

## Achtung, Fälscher!

In den letzten Wochen war oft davon zu lesen, dass Wissenschaftler ihre Daten ein bisschen manipuliert haben. Sie passten damit besser zu der These, die gerade untermauert werden sollte. Das ist kein Gegenwartsphänomen, sogar einigen wirklich bedeutenden Forschern wie etwa Mendel und Michelson war das „*corriger la fortune*“ nicht unbekannt. Wie aber kann man Daten eigentlich ansehen, ob sie möglicherweise gefälscht worden sind?

In vielen Fällen reicht eine statistische Analyse der Anfangsziffern. Dabei macht man sich ein überraschendes Phänomen zunutze, das unter dem Namen *Benford-Gesetz* bekannt ist: Bei den meisten aus „wirklichen“ Daten entstandenen Tabellen sind nämlich die Anfangsziffern nicht gleichmäßig verteilt. Die Eins kommt mit Abstand am häufigsten vor, dann die Zwei, usw.; für die Neun bleibt nur ein geringer Anteil übrig. Das lässt sich immer dann beobachten, wenn das Endergebnis durch eine Überlagerung mehrerer unabhängiger Einflüsse entstanden ist. Bei den Tabellen des Stromverbrauchs der Kraftwerkskunden, von Flusslängen und bei einer überwältigenden Fülle weiterer Situationen kann das nachgeprüft werden.

Das Gesetz ist nach dem amerikanischen Physiker Frank Benford (1883-1948) benannt. Seine Begründung ist in der Wahrscheinlichkeitsrechnung zu suchen. Hier geht es um das Thema „Der Zufall verwischt alle Spuren“, ein Aspekt, den sich übrigens auch die Kryptographie zunutze macht. Als Beispiel nehme man einen Würfel. Wenn der 1000 Mal geworfen und dann die Augensumme addiert wird, kommt irgendetwas zwischen 1000 und 6000 heraus. Dabei werden am Ende die Ziffern von Null bis Neun annähernd mit gleicher Wahrscheinlichkeit zu sehen sein. Wenn man statt zu Addieren nun multipliziert und an der ersten Ziffer des Produkts interessiert ist, so kann man das nach Übergang zu Logarithmen wieder auf Summen zurückführen. Und deswegen müssen die ersten Ziffern in guter Näherung so verteilt sein, wie es den gegenseitigen Abständen der Logarithmen von 1, 2, ..., 9 entspricht: 30 Prozent für die Anfangsziffer „1“, 17 Prozent für die „2“ usw. bis herunter zu 4 Prozent für die „9“.

Skeptiker sind eingeladen, das GOOGLE-Benford-Experiment durchzuführen. Man denkt sich eine beliebige vierstellige Zahl aus, etwa 3141. Dann schreibt man eine „1“ davor und fragt GOOGLE, wie viele Treffer es dafür gibt. Dann sucht man nach der 23141 (es wird also eine „2“ vor die Zahl der Wahl geschrieben), anschließend kommt die 33141 dran usw. Man kann bei diesem Experiment ziemlich sicher sein, dass die Trefferzahlen mit höher werdender Anfangsziffer rapide abnehmen. Fälschen zur Untermauerung des Gesetzes ist hier nicht erforderlich.

P.S.: Mehr dazu findet man in [http://www.nigrini.com/Benford's\\_law.htm](http://www.nigrini.com/Benford's_law.htm)