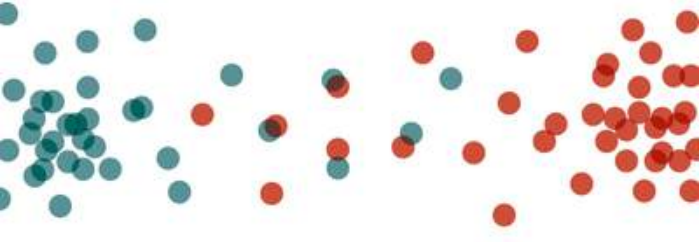
REDLINE | VERLAG

David Spiegelhalter

Die KUNST der STATISTIK



Was uns Daten wirklich sagen und wie wir dies im Alltag nutzen können

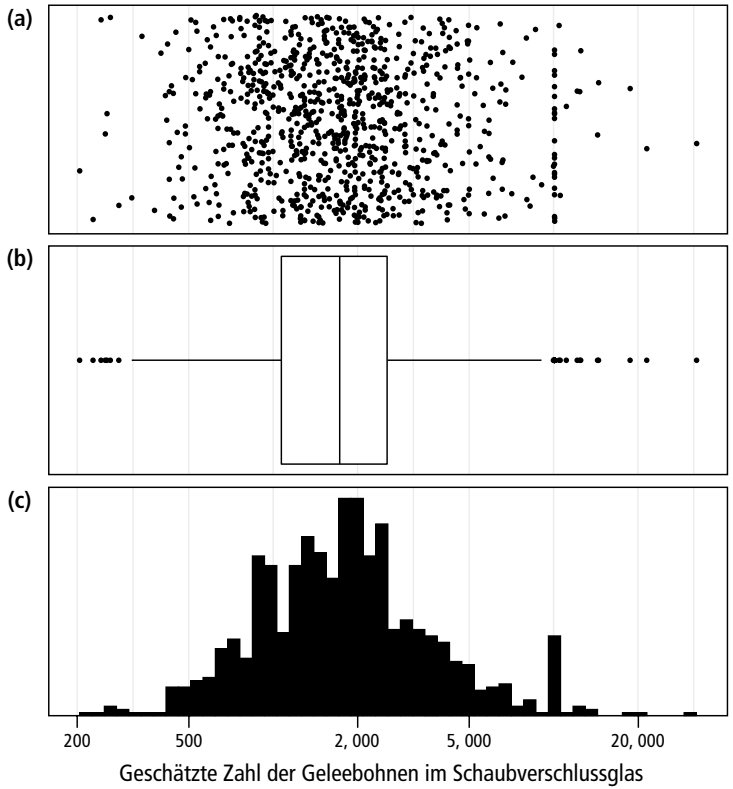


Abbildung 2.3
Grafische Veranschaulichung der Schätzwerte für die Zahl der Geleebohnen entlang einer logarithmischen Skala. (a) Punktdiagramm; (b) Box-Whisker-Plot; (c) Histogramm. Die Diagramme zeigen ein annähernd symmetrisches Muster.

Variablen können unterschiedliche Zahlenformate annehmen:

- Diskrete oder **Zählvariablen**: Wenn die Messwerte auf die natürlichen Zahlen 0, 1, 2 … beschränkt sind, wie beispielsweise die Zahl der Morde in einem Jahr oder die Zahl der Geleebohnen in einem Glasbehälter.

- **Stetige Variablen:** Messungen lassen sich zumindest theoretisch beliebig präzise durchführen – zum Beispiel Körpergröße und -masse variierend nach Per-

son und Zeitpunkt. Diese Angaben lassen sich natürlich auf ganze Zentimeter bzw. Kilogramm runden.

Wenn eine Reihe von Zähl- oder stetigen Variablen auf eine einzige Zahl reduziert wird, sprechen wir hier in der Regel vom **Durchschnitt**. Diesen Begriff kennen wir alle, beispielsweise in Ausdrücken wie Durchschnittslohn, Durchschnittsnoten oder Durchschnittstemperaturen, aber häufig ist unklar, wie diese Zahlen zu interpretieren sind (besonders, wenn derjenige, der sie zitiert, es selbst nicht verstanden hat).

Es gibt im Wesentlichen drei Interpretationen des Begriffs »Durchschnitt«:

- **Mittelwert** oder Mittel: die Summe der Zahlen geteilt durch ihre Anzahl.

- **Medianwert** oder Median: die Zahl, die in der Mitte steht, wenn die Zahlen zuvor ihrer Größe nach geordnet wurden. Das ist der »Durchschnitt«, mit dem Galton die Stimmen der vielen zusammenfasste.*

- **Modalwert** oder Modus: die am häufigsten vorkommende Zahl.

Man spricht hier auch von Lagemaßen der Datenverteilung.

Die Interpretation von »Durchschnitt« als Mittelwert liefert auch die Basis für so alte Kalauer wie die, dass fast alle Menschen überdurchschnittlich viele Beine haben (wobei der Durchschnitt vermutlich bei 1,99999 liegt) oder dass die Menschen im Schnitt einen Hoden haben. Aber nicht nur für Beine und Hoden kann der Durchschnitt im Sinne von Mittelwert ungeeignet sein. Die mittlere angegebene Zahl von Sexualpartnern und das mittlere Einkommen in einem Land haben möglicherweise mit der Situation der meisten Menschen wenig gemeinsam. Der Grund ist, dass wenige extreme Werte das Mittel ungebührlich in die Höhe treiben: Denken Sie an Warren Beatty oder Bill Gates (was Sexualpartner beziehungsweise Einkommen betrifft, sollte ich vielleicht hinzufügen).

* Noch im selben Jahr 1907 hinterfragte ein Korrespondent in der *Nature* Galtons Wahl des Medians und behauptete, der Mittelwert hätte eine bessere Schätzung ergeben.

Durchschnitte als Mittelwerte können höchst irreführend sein, solange die Ausgangsdaten kein symmetrisches Muster rund um einen zentralen Wert bilden, sondern schief verteilt sind wie die Schätzungen zur Zahl der Geleebohnen – typischerweise mit einer starken Häufung und einem Ende aus entweder sehr hohen Werten (zum Beispiel Einkommen) oder niedrigen (zum Beispiel Beine). Ich kann Ihnen so gut wie garantieren, dass Sie verglichen mit Menschen Ihres Alters und Geschlechts ein weit unter dem Durchschnitt (Mittelwert) liegendes Risiko tragen, binnen des nächsten Jahres zu sterben. Aus den britischen Sterbetafeln beispielsweise geht hervor, dass jedes Jahr von den Männern, die 63 Jahre alt werden, jeder Hundertste seinen 64. Geburtstag nicht mehr erleben wird. Aber weil von denen, die sterben werden, viele bereits ernsthaft erkrankt sind, können die Gesunden oder leidlich Gesunden von einem geringeren Sterberisiko ausgehen.

Leider sagen die Medien, wenn sie vom »Durchschnitt« sprechen, häufig nicht, ob er als Mittel- oder als Medianwert zu interpretieren ist. Das britische Office for National Statistics etwa berechnet den Durchschnittswochenverdienst, der ein Mittelwert ist, gibt zugleich aber auch von lokalen Behörden gemeldete Wochenverdienste an, die als Medianwerte zu verstehen sind. In diesem Fall könnte es hilfreich sein, zwischen »Durchschnittseinkommen« (Mittelwert) und dem »Einkommen des Durchschnittsbürgers« (Medianwert) zu unterscheiden. Haus- und Grundstückspreise weisen eine äußerst schiefe Verteilung mit einem langen Ende zur hochpreisigen Seite hin auf, weshalb offizielle Immobilienindizes Mediane verwenden.

Es ist jetzt an der Zeit, dass wir die Ergebnisse unseres Experiments in kollektiver Intelligenz mit den Geleebohnen bekannt geben: nicht so aufregend wie das Gewicht eines Ochsen, aber dafür mit etwas mehr Stimmen, als Galton zur Verfügung standen.

Aufgrund der Datenverteilung, die ein langes Ende zur rechten Seite hin aufweist, ist der Mittelwert in Höhe von 2408 ein schlechter Schätzer, und auch der Modalwert von 10 0000 scheint einer speziellen Vorliebe für runde Zahlen zu entspringen. Und so ist es vermutlich besser, Galtons Beispiel zu folgen und den Medianwert in Höhe von 1775 als Gruppenmeinung anzusetzen. Der wahre Wert betrug … 1616.[2] Nur ein Teilnehmer schlug genau diesen Wert vor, während 45 Prozent einen Wert darunter und 55 Prozent einen Wert darüber schätzten. Eine generelle Neigung, zu hoch oder zu tief zu schätzen, lässt sich demnach nicht

erkennen – wir sagen, dass der wahre Wert beim 45. **Perzentil** der empirischen Datenverteilung lag. Der Median, der das 50. Perzentil bezeichnet, überschätzte den wahren Wert um 1775 – 1616 = 159 und damit um rund 10 Prozent. Nur jeder zehnte Teilnehmer kam so dicht oder noch dichter an den wahren Wert heran. Die kollektive Intelligenz lieferte also ein ziemlich gutes Ergebnis und kam der Wahrheit näher als 90 Prozent der einzelnen Teilnehmer.

Die Streuung einer Datenverteilung beschreiben

Es reicht nicht, eine Verteilung auf einen einzigen Wert zu reduzieren; wir brauchen auch eine Vorstellung von der Streuung, auch Dispersion genannt. Die Kenntnis der durchschnittlichen Schuhgröße eines erwachsenen Mannes beispielsweise ermöglicht es einer Schuhfirma noch lange nicht zu entscheiden, wie viele Exemplare der unterschiedlichen Größen sie herstellen sollte. Es gibt keine Größe, die für alle richtig ist, wie die Passagiersitze in den Flugzeugen eindrücklich unter Beweis stellen.

Tabelle 2.1 listet verschiedene zusammenfassende Zahlen zu den Geleebohnenschätzungen auf, wovon sich drei auf die Streuung beziehen. Die **Spannweite** erklärt sich von selbst, reagiert aber offensichtlich sehr empfindlich auf extreme Werte wie die recht bizarre Schätzung von 31 337 Bohnen.* Auf den **Quartilsabstand** (*inter-quartile range*, IQR) wirken sich die Extreme hingegen nicht aus. Das ist der Abstand zwischen dem 25. und dem 75. Perzentil; hier liegen also die mittleren 50 Prozent der Werte – in unserem Fall alle Schätzungen von 1109 bis 2599 Bohnen, was der »Box« im Box-Whisker-Plot entspricht (siehe Abbildungen 2.2 und 2.3). Und dann ist da noch die **Standardabweichung**, ein häufig verwendetes Streuungsmaß. Sie ist das technisch anspruchsvollste Maß, eignet sich jedoch in Wahrheit nur für hinreichend symmetrische Datenverteilungen,** denn auch sie

* Es handelte sich fast sicher um einen Verschreiber, wobei 1337 gemeint war – eine numerische Wiedergabe des Wortes »leet«, das im Internetslang für »fähig« steht. (Unter »Leetspeak« versteht man das Ersetzen von Buchstaben durch Ziffern und Sonderzeichen; vgl. für einen ersten Überblick: https://de.wikipedia.org/wiki/Leetspeak.)

** Der Gini- oder Konzentrationskoeffizient ist ein Maß für die Streuung hochgradig schiefer Datenverteilungen wie beispielsweise Einkommen und wird häufig als Ungleichheitsmaß genutzt. Seine Form ist komplex und wenig intuitiv.

Zusammenfassende statistische Werte zur Beurteilung der Zahl der Geleebohnen in einem Glasbehälter

Mittelwert	2408
Medianwert	1775
Modalwert	10000
Spannweite	219 bis 31337
Quartilsabstand	1109 bis 2599
Standardabweichung	2422

Tabelle 2.1
Zusammenfassende statistische Werte für 915 Schätzwerte zur Anzahl der Geleebohnen im Glasbehälter. Die wahre Anzahl betrug 1616.

wird von Ausreißern übergebührlich beeinflusst. So genügt es beispielsweise, den (fast sicher irrtümlichen) Wert von 31337 aus den Daten zu entfernen, damit die Standardabweichung von 2422 auf 1398 sinkt.*

In unserem kleinen Experiment zeigte das Kollektiv ein beträchtliches Maß an Intelligenz, obgleich einige Antworten recht bizarr waren. Das zeigt, dass unter den Daten häufiger einmal Fehler, Ausreißer und andere seltsame Werte sein können, die nicht notwendigerweise einzeln identifiziert und ausgeschlossen zu werden brauchen. Es zeigt auch, wie nützlich es sein kann, zusammenfassende Maße zu verwenden, die vergleichsweise unempfindlich gegenüber Ausreißern wie der Zahl 31337 sind – wir sprechen hier von robusten Maßen, zu denen der Median und der Quartilsabstand zählen. Und es zeigt, wie wertvoll es sein kann, sich die Daten unbefangen anzuschauen – eine Lektion, die das nächste Beispiel noch einmal unterstreichen wird.

* Das Quadrat der Standardabweichung wird als **Varianz** bezeichnet. Sie lässt sich schwer unmittelbar interpretieren, ist aber mathematisch nützlich.