

Thomas Spiegler

Statistik schön einfach

Eine Einführung in 50 Bildern

BELTZ JUVENTA

Leseprobe aus Spiegler, Statistik schön einfach, ISBN 978-3-7799-6306-6

© 2020 Beltz Juventa in der Verlagsgruppe Beltz, Weinheim Basel

<http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-7799-6306-6>

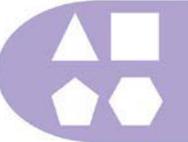
ein Wort zuvor

Dieses Buch ist für diejenigen geschrieben, die eher ungern ein Statistiklehrbuch in die Hand nehmen. In den üblichen Vorlesungen und Lehrbüchern zur Statistik wimmelt es oft von mathematischen Begriffen und Formeln. Wer Mathematik nicht gerade als eigene Stärke empfindet, wird davon schnell abgeschreckt. Das ist verständlich, aber unnötig. Denn die Grundideen der Statistik sind einfach, funktionieren sogar ohne Zahlen und werden von uns im Alltag immer wieder angewandt. Daher liegt diesem Buch die These zugrunde, dass jeder Statistik ‚kann‘. Es ist ein Buch für Menschen, die Statistik brauchen, aber sie (noch) nicht mögen, und auch für jene, die solchen Menschen Statistik verständlich machen möchten. Die üblichen Lehrbücher will und kann diese Einführung nicht ersetzen, aber kreativ ergänzen und gewaltfrei dazu hinführen.

Das vorliegende Buch basiert auf langjährigen Statistik-Lehrerfahrungen mit Studierenden, die eigentlich ‚etwas mit Menschen‘ und nicht mit Zahlen machen wollen. Ihr ehrliches Nachfragen forderte mich heraus, nach möglichst verständlichen Erklärungen zu suchen. Neben diesem essenziellen Input waren noch weitere Personen am Entstehen dieses Buches beteiligt. Minh Hang Le hat bei mehreren Fotos geholfen. Janna, Vivian, Sebastian und Tobias haben eine vorläufige Fassung kritisch gelesen und hilfreiche Hinweise gegeben. Andrea Cramer hat meine Kommata dorthin befördert, wo sie sein sollten und sprachliche Unsauberkeiten bereinigt. Frank Engelhardt vom Beltz Verlag begleitete vom ersten Gespräch an die Idee und Entstehung dieses Experiments mit motivierendem Interesse und gutem Rat, und Jakob Zey unterstützte die Arbeit am Layout und die Fertigstellung. All ihnen gilt mein herzliches Dankeschön.

Thomas Spiegler, Brandenburg an der Havel, im März 2020

Inhalt

jeder kann Statistik Was Statistik eigentlich macht, warum sie nicht schwer sein muss und wie wir sie im Alltag mühelos anwenden	9	
Häufigkeiten zählen Das Wichtigste ist nicht kompliziert: Häufigkeiten und Prozente	19	
die Mitte finden Der Durchschnitt macht alle gleich, in der Mitte steht der Median	27	
Verschiedenheit sehen Unterschiede treffend beschreiben: Quartile, Boxplot, Perzentile, mittlere Abweichung, Varianz, Standardabweichung, Variationskoeffizient, Gini-Koeffizient, z-Werte	43	
Zusammenhänge erkennen Erkennen, wie die Dinge miteinander verknüpft sind: Kreuztabellen, Streudiagramme, Korrelationskoeffizienten und Regressionsanalyse	79	
gut zu wissen Hilfreiche Begriffe kurz erklärt: Skalenniveaus, Stichproben, Ausblick auf die schließende Statistik	109	
Anhang, Anmerkungen, Quellen und Register Vertiefende Details, noch mehr Beispiele, Hinweise auf die Quellen und ein Stichwortregister	120	

jeder kann Statistik

Diese Überzeugung ist die Grundlage für das vorliegende Buch. Das kleine Kind, das noch nicht Lesen und Schreiben und gerade einmal die eigenen Finger abzählen kann, ist trotzdem in der Lage, die Holzringe richtig zu sortieren und zu sagen, wie viele von jeder Farbe vorhanden sind. Das ähnelt sehr dem, was das Statistische Bundesamt tut, wenn es seine Jahrbücher und Statistiken veröffentlicht.¹ Im Alltag nutzen wir ständig die Ideen, die der Statistik zugrunde liegen. Genau dort setzt dieses Buch an. Mit zahlreichen Beispielen zeigt es, wie wir im Alltag mühelos genau das tun, was Statistik auch leistet. Ausgehend von dieser ‚Statistik‘ im Alltag, die jeder kann, schlägt dieses Buch Brücken zu den Verfahren der wissenschaftlichen Statistik. Auch sprachlich werden hier weitgehend die Formulierungen verwendet, die auch im Alltag genutzt werden.² Das Buch macht deutlich, dass es oft nur ein kleiner Schritt ist von den vertrauten Alltagskonzepten hin zu der Statistik, die manchmal so kompliziert aussieht.

Zugegeben, für genaue Berechnungen ist etwas Mathematik in der Statistik sehr hilfreich. Aber um zu verstehen, wie Statistik funktioniert, benötigt man nur ganz wenig davon.



wechselhaft im Norden und gleichmäßig warm im Süden

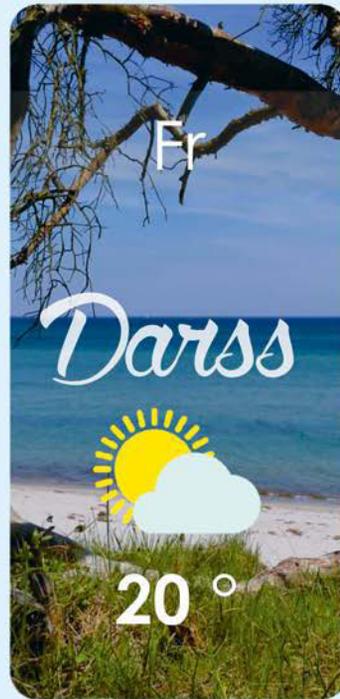
Yuna studiert in Berlin und fährt über ein langes Wochenende im Frühjahr an die Ostsee auf ihre 200 km nördlich gelegene Lieblings-Halbinsel Darß. Ihre Freundin Luisa studiert in Konstanz und fährt zur gleichen Zeit nach Italien an den 200 km südlich gelegenen Comer See.

An der Ostsee erlebt Yuna erst noch einen kühlen Regentag, aber dann einen für die Jahreszeit außergewöhnlich warmen Samstag und Sonntag. Am Comer See dagegen hat Luisa vier nahezu gleichmäßig warme und sonnige Tage.

Wenn Yuna das Wetter bei sich als wechselhaft beschreibt, und Luisa sagt, dass es gleichmäßig warm war, dann sind das beides Aussagen über die Streuung. Wir drücken mit diesen Begriffen aus, wie stark sich die Werte einzelner Tage unterscheiden.

Auf den folgenden Seiten wird mit diesem Beispiel gezeigt, wie man die Streuung der Tageshöchsttemperaturen genauer berechnen und in einer Maßzahl zusammenfassen kann.

▷ Streuung



Drei Grad mittlere Abweichung sind ziemlich wechselhaft

Die mittlere absolute Abweichung ist ein einfaches Maß, um die Streuung (die Verschiedenheit) in einer Gruppe von Werten zu bestimmen. Sie gibt an, wie weit die Werte durchschnittlich vom Mittelwert entfernt liegen. Das Bild zeigt die drei Schritte, wie sie berechnet wird.

Zuerst wird das arithmetische Mittel aller Werte berechnet: für die vier Tage auf dem Darß sind das 22 °C. Die einzelnen Tageswerte lagen darüber oder darunter. So war der Sonntag zum Beispiel 4 Grad wärmer als der Durchschnitt. Die mittlere absolute Abweichung ist der Durchschnitt aller Abstände, die die einzelnen Tageswerte zu diesen 22 °C haben. Im Beispiel sind dies 3 Grad. Man kann also sagen, dass an diesem langen Wochenende die Tageswerte im Mittel um 3 Grad von den 22 °C Durchschnittstemperatur abwichen.

Alternativ kann man die mittlere absolute Abweichung auch basierend auf den Abständen zum Median berechnen (in diesem Beispiel wäre dies der gleiche Wert, da Median und Durchschnitt der vier Tagestemperaturen gleich sind).

Bei dem zweiten Schritt der Berechnung macht es keinen Unterschied, ob ein Wert nach oben oder unten abweicht. Es wird mit dem positiven Betrag des Abstands gerechnet (absoluter Betrag genannt, daher auch mittlere *absolute* Abweichung).

Die mittlere absolute Abweichung ist leicht zu berechnen und kann gut verständlich interpretiert werden. Weitaus häufiger wird in der Statistik jedoch die Standardabweichung genutzt, da viele der komplexeren Analyseverfahren auf dieser basieren. Mehr zum Vergleich beider findet sich im Anhang. Die nächste Seite zeigt, wie die Standardabweichung berechnet wird.

▷ mittlere absolute Abweichung

1

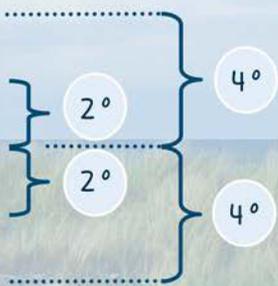
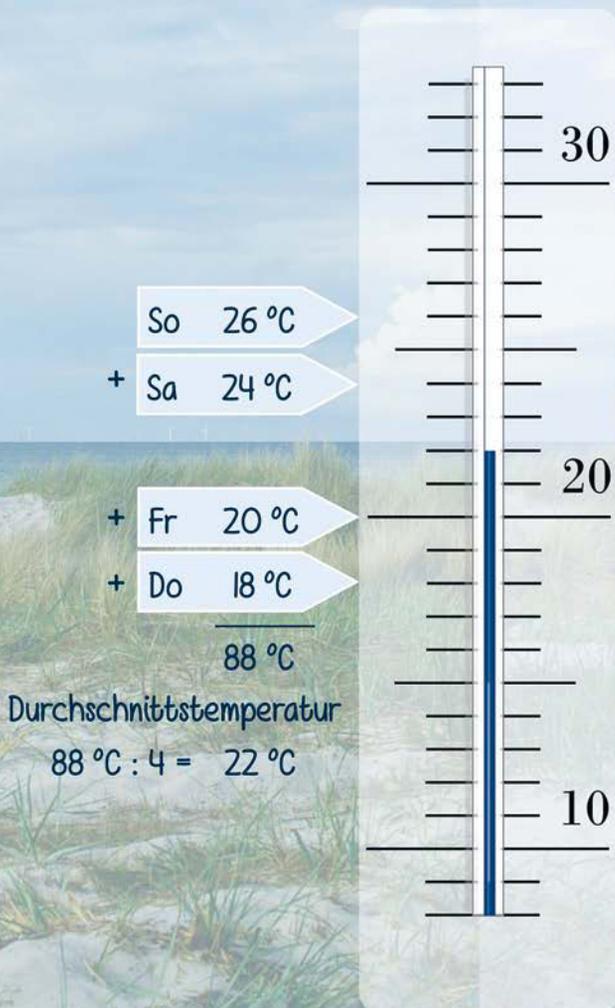
Durchschnitt aller Werte berechnen

2

für jeden Wert den Abstand zum Durchschnitt bestimmen

3

Durchschnitt dieser Abstände ermitteln



$$\begin{array}{r} 4^\circ \\ + 2^\circ \\ + 2^\circ \\ + 4^\circ \\ \hline 12^\circ \end{array}$$

Durchschnittsabstand:
 $12^\circ : 4 = 3^\circ$

mittlere absolute
Abweichung = 3 Grad



Bei Luisa am Comer See betrug die mittlere Abweichung nur 0,5 Grad

Anhang

Dieser Anhang gibt zu einigen Themen ergänzende Informationen und weitere Beispiele. Auch hier wird darauf verzichtet, die mathematischen Formeln in vollständiger Schreibweise aufzulisten und zu erklären. Dafür stehen auf dem Markt zahlreiche Statistiklehrbücher zur Verfügung, von denen viele auch auf die Anwendung in bestimmten Fachrichtungen eingehen (Wirtschaft, Soziologie, Psychologie, Naturwissenschaft ...). Die Ergänzungen in diesem Anhang folgen der gleichen Reihenfolge, in der die Themen im Hauptteil dargestellt wurden.

Häufigkeiten zählen

Prozentrechnungen lösen

Im Alltag begegnen uns relative Häufigkeiten meistens in Form von Prozentangaben. Wer die Prozentrechnung mit Prozentwert, Gesamtwert und Prozentsatz noch sicher aus Schulzeiten beherrscht oder sich eine andere gute Strategie zum Lösen von Prozentaufgaben eingeprägt hat, kann diesen Abschnitt getrost überblättern. Den anderen hilft vielleicht die folgende, einfache Methode, um Prozentaufgaben aller Art sicher zu lösen.

Es handelt sich um eine verkürzte Form des Dreisatzes. Fast alle Rechenaufgaben mit Prozent haben gemeinsam, dass drei Zahlen gegeben sind und eine vierte gesucht wird. Beispiel:

- Um wie viel sinkt der Preis einer 80-€-Hose wenn es 15 % Rabatt gibt?
Gegeben: 80 € entsprechen 100 %, gefragt ist, wie viel Euro 15 % sind.
- Wie viel Prozent haben für Person A gestimmt, wenn sie von 300 abgegebenen Stimmen 120 bekommen hat? Gegeben: 300 Stimmen entsprechen 100 %, gefragt ist, wie viel Prozent 120 Stimmen sind.

Mit den drei gegebenen Zahlen und der einen, die gesucht wird, kann man eine Verhältnisgleichung aufstellen. Für das zweite Beispiel heißt diese:

100 % $\hat{=}$ **300** (100 % entsprechen 300 abgegebenen Stimmen)

x % $\hat{=}$ **120** (gesucht ist, wie viel Prozent dann 120 Stimmen entsprechen)

Es ist egal, mit welchem Wert man beim Aufschreiben dieser Verhältnisgleichungen beginnt. Wichtig ist nur, dass die Prozente auf der einen Seite und die andere Einheit (hier Wahlstimmen) auf der anderen Seite untereinanderstehen. Dann funktioniert folgende einfache Lösung:

Um x zu erhalten, werden die beiden Zahlen, die sich diagonal gegenüberstehen (hier blau markiert), multipliziert und dann wird durch die Zahl, die übrig bleibt, geteilt. Dies ergibt:

$$x = \frac{120 \cdot 100 \%}{300} = (\text{Nullen wegkürzen}) = \frac{120 \cdot 1\%}{3} = 40 \%$$

Egal, ob man gleich zum Rechner greift oder erst den Bruch kürzt und dann eventuell im Kopf lösen kann, das Ergebnis ist hier 40 %. Die 120 Stimmen für Person A entsprechen 40 % der abgegebenen Stimmen.

Die Mitte finden

Der oft erwähnte und selten genutzte Modus

Neben dem arithmetischen Mittel und dem Median wird in der Statistikkultur oft auch der Modus (oder Modalwert) als weiteres Maß zentraler Tendenz genannt. Dieser Wert hat aber nicht wirklich etwas mit einer Mitte zu tun, da er einfach die Ausprägung benennt, die in einer Datengruppe am häufigsten vorkommt.

