

 WILEY-VCH

Andreas Heuer

Der perfekte Tipp



Statistik des Fußballspiels



ERLEBNIS
wissenschaft



Andreas Heuer
Der perfekte Tipp

Weitere Titel aus der Reihe »Erlebnis Wissenschaft«

Groß, M.
Von Geckos, Garn und Goldwasser
Die Nanoweltwelt lässt grüßen
2012
ISBN: 978-3-527-33272-4

Lutzke, D.
Surfen in die digitale Zukunft
2012
ISBN: 978-3-527-32931-1

Kricheldorf, H. R.
Menschen und ihre Materialien
Von der Steinzeit bis heute
2012
ISBN: 978-3-527-33082-9

Ganteför, G.
Klima – Der Weltuntergang findet nicht statt
2012
ISBN: 978-3-527-32863-5

Hüfner, J./Löhken, R.
Physik ohne Ende
Eine geführte Tour von Kopernikus bis Hawking
2012
ISBN: 978-3-527-41017-0

Roloff, E.
Göttliche Geistesblitze
Pfarrer und Priester als Erfinder und Entdecker
2012
ISBN: 978-3-527-32864-2

Zankl, H.
Kampfhähne der Wissenschaft
Kontroversen und Feindschaften
2012
ISBN: 978-3-527-32865-9

Al-Shamery, K. (Hrsg.)
Moleküle aus dem All?
2011
ISBN: 978-3-527-32877-2

Bergmann, H.
Wasser, das Wundererelement?
Wahrheit oder Hokuspokus
2011
ISBN: 978-3-527-32959-5

Schwedt, G.
Die Chemie des Lebens
2011
ISBN: 978-3-527-32973-1

Groß, M.
Der Kuss des Schnabeltiers
und 60 weitere irrwitzige Geschichten aus Natur und Wissenschaft
2011
ISBN: 978-3-527-32738-6

Groß, M.
9 Millionen Fahrräder am Rande des Universums
Obskures aus Forschung und Wissenschaft
2011
ISBN: 978-3-527-32917-5

Köhler, M.
Vom Urknall zum Cyberspace
Fast alles über Mensch, Natur und Universum
2011
ISBN: 978-3-527-32739-3

Schatz, G.
Feuersucher
Die Jagd nach dem Geheimnis der Lebensenergie
2011
ISBN: 978-3-527-33084-3

Schwedt, G.
Lava, Magma, Sternenstaub
Chemie im Inneren von Erde, Mond und Sonne
2011
ISBN: 978-3-527-32853-6

Synwoldt, C.
Alles über Strom
So funktioniert Alltagselektronik
2011
ISBN: 978-3-527-32741-6

Andreas Heuer

Der perfekte Tipp

Statistik des Fußballspiels



WILEY-
VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Autor

Prof. Dr. Andreas Heuer
Universität Münster
Institut für Physikalische Chemie
Corrensstr. 29/30
48149 Münster

Cover Design Simone Benjamin, McLeese
Lake, Canada

Satz Mitterweger & Partner, Plankstadt

Druck und Bindung Ebner & Spiegel GmbH,
Ulm

Gedruckt auf säurefreiem Papier

1. Auflage 2012

Alle Bücher von Wiley-VCH werden sorgfältig erarbeitet. Dennoch übernehmen Autoren, Herausgeber und Verlag in keinem Fall, einschließlich des vorliegenden Werkes, für die Richtigkeit von Angaben, Hinweisen und Ratschlägen sowie für eventuelle Druckfehler irgendeine Haftung

**Bibliografische Information
der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© 2012 Wiley-VCH Verlag & Co. KGaA,
Boschstr. 12, 69469 Weinheim, Germany

Alle Rechte, insbesondere die der Übersetzung in andere Sprachen, vorbehalten. Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Photokopie, Mikroverfilmung oder irgendein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsmaschinen, verwendbare Sprache übertragen oder übersetzt werden. Die Wiedergabe von Warenbezeichnungen, Handelsnamen oder sonstigen Kennzeichen in diesem Buch berechtigt nicht zu der Annahme, dass diese von jedermann frei benutzt werden dürfen. Vielmehr kann es sich auch dann um eingetragene Warenzeichen oder sonstige gesetzlich geschützte Kennzeichen handeln, wenn sie nicht eigens als solche markiert sind.

Print ISBN: 978-3-527-33103-1

ePDF ISBN: 978-3-527-65079-8

ePub ISBN: 978-3-527-65078-1

mobi ISBN: 978-3-527-65077-4

oBook ISBN: 978-3-527-65076-7

Inhalt

- I Einführung 1**
 - 1 Fußball: Hobby und Wissenschaft 3
 - 2 Statistik der Tore 9

- II Statistische Eigenschaften der Bundesliga 15**
 - 3 Basis zur Prognose: Tore, Chancen, Siege 17
 - 4 Der Marktwert: Geld schießt viele Tore 49
 - 5 Zufall, Leistung und Fluktuationen 55
 - 6 Mythen und Erstaunliches 91

- III Prognose in der Praxis: Chancen und Grenzen 107**
 - 7 Bestimmung der Leistungsstärke 109
 - 8 Prognose einzelner Fußballspiele 131
 - 9 Saisonvorhersagen: Wahrscheinlichkeiten für Meisterschaft und Abstieg 167
 - 10 Prognose von internationalen Turnieren 179

- IV Das Spiel und das Drumherum unter der Statistik-Lupe 187**
 - 11 Trainerentlassung: Rettungshalm oder Ausdruck von Verzweiflung? 189
 - 12 Spieldaten: interessante oder unnütze Details? 201
 - 13 Sportnoten (kicker und Impire): Spiegelbild der gezeigten Leistung? 213
 - 14 Spielgeschehen: Dramatik der letzten Minuten 225

- V Zusammenfassung und Ausblick 251**
 - 15 Zum Zufallsbegriff im Fußball 253
 - 16 Vergleich der Bundesliga mit anderen Ligen und Sportarten 259
 - 17 Was bleibt? 275

| | | |
|-----------|-------------------------------------|------------|
| VI | Anhang, Literaturverzeichnis | <i>277</i> |
| | Anhang | <i>279</i> |
| | Literatur | <i>313</i> |
| | Sachregister | <i>317</i> |
| | Namensregister | <i>321</i> |

Geleitwort

Dumm gelaufen für die Bayern. Im November 2007 verpasste die Mannschaft den vorzeitigen Einzug in die nächste UEFA-Pokal-Runde. Mit einem Sieg gegen Bolton Wanderers hätte man alles klar machen können. Doch das Spiel endete 2:2 – wohl auch, weil Trainer Ottmar Hitzfeld wichtige Spieler wie Philipp Lahm oder Franck Ribéry schonte. Bayern-Vorstand Karl-Heinz Rummenigge war sauer auf Hitzfeld – ein ausgebildeter Mathematiklehrer – und erklärte: »Fußball ist keine Mathematik.« Das klang plausibel. Warum sollten ausgerechnet Mathematiker eines der größten Mysterien der Neuzeit, den Fußball, erklären können?

Ich selbst hatte lange Zeit Zweifel, ob man dem Spiel auf dem grünen Rasen wirklich mit Mathematik beikommen kann. Natürlich arbeiten Trainer und Sportreporter schon seit längerem mit Statistiken. Da wird Stürmern ein guter Lauf attestiert oder Mannschaften eine ausgewiesene Fähigkeit zum Drehen von Spielen angedichtet. Dabei weiß jeder: Irgendwann endet auch die schönste Serie, Prognosen sind auf dieser Basis kaum möglich.

Anfang 2008 hatte ich zum ersten Mal Kontakt mit Andreas Heuer. Er behauptete damals, ein Modell gefunden zu haben, mit dem er die Zahl der Tore in einem Spiel vorhersagen kann. Zwei Jahre später hatte Heuer das Modell so weit entwickelt, dass er damit ausrechnen konnte, mit welcher Wahrscheinlichkeit ein Bundesliga-Team Meister wird oder absteigt. Irgendwann hat der Physiker aus Münster die Prognosen seines Modells auch mit den Quoten von Wettanbietern verglichen. Dabei stellte sich heraus, dass seine Vorhersagen im Mittel etwas besser waren als jene, die sich aus den Quoten ergaben.

Das Grundprinzip von Heuers Modell finde ich vollkommen einleuchtend: Es gibt gute Mannschaften und schlechte. Die Guten erarbeiten sich mehr Torchancen als die Schlechten. Ob aus einer Mög-

lichkeit ein Tor wird oder nicht, ist dann jedoch Glückssache. Im Grunde vergleicht er das Tore Schießen mit dem Werfen von Würfeln: Nur wenn eine Sechs fällt, zappelt der Ball im Netz. Wie oft ein Team würfeln darf, hängt von der eigenen Spielstärke und der des Gegners ab. Damit ist auch klar: Wenn Bayern 20 Mal würfelt, aber keine Sechs darunter ist, dann kann auch ein Drittligist ein Spiel gegen die Münchner gewinnen. Er braucht bei seinen drei, vier Würfelversuchen, die er höchstens hat, eben etwas Glück.

Bei SPIEGEL ONLINE habe ich mehrfach über das Bundesligamodell berichtet, ein mehrseitiger Artikel erschien in »Spektrum der Wissenschaft«. Inzwischen liefert Andreas Heuer mehrmals im Laufe einer Saison auf SPIEGEL ONLINE seine Prognosen für die Tabelle am 34. Spieltag. Sie werden im Laufe einer Saison immer präziser.

Ich finde es toll, wenn Forscher sich so in ein Thema hineinknien wie Andreas Heuer. Er verbindet die Begeisterung für den Fußball mit seiner wissenschaftlichen Arbeit. Denn sie liefert die raffinierten statistischen Methoden für das Bundesligamodell. So ist aus einer Leidenschaft, die Hunderttausende weltweit jede Woche ins Stadion zieht, dieses spannende Buch entstanden.

Zum Schluss noch eine Bemerkung zu Karl Heinz Rummenigge: Natürlich kommen Trainer und Profis wunderbar ohne Algebra, Differentialgeometrie und Wahrscheinlichkeitsrechnung aus. Wenn man das Spiel aber so gründlich analysiert wie in diesem Buch, dann kann es keinen Zweifel mehr geben: Fußball ist Mathematik. Ottmar Hitzfeld formulierte das etwas diplomatischer, als er Rummenigge entgegnete: »Ich hoffe, dass ich das Fußball-Einmaleins kann.«

Holger Dambeck
(Wissenschaftsredakteur bei SPIEGEL ONLINE)

Der perfekte Tipp – die Statistik des Fußballspiels

Danksagung

Es gibt viele Menschen, die mich auf dem Weg vom Fußball-Hobby über die immer ernsthaftere Beschäftigung mit den statistischen Eigenschaften bis hin zum Schreiben dieses Buchs begleitet haben. Zunächst möchte ich Oliver Rubner meinen besonderen Dank aussprechen für die vielen Diskussionen und seine vielfältigen Beiträge, die zum Gelingen und zur inhaltlichen Breite dieses Buchs wesentlich beigetragen haben. Ebenso sind vielfältige sportstatistische Auswertungen von Christian Müller (Trainerwechsel), Dennis Riedl (Spieldaten, internationale Ligen) und Jens Smiatek (Notenqualität, Handball-Vergleich) in dieses Buch eingeflossen, wofür ich ebenfalls sehr dankbar bin. Die Zusammenarbeit mit meinem Kollegen Bernd Strauss aus den Sportwissenschaften hat mir sehr geholfen, Theorie und Praxis noch weiter zusammenzuführen. Viele Verbesserungsvorschläge bei der Korrektur des Manuskripts kamen von Andreas Christian, Stefan Hopp, Christian Rehwald und Dennis Riedl. Der WWU Münster möchte ich für die Unterstützung des Projekts danken und hier insbesondere die umsichtige und vielfältige Hilfe von Andrea Staubermann erwähnen. Dem Verlag Wiley-VCH möchte ich herzlich für die gute Zusammenarbeit danken und Holger Dambeck für die Erstellung des Geleitworts und für sein wissenschaftsjournalistisches Interesse an unseren Projekten.

Mein Vater hat mich durch den gemeinsamen Besuch meines ersten Spiels (1:5 Heimmiederlage meiner anschließenden Lieblingsmannschaft) und, solange es ihm möglich war, durch das Sammeln vieler interessanter Informationen beim Fußball-Thema immer wieder begleitet. Ganz besonders möchte ich zum Schluss meiner Frau Susan für ihre liebevolle Geduld und die konstruktiven Vorschläge in den verschiedenen Phasen der Manuskripterstellung herzlich danken.



Einführung

1

Fußball: Hobby und Wissenschaft

1.1 Warum dieses Buch?

Fußball als vielleicht liebste Nebenbeschäftigung ist scheinbar allgegenwärtig: auf den Titelseiten der Zeitungen, als wichtiges Thema bei Fernsehsendern und in der Werbung, beim mittäglichen Gespräch mit Kollegen und für viele natürlich am Wochenende im Stadion oder Wohnzimmer. Fußball führt auch zu einem speziellen Montags-Phänomen: Menschen, die sich über die aktuelle Tabelle der Bundesliga beugen und sich mittels Anwendung intuitiver statistischer Verfahren fragen, ob ihr Lieblingsverein in dieser Saison noch Meister werden kann. Doch was würde eine objektive Sichtweise liefern? Welchen Informationsgehalt besitzen Tore und Punkte für die Wahrscheinlichkeit der Meisterschaft? Eine Mannschaft, die wie Mainz in der Saison 2011/12 nach zwei Spieltagen an der Tabellenspitze steht, ist sicherlich noch kein Meisterschaftsfavorit.

Zur sprachlichen Vermittlung hat sich ein fußballspezifisches Vokabular gebildet. Eigenschaften wie »Lauf«, »Heimstärke« oder »Angstgegner« werden Mannschaften in speziellen Situationen zugewiesen. Das spiegelt eine allzu menschliche Tendenz wider, Beobachtungen in ein mehr oder weniger sinnvolles Raster einzuordnen. Doch gibt es wirklich einen Lauf, der eine Mannschaft an die Tabellenspitze spült? Oder ist eine Siegesserie vielmehr wie ein mehrfaches Würfeln derselben Zahl aufzufassen, also reiner Zufall? Kurzum: Was sind Mythen und was sind Tatsachen?

Wenn es keine Zufälligkeiten im Fußballspiel gäbe, könnten diese Fragen leicht beantwortet werden. Wenn eine Mannschaft zum Beispiel zu Hause weniger Punkte als auswärts erreicht hat, dürfte man diese Mannschaft durchaus als heimschwach bezeichnen. Nun kann aber ein Ball von der Unterkante der Latte entweder vor oder hinter der Linie landen – der Paradefall eines wembleyartigen Zufallspro-

zesses. Das letztliche Ergebnis kann deswegen manchmal so ungerecht sein. Das gilt für ein Fußballspiel viel mehr als zum Beispiel für ein Handballspiel. Ein Aluminiumtreffer mehr und dafür ein Tor weniger haben im Handball bei den wenigsten Spielen einen relevanten Effekt. In der Fußball-Bundesliga hingegen werden wir sehen, dass ein typischer Spielausgang zu 86 % durch Zufall bestimmt wird, also nur schwer vorhersagbar ist. Aber letztlich ist es gerade die daraus resultierende Spannung, deretwegen sich so viele Menschen für Fußball begeistern lassen. Oder mit den Worten von Gerd Delling: »40.000 im Häuschen sind aus demselben«.

In diesem Buch geht es insbesondere um die 14 % des Spielergebnisses, die nicht Zufall sind und eine Top-Mannschaft von einem Abstiegskandidaten unterscheiden lassen. Durch geeignete statistische Betrachtungen können wir genau diesen Anteil identifizieren. Dieses Buch möchte nun versuchen, mit möglichst wenigen Formeln und durch Betonung der unterliegenden Konzepte für viele Leser eine neue Sicht auf das Phänomen Fußball zu vermitteln. Viele der Schlussfolgerungen erscheinen dann hoffentlich ähnlich eindeutig wie die tiefeschürfende Erkenntnis vom Bundesligatrainer Reinhold Fanz, dass es ganz schwer zu gewinnen ist, wenn man keine Tore macht. Machen Sie sich darauf gefasst, althergebrachte Vorstellungen zu revidieren – zumindest erging es mir gelegentlich so.

Bei Zufallsexperimenten wie dem Münzwurf sind keine *sicheren* Vorhersagen (»als nächstes kommt die Zahl«), aber immerhin *perfekte* Vorhersagen (»als nächstes kommt die Zahl mit 50 % Wahrscheinlichkeit«) möglich. Ist der perfekte Tipp, so wie im Buchtitel angekündigt, auch beim Fußball realisierbar? Tatsächlich werden Sie sehen, wie weit man sich der perfekten Fußball-Vorhersage annähern kann. Mindestens genauso spannend ist aber der Weg zum Ziel. Dort kann der Leser viel über die Eigenschaften eines Fußballspiels und dessen Mythen kennenlernen. So wie auch Christoph Biermann in der »Fußball-Matrix« (2009) auf der »Suche nach dem perfekten Spiel« viele interessante Bausteine des Fußballgeschehens beleuchtet hat.

Wieso schreibt ein Naturwissenschaftler, der sich ansonsten mit der Theorie komplexer Systeme im physikalischen und chemischen Kontext beschäftigt, ein Buch über Fußballstatistik? Es begann vor einigen Jahren mit einer Frage, die mein Kollege Metin Tolan (»So werden wir Weltmeister« – hoffentlich zumindest 2014) gestellt hatte.

Woher wissen wir überhaupt, dass der Meister besser ist als der Absteiger? Wird nicht vielleicht alles durch Zufall bestimmt? Diese Frage ist ausbaufähig: Kann man grundsätzlich zufällige und leistungsabhängige Anteile trennen? Um wie viel besser sind die guten im Vergleich zu den schlechten Mannschaften? Es eröffnete sich eine riesige »Spielwiese«, auf der ich zusammen mit Kollegen aus Münster meine beruflichen Erfahrungen bei der Aufarbeitung und Interpretation von Zahlenkolonnen einbringen konnte. Es entstand ein Hobbyprojekt, in dem wir unser Dasein als ganz normale Fußballfans mit unserer Tätigkeit als Wissenschaftler verbinden konnten. Inzwischen haben wir einige wissenschaftliche Publikationen zu diesem Themenbereich veröffentlicht. Zudem habe ich die Möglichkeit, die Ergebnisse der Saisonvorhersage regelmäßig bei SPIEGEL ONLINE zu veröffentlichen. Schließlich entstand die Idee, die vielen einzelnen Resultate in Form eines Buches zusammenzuführen. Die meisten Themenbereiche werden durch konkrete Analysen untermauert. Im Anhang finden sich dann für interessierte Leser weitergehende Details. Fußball lebt von den Emotionen. Nichtsdestotrotz kann eine Besinnung auf Fakten sowohl für Fußball-Interessierte als auch für die dort handelnden Personen durchaus nützlich sein.

1.2 Zielgruppen

Zunächst einmal richtet sich dieses Buch an die vielen Leser mit Interesse am Sportgeschehen und insbesondere natürlich am Fußball. Mit vielen Beispielen und ausführlichen Erklärungen können die verschiedenen Aspekte, die mit dem Fußballspiel verbunden sind, ganz neu erfahren werden. Ist die Meisterschaft aufgrund der unterschiedlichen Marktwerte der Mannschaften schon zu Saisonbeginn entschieden? Wie real ist der Bayern-Dusel? Wie wirken sich Tore auf den Spielverlauf aus? Sollten ab und zu die Trainer entlassen werden, oder ist das ehemalige Freiburger Finke-Modell zu bevorzugen? Wie gut kann man grundsätzlich Fußballspiele vorhersagen?

Die Struktur dieses Buches soll aber auch dazu dienen, noch spezifischere Zielgruppen anzusprechen:

- Leser mit Interesse an Mathematik und Statistik: Durch zahlreiche Graphiken und vom Haupttext getrennte Anhänge mit den genauen mathematischen bzw. statistischen Analysen können die

einzelnen Überlegungen Schritt für Schritt nachvollzogen werden. Tatsächlich entsprechen viele der Ansätze denen, die bei psychologischen Testverfahren angewandt werden.

- Leser, die z. B. in Tippgemeinschaften aktiv sind: Ein Schwerpunkt dieses Buches besteht darin, die Leistungsstärke von Mannschaften zu charakterisieren und entsprechend abzuschätzen. Mit solchen Informationen können dann die nächsten Spiele oder gar die gesamte Restsaison prognostiziert werden. Welche Prognosequalität ist erreichbar, gerade auch im Vergleich zu Buchmachern? Wie sieht das konkrete Prognoseverfahren aus?
- Leser mit Interesse an Managementfragen: Einige der hier untersuchten zentralen Themenbereiche (z. B. Nutzen von Trainerentlassungen, Auswirkung zusätzlicher Investitionen oder Beurteilung des Einflusses konkreter Handlungen) können auch auf andere Bereiche wirtschaftlichen und personellen Handelns übertragen werden.
- Lehrer und Schüler: Dieses Buch kann für viele Schüler einen hoch motivierenden Einstieg in statistische Themen bieten. Da fast alle verwendeten Zahlen zum Themenbereich Fußball direkt im Internet zur Verfügung stehen, können viele der Untersuchungen selbst reproduziert werden. Auch ich verwende für die Mathematik-Ausbildung an der Universität einige der Ergebnisse zur Motivation von Statistik-Themen.

1.3 Aufbau

Nach der Einführung in einige grundlegende statistische Ideen in Teil I werden im Teil II verschiedene Größen untersucht, die zur Leistungsdiagnostik einer Mannschaft verwendet werden können. Welche frei verfügbaren Daten (Tore, Punkte, Torchancen, Marktwerte, ...) besitzen den größten Informationsgehalt über die Qualität einer Mannschaft? Wie bestimmt man überhaupt den Informationsgehalt? Und gibt es tatsächlich Mannschaften, die viel mehr Torchancen als andere versieben? Im Kapitel »Mythen und Erstaunliches« werden Sie sehen, dass alle Mannschaften praktisch die identische Chancenverwertung aufweisen und genau deswegen die Kenntnis

der Torchancen der vergangenen Spiele so wertvoll für gute Vorhersagen sind. Ebenso werden Sie dort erfahren, dass der Begriff des »Laufs« für eine koordinierte Beinbewegung, aber nicht für die Beschreibung einer überlangen Siegesserie dienen sollte und dass es Neulinge erstaunlich schwer haben, sich von ihrer Außenseiterrolle zu lösen.

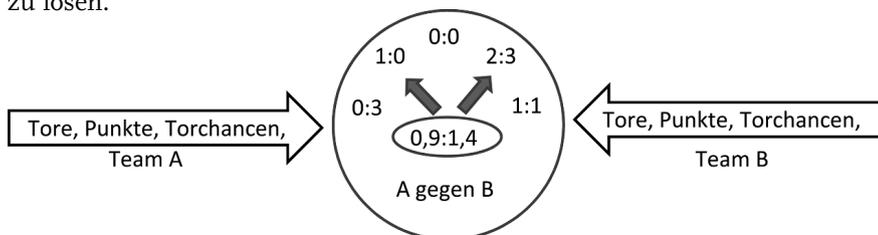


Abb. 1.1 Der Weg zur Spielvorhersage im Teil III (s. Haupttext).

In Teil III wird dann aus den vorliegenden Informationen über zwei Mannschaften mit Hilfe der explizit hergeleiteten Fußball-Formel eine möglichst gute Prognose von einzelnen Spielen durchgeführt. So werden Sie sich dem perfekten Tipp am Beispiel des Spiels Köln gegen Dortmund aus der Saison 2009/10 in sieben Schritten nähern. Das Ergebnis 0,9 zu 1,4 ist kein Rechenfehler sondern stellt das eigentlich faire Resultat dar, basierend auf den 14% zufallsfreien Beiträgen zum Spielergebnis. Im 8. Schritt kommen noch die 86% Zufall hinzu, womit dann Aussagen über das tatsächliche Endergebnis formuliert werden können (»Wie wahrscheinlich ist ein 2:3?«, was übrigens das tatsächliche Ergebnis war). Auf diese Art kann man auch Endtabellen abschätzen und ziemlich genau die Wahrscheinlichkeit bestimmen, dass der Lieblingsverein in dieser Saison Meister wird. In Teil IV werden dann viele Fragen geklärt, die Sie sich in der einen oder anderen Weise sicherlich schon gestellt haben. Was bringt der Trainerwechsel? Welche Aussagekraft haben eigentlich die Sportnoten? Sie werden sehen, dass durch kreative Datenanalysen teilweise überraschende Antworten gefunden werden können. In Teil V schließlich wird die Fußball-Bundesliga mit anderen Fußball-Ligen, aber auch mit der viel torreicheren Handball-Bundesliga verglichen und der immer wiederkehrende Begriff des "Zufalls" näher beleuchtet.

Viel Spaß bei dieser Lektüre!

2

Statistik der Tore

2.1 Wie viele Tore schießt eine Mannschaft?

Tore als das zentrale Element eines Fußballspiels sind natürlich das ideale Beispiel um einige Begriffe kennenzulernen, die in der Fußballstatistik vorkommen werden. Wie auch im weiteren Verlauf des Buches stehen die Details im Anhang, während im Haupttext die wesentlichen Ideen beschrieben werden.

Wir beginnen zunächst mit der Frage, wie viele Tore eine Mannschaft in einem Spiel geschossen hat. In Abb. 2.1 ist die Häufigkeit dargestellt, mit der genau kein Tor oder genau ein Tor oder genau zwei Tore usw. vorkamen. Das Kreuz bei 0 Toren drückt zum Beispiel aus, dass in 20,4% aller Spiele die Heimmannschaft kein Tor geschossen hat.

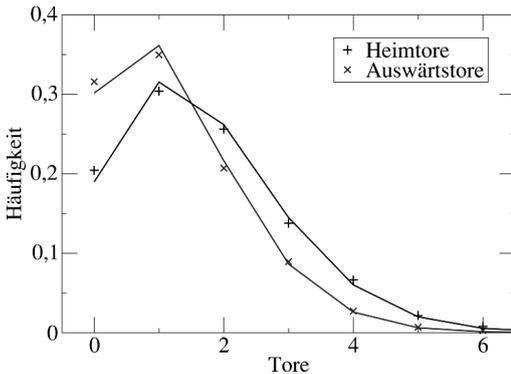


Abb. 2.1 Verteilung der Tore, die eine Mannschaft in einem Spiel geschossen hat. Unterschieden wird zwischen Heim- und Auswärtsspielen. Die Bedeutung der durchgezogenen Linien wird im Text erläutert. Berücksichtigt werden alle Spiele zwischen den Saisons 1995/96 und 2010/2011.

Zunächst einmal folgt nicht sehr überraschend, dass im Mittel Heimmannschaften mehr Tore schießen. Größere Torzahlen kommen dort häufiger vor als bei Auswärtsmannschaften. Hier spiegelt sich der allgegenwärtige Heimvorteil wider. Konkret folgt nach Berechnung der

Mittelwerte, wie im Anhang A2.1 beschrieben, dass im Durchschnitt 1,66 Heim- und 1,20 Auswärtstore pro Spiel gefallen sind.

Die Anzahl tatsächlich geschossener Tore ist allerdings breit verteilt um diese Mittelwerte. Wir führen nun eine ganz zentrale Größe ein, mit der die Abweichungen der tatsächlich geschossenen Tore von der mittleren Toranzahl erfasst werden kann. Dazu verwenden Statistiker den Begriff der *Varianz*. Sie erfasst den mittleren quadratischen Abstand vom Mittelwert. Hier ein einfaches Beispiel: Nehmen wir an, dass in der Hälfte aller Spiele die Heimmannschaft ein Tor und in der anderen Hälfte genau fünf Tore schießt. Im Mittel schießt die Heimmannschaft also drei Tore. Die Abweichung vom Mittelwert beträgt im ersten Fall (-2) und im zweiten Fall $+2$ Tore. Die quadratische Abweichung ist somit in beiden Fällen 4, resultierend aus $(-2) \cdot (-2)$ bzw. $(+2) \cdot (+2)$. Daher ist insbesondere auch die mittlere quadratische Abweichung, also die Varianz, gleich vier. Wie man die Varianz für vorgegebene Messwerte abschätzen kann, wird in A2.1 gezeigt. Für die beiden obigen Beispiele aus Abb. 2.1 findet man Varianzen von 1,77 und 1,30. Schließlich führt man noch die *Standardabweichung* als die Wurzel der Varianz ein. Anschaulich drückt die Standardabweichung aus, wie weit typischerweise ein konkreter Messwert vom Mittelwert entfernt ist. Im obigen Beispiel ist die Standardabweichung gerade 2 – tatsächlich sind die konkreten Torzahlen ja jeweils 2 vom Mittelwert entfernt. Für die Verteilung der Heimtore ergibt sich eine Standardabweichung von 1,33 und für die Auswärtstore 1,14.

Gibt es wirklich den Heimvorteil? Vielleicht ist der beobachtete Unterschied zwischen Heim- und Auswärtsmannschaften ja nur eine Folge der endlichen Anzahl von Spielen. Schließlich würde man nach dem dreimaligen Würfeln einer 6 auch nicht behaupten, man würde auch weiterhin nur diese eine Ziffer werfen. Hier haben wir das Problem der endlichen Stichprobe. Je kleiner sie ist, umso weniger relevant sind die möglichen Schlussfolgerungen. Angenommen, wir hätten zufällig den 6. Spieltag der Saison 2010/11 ausgewählt. Dort fielen 13 Heimtore und 15 Auswärtstore. Intuitiv hätte man wohl gesagt, dass nur neun Spiele nicht besonders aussagekräftig sind, während die für die Daten in Abb. 2.1 verwendeten 4896 Spiele deutlich informativer sein sollten. Diese Intuition kann zum Glück auch quantitativ erfasst werden. Zunächst einmal muss man sich Rechenschaft ablegen, wie genau beide Mittelwerte, also 1,66 und 1,20, überhaupt bestimmt werden können. In A2.2 wird gezeigt, dass man mit

Hilfe der Standardabweichung und der Anzahl der verwendeten Spiele den *statistischen Fehler* bei der Schätzung eines Mittelwertes bestimmen kann. Daraus ergibt sich für den Mittelwert der Heimtore $1,66 \pm 0,02$. Was bedeutet diese Aussage? Der wahre Wert kann durchaus von 1,66 verschieden sein, es ist aber praktisch ausgeschlossen, dass eine Auswahl von anderen Spielen aus diesem Zeitbereich z. B. einen Wert von 1,76 ergeben hätte. Vielmehr (s. A2.2) kann man folgern, dass mit 68% Wahrscheinlichkeit die wahre Anzahl zwischen 1,64 und 1,68 liegt und mit 95% Wahrscheinlichkeit zwischen 1,62 und 1,70. Für die Auswärtstore folgt $1,20 \pm 0,02$. Daraus resultiert ein Unterschied zwischen Heim- und Auswärtstoren von $0,46 \pm 0,03$. Die Berechnung dieses statistischen Fehlers von 0,03 wird in A2.3 diskutiert. Damit ist gezeigt, dass der Heimvorteil auch unter Berücksichtigung des statistischen Fehlers definitiv vorliegt. Man spricht hier auch von einem statistisch signifikanten Ergebnis. Tatsächlich hätte sich für den oben genannten 6. Spieltag für den Unterschied von Heim- und Auswärtstoren das Ergebnis $-0,22 \pm 0,61$ ergeben. Der statistische Fehler ist somit so groß, dass aus diesen neun Spielen keine relevante Aussage über die Existenz oder Abwesenheit eines Heimvorteils getroffen werden kann.

Wir hatten ja schon erwähnt, dass in 20,4% aller Fälle (in absoluten Zahlen in 1001 Fällen) die Heimmannschaft kein Tor schießt. Angenommen, eine Theorie würde vorhersagen, dass diese Zahl 18,9% lauten sollte. Wäre die Theorie dann falsch, da der theoretische Wert um 1,5% kleiner ist? Oder ist die Verlässlichkeit des Wertes von 20,4% wegen der endlichen Größe der Stichprobe so klein, dass der Unterschied von 1,5% im Rahmen des statistisch Erwarteten liegt? Im Anhang A2.5 wird beschrieben, wie man den statistischen Fehler eines Messpunktes bestimmen kann. Dazu werden die Eigenschaften der sogenannten *Binomial-Verteilung* (s. A2.4) verwendet. Für dieses Beispiel ergibt sich $0,204 \pm 0,006$. Somit wird mit 95% Wahrscheinlichkeit der wahre Wert zwischen 0,192 und 0,216 liegen. Ein Wert von 0,189 läge somit außerhalb dieses Bereichs und ist somit mit großer Wahrscheinlichkeit nicht vereinbar mit dem Messwert von 0,204. Zur graphischen Darstellung des statistischen Fehlers werden wir immer wieder sogenannte *Fehlerbalken* verwenden, hier wäre es ein senkrechter Strich zwischen den Werten $0,204 - 0,006 = 0,198$ und $0,204 + 0,006 = 0,210$.

Von verschiedenen Autoren im Bereich der Sportstatistik wurde schon früh erkannt, dass die Verteilung der Tore einer sogenannten

Poisson-Verteilung ähnelt [Mor56, Ree68, Hil74]. Man kennt Poisson-Verteilungen aus ganz anderen Bereichen. Wenn man sich zum Beispiel fragt, wie viele Blitze pro Jahr in einem speziellen Wald einschlagen, wird diese Anzahl gemäß der Poisson-Verteilung beschrieben. Die mathematische Darstellung der Verteilung wird in A2.4 erläutert. Die Poisson-Verteilung besitzt eine ganz wichtige Eigenschaft: Wenn man weiß, dass zum Beispiel im Mittel zwei Blitze pro Jahr einschlagen, dann ergibt sich daraus schon die gesamte Verteilung. Zum Beispiel kann man vorhersagen, dass in einem konkreten Jahr mit 18 % Wahrscheinlichkeit genau drei Blitze einschlagen werden. Es zeigt sich übrigens, dass für eine Poisson-Verteilung die Varianz identisch zum Mittelwert ist. Wir können nun die empirisch bestimmten Häufigkeiten in Abb. 2.1 mit einer Poisson-Verteilung vergleichen. Da sich Heim- und Auswärtstore unterschiedlich verhalten, führen wir den Vergleich jeweils getrennt durch. Dabei benutzen wir wieder die Eigenschaft, dass die Kenntnis des Mittelwertes die gesamte Verteilung bestimmt. Beide Verteilungen sind in Abb. 2.1 eingezeichnet. Man sieht tatsächlich eine gute Übereinstimmung. Bei genauerer Überprüfung hingegen stellt man fest, dass die empirisch gefundene Torverteilung etwas breiter als die Poisson-Verteilung ist. Tatsächlich stellt man sowohl bei Heim- als auch bei Auswärtstoren fest, dass die Varianz der tatsächlichen Daten um 0,10 höher ist. Dieser Unterschied wird noch wichtig werden.

2.2 Wie wichtig ist der Zufall in einem Spiel?

Wie kann man nun die Verteilung z. B. der Heimtore in Abb. 2.1 verstehen? Grundsätzlich sind zwei sehr unterschiedliche Szenarien möglich, so wie in Abb. 2.2 skizziert.

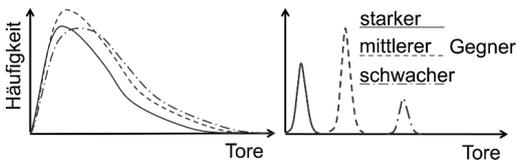


Abb. 2.2 Verschiedene Szenarien zur Erklärung der breiten Verteilung der Anzahl der Tore einer Mannschaft in einem Spiel. Links ist die tatsächliche Torzahl hauptsächlich durch Zufallsprozesse,

rechts durch den Unterschied der Leistungsstärken bestimmt. Die einzelnen Linien entsprechen Spielen einer Mannschaft gegen verschieden starke Gegner.

1. Zufallsdominiertes Spiel: In einem Fußballspiel erzielt eine Mannschaft pro Minute mit einer kleinen Wahrscheinlichkeit ein Tor (z. B. 2% Wahrscheinlichkeit pro Minute). Bei diesem Beispiel würden dann in 90 Minuten 1,8 Tore fallen. Wie oft allerdings im konkreten Spiel ein Tor erzielt wird, ist nicht vorhersehbar und somit durch den Zufall bestimmt. Es können kein Tor oder sechs Tore fallen. Dies ist analog zum sechsmaligen Würfeln, bei dem gar nicht oder im Extremfall jedes Mal eine bestimmte Zahl gewürfelt werden kann. In diesem Szenario sollten die resultierenden Tore sehr gut durch eine Poisson-Verteilung beschrieben werden. Ähnliche Konzepte verwendet man bei der Beschreibung des radioaktiven Zerfalls, bei dem es ja ebenfalls eine feste Wahrscheinlichkeit pro Zeitintervall gibt, dass ein Kern zerfällt.
2. Leistungsdominiertes Spiel: Aufgrund der individuellen Leistungsstärken der Mannschaften in einem Spiel ist z. B. die Anzahl der zu erwartenden Anzahl von Heimtoren relativ gut vorbestimmt. Der Zufall, also die Unsicherheit bei der Vorhersage, spielt nur eine untergeordnete Rolle. Dieses wird durch die recht schmalen Linienbreiten in Abb. 2.2 (rechts) ausgedrückt. Die breite Verteilung der Toranzahl in Abb. 2.1 resultiert daher, dass die Leistungsstärken der Teams in den verschiedenen Spielen sehr unterschiedlich sein können. Anschaulich könnte man sich das ähnlich wie bei einer Mathematik-Klausur vorstellen. Auch dort ist der mögliche Notenhorizont eines Schülers häufig relativ gut vorbestimmt und die breite Notenverteilung spiegelt im Wesentlichen die unterschiedlichen Mathematik-Fähigkeiten der Schüler wider.

Basierend auf Abb. 2.1 ist zwischen beiden Szenarien nicht zu unterscheiden. Wir werden uns aber später eindeutig für Szenario (1) entscheiden.

2.3 Die Glockenkurve im Fußballspiel

Seit einigen Jahren werden mehr und mehr Parameter während eines Spiels erfasst, um ein Fußballspiel in seiner ganzen Breite statistisch beschreiben zu können. Davon wird in Kap. 12 noch ausführ-

lich die Rede sein. So kann man zum Beispiel die Pässe in einem Spiel zählen. Im Mittel der Saisons 2009/10 und 2010/11 ergeben sich ca. 730 Pässe pro Spiel. Allerdings kann der konkrete Wert von Spiel zu Spiel durchaus variieren wie in Abb. 2.3 zu sehen ist. Diese Verteilung ähnelt einer Glockenkurve, die auch *Gauß-Verteilung* genannt wird. Deren mathematische Darstellung finden Sie im Anhang A2.4. In Abb. 2.3 ist diejenige Gauß-Verteilung gezeigt, die möglichst gut die empirische Verteilung wiedergibt. Gauß-Verteilungen finden sich häufig bei solchen beobachteten Größen, die sich als Summe verschiedener Effekte beschreiben lassen. Ein historisches Beispiel aus den Anfängen der angewandten Statistik betrifft die Verteilung des Brustumfangs, der für mehrere tausend Soldaten bestimmt wurde und sehr gut durch eine Gauß-Verteilung beschrieben werden kann. Wegen der Wichtigkeit hatte es diese Verteilung ja immerhin bis auf den 10 DM-Schein geschafft.

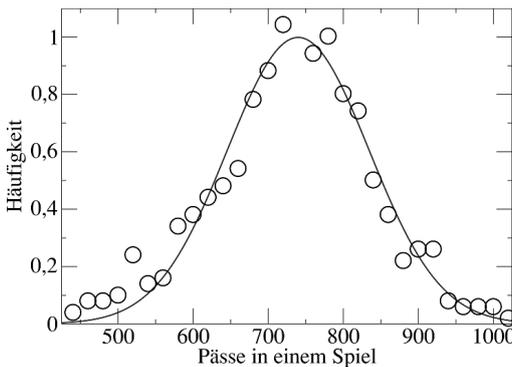


Abb. 2.3 Die Anzahl der Pässe pro Spiel während der Saisons 2009/10 und 2010/11. Zudem ist eine an die Daten bestmöglich angepasste Gauß-Verteilung gezeigt.



Statistische Eigenschaften der Bundesliga

3

Basis zur Prognose: Tore, Chancen, Siege

In diesem Kapitel wenden wir uns den Grundlagen der Prognose zu. Dabei sehen wir, welche Informationen für eine erfolgreiche Prognose relevant sind und welche nicht.

3.1 Extremereignisse

Einige mehr oder weniger skurrile Fakten

Längste Sieges- und Niederlagenserien, größte Zuschauerzahlen, häufigste Auf- und Abstiege. Aus dem Fußballgeschehen des letzten Jahrhunderts lassen sich viele schöne und tragische Geschichten erzählen, die mit Extremereignissen zu tun haben. Eine interessante Zusammenstellung einiger dieser Fakten findet sich in dem informativen Buch »Fast alles über Fußball« von Christoph Biermann (2006). Hier ein paar Beispiele:

- »In der Pause hatte ich schon gefragt: Wo ist das nächste Arbeitsamt?«. Das sagte Kevin Keegan als Coach von Manchester-City nach dem 4:3-Sieg seiner Mannschaft bei Tottenham Hotspur nach einem 0:3-Halbzeit-Rückstand. Noch spannender machte es Bayern München in der Saison 1976/77 beim VfL Bochum. Dem 0:3-Rückstand zur Halbzeitpause folgt kurz darauf ein 0:4. Dann langten jedoch Müller (2×), Hoeneß, Rummenigge und Schwarzenbeck zu. Die 5:4-Führung konnte Bochum noch ausgleichen, bevor kurz vor Schluss noch einmal Hoeneß einen denkwürdigen 6:5-Sieg sicherstellte. Eine Aufholjagd für die Fußball-Ewigkeit, denn nie mehr wurde in der Bundesliga eine höhere Führung verspielt.
- Auch in zwei anderen Spielen hat eine Mannschaft fünf Tore geschossen, aber trotzdem verloren (jeweils 5:6).

- Es gibt keine Mannschaft, die bislang gegen Bayern München eine positive Spielbilanz aufzuweisen hat. Spitzenreiter bei dieser Spezialdisziplin sind übrigens die Stuttgarter Kickers mit zwei Siegen und zwei Niederlagen, also einer immerhin ausgeglichenen Bilanz.
- Der »unverdienteste« Verlierer der letzte 15 Jahre war Leverkusen bei der 1:2-Heimniederlage gegen Rostock in der Saison 2000/01. 16:2 Torchancen standen am Ende für Leverkusen in der Statistik, aber eben nur ein Tor. Bei keinem Spiel war das Missverhältnis zwischen Torchancen und Lohn größer. Zitat aus dem kicker-Spielbericht: »Ohne den verletzten Kirsten überbot sich der Bayer-Angriff im Auslassen selbst bester Chancen, wobei sich vor allen Dingen Neuville negativ hervor tat. Hansa beschränkte sich auf das Zerstören und hatte letztendlich Glück und in Pieckenhagen einen souveränen Keeper.« Ähnlich ist sicherlich das Ausscheiden von Barcelona im Champions-League Halbfinale 2011/2012 gegen Chelsea bei insgesamt 18:4 Torchancen zu sehen.

Bei Extremereignissen denken viele Fußball-Fans sicherlich insbesondere an Fußballspiele, in denen eine Mannschaft demontiert wird. Es gibt hier einige Spiele, die bei diesem Thema eng mit Borussia Dortmund verbunden sind.

Ein schauriger Höhepunkt spielte sich am letzten Spieltag der Saison 1977/78 ab. Zu diesem Zeitpunkt lagen Köln und Mönchengladbach punktgleich an der Tabellenspitze. Köln wies aber ein um 10 Tore besseres Torverhältnis auf. Im Falle eines Kölner Siegs bei St. Pauli hätte die Meisterschaft also in trockenen Tüchern sein sollen. Köln gewann auch sein Spiel. Doch es sollte trotzdem noch spannend werden. Schon nach gut 20 Minuten stand es nämlich 4:0 bei der Partie zwischen Mönchengladbach und Dortmund. Und es ging in ähnlichem Tempo weiter. Am Ende stand ein 12:0-Sieg – eingefangen in jenem unvergesslichen Bild der Anzeigetafel um 17:15 Uhr, deren begrenzte Ausmaße ein 13. Tor nicht mehr verkraftet hätte.

Zum Glück gewann Köln mit fünf Toren Unterschied, so dass diese 12 Tore keinen Einfluss mehr auf die Meisterschaft hatten und Köln die bis heute letzte Meisterschaft gewann. Es wurden sogar kurzzeitig Ermittlungen seitens des DFB durchgeführt, die nach Anhörung der beteiligten Dortmunder Spieler aber schnell wieder eingestellt wurden. Dortmunds Trainer Otto Rehhagel wurde nach