

Auch dieses Vorwort sollte man lesen

Wer macht sich eigentlich die Mühe, all die unsinnigen Wahrscheinlichkeiten auszurechnen? Die für den Tod durch einen umstürzenden Getränkeautomaten (sie beträgt angeblich 1:112.000). Die, von einem Meteorit erschlagen oder die, vom Blitz getroffen zu werden. Würde es unser Lebensgefühl verbessern, wenn die Wahrscheinlichkeit, im Lotto den Hauptgewinn abzusahnen, nicht 1:13 Millionen, sondern mit 1:6,5 Millionen doppelt so hoch wäre? Beruhigt es irgendjemanden, dass es 1888mal wahrscheinlicher ist im Lotto zu gewinnen statt ein zweites Mal vom Blitz getroffen zu werden oder gar beides (ein Kanadier wurde angeblich zweimal vom Blitz getroffen und gewann im Lotto 1 Million, Wahrscheinlichkeit 1:2,6 Billionen)? Oder dass es 14000mal wahrscheinlicher ist, auf der Toilette zu sterben als durch einen Blitz? Und dass der Tod durch ein Produkt für Rechtshänder dreimal so tödlich für Linkshänder sein kann (das liegt vor allem an den Motorsägen)? Oder wie wäre es mit der vierfachen Wahrscheinlichkeit, ins Krankenhaus zu kommen, weil man Schwierigkeiten beim Anziehen der Hose hatte (beim Hose anziehen zu sterben ist mit 1:30 Millionen auch nicht zu unterschätzen)? Solche extremen Wahrscheinlichkeiten zu berechnen ist vielleicht spaßig, aber letztendlich unsinnig. Alles dermaßen unwahrscheinlich, dass sich eine Berechnung eigentlich verbietet.

Die Antwort auf die Frage, warum man die in Band 1 dargestellte Beschreibende Statistik braucht, fiel mir nicht schwer. Die dort erworbenen Kenntnisse ließen uns viele Aussagen sinnvoll hinterfragen. Ein Beispiel war die Berechnung des Preisniveaus. Eine Kennziffer der Lebenshaltungskosten in Deutschland sagt etwas darüber aus, was unser Geld am Ende des Jahres noch wert ist, wie viel wir also real verlieren. Ein arithmetisches Mittel, das die durchschnittlichen Geburten pro Frau angibt, kann für Prognosen verwendet werden, wie viele von uns es (bei sonst gleichen Rahmenbedingungen, also ohne Krieg, Pest, Pandemien) in 30 Jahren geben wird und ob es für die Rentenzahlungen reicht. Und ob es einen Zusammenhang zwischen dem Austritt von Ozon aus Laserdruckern und Lungenkrebs geben könnte, interessiert vielleicht viele Büroangestellte.

Warum man aber die Schließende Statistik braucht, konnte selbst ich nicht immer ausreichend beantworten. Vielleicht lag es an den abstrusen Rechenbeispielen. Im Ernst: Wer interessiert sich dafür, ob er aus einer Lostrommel eine Primzahl zieht? Wer möchte wissen, wie groß das Intervall ist, das die jährliche Nutzungsdauer im Internet eingrenzt? Reicht es nicht zu wissen, dass sie bei durchschnittlich 34000 Stunden liegt? Wohl nicht. Sonst gäbe es ja nicht so viele Veranstaltungen und so viele Lehrbücher, die sich mit der Schließenden (oder Analytischen oder Induktiven) Statistik beschäftigen würden. Vielleicht liegt es an der Aufbereitung des Stoffes? An den immer gleichen Beispielen, in denen gewürfelt wird, in denen Münzen auf Wappen oder Zahl fallen oder in denen

irgendwelche Warenlieferungen zum Teil fehlerhaft sind. Vielleicht wäre es interessanter, darüber nachzudenken, wie hoch die Überlebenswahrscheinlichkeit ist, wenn Krebs diagnostiziert wurde (30 % von uns, mit steigender Tendenz, werden vermutlich an Krebs sterben). Natürlich hilft diese Wahrscheinlichkeit einem einzelnen Kranken nicht abschließend weiter; was heißt schon 30 %? Aber dass die Mehrheit überlebt, gibt sicher ein bisschen Hoffnung. Läge die Überlebenswahrscheinlichkeit sogar bei 95 %, beruhigt das vielleicht mehr als jedes Beruhigungsmittel es tun könnte. Ein Politiker, der für das Gesundheitswesen zuständig ist, könnte sich durchaus mit der Frage beschäftigen: Ist es besser, die Tabaksteuer drastisch zu erhöhen, so hoch, dass fast keiner mehr raucht? Mit der Folge, dass enorme Defizite im Haushalt entstehen. Oder lassen wir es bei der bisherigen Politik und akzeptieren wir die vielen Todesfälle durchs Rauchen (die Wahrscheinlichkeit, durchs Rauchen zu sterben, beträgt jedes Jahr bis zu 0,5 Prozent; jeder 7. in Deutschland Lebende, im weltweiten Vergleich überdurchschnittlich viele, ist davon betroffen)? Helfen wir mit den Milliarden aus der Tabaksteuer lieber anderen Bedürftigen (an dieser Stelle wollen wir bitte keine Diskussion darüber aufmachen, ob die Einnahmen aus der Tabaksteuer zweckdienlich verwandt werden, das steht auf einem anderen Blatt)?

Schon in der Beschreibenden Statistik gilt, dass alle Erhebungen nur Momentaufnahmen darstellen. Wenn 30 % die Partei XY gewählt haben, dann könnte das Ergebnis am nächsten Tag schon anders aussehen. Wenn auf deutschen Straßen im Jahr z vor allem silberne Autos unterwegs waren, kann das im Jahr $z+1$ anders aussehen. Wenn die Inflationsrate oder die Arbeitslosenquote im März 2,3 % betrugen, wird das im April nur selten wieder so sein. In der Beschreibenden Statistik lebten wir also mit Unsicherheiten. Und nicht anders ist es in der hier vorgestellten Schließenden Statistik, auch hier sind die Ergebnisse zwar sehr wahrscheinlich, aber nicht sicher. Deshalb könnte man auch hier generell kritisieren: Was nützen mir Aussagen über Ereignisse, die nur mit einer Wahrscheinlichkeit von X % eintreffen werden? Weil eine wahrscheinliche Aussage besser ist als gar keine! Wenn wir berechnen können, dass die Ausfallwahrscheinlichkeit eines Rechners bei 10 % liegt, dann sollten wir einen zweiten als Backup vorsehen, liegt sie bei 0,1 %, kann man auf einen langen fehlerlosen Betrieb hoffen.

Dass ein Interesse an der Schließenden Statistik durchaus Millionen einbringen kann, zeigt das Beispiel »Sell in May and go away«, ein Ausspruch, den alle Börsianer kennen. Aber ist da wirklich etwas dran? In einer Studie kamen Wissenschaftler zu einem Ergebnis, das viele Aktienhändler überrascht hat. Denn danach gibt es in den meisten Jahren seit Einführung des Deutschen Aktienindex (DAX, seit 1988) nur einen unbedeutend kleinen Unterschied zwischen der Aktienentwicklung von November bis April und der zwischen Mai bis Oktober. Im Winter Aktien zu erwerben, um diese dann im Mai zu verkaufen, bringt also nichts. Mit einer Ausnahme: Der Effekt trat (sogar seit 1927) in den Jahren tatsächlich ein, denen die Vereidigung des neuen US-Präsidenten vorausging. Dann hat beispielsweise der Dow Jones im Winterhalbjahr um durchschnittlich 3,82 Prozent zugelegt, im darauffolgenden Sommer aber um 2,58 Prozent nachgegeben. So etwas zu wissen, und ist es nur etwas wahrscheinlicher als andere Entwicklungen, kann einen Gewinn von Millionen ausmachen. Zumal dieser sogenannte Halloween-Effekt mit schöner Regelmäßigkeit alle 4 Jahre auftritt.

Das im ersten Band Gesagte gilt natürlich auch hier: Wir werden mit Statistiken überhäuft; es gibt scheinbar nichts, für das es keine Statistik gibt, und kein Medium, in dem nicht zahlreiche Statistiken zitiert oder interpretiert werden. Und oft haben wir Nutzer den Eindruck, dass eine Statistik nicht zu uns passt oder willkürlich ist. Die Glaubwürdigkeit der Statistik leidet unter dieser Masse an Informationen. Die Zahlen, die uns angeboten werden, erscheinen uns immer weniger glaubhaft. Wer nicht raucht, täglich Sport treibt und sich gesund ernährt, findet sich in einer Statistik nicht wieder, in der die Wahrscheinlichkeit ausgerechnet wurde, in genau diesem Jahr zu sterben. Außerdem interessieren wir uns häufig erst dann intensiver für Statistiken, wenn wir in einer aktuellen Situation Orientierung suchen. Eine Statistik über die Überlebenswahrscheinlichkeit im Falle einer schweren Krankheit beschäftigt uns. Die Wahrscheinlichkeit, das Rentenalter zu erreichen, interessiert uns (und die Rentenkasse) ebenfalls. Zwar ist uns allen bekannt, dass wir etwa 70, 80 Jahre leben werden, also ca. 10 bis 15 Jahre Rente beziehen werden. Aber lohnt sich die private Altersvorsorge, wenn wir unser Leben lang in einem Bergwerk geschuftet haben (wie z. B. unsere Vorgängergeneration im Ruhrgebiet oder im Saarland) und die Wahrscheinlichkeit, früh an einer Staublung zu sterben, hoch ist? Bei welchem Professor sind die Klausuren einfach und die Wahrscheinlichkeit zu bestehen deshalb hoch? Das sind durchaus Statistiken, die unser Interesse wecken.

Es gibt einen weiteren Grund, sich intensiver mit der Schließenden Statistik auseinander zu setzen. Offensichtlich können viele von uns Wahrscheinlichkeiten kaum korrekt einschätzen. Wir können einfach nicht beurteilen, ob es sinnvoller ist, Maßnahmen zu ergreifen, die unsere Überlebenschance von 98 % auf 100 % steigern oder, ob wir lieber Maßnahmen ergreifen, die unsere Chance von 70 % auf 90 % verbessern. 100 % scheint erstrebenswert, aber eine Steigerung der Wahrscheinlichkeit um 20 % bringt eindeutig mehr.

Deshalb soll auch dieser 2. Band ein einfach gehaltenes Buch zur Statistik sein, das ohne fundierte Kenntnisse der Mathematik lesbar und verständlich bleibt. Eine Einführung zur Schließenden Statistik also, die sich nicht nur an die Studierenden aller Fachrichtungen richtet, sondern auch den interessierten Laien ansprechen möchte, der bisher vergeblich versucht hat, Statistik zu verstehen. Gleichzeitig soll vor allem die oben angedeutete kritische Haltung aufgegriffen werden. Denn da gerade die Schließende Statistik oftmals sehr abstrakt und sogar mit Integralen daherkommt, möchte mancher sich lieber gar nicht mit ihr beschäftigen. Daran ändert auch nichts, dass in den letzten 5 Jahren 40 % mehr Studierende gezählt wurden. Die, die sich gerne für die Mathematik interessieren, sind leider nicht mehr geworden. Oder anders ausgedrückt: Die Wahrscheinlichkeit, dass ein Studierender sich für Mathematik und Statistik interessiert, ist gleichgeblieben (oder eher sogar gesunken). Und natürlich gibt es auch zur Schließenden Statistik passende Aussagen, die uns den Spaß an ihr verderben wollen. »Wenn Sie mal eine Telefonnummer vergessen haben fragen Sie einen Statistiker, er kann Ihnen eine gute Schätzung liefern« ist so ein Ansatz. Davon wollen wir uns nicht leiten, und schon gar nicht beirren lassen.

Ich versuche daher, in dieser Einführung zu zeigen, dass für das Verständnis der statistischen Methodenlehre die Beschäftigung mit komplizierten mathematischen

Ansätzen, vielen Formeln und Rechenvarianten nicht zwingend notwendig ist. Wieder fast ohne Formeln und ohne Taschenrechner werden wir uns mit dem Thema auseinandersetzen; es geht uns um das grundsätzliche Verstehen, nicht um das Rechnen. Erst im Anhang werden zusammenfassende Aufgaben besprochen, die (überwiegend) nur mit Hilfe eines Taschenrechners gelöst werden können.

Mein ältester Sohn hat das Ganze wieder Korrektur gelesen und mir zahlreiche Hinweise gegeben, wie man etwas besser formulieren könnte. Meine Frau und mein kleiner Sohn haben mir auch diesmal die ungestörte Ruhe in meinem Arbeitszimmer ermöglicht, die man für so ein Werk braucht. Streit darüber gab es zum Glück keinen. Ein Dank geht auch an Alexander Hau für die Erstellung der zahlreichen Grafiken. Und wieder möchte ich mich besonders bei Herrn Dr. Fliegauf vom Verlag Kohlhammer bedanken, der auch diesen zweiten Band intensiv gelesen hat und wieder zahlreiche Anregungen und Verbesserungsvorschläge einbrachte. Schade, dass unsere Zusammenarbeit damit erst einmal zu Ende geht.

Bochum, im Mai 2017

Frank Siegmann