Körpergröße zwischen Walisern und Engländern.

Dann sehen wir auf die Messungen, d.h. auf die Daten der Stichprobe, machen etwas Mathematik und sagen dann: Falls es wirklich keinen Unterschied zwischen den durchschnittlichen Körpergrößen gibt, ist die Wahrscheinlichkeit, bei einer 200-elementigen Stichprobe den gemessenen Unterschied zu bekommen, nur etwa 20 %. Ob das nun bedeutet, dass Waliser tatsächlich größer sind als Engländer oder ob das bedeutet, dass zufälligerweise die walisische Stichprobe viele große Leute umfasste, das beantwortet die Statistik nicht. Statistische Analyse kann nur Wahrscheinlichkeiten liefern (in unserem Fall 20 %).

Es hat sich im Laufe der Jahre eingespielt, dass man Obacht geben sollte, wenn die Wahrscheinlichkeit für eine Beobachtung nur 5 % beträgt; man spricht dann von statistischer Signifikanz. Das bedeutet aber nicht, dass man eine Wahrheit bewiesen hätte. Es kommt in statistischen Untersuchungen immer wieder vor, dass eine Messung statistisch signifikant ist, ohne dass irgendetwas dahinter steckt. Auch das ist nicht die Schuld der Statistik.

4 Ursache und Wirkung

Ein weiterer Bereich, in dem Statistik gerne missverstanden wird, ist die Verwechslung von Korrelation mit Verursachung. So gab es eine Zeitungsschlagzeile

> „Reiche leben länger!“,

über einen Bericht über eine statistisch signifikante Korrelation zwischen Lebenserwartung und

Einkommen. Dieser Bericht war sicherlich korrekt, aber man darf nicht aus Korrelationen auf Ursache und Wirkung schließen.
Der Zeitungsbericht erweckte den Eindruck, dass man sich für Geld ein langes Leben kaufen könne: Mit Geld kann man sich bessere Behandlungen leisten oder schneller an ein Klinikbett kommen. Möglicherweise ist daran etwas Wahres, aber die Daten können auch auf ganz andere Arten interpretiert werden.
Die Fallstricke bei Korrelationen wurden vor einigen Jahren durch eine berühmte Studie über Störche und Babies deutlich. Ein Statistiker suchte zufällig 50 schwedische Städte aus und zählte dort die bewohnten Storchennester. Dann ging er zu den jeweiligen Einwohnermeldeämtern, um zu erfahren, wie viele Neugeborene es in dem betreffenden Jahr gab. Die Korrelation war erstaunlich: Die Städte mit den meisten Storchennestern hatten die höchste Geburtenrate und umgekehrt. Natürlich war die Erklärung naheliegend: Beide Größen (Anzahl der Nester und Anzahl der Neugeborenen) standen unabhängig voneinander mit der Größe der jeweiligen Stadt in enger Beziehung. Je größer die Stadt, um so größer war die Anzahl der Schornsteine, die für Storchennester geeignet waren. (Siehe auch Matthews 2001.)

5 Das Gesetz der Mittelwerte

Auch Mittelwerte geben zu Verwirrungen Anlass. So empört man sich häufig darüber, dass so viele Leute weniger als der Durchschnitt verdienen. Vor einigen Jahren gab es in *The Times* dazu mehrere Leserbriefe. Ein Politiker hatte gesagt, dass jeder, der weniger als den Durchschnittsverdienst erhalte, arm sei. (In Wahrheit wird Armut so definiert: Man ist arm, wenn man weniger als die Hälfte des Durchschnittsverdienstes bekommt.) Es gab Leserbriefe mit der Meinung, dass man dem nur entgehen könne, wenn jeder gleich viel verdiene. Es gab auch einen ebenso dummen Leserbrief mit der Meinung, dass bei jeder Einkommensverteilung, sei sie auch noch so gerecht, immer 50 % der Leute weniger als den Durchschnittsverdienst erhalten würden. Dieser Schreiber hatte natürlich die Begriffe „arithmetisches Mittel“ und „Median“ verwechselt.
Die Auseinandersetzung wurde abrupt mit dem folgenden Leserbrief beendet:

> Angesichts der Tatsache, dass es einige Waliser gibt, die leider nur ein Bein haben, und dass es eine noch etwas kleinere Zahl von Walisern gibt, die gar keine Beine haben, kommt man zu der Aussage, dass jeder Waliser im Durchschnitt 1,9995 Beine hat. Daraus lässt sich schließen, dass 99,95 % aller Waliser mehr als die durchschnittliche Zahl von Beinen haben.

6 Über Münzwürfe

Wie groß ist die Wahrscheinlichkeit, dass es bei 5 Münzwürfen einen Fünfer-Pasch (d.h. ZZZZZ oder WWWWW) gibt? Die Antwort ist natürlich 1/16. Die Wahrscheinlichkeit für WWWWW ist 1/32. Was wird man nun vom 6. Münzwurf erwarten, wenn man schon WWWWW erhalten hat? Natürlich stehen die Chancen 50 zu 50 für erneutes W. Aber was wird man vom 11. Münzwurf erwarten, wenn ihm WWWWWWWWWW vorausgegangen ist? Immer noch eine 50-zu-50-Chance für W?

7 Bayes im Gericht

Klassische statistische Methoden sagen, dass 10-mal Wappen in Serie mit der Annahme, die Münze sei nicht gezinkt, schwer verträglich ist. Klassische statistische Methoden sagen einem nicht, mit welcher Wahrscheinlichkeit man nach 10-mal Wappen wieder Wappen zu erwarten hat. Während der letzten 40 Jahre haben die Statistiker eine ganz andere Vorgehensweise populär gemacht: Es wird nicht mehr untersucht, mit welcher Wahrscheinlichkeit Daten zu erwarten sind, wenn eine bestimmte Hypothese zutrifft, sondern es wird die viel interessantere Frage beantwortet, mit welcher Wahrscheinlichkeit eine gewisse Hypothese zutrifft, wenn die und die Daten vorliegen. Da die Mathematik dazu auf dem Satz von Bayes (Reverend Thomas Bayes lebte im 18. Jahrhundert) beruht, wird dieser Zweig der Statistik nach Bayes benannt.

Ein gutes Beispiel für die Anwendung der Bayes-Statistik sind Gerichtsentscheidungen, die mit DNA zu tun haben.
Wenn man am Ort eines Verbrechens Fingerabdrücke einer bestimmten Person findet, kann man sich zu 100 % sicher sein, dass diese Person am Tatort war, da Fingerabdrücke absolut einzigartig sind: Es gibt keine zwei Personen mit denselben Fingerabdrücken. Eine ebenso wirkungsvolle Methode benutzt die DNA, die aus Blutspuren oder anderen Körperausscheidungen identifiziert werden kann. Mittlerweile ist die DNA-Methode ebenso zuverlässig wie die Fingerabdrucks-Methode. Allerdings war das vor 10 Jahren noch nicht so. Damals konnte ein (noch nicht so ausgeschärftes) DNA-Profil etwa in Großbritannien noch von 10 Leuten geteilt werden. Daher war eine DNA-Übereinstimmung zwar sehr selten, aber noch kein sicherer Beweis vor Gericht.

Die damit zusammenhängende statistische Vorgehensweise wird am besten an einem Beispiel deutlich. Eine Frau wurde vergewaltigt und das (damals noch unscharfe) DNA-Profil des Täters ermittelt. Mr. Jones hatte dasselbe DNA-Profil und wurde der Tat verdächtigt, obwohl die Frau ihn nicht als Angreifer identifizieren konnte. Da die Wahrscheinlichkeit, dass Mr. Jones unschuldig war und nur zufällig das passende DNA-Profil hatte, nur 1 zu 3 Millionen betrug, wurde er verurteilt.

Ist der Trugschluss deutlich? Die Wahrscheinlichkeit, dass ein Unschuldiger zufälligerweise die passende DNA hatte, betrug 1 zu 3 Millionen, und dies wurde als die Wahrscheinlichkeit interpretiert, dass Mr. Jones unschuldig war. Aber das ist ein komplettes Durcheinanderbringen der Verhältnisse:

Die Wahrscheinlichkeit, dass ein Unschuldiger das passende DNA-Profil hat, ist *nicht* die Wahrscheinlichkeit, dass jemand mit dem passenden DNA-Profil unschuldig ist.

Die Wahrscheinlichkeit, dass bei Zutreffen der Hypothese H die Daten D zu beobachten sind, ist *nicht* dasselbe wie die Wahrscheinlichkeit, dass bei den Daten D die Hypothese H zutrifft.

Man muss also bei Gericht zusätzliche Argumente anführen, um die Schuld von Mr. Jones zu beweisen.

Man hat beim Würfeln analoge Verhältnisse: Die Wahrscheinlichkeit, bei einer ungezinkten Münze 10-mal hintereinander Wappen zu bekommen, beträgt 1/1024. Aber die Wahrscheinlichkeit, dass es sich bei einer Wurfserie von 10-mal Wappen um einen ungezinkten Würfel handelt, beträgt nicht 1/1024.

8 Zum Ziegenproblem

Auch beim Ziegenproblem geht es darum, die verfügbaren Informationen richtig zu verarbeiten.

Es gibt drei Türen (A, B und C); hinter einer davon steht ein Auto, und hinter den beiden anderen steht jeweils eine Ziege. Der Kandidat muss raten, hinter welcher Tür das Auto steht. Falls er richtig wählt, bekommt er das Auto, andernfalls muss er mit einer der beiden Ziegen vorlieb nehmen. Natürlich ist die Wahrscheinlichkeit, dass er die richtige Tür wählt, so groß wie 1/3.

Nehmen wir an, dass er Tür A gewählt hat. Der Moderator geht dann zu Tür B und sagt: „Sie haben Tür A gewählt. Ich sage Ihnen nicht, ob hinter Tür A ein Auto oder eine Ziege steht. Aber ich zeige Ihnen jetzt, das hinter Tür B *kein* Auto steht.“ Dabei öffnet er Tür B, hinter der tatsächlich eine Ziege steht. Dann bietet er dem Kandidaten an, sich (von A nach C) umentscheiden zu dürfen.

Lohnt sich eine Umwahl oder nicht? Das ist die entscheidende Frage. Sollte der Kandidat bei seiner Wahl (Tür A) bleiben, oder ist es für ihn vorteilhafter, zu Tür C zu wechseln?

Wenn man das Ziegenproblem nicht kennt, gibt es zwei offensichtliche Arten, es zu lösen, und beide führen zu unterschiedlichen Ergebnissen.

Der erste Weg: Der Moderator hat nur das 3-Türen-Problem in ein 2-Türen-Problem verwandelt, er hätte gleich mit nur zwei Türen anfangen können, so dass es völlig egal ist, ob man umwählt oder nicht: Die Wahrscheinlichkeit, das Auto zu gewinnen, beträgt in jedem Fall 50 %.

Der zweite Weg: Bei der ursprünglichen Wahl des Kandidaten gab es ein Gewinnchance von 1/3 und eine Verlustchance von 2/3. Mit Wahrscheinlichkeit 1/3 war das Auto hinter Tür A und mit Wahrscheinlichkeit 2/3 war das Auto hinter B oder C. Das einzige, was der Moderator getan hat, war, die Möglichkeit, dass das Auto hinter Tür B sein könne, zu eliminieren, so dass sich die Wahrscheinlichkeit, dass das Auto nicht hinter Tür A ist, nunmehr auf Tür C konzentriert. Kurz: Mit Wahrscheinlichkeit 2/3 befindet sich das Auto hinter Tür C, so dass sich ein Umwählen lohnt.

Dies ist die populäre Lösung, die von praktisch allen vernünftigen Autoren vertreten wird.

Aber: Diese populäre Lösung ist falsch.

Manche Leute vertreten den ersten Weg. Aber auch diese Lösung ist falsch.

Sowohl diejenigen, die „Auf jeden Fall umwählen!“ sagen, als auch die anderen, die sagen: „Umwählen lohnt sich nicht!“ haben Unrecht!

Die große Unbekannte ist: Warum hat der Moderator gerade Tür B geöffnet? Was war seine Strategie? Er muss gewusst haben, hinter welcher Tür das Auto stand (sonst hätte er womöglich die Auto-Tür geöffnet). Hat er sich spontan für Tür B entschieden, oder hatte er Anweisung, auf jeden Fall Tür B zu öffnen?

Nehmen wir den zweiten Fall an: Der Moderator wusste, dass hinter Tür B eine Ziege stand und hatte die Anweisung, unabhängig davon, was der Kandidat sagen würde, auf jeden Fall Tür B zu öffnen. Wenn der Kandidat ursprünglich Tür B gewählt hätte, so hätte der Moderator auch Tür B geöffnet und gesagt: „Tut mir leid, aber Sie haben die falsche Tür gewählt.“ Hätte der Kandidat A oder C gewählt, so hätte der Moderator auch Tür B geöffnet und dem Kandidaten die Möglichkeit der Umwahl eingeräumt.

Falls dies die Strategie des Moderators gewesen sein sollte, so hätte er keinerlei neue Information

durch das Öffnen von B hinzugefügt, und es wäre in der Tat einerlei, ob der Kandidat umwählt oder nicht.
Eine andere Strategie des Moderators hätte die folgende sein können: Egal, was der Kandidat gewählt hat, immer eine Ziegentür zu öffnen. In diesem Fall hat er tatsächlich neue Information hinzugefügt und hat damit die Chancen des Kandidaten erhöht, durch Umwählen zum Auto zu kommen.
Aber es gibt auch noch zwei weitere Strategien, die der Moderator hätte einschlagen können. Er hätte sich auf folgende Art und Weise kandidatenfreundlich verhalten können: Wenn der Kandidat die Auto-Tür richtig rät, tue nichts. Anderenfalls öffne eine Ziegen-Tür und gib dem Kandidaten die Chance zur Umwahl. Bei dieser Strategie sollte der Kandidat natürlich auf jeden Fall umwählen.
Eine analoge, aber kandidaten*un*freundliche Strategie wäre: Öffne eine Ziegen-Tür mit der Chance zur Umwahl nur dann, wenn der Kandidat die Auto-Tür richtig geraten hat. In diesem Fall sollte der Kandidat auf gar keinen Fall umwählen.

Es gibt somit vier mögliche Strategien für den Moderator. Die richtige Antwort darauf, ob man umwählen solle oder nicht, hängt von der Strategie des Moderators ab, *und diese ist nicht bekannt*. So, wie das Problem präsentiert worden war, ist es kein mathematisches, sondern eher ein psychologisches Problem.

9 Hinterher ist man immer klüger

Auch das Durchforsten von allen möglichen Statistiken, Finden von unerwarteten Beziehungen und das „Beweisen", dass diese Zusammenhänge unwahrscheinlich und daher von Bedeutung sind, kann zur Vermengung von Statistik und Lügen führen. Ein Leserbrief an *The Times* aus dem Jahre 1963 ist dafür ein gutes Beispiel: Jemand hatte sich die tägliche Niederschlags-Statistik seit dem Ende des Zweiten Weltkriegs angesehen und festgestellt, dass der regenreichste Tag der Woche meistens der Donnerstag war, der Tag von Thor, dem Gott des Donners, während der trockenste Tag meistens der Sonntag war, der Tag der Sonne. Der Leserbriefschreiber behauptete, dass sein Fund statistisch signifikant sei, weil die Wahrscheinlichkeit, dass Donnerstag der feuchteste und Sonntag der trockenste Tag sind, nur 1/42 betrage – jede Wahrscheinlichkeit unter 1/20 wird als statistisch signifikant bezeichnet.
Im Jahr 1995 gab es wiederum einen Leserbrief an *The Times* mit derselben Thematik: Auch die Wetterbeobachtungen der nunmehr letzten 50 Jahre zeigten, dass Donnerstag der Tag des Regens und Sonntag der Tag der Sonne seien.
Hätte aber der zweite Leserbriefschreiber nur die Daten zwischen 1963 und 1993 beachtet, hätte er festgestellt, dass Donnertag nur der viertnasseste Tag war und auch Sonntag nicht mehr der trockenste.

Was man niemals machen darf, ist, eine Beziehung zwischen den Daten festzustellen und *danach* den Statistiker zu fragen, wie wahrscheinlich diese Beziehung sei. Wenn ein Ereignis bereits stattgefunden hat, macht es keinen Sinn, nach der Wahrscheinlichkeit dieses Ereignisses zu fragen. Die Wahrscheinlichkeit dafür, dass etwas sich ereignet, was sich bereits ereignet hat, beträgt immer 100 %. Etwas anderes ist es, wenn man nach der Wahrscheinlichkeit eines Ereignisses *vor* dessen Eintreffen fragt.
Wenn einem etwas Bemerkenswertes an einem Datensatz aufgefallen ist, darf man nicht denselben Datensatz benutzen, um die Signifikanz der Auffälligkeit zu testen, sondern darf das nur anhand eine neuen Datensatzes tun.

10 Was für eine merkwürdige Koinzidenz!

Alles, was in der Welt passiert, ist bemerkenswert und ein Ergebnis von mehreren einzigartigen Ereignissen. Schon die Tatsache, dass wir alle auf die Welt gekommen sind, ist das Ergebnis vieler seltener Voraussetzungen.
Sieht man zurück, so gibt es viele bemerkenswerte Zusammentreffen von Ereignissen. Zum Beispiel hat Mrs. Evelyn Adams aus New Jersey innerhalb eines Jahres zweimal den Hauptgewinn der staatlichen Lotterie gewonnen. Daraufhin rief ein Reporter bei der nächsten Universität an und fragte den dortigen Statistik-Professor einiges über Wahrscheinlichkeiten. Der Zeitungsbericht besagte dann, dass so etwas wie der Doppelgewinn von Frau Adams nur eine Wahrscheinlichkeit von 1/17.640.000.000.000 habe.
In Wahrheit war der Statistikprofessor gefragt worden, wie groß die Wahrscheinlichkeit ist, dass eine spezielle Zahlenkombination (die von Frau Adams) an zwei speziellen Tagen (den Gewinntagen von Frau Adams) vorkommt. Für diese Zahlenkombination ist die Wahrscheinlichkeit 1/4.200.000, und man muss diesen Wert quadrieren, um auf die Wahrscheinlichkeit zu kommen, dass diese Kombination an zwei speziellen Tagen vorkommt. Aber dieser Lösungsweg hat mit der ursprünglichen Aufgabe nichts zu tun:

Zunächst einmal hat Frau Adams nicht ein einziges Los gekauft, sondern sie kaufte jede Woche etwa 100 Lose. Dann müssen wir die Tatsache, dass sie das erste Mal gewonnen hat, für die Berechnung der Wahrscheinlichkeiten ignorieren: Irgendjemand gewinnt immer. Die richtige Frage ist vielmehr: Wie groß ist die Wahrscheinlichkeit, dass jemand, der jede Woche 100 Lose kauft, innerhalb eines Jahres zweimal den Hauptgewinn erhält?

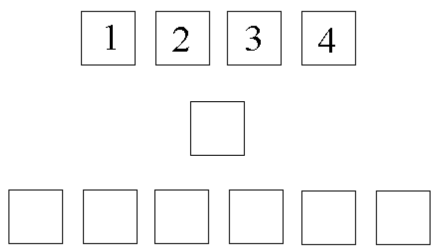
Es gibt viele merkwürdige Koinzidenzen. Als ich in den 1960-er Jahren in Liverpool studierte, wohnte ich mit 12 anderen Studenten in einem Wohnheim. In den Sommerferien traf dann rein zufällig einer der Studenten einen anderen der zwölf in einem Bus in Salt Lake City. Dann haben sich zwei meiner späteren Arbeitskollegen rein zufällig in einem Fahrstuhl am Hoover-Damm getroffen, oder zwei haben sich rein zufällig am Strand von Menorca getroffen oder in einem einsamen Fischerdorf in Portugal. Es gibt häufig solche unerwarteten Zusammentreffen. Ich teile sie in 3 Typen ein:
Bei Typ 1 ist die wahre Wahrscheinlichkeit für ihr Eintreffen bedeutend höher als man intuitiv meint. Ein gutes Beispiel ist das klassische Geburtstagsproblem, bei dem sich die meisten Leute nicht vorstellen können, dass man nur 23 Leute braucht, damit die Wahrscheinlichkeit für einen gemeinsamen Geburtstag über 50 % liegt.
Beispiele von Typ 2 sind die oben genannten Treffen in Salt Lake City usw. Sie erscheinen deswegen als so unwahrscheinlich, weil sie nach dem Eintreffen aus den vielen Möglichkeiten, die vorkommen können, ausgewählt wurden. Die obigen Typ-2-Beispiele haben auch etwas vom Typ 1, da die Leute, denen so etwas zustieß, von ähnlichem Typ und von ähnlicher Lebensgestaltung waren.
Typ 3: Das sind die Beispiele, die wirklich überraschend sind und bei denen man an übernatürliche Kräfte oder Betrügereien glauben möchte.

11 Spiel mit dem Zufall

Man wird schon gemerkt haben, dass ich *kein* Spieler bin – jeder Spieler kann auf Dauer nur verlieren! Ein guter Rat an alle Leser: Spielen Sie niemals in Casinos! Diese verdienen sehr viel Geld, und zwar dadurch, dass sie es den Verlierern aus der Tasche ziehen.
Dies gilt auch für das Lotto: Die beste Strategie besteht darin, niemals ein Los zu kaufen.
Dies sage ich als logisch argumentierender Mathematiker. Aber wenn ich kein Los kaufe, habe ich auch nicht die Chance, durch Lotto reich zu werden. Insofern ist es nicht irrational, jede Woche ein Los zu kaufen, obgleich es nicht logisch ist. Es ist aber natürlich irrational, jede Woche viel Geld für Lose auszugeben.
Aber wenn man schon Lotto spielen möchte: Welchen Ratschlag kann der Statistiker geben?
Natürlich kann ich nicht dabei helfen, die richtigen Zahlen anzukreuzen. Ich kann also nicht helfen, zu gewinnen. Aber ich kann dabei helfen, im Fall eines Gewinns mehr zu gewinnen.
Der Schlüssel liegt darin, *wie* Leute zufällige Zahlen ankreuzen. Wenn jedermann in England seine Zahlen wirklich zufällig ankreuzen würde, gäbe es jede Woche zwei bis drei Haupt-Gewinner, wobei die Anzahl der Haupt-Gewinner selten über 15 liegen würde, und es hätte in den letzten 10 Jahren nur in 40 Fällen gar keinen Haupt-Gewinner gegeben. In Wirklichkeit gab es mehr als 130-mal keinen Gewinner, es gab mehrere Male mehr als 20 Gewinner, einmal sogar 133 Haupt-Gewinner.
Wir sehen also, dass es bei vielen Ziehungen entweder viel mehr oder viel weniger Haupt-Gewinner gibt, als zu erwarten wären, wenn jeder ganz zufällig ankreuzen würde. Daraus folgt: Die meisten Leute, die einen Haupt-Gewinn erzielt haben, müssen ihn mit vielen anderen teilen. Der Grund besteht darin, dass Leute nicht gut Zufallszahlen ankreuzen können.

12 Denke dir eine Zahl

Mit dem folgenden Experiment kann man testen, wie gut man im Erstellen von Zufallszahlen ist:



1. Man soll in der oberen Zeile eine Zahl markieren.
2. Man soll in das Quadrat der zweiten Zeile eine einstellige Zufallszahl (zwischen 0 und 9) hineinschreiben.
3. Man soll (wie beim Lotto) sich sechs Zufallszahlen zwischen 1 und 49 ausdenken, sie anordnen und in die Quadrate der dritten Zeile schreiben.

Wenn man das Experiment mit vielen Leuten durchführt, wird man mit hoher Wahrscheinlichkeit folgendes feststellen:
In der ersten Zeile sollten zwar alle vier Zahlen mit etwa gleicher Häufigkeit vorkommen, wenn wirklich eines der Quadrate zufällig gewählt worden wäre. Die Erfahrung zeigt aber, dass die Zahl 3 mit etwa 50 %-iger relativer Häufigkeit gewählt wird.
Bei der zweiten Zeile zeigt die Erfahrung, dass die Zahl 7 von fast einem Drittel aller Leute gewählt wird; bei wirklich zufälliger Wahl hätte man für die eine relative Häufigkeit von etwa 1/10 erwarten müssen.
Bei wirklich zufälliger Wahl würde auch die Kombination „3 in Zeile 1 und 7 in Zeile 2“ mit relativer Häufigkeit von etwa 1/40 auftreten und nicht, wie in realen Experimenten beobachtet, mit etwa 1/10.
Nun zur dritten Zeile. Es zeigt sich, dass deutlich weniger als die Hälfte aller Leute ein Sextupel gewählt hat, bei denen zwei (oder mehr) aufeinander folgende Zahlen vorkommen. In Wahrheit ist bei 6 zufällig gewählten Zahlen im Bereich von 1 bis 49 in etwa der Hälfte aller Fälle zu erwarten, dass zwei konsekutive Zahlen vorkommen. Dies kann man schon der Lotto-Statistik entnehmen: Im (englischen) Lotto gab es zwischen 1994 und 2004 in 402 von 845 Ziehungen (also in etwa 48 %) aufeinander folgende Zahlen. In diesen Fällen gab es auch jeweils weniger Leute, die den Haupt-Gewinn erzielten, und die mussten sich die Gewinnsumme folglich mit weniger Leuten teilen.

Um möglichst viel zu gewinnen, wenn man gewinnt, sollte man auf Zahlen setzen, die von nur wenigen Menschen getippt werden. Die Lotterien in der Schweiz oder in Kanada halten das Tippverhalten ihrer Kunden nicht geheim. Man macht dabei folgende Beobachtungen:
Die beliebteste Zahl ist die 7; danach kommt die 11. Ungerade Zahlen sind beliebter als gerade Zahlen. Zahlen über 40 werden seltener getippt.

Hier ist mein Ratschlag: Man sollte die Zahlen wirklich zufällig wählen. Das geschieht am besten, indem man 6 von 49 Zettelchen aus einem Hut zieht, dabei aber folgendes beachtet:
1. Es sollte mindestens einmal ein Paar aufeinander folgender Zahlen geben.
2. Es sollte weder eine 7 noch eine 11 dabei sein.
3. Es sollten nicht mehr als 3 ungerade Zahlen dabei sein.
4. Mindestens eine Zahl sollte größer sein als 41.

Wenn eines dieser Kriterien nicht erfüllt ist, sollte man erneut 6 von 49 Zettelchen aus einem Hut ziehen.
Durch diese Maßnahmen werden die Gewinnchancen nicht vergrößert. Aber wenn man gewinnt, braucht man sich den Gewinn nicht mit so vielen anderen Leuten zu teilen. (Anm. des Hrsg.: Weiteres zu dem Thema findet sich bei Henze / Riedwyl.)

Literatur

Henze, N. und Riedwyl, H.: How to win more. Strategies for increasing a lottery win. 1998: A. K. Peters.

Matthews, R. (2001). Der Storch bringt die Babys zur Welt (p = 0.008). SiS **21** (2001) 2, 21 – 23.

Bei der Übersetzung hat der Heftherausgeber die Originalquelle gekürzt und bearbeitet.

Originalquelle:
Frank Duckworth:
The Royal Statistical Society Schools Lecture 2004: ‘Lies and Statistics’
Teaching Statistics, vol. 28 (2), Summer 2006, 34 – 39 und vol. 28 (3), Autumn 2006, 84 – 89.

Anschrift des Verfassers:
Frank Duckworth
Stinchcombe, Gloucestershire, England.
e-mail: `f.duckworth@rss.org.uk`