

STATISTIK

IN CARTOONS



LARRY GONICK
& WOOLCOTT SMITH

Leseprobe

VAHLEN

Erstmals in deutscher Übersetzung: Der Statistik-Bestseller aus Amerika

Leserstimmen zur Originalausgabe:

»A great icebreaker to an often cold subject«

»Excellent and practical«

»GREAT BOOK (covers probability theory painlessly & well)«

»Clear, fun presentation of statistics«

»Fantastic learning tool. Makes light reading of statistics.«

»Anschaulicher kann man Statistik nicht darstellen«



Die Autoren:

Larry Gonick ist Grafiker, Autor und Co-Autor verschiedener Cartoons. Seit 1972 verpackt er wissenschaftliche Inhalte in leicht verständliche und unterhaltsame Form. Seine Cartoons werden unter anderem in Harvard, Stanford und Edinburgh zur Ausbildung eingesetzt. Das populärste seiner Bücher ist »The Cartoon Guide to Statistics«.

Woolcott Smith ist Professor für Statistik an der Fox School of Business and Management der Temple University.

Übersetzung und Adaption

Prof. Dr Erhard Cramer, Leiter des Instituts für Statistik und Wirtschaftsmathematik an der RWTH Aachen, und Dr. Katharina Cramer, Brauweiler

INHALTSVERZEICHNIS*

KAPITEL 1:	
WAS IST STATISTIK?	1
KAPITEL 2:	
BESCHREIBENDE STATISTIK	7
KAPITEL 3:	
WAHRSCHEINLICHKEITSRECHNUNG	27
KAPITEL 4:	
ZUFALLSVARIABLEN	53
KAPITEL 5:	
ZWEI VERTEILUNGEN?!	73
KAPITEL 6:	
STICHPROBEN-	89
KAPITEL 7:	
KONFIDENZINTERVALLE	111
KAPITEL 8:	
HYPOTHESENTESTS	137
KAPITEL 9:	
POPULATIONSVERGLEICHE	157
KAPITEL 10:	
VERSUCHSPLANUNG	181
KAPITEL 11:	
REGRESSIONSANALYSE	197
KAPITEL 12:	
ZUSAMMENFASSUNG	211
LITERATURVERZEICHNIS	221
SACHVERZEICHNIS	224

*Diese Leseprobe umfasst die ersten 14 Seiten des gesamten Buches.

Authorized translation from the English language edition, entitled
THE CARTOON GUIDE TO STATISTICS
by Larry Gonick and Woolcott Smith,
published by arrangement with Collins, an imprint of HarperCollins Publishers, LLC.

ISBN 978 3 8006 3589 6
© 2009 Verlag Franz Vahlen GmbH
Wilhelmstraße 9, 80801 München

Kapitel 1

WAS IST STATISTIK?

*Wir wurschteln uns durch's Leben,
indem wir Entscheidungen
ohne vollständige Information treffen...*



Die meisten von uns leben mit einem gewissen Grad an Ungewissheit recht gut.



Was die Statistik so einzigartig macht, ist ihre Fähigkeit, Ungewissheit zu **quantifizieren**, d. h. diese zu präzisieren. Das erlaubt Statistikern, **grundlegende Aussagen** mit völliger Zuversicht zu treffen - jedenfalls bzgl. des Grades an Unsicherheit!

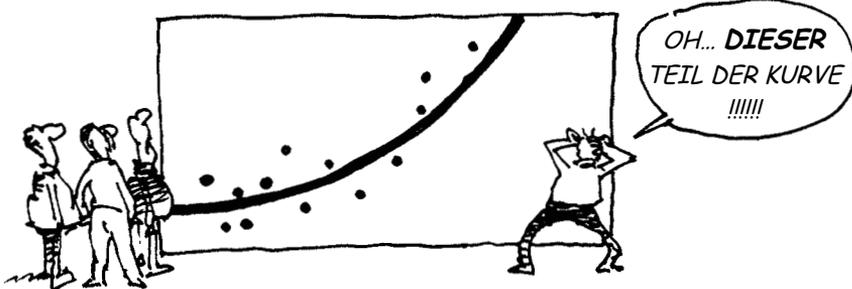


Hier geht's aber nicht nur um eine Suppenbestellung! Statistik spielt auch bei Fragen über **Leben und Tod** eine Rolle ...

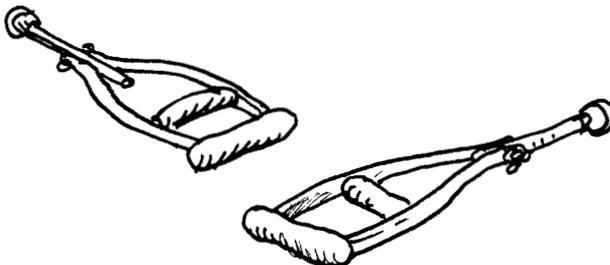
HEY, HATTEN SIE DIE SUPPE HIER SCHON 'MAL, ALS DER CHEF NICHT DA WAR?



Bei der Explosion der Raumfähre **Challenger** beispielsweise wurden 1986 sieben Astronauten getötet. Die Entscheidung, bei -2°C Außentemperatur zu starten, wurde ohne eine einfache Analyse der Messdaten bei niedrigen Temperaturen getroffen.



Ein positiveres Beispiel ist der von **Salk** entwickelte **Impfstoff gegen Kinderlähmung**. Der Impfstoff wurde unter strengen Kontrollen - um verzerrte Ergebnisse auszuschließen -, an mehr als 400 000 Kindern getestet. Eine umfassende statistische Analyse der Ergebnisse belegte die Wirksamkeit des Impfstoffes. Heutzutage ist Kinderlähmung daher nahezu unbekannt.



Um ihre großartigen mathematischen Zauberstücke zu vollbringen, stützen sich Statistiker auf drei verwandte Bereiche:

Daten- analyse

Das Sammeln, Visualisieren und Zusammenfassen von Daten;

Wahrscheinlich- keitsrechnung

Die Gesetze des Zufalls, in und außerhalb von Casinos;

Inferenzstatistik

Die Wissenschaft, mit Hilfe der Wahrscheinlichkeitsrechnung statistische Schlüsse aus Daten zu ziehen.



In diesem Buch werden alle drei Themenbereiche betrachtet und auf verschiedenste Situationen angewendet, in denen Statistik in der modernen Welt eine zentrale Rolle spielt.



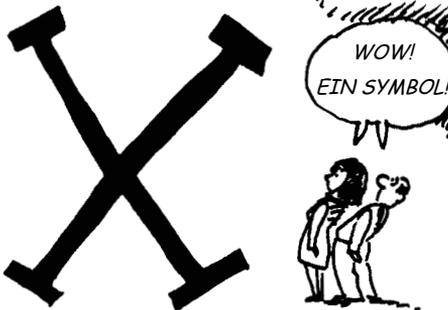
In Kapitel 2 betrachten wir einen einfachen Datensatz: die Körpergewichte in einem Haufen Studierender.



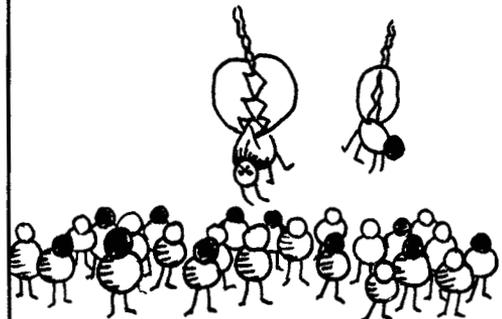
In Kapitel 3 studieren wir die Gesetze der Wahrscheinlichkeitsrechnung an ihrem Geburtsort, dem Spieltisch.



Kapitel 4 und 5 zeigen, wie sich die Welt unter Verwendung des Konzepts der Zufallsvariable durch Wahrscheinlichkeitsmodelle beschreiben lässt.



In Kapitel 6 wird eine der grundlegenden Methoden der Statistik vorgestellt: das Ziehen von Stichproben aus einer großen Grundgesamtheit.



In Kapitel 7 und folgende beschreiben wir, wie Statistik in Alltagssituationen - etwa bei Wahlumfragen, in der Qualitätskontrolle einer Produktion, bei medizinischen Studien, in der Umweltüberwachung und in der Rechtsprechung - eingesetzt wird.

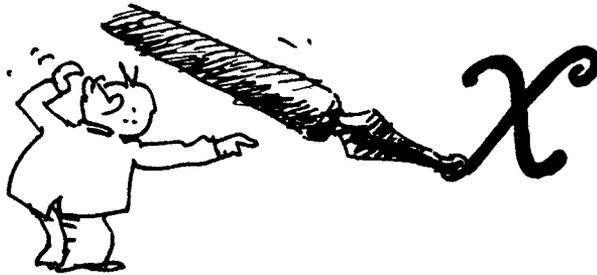


Abschließend muss noch ein Aspekt in der Wahrnehmung von Statistik erwähnt werden: Das in der heutigen Welt weit verbreitete **Misstrauen** gegenüber der Statistik. Das Thema „Lügen mit Statistik“ wird allseits diskutiert, wohingegen gute statistische Analysen im Alltag fast nie zu finden sind. Was ist zu tun?

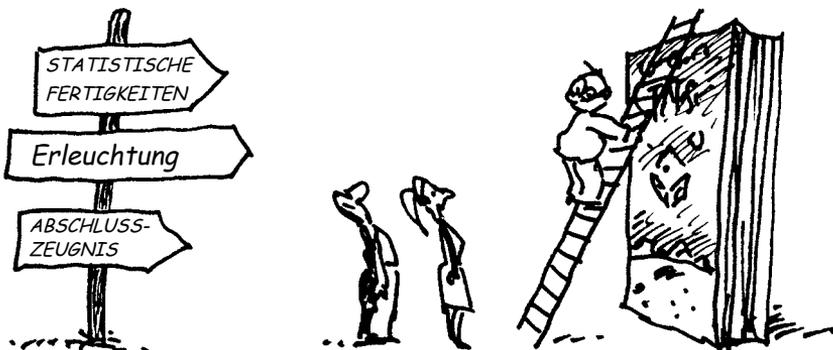
DREI VON VIER ÄRZTEN EMPFEHLEN,
ÄUSSERUNGEN BEGINNEND MIT
„DREI VON VIER ÄRZTEN...“
KEINEN GLAUBEN ZU SCHENKEN...



Unserer bescheidenen Meinung nach ist es keine so schlechte Idee, **etwas mehr über dieses Thema zu wissen...** Und daher haben wir dieses Buch geschrieben!



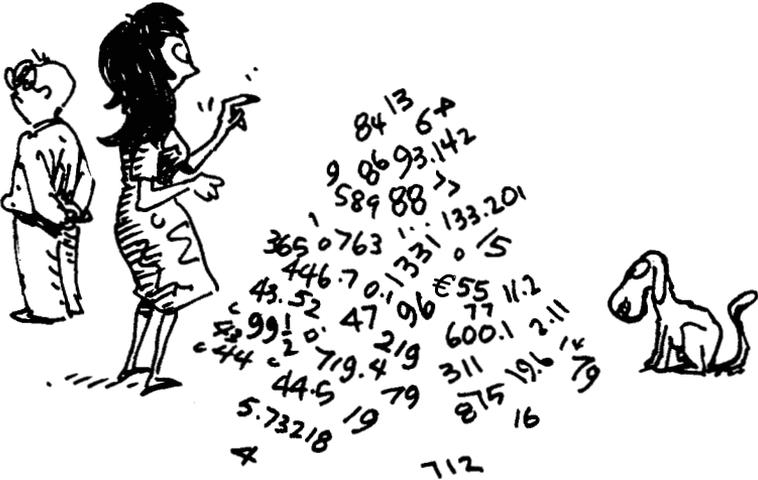
Im Folgenden versuchen wir, die Bausteine der Statistik so anschaulich und einfach wie möglich zu präsentieren. Alles was man dazu braucht, ist etwas **Geduld**, ein bisschen **Nachdenken** und eine gewisse Toleranz der **Mathematik** gegenüber - oder, falls nicht vorhanden, empfiehlt sich möglicherweise der **Besuch eines Kurses!!**



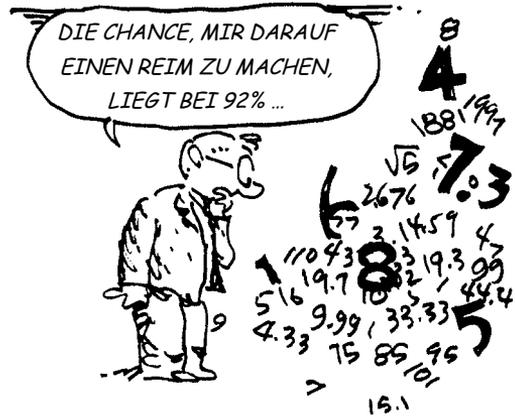
Kapitel 2

BESCHREIBENDE STATISTIK

HMM... TJA...
ES KÖNNTE...
SIE SIND...
ÄH...
HÜST....



Daten sind die Rohstoffe der Statistiker, die Zahlen, die wir verwenden, um die Realität zu erklären. Alle statistischen Probleme betreffen entweder das Sammeln, Beschreiben und Analysieren von Daten oder das **Nachdenken** über das Sammeln, Beschreiben und Analysieren von Daten.



Dieses Kapitel befasst sich mit der **Beschreibung** von Daten. Wie können wir Daten adäquat zusammenfassen? Wie können wir Strukturen in einem Haufen bloßer Zahlen entdecken? Wie können wir die Verteilung der Daten darstellen?



Nun gut, um Daten zu beschreiben, braucht man natürlich zunächst reale Daten, die beschrieben werden sollen ... Also, lasst uns Daten sammeln!



Hier sind ein paar echte Daten:
 Als Teil eines Hörsaalexperiments
 gaben 92 Studierende der Penn
 State Universität ihr Gewicht an:



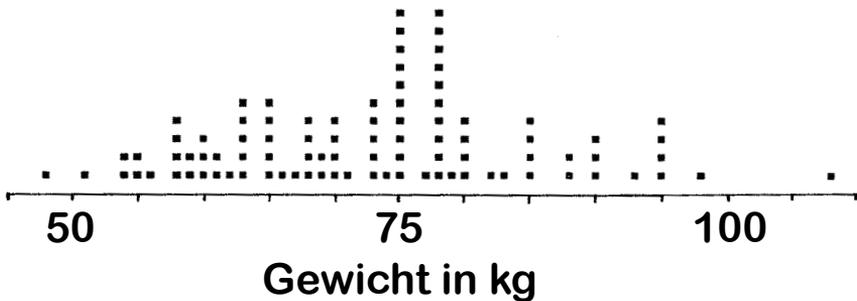
Männer

70 73 80 95 78 83 75 95 98 69 80 78 77 73 85 88 88 85 90 68 85 79 65 93 95
 78 85 78 108 75 73 78 78 75 78 75 90 80 68 80 65 78 75 74 78 75 70 90 95
 73 75 82 70 71 68 62 78

Frauen

70 60 65 69 61 63 58 73 75 56 63 65 60 65 66 60 59 63 68 63 59 61 58 51 58
 75 55 58 54 48 63 67 55 75 54

Kommen wir zur Sache und zeichnen ein **Punktendiagramm**: Pro Studierendem wird jeweils ein Punkt oberhalb seines Gewichts eingezeichnet.



Fällt Ihnen hier ein **Problem** auf - etwa die Häufungen bei 70, 75 und 85 kg? Die Studierenden tendieren offenbar dazu, ihr Gewicht in gewissen Schritten anzugeben. In Alltagssituationen wie dieser kann solches Runden allgemeine Muster verdecken ... Aber für den Augenblick wollen wir das vernachlässigen.

Wir können die Daten in einer **Häufigkeitstabelle** zusammenfassen. Dazu teilen wir den Zahlenstrahl in Intervalle ein und zählen die in jedem Intervall liegenden Gewichte der Studierenden. Die **Häufigkeit** ist die in jedem betrachteten Intervall festgestellte Anzahl. Die **relative Häufigkeit** ist der Anteil an Gewichtsangaben in einem Intervall, d. h. die Häufigkeit geteilt durch die Gesamtzahl der Studierenden.

Klassenintervall	Intervallmitte	Häufigkeit	Relative Häufigkeit
43,75 – 51,2	47,5	2	0,022
51,25 – 58,7	55	9	0,098
58,75 – 66,2	62,5	19	0,207
66,25 – 73,7	70	17	0,185
73,75 – 81,2	77,5	27	0,293
81,25 – 88,7	85	8	0,087
88,75 – 96,2	92,5	8	0,087
96,25 – 103,7	100	1	0,011
103,75 – 111,2	107,5	1	0,011
Gesamtsumme		92	1,000

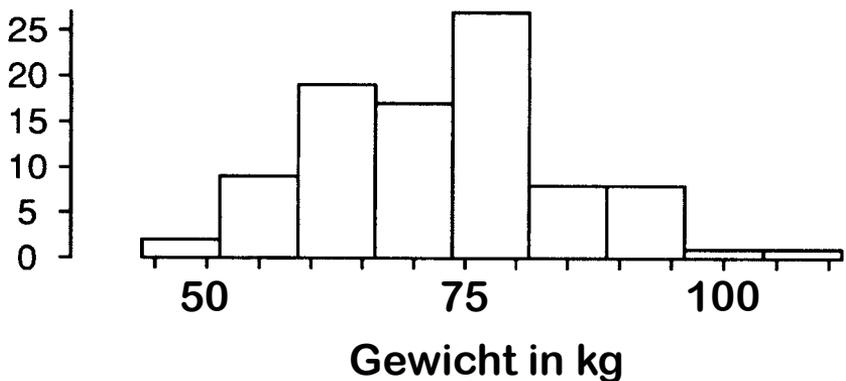
Achtung: Wir haben die Intervallgrenzen so gewählt, dass die störenden kg-Vielfachen nicht getroffen werden. So lässt sich die von den Studierenden erzeugte Verzerrung begrenzen.

Faustregel zur Festlegung der Klassenintervalle:

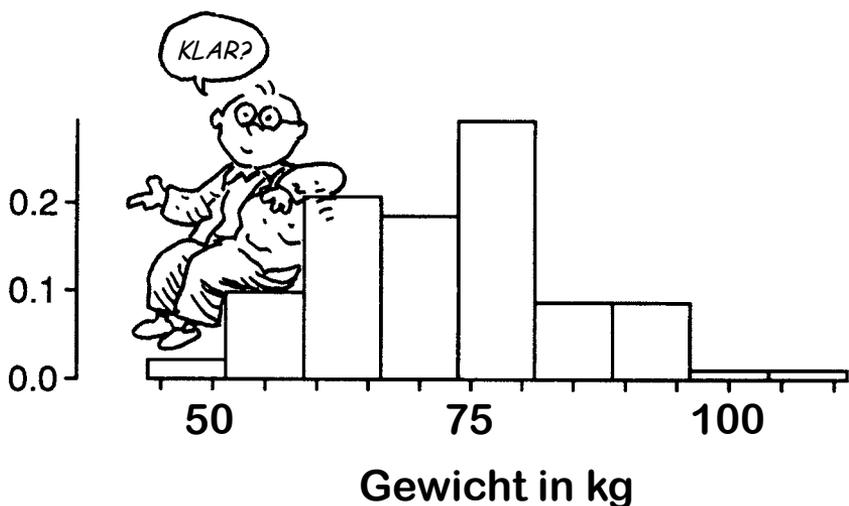
- 1)** Wähle Intervalle gleicher Länge mit Mittelpunkten an geeigneten „runden“ Werten.
- 2)** Bei kleinen Datensätzen: Wähle eine kleine Anzahl Intervalle.
- 3)** Bei großen Datensätzen: Wähle mehr Intervalle!



In der Häufigkeitstabelle wird dargestellt, wie viele Datenpunkte „in der Nähe“ eines beliebigen Wertes liegen. Grafisch können diese Information in einem speziellen Balkendiagramm dargestellt werden: dem **Histogramm**. Jeder Balken deckt ein Klassenintervall ab und ist an dessen Mittelpunkt zentriert. Die Balkenhöhe ist gleich der Anzahl Datenpunkte im betrachteten Intervall.



Wir können ebenso ein **Histogramm der relativen Häufigkeiten** zeichnen, indem die relative Häufigkeit gegen das Gewicht aufgetragen wird. Mit Ausnahme der vertikalen Achse sieht es genau gleich aus.

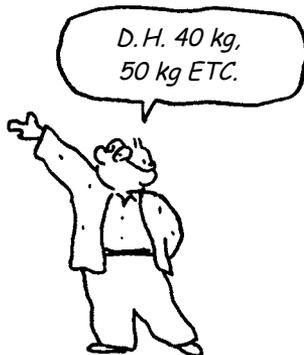


Der Statistiker **John Tukey** ersann ein Schnellverfahren, um Daten zusammenzufassen und die einzelnen Datenpunkte trotzdem zu erhalten: das **Stamm-Blatt-Diagramm**.



Im Beispiel der Gewichtsdaten ist der Stamm eine Spalte von Zahlen, die aus den Zehnerstellen der Messwerte gebildet wird (d. h. die Einerstelle wird gestrichen).

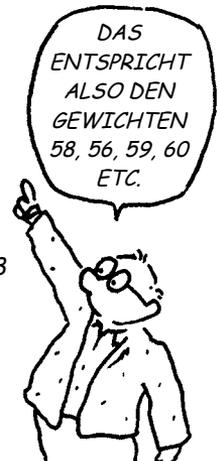
- 4
- 5
- 6
- 7
- 8
- 9
- 10



Füge nun die Einerstellen der Gewichte in den zugehörigen Zeilen hinzu:

Stamm: Blätter

- 4:
- 5: 869
- 6: 0133003595568
- 7: 035
- 8:
- 9:
- 10:



Fertig ausgefüllt, sieht es folgendermaßen aus:

- 4: 8
- 5: 144869988585
- 6: 013320033139855985855687
- 7: 03303340301858795885885858585855
- 8: 030002588555
- 9: 030055855
- 10: 8

Zuletzt: sortiere die „Blätter“ vom Kleinsten zum Größten.

- 4: 8
- 5: 144556888899
- 6: 000112333335555567888899
- 7: 00001333334555555555788888888899
- 8: 000023555588
- 9: 000355558
- 10: 8



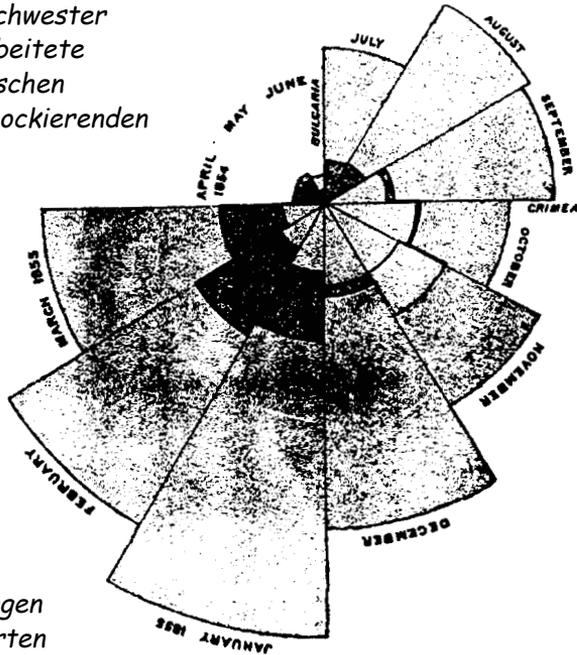
Die vielen Mehrfachnennungen illustrieren deutlich die Verzerrung in den Angaben der Studierenden!

Eine gute statistische Grafik
ist sowohl Kunst als auch
Wissenschaft

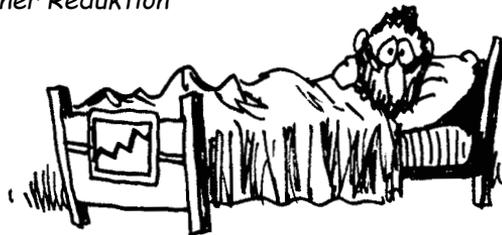


UND MANCHMAL AUCH
POLITIK!

Die kämpferische Krankenschwester
Florence Nightingale verarbeitete
Sterbestatistiken von britischen
Militärhospitälern zu so schockierenden
Histogrammen wie diesem:
Die radiale Achse zeigt
die Todeszahlen britischer
Soldaten im Krimkrieg,
unterschieden nach
Hospitälern und
Schlachtfeldern.



Ihre statistischen Bemühungen
führten direkt zu verbesserten
hygienischen Bedingungen in den
Hospitälern und zu einer Reduktion
der Sterberate.



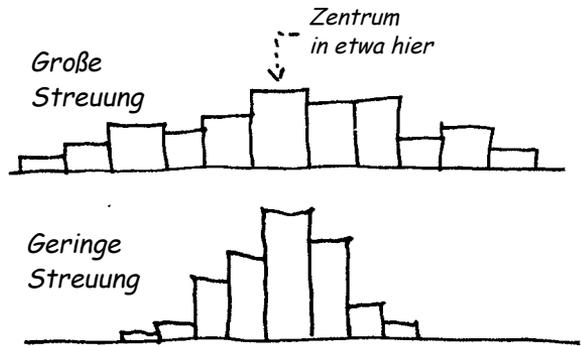
GERETTET
DURCH
STATISTIK!

STATISTISCHE KENNGRÖSSEN

Nun machen wir den Schritt vom Bild zur Formel. Unser Ziel ist es, simple Eigenschaften eines Datensatzes durch einfache Kenngrößen zu beschreiben...



Jeder Datensatz weist zwei wichtige Charakteristika auf: den **Zentralwert** und die **Streuung** um diesen Wert. Diese Konzepte werden durch diese konstruierten Histogramme illustriert.



Wir fahren fort mit ein wenig Notation. Angenommen, wir machen ein paar Beobachtungen ... n , um exakt zu sein ... Dann schreiben wir

$$x_1, x_2, x_3, \dots, x_n$$

für die Beobachtungswerte. n ist also die Gesamtzahl an Beobachtungen und x_4 ist beispielsweise der Wert der vierten Beobachtung.

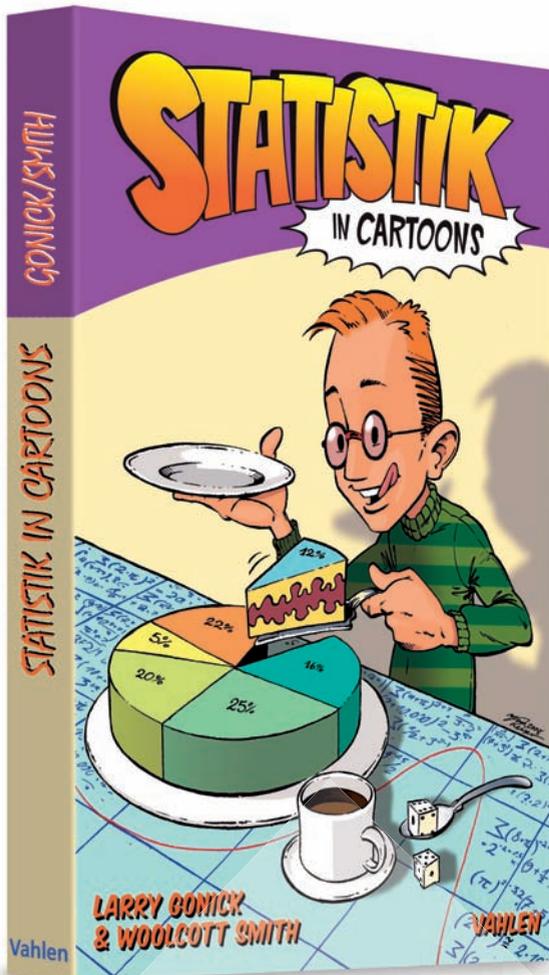
Eine **Datenreihe** ist eine Tabelle der Daten:

Beobachtung	1	2	3	4	...	n
Messwert	x_1	x_2	x_3	x_4	...	x_n



Deutsche Erstauflage März 2009.

Anschaulicher kann man Statistik nicht darstellen



Das ideale Buch für Studierende aller Fachrichtungen, bei denen die Statistik zentraler Bestandteil des Curriculums im Bachelor ist (vom Bioinformatiker bis zum Wirtschaftswissenschaftler) sowie für Schüler, die sich mit Statistik in der Oberstufe beschäftigen müssen.

Gonick/Smith

Statistik mit Cartoons

Von Larry Gonick und Prof. Dr. Woollcott Smith.
Übersetzung und Adaption durch Prof. Dr. Erhard Cramer, Aachen 2009.

Rund 230 Seiten. Kartoniert ca. € 19,80
ISBN 978-3-8006-3598-6

