

DAS

DECISION

MAKER

PLAYBOOK

12 TAKTIKEN, UM KLARER ZU DENKEN,
UNSICHERHEIT ABZUBAUEN UND
BESSERE ENTSCHEIDUNGEN ZU TREFFEN

SIMON MUELLER UND JULIA DHAR

DAS

DECISION

MAKER

PLAYBOOK

12 TAKTIKEN, UM KLARER ZU DENKEN,
UNSIKERHEIT ABZUBAUEN UND
BESSERE ENTSCHEIDUNGEN ZU TREFFEN

SIMON MUELLER UND JULIA DHAR

triebsmitarbeiter für das laufende Geschäftsjahr vor. Einige von ihnen erhalten Bonuszahlungen, weil sie zusätzliche Aufträge an Land gezogen haben, während andere ihr normales festes Gehalt bekommen. Sie wollen wissen, ob die Bonuszahlungen gerechtfertigt sind, und notieren sich einige Hypothesen, die sich aufdrängen (stellen Sie sich diese als sehr spontan auftauchende Gedanken vor). Eine solche Liste könnte folgendermaßen aussehen:

- H1: Die Vertriebsmitarbeiter verkaufen mehr, wenn ihnen Bonuszahlungen angeboten werden.
- H2: Die Umsätze steigen proportional zu den potenziellen Bonuszahlungen.
- H3: Die daraus resultierenden Umsatzsteigerungen sind höher als die Kosten für die Bonuszahlungen.

Wenn wir als Mentoren junge oder neue Mitarbeiter betreuen, stellen wir häufig fest, dass sie zögern, eine Hypothese aufzustellen. Sie haben Angst, dass es negative Auswirkungen hätte, sollte die Hypothese sich als falsch erweisen. Wir hören z.B.: „Ich will mich nicht mit meiner Meinung aus dem Fenster lehnen, bevor ich die Daten nicht gründlich unter die Lupe genommen habe“, oder „Es ist noch zu früh, etwas dazu zu sagen“. Wir verstehen das, aber eine Hypothese ist nur eine grobe Idee, keine Antwort oder Schlussfolgerung. Indem Sie eine solche Hypothese aufstellen, strukturieren Sie Ihre Analyse und versehen sie mit Prioritäten. Ohne einen solchen Schritt werden Sie womöglich von Ihren Daten überrollt. Dieser Schritt ist umso wichtiger, wenn Sie die Analyse nicht selbst durchführen, sondern der Endverbraucher sind. Wenn jemand aus Ihrem Team oder ein Berater sich darauf vorbereitet, diese Arbeit zu erledigen, dann ist es von entscheidender Bedeutung, dass Sie selbst in die Formulierung und Überarbeitung Ihrer Hypothese involviert sind, sodass Sie auch etwas von der Übung haben.

Schöpfen Sie Ihre Daten aus

Auch wenn Sie bereits eine Reihe wohlüberlegter Hypothesen aufgestellt haben, sollten Sie nicht direkt anfangen, sie zu überprüfen. Vorher müssen Sie sicherstellen, ob die Datenbasis sich für Ihren Job überhaupt eignet. Die folgende Checkliste gibt Ihnen einen Überblick über die Fragen, die wir üblicherweise stellen, bevor wir in Datensätze abtauchen.

Fragestellung	Warum ist das wichtig?	Beispiel
<p>Ist dieser Datensatz in vollem Umfang repräsentativ?</p>	<p>Ihre Daten müssen Annäherungswerte an die Realität darstellen. Das kann auf zweierlei Weise geschehen:</p> <ul style="list-style-type: none"> • Datensatz besteht aus Grundgesamtheit. • Datensatz enthält Stichproben. <p>Wenn Ihre Daten nicht die Grundgesamtheit aller Ereignisse oder eine Stichprobe enthalten, dann sind sie nicht repräsentativ. Das bedeutet, dass es nicht möglich ist, statistisch relevante Schlussfolgerungen aus den Daten herzuleiten. Es kann trotzdem sein, dass Sie aus den vorhandenen Daten wertvolle Rückschlüsse ziehen können, aber Sie sollten vorsichtig sein.</p>	<p>Sie verfügen wahrscheinlich über die Daten sämtlicher Transaktionen, die ein Kunde jemals mit Ihrem Unternehmen vorgenommen hat. Aus Ihrer Sicht kann es sich dabei um eine Grundgesamtheit handeln. Wenn sich die Daten allerdings nur auf Geschäfte beziehen, die mit einer Kreditkarte beglichen wurden, nicht jedoch auf Bargeschäfte, dann wäre der Datensatz nicht repräsentativ.</p> <p>Manchmal ist es unhandlich, unpraktisch oder zu teuer, Daten über die Grundgesamtheit zusammenzutragen. Nehmen wir an, Sie wollen wissen, wie zufrieden all Ihre Mitarbeiter im Unternehmen sind. Sie könnten jeden Einzelnen fragen – oder Sie könnten zufällig ausgewählte Mitarbeiter befragen. Wichtig daran ist, dass die Stichprobe sich im Hinblick auf relevante Kriterien der Grundgesamtheit annähert (dass sie beispielsweise den gleichen Anteil an Frauen und Männern erfasst, der in der gesamten Belegschaft vorhanden ist, ebenso wie den gleichen Anteil an alten und jungen Mitarbeitern oder dass die Stichprobe Mitarbeiter aus allen Niederlassungen enthält).</p>

Tabelle 3.1: Repräsentativität des Datensatzes

Fragestellung	Warum ist das wichtig?	Beispiel
Aus welchen Quellen stammen die Datensätze? Wie wurden sie erfasst?	<p>Warnsignale, die darauf hinweisen, dass die Daten Verzerrungen aufweisen oder nicht repräsentativ sind:</p> <ul style="list-style-type: none"> • Antworten von Selbsternannten: Gibt es eine bestimmte (Unter-)Gruppe von Mitarbeitern, die eher bereit sind, Fragen zu beantworten als andere? • Ichbezogene Antworten: Menschen haben eine Neigung dazu, Fragen so zu beantworten, dass es ihnen zum Vorteil gereicht. • Antworten Selbstinteressierter Analysten: Datensätze, bei denen die Person, die die Analyse durchführt, ein Interesse am Ergebnis ihrer Arbeit hat, sollten ebenfalls mit Vorsicht behandelt werden. Wenn ein Hersteller von Reinigungsmitteln eine Studie durchführt, die mit der Schlussfolgerung endet, dass ein durchschnittlicher Haushalt vor gefährlichen Bakterien nur so wimmelt und dem nur durch den entschlossenen Einsatz von Reinigungsmitteln entgegengewirkt werden kann, dann sollten Sie stutzig werden. • Absichtserklärungen versus beobachtete Verhaltensweisen: Wenn Sie die Mitglieder eines Fitness-Studios befragen, ob sie beabsichtigen, in der darauffolgenden Woche zu trainieren, ist das weit weniger aussagekräftig, als wenn Sie das Einbuchungen der Mitglieder an den Geräten untersuchen (offenbarte Präferenzen). 	<p>Wir greifen noch einmal unsere Mitarbeiterbefragung auf. Wenn Sie einen Umfragebogen an alle Mitarbeiter verschicken, um sie nach ihrer Zufriedenheit zu befragen, dann sollten Sie annehmen, dass die Antworten repräsentativ sind. Wer beantwortet solche Fragebögen? Diejenigen, die sehr zufrieden sind, diejenigen, die sehr unzufrieden sind, und diejenigen, die ein Interesse daran haben, dass die Ergebnisse besonders gut oder schlecht ausfallen.</p> <p>Zweitens: Wenn Sie Menschen die Frage stellen, was sie tun, statt zu beobachten, wie sie sich verhalten, dann seien Sie vorsichtig. Als Verhaltensökonomin lebt Julia nach der Devise „jede/-r täuscht ohne böse Absicht“. Menschen neigen dazu, ihr positives Verhalten überzubewerten und ihr weniger positives herunterzuspielen (das wird als selbstwertdienliche Strategie bezeichnet). Jedes Mal, wenn Sie auf ichbezogene Antworten vertrauen müssen, sollten Sie sich der Gefahr der selbstwertdienlichen Verzerrung bewusst sein.</p> <p>Wenn Sie beispielsweise heterosexuelle Paare befragen, welchen Anteil an der Hausarbeit sie leisten, dann ist die Summe in der Regel höher als 100 Prozent. Es liegt auf der Hand, dass es aussagekräftiger wäre, das Verhalten zu beobachten oder Haushaltstagebücher auszuwerten.</p>

Tabelle 3.2: Quellen und Erfassung des Datensatzes

Fragestellung	Warum ist das wichtig?	Beispiel
Wer wird durch diesen Datensatz nicht erfasst? ⁵	<p>Zu wissen, wer oder was in Ihren Datensätzen nicht auftaucht, ist wirklich wichtig. Sie sollten sich über zwei Fehlerquellen im Klaren sein:</p> <ul style="list-style-type: none"> • Zufällig fehlende Daten: Stellen Sie sich diese als Ereignisse vor, die aus Ihren Datensätzen ausgeschlossen wurden, ohne dass dies die Variablen beeinflussen würde, an denen Sie interessiert sind. • Nicht zufällig fehlende Daten: Ereignisse, die in Ihrem Datensatz fehlen und durch die die Variablen, an denen Sie interessiert sind, tatsächlich verfälscht werden. Hierbei handelt es sich um systematische Verzerrungen, durch die Erkenntnisse aus Ihren Datensätzen signifikant verfälscht werden könnten. 	<p>Wie sehen diese beiden Typen fehlender Daten aus?</p> <ul style="list-style-type: none"> • Zufällig fehlende Daten: Sagen wir, die Kasse in einer bestimmten Filiale hat an einem bestimmten Tag im Mai nicht richtig funktioniert. Für diese Filiale liegen Ihnen also nur die Datensätze für 364 Tage vor. Solange dieser Tag nicht aus irgendeinem Grund von besonderem Interesse ist (z.B. der Black Friday), dann spielt das Fehlen der Daten vermutlich keine große Rolle, wenn Sie untersuchen wollen, welche Ihrer Filialen am erfolgreichsten war. Sie können die fehlenden Daten eventuell durch den Datensatz vom gleichen Tag des vorangegangenen Jahres ersetzen. • Nicht zufällig fehlende Daten: In unserer Vertriebsdatenbank werden die Onlineumsätze bestimmter Bundesstaaten nicht erfasst, sondern nur die direkten Verkäufe, die von Vertriebsmitarbeitern abgeschlossen werden – und die Onlineumsätze in anderen Bundesstaaten. In diesem Fall müssen Sie einen anderen Weg finden, um die fehlenden Daten zusammenzutragen, bevor Sie mit der Analyse fortfahren. Ansonsten können Sie die Daten nur benutzen, um Schlussfolgerungen über Ihre Vertriebsmitarbeiter zu gewinnen, nicht jedoch über Ihre Onlinekanäle.

Tabelle 3.3: Fehlende Daten beim Datensatz

Bauen Sie Säulen aus Daten

Nun haben Sie die Tabellen auf den letzten Seiten durchgearbeitet und sind mit Ihren Datensätzen zufrieden. Sie erscheinen Ihnen repräsentativ, weisen relativ geringe Verzerrungen auf und dienen dem gewünschten Zweck. Fangen wir also an, sie auszuwerten.

Unser Gehirn liebt das, was Ökonomen als Referenzpunkte bezeichnen. Das sind einzelne Daten, auf die wir zugreifen und die wir nutzen, um unsere Gedanken um sie herum zu organisieren. Ein Durchschnittswert ist ein solcher Referenzpunkt. Wir haben alle in unserem Mathematikunterricht gelernt, dass ein Durchschnittswert

5 Das Problem fehlender Daten ist von großem Interesse, sowohl für Wissenschaftler in der Forschung als auch für Datenanalysten und Systemingenieure. Letztendlich geht das Thema uns alle etwas an. Zur intensiveren Beschäftigung mit fehlenden Daten siehe Raghunathan, T. (2015), *Missing Data Analysis in Practice*. Chapman and Hall/CRC.

berechnet wird, indem wir die Summe aller Werte durch die Anzahl der Werte dividieren. In diesem Sinne ist der Durchschnittswert sinnvoll. Er lässt Rückschlüsse darauf zu, was ungefähr richtig ist.

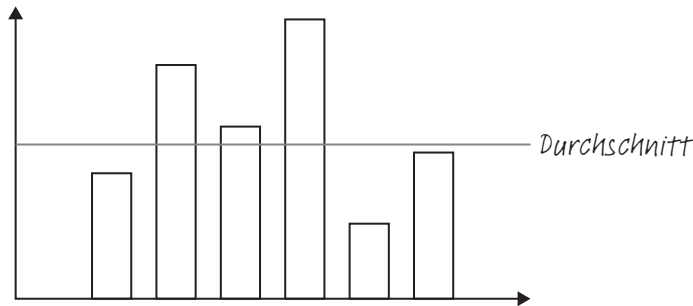


Abbildung 3.2: Darstellung eines Durchschnittswerts

In manch anderer Hinsicht ist ein Durchschnittswert jedoch überhaupt keine Hilfe und die Fixierung vieler Führungskräfte, Lehrer, Politiker und Analysten auf durchschnittliche Leistungen kann in die Irre führen oder sogar gefährlich werden. Andererseits können Durchschnittswerte enorm zur Motivation beitragen. Viele Menschen vergleichen sich gern mit dem Durchschnitt und versuchen, diesen zu übertreffen.

Stellen wir uns einen Biologielehrer vor, der eine Klasse mit 30 Schülern unterrichtet. Diese Schüler legen regelmäßig schriftliche Prüfungen ab, um zu überprüfen, ob sie die Unterrichtsinhalte verstanden haben. Natürlich gibt es einige Schüler, die besser abschneiden als andere. Im Folgenden finden Sie ein Streudiagramm, das die Leistungen der Schüler in ihrer letzten Klausur enthält: die Zahl der Schüler mit einer bestimmten Punktzahl auf der y-Achse und die Wertung (bis zu 100 Punkte) auf der x-Achse.

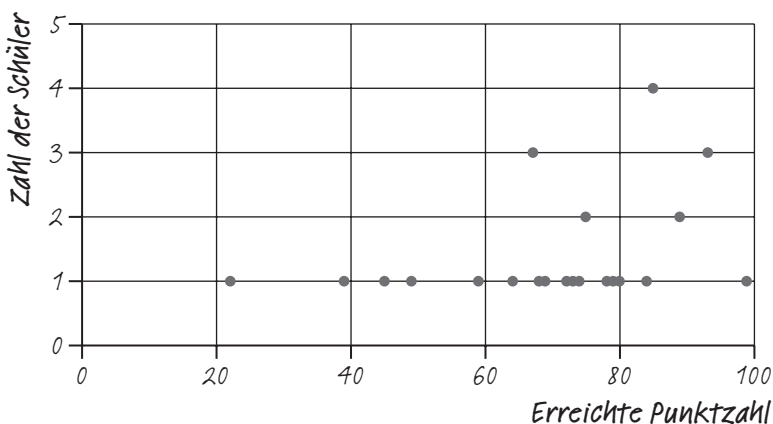


Abbildung 3.3: Leistungsverteilung von Schülern im Streudiagramm

Als Erstes können wir einige Angaben dazu machen, wie unsere Daten verteilt sind. Wir haben eine geringe Anzahl an Schülern, die weniger als 50 von 100 möglichen Punkten erreicht haben, recht viele Schüler, die zwischen 60 und 80 Punkten erreicht haben, und ein paar wenige Schüler mit 80 bis 100 Punkten. Um wirklich aussagekräftige Rückschlüsse aus irgendeinem neuen Datensatz zu gewinnen, übertragen wir die Daten als Erstes in ein Säulendiagramm. Wir glauben, dass Säulen- oder auch Balkendiagramme ungeliebte, aber sehr aussagekräftige Grafiken sind. Ein Säulendiagramm ermöglicht Ihnen, auf den ersten Blick Wahrscheinlichkeiten zu erkennen. Es erlaubt Ihnen, sofort zu erfassen, wie wahrscheinlich ein bestimmtes Ereignis (in diesem Fall die erreichte Punktzahl) ist. Sehen Sie sich die Testergebnisse als Säulendiagramm an.

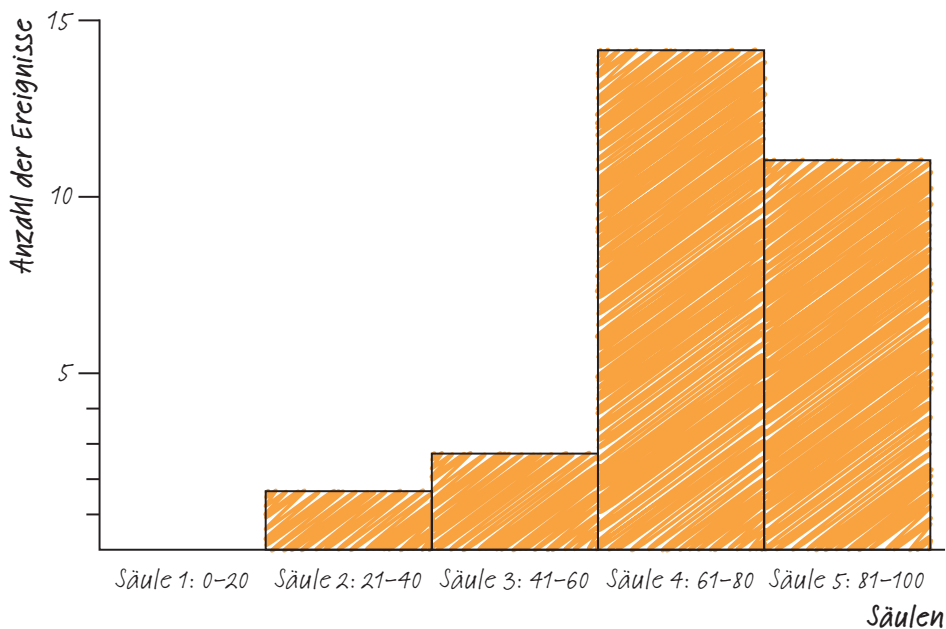


Abbildung 3.4: Leistungsverteilung von Schülern im Säulendiagramm

Wenn wir das Säulendiagramm erstellt haben, können wir auf den ersten Blick über- und unterdurchschnittliche Leistungen erkennen. Wir können auch damit beginnen, uns die Verteilung der Ergebnisse anzusehen. Werfen Sie einen Blick auf das folgende Säulendiagramm, um eine grobe Skizze der Datenverteilung zu sehen. Sie haben eine Klasse, die ziemlich gut abgeschnitten hat, in der aber einige Schüler sind, die ungenügende Leistungen gezeigt haben. Stellen Sie sich vor, Sie wiederholen diese Übung mit den Datensätzen Ihrer Vertriebsmitarbeiter und deren Umsätzen oder mit Informationen über Ihren IT-Support und die dort bearbeiteten Anfragen. Sie erkennen sofort, wie aussagekräftig ein Säulendiagramm für einen ersten Eindruck zu Ihren Daten ist. Wir können eine beachtliche Zahl von Schülern sehen, die sehr gut abgeschnitten haben, eine große Anzahl, die gute Leistungen erbracht hat, und eine kleine Gruppe, die zu kämpfen hat.

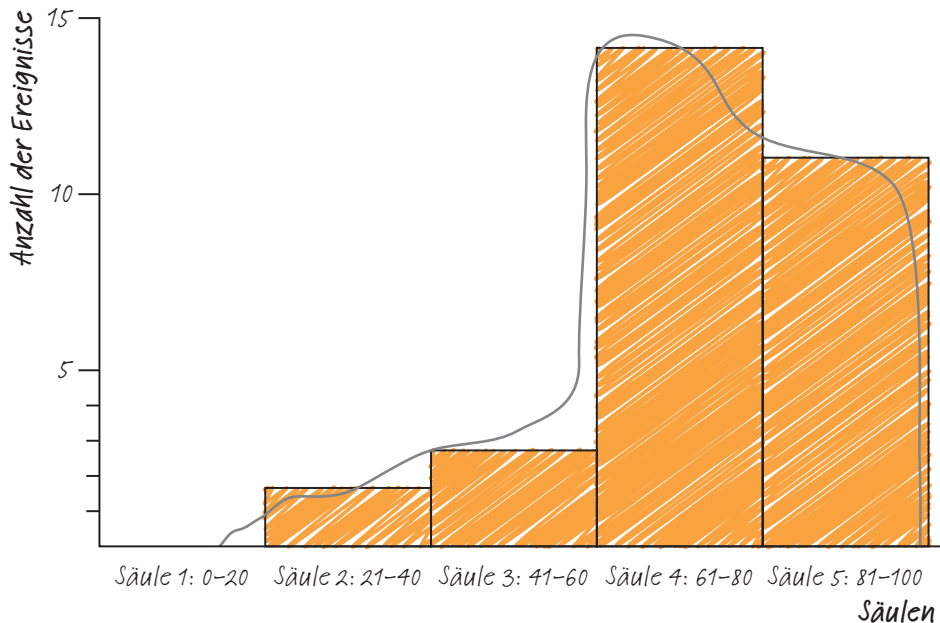


Abbildung 3.5: Ermittlung des Durchschnittswerts der Punkte

Wenn Sie sich die Grafik oben ansehen, sollten Sie sofort erkennen, dass eine durchschnittliche Punktzahl hier nicht sonderlich hilfreich ist. Aber es gibt einige andere Werte, die Sie vielleicht weiterbringen. Vielleicht wollen Sie sich einen Überblick über die Bandbreite der Leistungen verschaffen und schauen auf die Mindest- und Höchstpunktzahl. Sobald wir unser Säulendiagramm erstellt haben, bedienen wir uns mehrerer Verfahren aus der deskriptiven Statistik, um herauszufinden, welche Geschichte die Daten zu erzählen haben. Genau das spielen wir mit unserem Datensatz auf den folgenden Seiten durch. Das Ziel dieses Abschnittes ist es, Ihnen ein Gefühl für die Daten zu vermitteln und nicht unbedingt irgendwelche Schlussfolgerungen zu ziehen. Beispielsweise können wir anhand des Säulendiagramms nicht erkennen, ob die Klausur einfach oder schwierig war, ob die Schüler, die schlecht abgeschnitten haben, irritiert oder krank oder nach der Mittagspause einfach abgelenkt waren. Wenn Ihr Säulendiagramm steht, sollten Sie als Erstes eine „Fünf-Zahlen-Zusammenfassung“ erstellen.

Nutzen Sie Methoden aus der deskriptiven Statistik

Die „Fünf-Zahlen-Zusammenfassung“ ist ein einfaches Tool, mit dem Sie Ihre Daten noch ein wenig genauer erforschen können. In unserem Datensatz sieht die Zusammenfassung folgendermaßen aus:

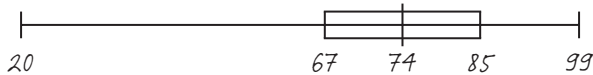


Abbildung 3.6: Fünf-Zahlen-Zusammenfassung

Das Diagramm wird auch als „Box-and-Whisker-Plot“ bezeichnet. Dabei handelt es sich um eine weitere, völlig unterschätzte Grafik, die Ihre Daten wirkungsvoll darstellt. Ein Box-and-Whisker-Plot enthält fünf Werte:

- Den Median, oder auch Mittelwert (Zentralwert), das ist die erreichte Punktzahl, die genau in der Mitte der (Punkt-)Verteilung liegt. Es ist immer hilfreich, sowohl den Median als auch den Durchschnittswert einer Verteilung zu kennen, weil der Median durch die Ausreißer (z.B. ein sehr niedriges oder ein sehr hohes Ergebnis) weniger beeinträchtigt wird.

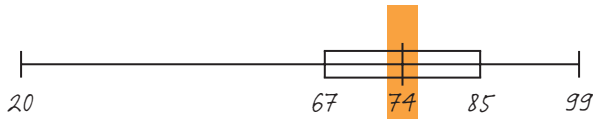


Abbildung 3.7: Median

- Das Ereignis mit dem niedrigsten Wert des Datensatzes (in diesem Fall das Ergebnis unseres unglücklichen Schülers mit der Punktzahl 20).
- Das Ereignis mit dem höchsten Wert des Datensatzes (in diesem Fall unsere Schlauberger, die 99 Prozent der Punkte erreicht haben).

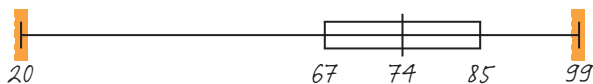


Abbildung 3.8: Niedrigster und höchster Wert des Datensatzes

- Das erste Quartil unseres Datensatzes (Q1) und das dritte Quartil des Datensatzes (in diesem Fall 67 bzw. 85).

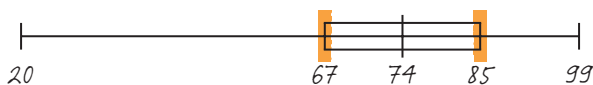


Abbildung 3.9: Erstes und drittes Quartil des Datensatzes

Diese ersten Spielereien mit Ihren Daten nehmen so gut wie keine Zeit in Anspruch. Wenn Sie einen überschaubaren Datensatz zur Verfügung haben, können Sie die Berechnungen von Hand erledigen, wie hier gezeigt. Wenn es sich um einen größeren Datensatz handelt, greifen Sie auf ein Datenanalyse-Paket (Microsoft Excel, SPSS, Stata, R) zurück. Wenn Sie ein Team von Mitarbeitern haben, die sich mit der Datenanalyse beschäftigen, dann bitten Sie sie darum, dass Ihnen ein solcher statistischer Überblick zur Verfügung gestellt wird, damit Sie sich ein realistisches Bild von den Daten machen können, mit denen Sie arbeiten.

Diese drei Werkzeuge (Streudiagramm, Säulendiagramm und Box-and-Whisker-Plot) können sehr hilfreich sein, um Ihre Daten in den Griff zu bekommen. Stellen Sie sich einen Moment lang vor, dass Sie Callcenter-Manager sind und die Daten, die Sie oben verwendet haben, die durchschnittlichen Minuten ungenutzter Zeit darstellen, die jedes Ihrer Teammitglieder zwischen zwei Anrufen verbringt. Jetzt ist es sehr einfach, die Spitzenreiter und die Schlusslichter innerhalb des Teams zu identifizieren und Leistungsabweichungen zu erkennen. In unserem Box-and-Whisker-Plot hat unser Lehrer seine Daten sofort unterteilt. Er hat nun nicht mehr 30 Schüler, sondern vier Gruppen, in denen unterschiedliche Leistungen erbracht werden und die jeweils unterschiedliche Lernansätze benötigen.

Verteilen Sie Ihre Daten

Sie haben oben gesehen, wie die Werte aus unserem Beispiel verteilt wurden: Nun werden wir Ihnen einige weitere gängige Verteilungsfunktionen aus der Statistik vorstellen (also zeigen, welche Formen Ihre Daten annehmen können) und Ihnen helfen zu antizipieren, wo Sie diese finden können.⁶

Wenn Sie diese Verteilungsfunktionen kennen und wissen, unter welchen Bedingungen sie mit großer Wahrscheinlichkeit auftreten werden, können Sie aus dem, was Sie tatsächlich beobachtet haben, Rückschlüsse über die Wahrscheinlichkeit der Ereignisse ziehen. Und mehr noch, wenn Sie die gängigen Verteilungsfunktionen kennen, hilft Ihnen das, die Datenpunkte, die Sie noch nicht erfasst haben, besser einzuschätzen. Wenn Sie die Verteilung Ihrer Daten verstehen, erkennen Sie Muster hinter den einzelnen Ereignissen.

6 Mehr zum Thema Verteilung finden Sie bei Tegmark, M. (2014) *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. Vintage.

Normalverteilung

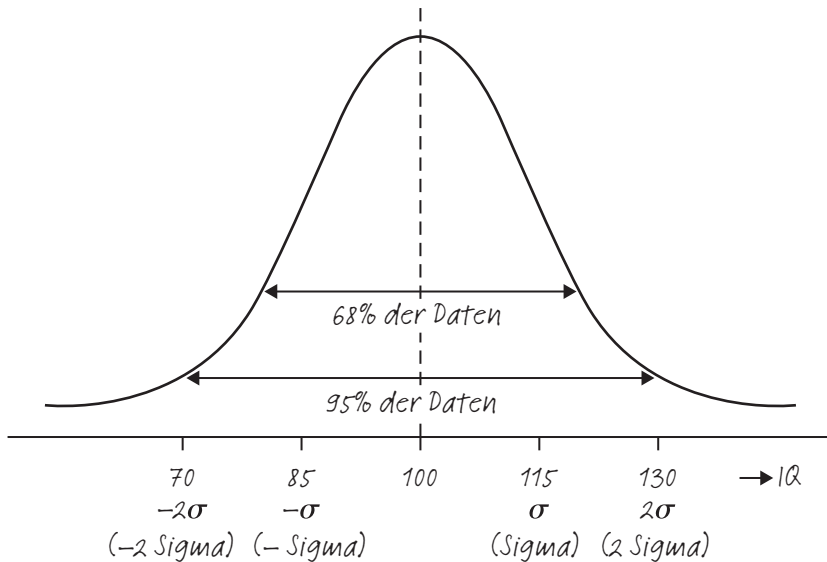


Abbildung 3.10: Normalverteilung am Beispiel des IQ einer Gruppe von Menschen

Mit dieser Verteilungsfunktion sind Sie höchstwahrscheinlich vertraut (sie wird ihrer Form wegen üblicherweise als „Glockenkurve“ oder auch als Gauß-Verteilung nach dem berühmten Mathematiker und Physiker Carl Friedrich Gauß bezeichnet). Es handelt sich um eine der häufigsten Verteilungsfunktionen in der Naturgeschichte. Wenn Sie die Größe oder auch die IQs einer Gruppe von Menschen grafisch darstellen wollen, dann nimmt die Verteilung die Form einer Glockenkurve an. Der durchschnittliche IQ einer Population ist der Standard von 100. Das bedeutet, dass Sie erwarten können, dass 68 Prozent der Population einen IQ zwischen 85 und 115 haben und 95 Prozent der Population einen IQ zwischen 70 und 130 aufweisen. Mit anderen Worten, es ist extrem unwahrscheinlich, dass eine zufällig ausgewählte Person einen IQ von mehr als 130 oder weniger als 70 aufweist.

Sicher haben Sie schon von der 80:20-Regel gehört? Ihr liegt die Idee zugrunde, dass 80 Prozent aller Umsätze von 20 Prozent der Vertriebsmitarbeiter erzielt werden oder dass 80 Prozent der Reklamationen von 20 Prozent der Kunden stammen. Dieses verbreitete Phänomen wird durch das Pareto-Prinzip beschrieben. Wenn Sie eine solche Verteilung sehen, dann bedeutet das immer, dass eine relativ kleine Anzahl von Inputs für einen relativ großen Anteil des Outputs verantwortlich ist.

Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwort- und DRM-Schutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: **info@pearson.de**

Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten oder ein Zugangscode zu einer eLearning Plattform bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.** Zugangscodes können Sie darüberhinaus auf unserer Website käuflich erwerben.

Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

<https://www.pearson-studium.de>