Wer zahlt gewinnt

Michael Joswig

„Fußball ist keine Mathematik“. Diesen Satz musste sich der damalige Bayern-Trainer und vormalige Mathematiklehrer Ottmar Hitzfeld von seinem Vorstandschef Karl-Heinz Rummenigge nach einem unglücklichen Unentschieden im Herbst 2007 anhören. Hat Rummenigge recht? Zu jeder Bundesligasaison fragen sich Experten und Fans aufs Neue, was denn am Ende den Meister eigentlich ausmacht. Die eine Mannschaft hat einen starken Sturm, die nächste verteidigt geschickt, und die dritte hat einen international erfahrenen Trainer. Spätestens seit der Weltmeisterschafts-Euphorie schießen überall die Tippgemeinschaften und Expertenrunden aus dem Boden. Kann man sich mit etwas Mathematik einen Vorteil bei der Bundesligawette verschaffen?

Fußball ist ein Fest für die Statistik. Alles wird gemessen und gezählt. Wer wann wieviele Tore geschossen hat sowieso, aber eben auch, wann welche Ecke geschossen wurde, wie viele Torschüsse links vorbei gingen und wie viele rechts, welcher Spieler welche Lufstrecke zurücklegt und wie lang er kein Tor geschossen hat. Im Zeitalter des Internets haben sich zahlreiche spezialisierte Webseiten dem Fußball verschrieben. Zum Beispiel www.rsssf.com ist eine echte Schatzgrube. Wer wissen möchte, welcher Club in der englischen Southern League (1st Division) die Saison 1894/95 gewonnen hat (Antwort: Luton Town), ist hier genau richtig. Für die Frage nach der Zuschauerentwicklung speziell im deutschen Fußball ist man auf www.dsfs.de bestens aufgehoben. Daten ganz besonderer Art aber werden unter transfermarkt.de vorgehalten. Denn hier geht es um das Geld. Jeder Spielertransfer wird verzeichnet zusammen mit vielen Details und Gerüchten. Ob Ivan Rakitic vom FC Schalke 04 zu Iuventus Turin wechselt oder nicht, auf transfermarkt.de kann man es vermutlich zuerst erfahren. Vor allem aber bekommt jeder Spieler durch die Transfersummen einen numerischen Wert zugeordnet, der irgendwie seine Leistung auf dem Rasen reflektieren soll. Kann man aus diesen Informationen Hinweise auf das Abschneiden einer ganzen Mannschaft ziehen?

Die Bundesligasaison 2007/08

Am Ende der Saison 2007/08 stand wieder einmal der FC Bayern München als Meister fest, dicht gefolgt von Mannschaften, die auch in den Vorjahren stets auf den vorderen Rängen zu finden waren. Die Frage ist, inwieweit man das auch im Voraus hätte ahnen können. Hierzu lohnt es sich die Daten von transfermarkt.de vor Beginn der Saison heranzuziehen. In der Tabelle 1 sind die 18 Mannschaften der 1. Bundesliga absteigend geordnet nach dem durchschnittlichen Transferwert ihrer Spieler (angegeben in Tausend Euro), also der monetären Entsprechung der in der Sportschau oft thematisierten “individuellen Klasse”. Zusätzlich sind die Gesamtgröße des

Bundesliga 2007/08. Wert: Stand 10. Juli 2007, Platz: Saisonende. (Quelle: transfermarkt.de)

Verein	Kader	∅-Alter	Marktwert in TEuro	∅-Mw.	Rang	Platz	Differenz
FC Bayern München	27	26,2	233 650	8 654	1	1	0
Werder Bremen	33	24,2	134 400	4 073	2	2	0
FC Schalke 04	28	25,1	104 500	3 732	3	3	0
VfB Stuttgart	30	24,8	103 350	3 445	4	6	–2
Hamburger SV	29	24,4	93 650	3 229	5	4	1
Borussia Dortmund	26	26,3	83 200	3 200	6	13	–7
Bayer 04 Leverkusen	24	24,9	67 625	2 818	7	7	0
Hertha BSC	24	25,4	54 300	2 262	8	10	–2
1. FC Nürnberg	32	25,0	70 750	2 211	9	16	–7
Eintracht Frankfurt	25	25,9	50 450	2 018	10	9	1
VfL Wolfsburg	29	25,6	54 300	1 872	11	5	6
Hannover 96	30	27,3	52 650	1 755	12	8	4
VfL Bochum	27	26,3	32 850	1 217	13	12	1
Arminia Bielefeld	31	26,6	35 850	1 156	14	15	–1
Karlsruher SC	23	25,6	26 300	1 143	15	11	4
MSV Duisburg	31	26,6	27 400	884	16	18	–2
FC Energie Cottbus	30	26,4	22 250	742	17	14	3
Hansa Rostock	25	25,5	18 325	733	18	17	1
Gesamt	504	25,7	1 265 800	2 512			

„Fußball ist keine Mathematik“ (Quelle: www.afb-media.de)

Kaders, das Durchschnittsalter und der Gesamtwert der Mannschaft aufgeführt. Die letzten beiden Spalten zeigen die Platzierung am Ende der Saison und die Differenz zwischen Durchschnittswertrang und Platzierungsrang. Es fällt unmittelbar auf, dass die Differenzen zumeist sehr gering sind. Als ein weiteres interessantes Detail sei angefügt, dass bei den beiden Mannschaften, die zwischen Investitionen und Erfolg das größte Missverhältnis ausweisen, Dortmund und Nürnberg, gegen Ende der Saison die Trainer gefeuert wurden.

Rang-Korrelationen

Wir wollen die Tabelle 1 mit den Methoden der elementaren Statistik untersuchen. Der *Korrelationskoeffizient* zweier Stichproben x und y (nach Pearson) berechnet sich zu

$$r_{x,y} = \frac{s_{x,y}}{s_x s_y}$$

mit

$$s_{x,y} = \frac{1}{n-1} \sum_{i=0}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

und $s_x = \sqrt{s_{x,x}}$. Hier sind $\bar{x}$ und $\bar{y}$ die Mittelwerte der beiden Stichproben, und n ist die Größe beider Stichproben. Es gilt stets $-1 \leq r_{x,y} \leq 1$. Im Allgemeinen sind die Stichproben reellwertig; es macht aber auch Sinn die Berechnung der Korrelation auf Ordinalzahlen anzuwenden, die von Reihenfolgen herrühren. In diesem Fall spricht man von *Rang-Korrelation*. In unserem Fall könnte x_i der Spalte “Rang” der Tabelle 1 entsprechen und y_i dem Wert in der Spalte “Platz”. In der Tabelle 2 sind die Rang-Korrelationen der verschiedenen Merkmale aus Tabelle 1 mit dem Abschneiden am Ende der Saison aufgeführt. Die höchste Korrelation zum Erfolg besteht für den durchschnittlichen Wert der einzelnen Spieler.

Es bleibt zu klären, wie signifikant die berechneten Korrelationen sind. Ein naheliegendes Verfahren besteht darin, die berechnete Korrelation mit der Korrelation einer zufälligen Permutation zu vergleichen. Wollte man für das exakte Ergebnis alle 18!=6 402 373 705 728 000 Permutationen testen, so würde das auch auf einem schnellen Rechner ziemlich lange dauern. Daher begnügt man sich in der Praxis oft mit einem *approximativen Permutationstest*. Hierfür zieht man eine gewisse Anzahl von Permutationen zufällig und zählt, welcher Anteil unter diesen eine höchstens ebenso große Korrelation aufweist wie die auf Signifikanz zu testende Stichprobe. Die aus 100 000 zufälligen Permutationen geschätzten Signifikanzwerte sind ebenfalls der Tabelle 2 zu entnehmen.

Die Signifikanz der Korrelation zwischen der Platzierung am Ende der Saison und dem durchschnittlichen Transferwert der Spieler ist mit über 99,99 % überwältigend, dicht gefolgt von der Signifikanz der Korrelation zum Gesamtwert der Mannschaft. Es besteht aber auch eine fast ebenso starke Signifikanz der Korrelation zum Alter der Spieler (je jünger desto besser). Dagegen schien 2007/08 die Größe des Kaders gar keinen Einfluss zu haben.

Bundesliga 2007/08 – Korrelation zur Platzierung am Saisonende

	Kader	∅-Alter	Marktwert	∅-Mw.
Korrelation	−0,04438	0,47988	0,76058	0,80186
Signifikanz	0,43122	0,97767	0,99980	0,99997

Bundesliga 2008/09. Wert: Stand 11. August 2008, Platz: nach 17 Spieltagen. (Quelle: `transfermarkt.de`)

Verein	Kader	∅-Alter	Marktwert in TEuro	∅-Mw.	Rang	Platz	Differenz
FC Bayern München	23	26,3	252 900	10 996	1	2	−1
Werder Bremen	27	24,3	130 850	4 846	2	8	−6
FC Schalke 04	28	25,8	133 700	4 775	3	7	−4
Bayer 04 Leverkusen	26	24,0	97 850	3 763	4	5	−1
Hamburger SV	28	23,8	94 750	3 384	5	4	1
VfB Stuttgart	28	24,8	94 700	3 382	6	10	−4
Borussia Dortmund	27	25,9	82 750	3 065	7	6	1
VfL Wolfsburg	32	25,5	92 600	2 894	8	9	−1
Hannover 96	27	26,4	66 200	2 452	9	13	−4
Hertha BSC	28	23,9	64 100	2 289	10	3	7
Eintracht Frankfurt	31	25,1	62 550	2 018	11	12	−1
VfL Bochum	29	27,1	44 350	1 529	12	17	−5
Karlsruher SC	23	26,2	33 200	1 443	13	15	−2
1. FC Köln	29	26,1	36 500	1 259	14	11	3
Borussia M'gladbach	28	26,7	34 550	1 234	15	18	−3
TSG 1899 Hoffenheim	26	23,9	31 950	1 229	16	1	15
Arminia Bielefeld	27	25,9	29 700	1 100	17	14	3
FC Energie Cottbus	32	27,1	29 000	906	18	16	2
Gesamt	499	25,5	1 412 200	2 830			

Bundesliga 2008/09 – Korrelation zur Platzierung nach der ersten Saisonhälfte

	Kader	∅-Alter	Marktwert	∅-Mw.
Korrelation	−0,42828	0,72136	0,56037	0,56244
Signifikanz	0,03759	0,99947	0,99158	0,99179

Natürlich stellt sich die Frage, ob die Saison 2007/08 typisch war oder eben nicht. Die Analyse sämtlicher Bundesliga-Jahrgänge seit 1963/64 muss an dieser Stelle leider entfallen. Vergleichsweise wollen wir aber noch einen Blick auf die laufende Saison 2008/09 werfen. Die entsprechenden Daten stehen in Tabelle 3, wobei sich die Platzierung auf das Ergebnis nach 17 Spieltagen bezieht, also nach der ersten Saisonhälfte. Die zugehörigen Korrelationen und (wiederum durch approximativen Permutationstest geschätzten) Signikanzniveaus sind Tabelle 3 zu entnehmen. Der größte Unterschied liegt darin, dass sich für die Kadergröße eine zu etwa 96,2 % signifikante Anti-Korrelation ergibt. Dies würde suggerieren, dass ein kleinerer Kader potenziell schlagkräftiger wäre. Die Korrelation zwischen durchschnittlichem Wert der Spieler und dem Abschneiden fällt zwar schwächer aus als in der vorigen Saison, ist mit knapp 99,18 % aber immer noch ausgeprägt.

Versuch einer Interpretation

Mindestens in den beiden betrachteten Szenarien ist die Korrelation zwischen dem Durchschnittswert der Spieler einer Mannschaft und dem Abschneiden am Ende auffällig hoch. Korrelation und Kausalität sind zweierlei, aber der Einfluss des Geldes ist hier sicherlich keine Überraschung. Interessanter finde ich die außergewöhnlich hohe Signifikanz. Diese lässt vielleicht den Schluss zu, dass die teilweise astronomischen Transfersummen gerechtfertigt sind und mindestens im Mittel den wirklichen Wert der Spieler präzise wiedergeben. Die etwas geringere Signifikanz für die erste Hälfte der laufenden Saison mag damit zusammenhängen, dass sich die wahre Stärke einer Mannschaft erst in einem langen Turnier mit vielen Spielen zeigt. Dennoch ist natürlich klar, dass das Geld allein doch keine Tore schießen kann. Eine Mannschaft aus elf superteuren Mittelstürmern wird die Meisterschaft vermutlich auch nicht holen. Es bedarf zusätzlichen Sachverstands, um die Investitionen auch in Ergebnisse umzusetzen. Und hier wiederum lohnt ein Blick auf die jeweils letzte Spalte in den Tabellen 1 und 3. Je höher der Eintrag desto eher scheinen Trainer und Management ihre Sache im Griff zu haben – wie zum Beispiel Felix Magath in Wolfsburg (2007/08) und Ralf Rangnick in Hoffenheim (2008/09). Umgekehrt ergibt sich eine gewisse Rechtfertigung für die eingangs erwähnten Trainerentlassungen.

Wer jetzt immer noch fragt, warum eine schöne Nebensache wie Fußball einer so ausführlichen mathematischen Betrachtung bedarf, dem sei ein Auspruch des legendären Trainers des FC Liverpool, Bill Shankly, ans Herz gelegt: "Einige Leute halten Fußball für einen Kampf um Leben und Tod. Ich mag diese Einstellung nicht. Ich versichere Ihnen, dass es viel ernster ist!"

Prof. Dr. Michael Joswig, Fachbereich Mathematik, Optimierung, TU Darmstadt, Schlossgartenstraße 7, 64289 Darmstadt
joswig@mathematik.tu-darmstadt.de