

A Starting Point for Analyzing Basketball Statistics

Bachelor-Seminar: Statistik im Sport
im Studienfach Statistik



Datum: 05.12.2009
Vortrag: Stephan Hable
Betreuer: Manuel Eugster
Dozent: Prof. Dr. Friedrich Leisch

Gliederungspunkt

1. Sportart Basketball

Spielidee
Begriffserklärung
Grundregeln
Daten

Gliederung

- 1. Sportart Basketball**
- 2. Problem der großen Vielfalt von statistischen Möglichkeiten im Basketball**
- 3. Prinzip der gleichen Anzahl an Possessions pro Spiel**
- 4. Basisstatistiken im Basketball**
- 5. Zusammenfassung und Ausblick**

Gliederung

Stephan Hable

2

Spielidee

Zwei Teams mit jeweils bis zu 12 Spieler
5 gegen 5 auf dem Feld
Beliebiges Wechseln bei totem Ball

4 Spielabschnitte mit 12 Minuten reiner Spielzeit
Sprungball nur zu Beginn der Partie (und bei möglicher Overtime)
Viertelbeginn jeweils mit abwechselnden Ballbesitz

Freiwurf: 1 Punkt
Wurf aus dem Feld: 2 Punkte
Wurf hinter der Dreierlinie: 3 Punkte

Foul bei Korbaktion gibt Freiwürfe (erfolgreicher Wurfversuch zählt sogar noch)
Foul bei keiner Korbaktion gibt Einwurf
Ab dem 5. Mannschaftsfoul gibt es immer Freiwürfe
Offensivfoul gibt jedoch keine Freiwürfe, zählt aber zu den Mannschaftsfouls
6 Fouls pro Spieler

Begriffserklärung

- NBA ▶ Nordamerikanische Profibasketballiga
- Possessions ▶ Ballbesitz
- Play ▶ Spielzug
- Turnover ▶ Ballverlust eines einzelnen Spielers
- Team Turnover ▶ Ballverlust der ganzen Mannschaft
- Rebound ▶ gesicherter Abpraller nach einem missglückten Wurf
- Team Rebound ▶ Ball fliegt nach Fehlwurf ins Aus oder
▶ Freiwurf ohne Rebounderlaubnis wird verworfen
- Steal ▶ Stehlen des Balles oder Abfangen eines Passes
- Field Goal ▶ Wurf (u.a. Korbleger oder Dunk)
- Freiwurf ▶ ungehinderter Wurf auf dem Korb nach einem Foul

Grundregeln

Es gibt 7 Arten des technischen Fehlers:

- Schrittfehler
- Doppeldribbling
- Schauffeln
- Rückspiel
- 24 Sekunden Uhr (Shotclock)
- 8 Sekunden (um aus der eigenen Hälfte zu kommen)
- 5 Sekunden (um beim Einwurf zu passen)

Die letzten drei Aufzählungen werden auch als Team Turnover bezeichnet.

Wichtig: Eine Verletzung dieser Regeln führt zum Ballbesitz der Gegner

Daten: Box Score

DALLAS MAVERICKS (1-1)																	
		FIELD GOALS					REBOUNDS										
	POS	MIN	FGM-A	3PM-A	FTM-A	±	OFF	DEF	TOT	AST	PF	ST	TO	BS	BA	PTS	
S.Marion	F	33:45	8-15	0-0	2-2	37	4	2	6	1	2	1	1	1	0	18	
D.Nowitzki	F	38:11	5-14	1-4	10-11	32	0	10	10	1	2	1	1	1	0	21	
E.Dampier	C	17:39	3-5	0-0	2-4	13	3	7	10	1	2	1	2	1	0	8	
O.Ross	G-F	13:01	2-3	0-0	0-0	-4	0	0	0	1	4	0	0	0	0	4	
J.Kidd	G	33:31	1-4	0-2	0-0	24	1	3	4	11	1	1	1	6	0	2	
J.Terry		33:33	5-12	3-5	3-5	7	0	3	3	2	0	0	1	0	1	16	
D.Gooden		12:15	0-4	0-0	2-2	-5	0	2	2	0	2	0	1	0	0	2	
J.Barea		24:33	3-7	0-0	6-6	0	2	2	4	2	2	0	2	0	0	12	
K.Humphries		20:08	4-8	0-0	0-2	-14	3	4	7	0	5	2	3	0	0	8	
J.Singleton		07:55	1-3	1-3	0-0	-10	0	0	0	0	1	0	0	0	0	3	
R.Beaubois		02:49	0-1	0-0	0-0	-5	0	0	0	0	0	0	0	0	0	0	
M.Carroll		02:38	0-1	0-1	0-0	-5	0	0	0	0	0	0	0	0	0	0	
Total		240	32-77	5-15	25-32		13	33	46	19	21	6	17	3	1	94	
			41.6%	33.3%	78.1%		TEAM REBS: 18				TOTAL TO: 18						

Daten: Play-by-Play

START OF 1ST QUARTER		
(12:00) JUMP BALL BYNUM VS DAMPIER (ODOM GAINS POSSESSION)		
	11:43	Bynum Jump Shot: Missed
	11:42	Bryant Rebound (Off:1 Def:0)
	11:32	Bryant Jump Shot: Missed
Nowitzki Rebound (Off:0 Def:1)	11:31	
Dampier Foul : Offensive (1 PF)	11:20	
Dampier Turnover : Foul (1 TO)	11:20	
[LAL 2-0]	11:08	Fisher Jump Shot: Made (2 PTS) Assist: Odom (1 AST)
Ross Jump Shot: Missed	10:56	
	10:55	Bynum Rebound (Off:0 Def:1)
	10:43	Bryant Jump Shot: Missed
Marion Rebound (Off:0 Def:1)	10:42	
	10:35	Bryant Foul : Shooting (1 PF)
Dampier Free Throw 1 of 2 Missed	10:35	
Team Rebound	10:35	
Dampier Free Throw 2 of 2 (1 PTS)	10:35	
[DAL 1-2]	10:35	
Ross Foul : Shooting (1 PF)	10:18	
[LAL 3-1]	10:18	Bryant Free Throw 1 of 2 (1 PTS)
	10:18	Bryant Free Throw 2 of 2 Missed
Nowitzki Rebound (Off:0 Def:2)	10:17	
Kidd Pullup Jump shot: Made (2 PTS)	09:55	
[DAL 3-3]	09:55	
	09:45	Fisher Jump Shot: Missed
	09:44	Artest Rebound (Off:1 Def:0)
	09:38	Bryant Reverse Layup Shot: Made (3 PTS) Assist: Artest (1 AST)
[LAL 5-3]	09:38	

Daten: Eigene Resultate

Eigene Auswertungen auf dem Overhead

NBA Saison 2005-2006

Vergleich:

Dallas Mavericks und ihr Starspieler Dirk Nowitzki

Kobe Bryant und seine Los Angeles Lakers

1. Sportart Basketball

Stephan Hable

8 Gliederung

Stephan Hable

Problem der großen Vielfalt

In der NBA gibt es unzählige Statistiken zu Spielern, Mannschaften und Trainern:

- + kein Datenmangel
- + interessante Informationen zu jeder erdenklichen Situation
- + großer Marktplatz für neue Ideen

- mögliche Falschaussagen verschiedener Methoden
- teilweise schwierige Kommunikation verschiedener Gruppen untereinander
- Probleme der Zuschauer den Argumenten der Gruppen zu folgen

→ Einheitliches Konzept zur statistischen Analyse im Basketball
von D. Oliver und J. Hollinger

2. Problem

Stephan Hable

9 Gliederung

Stephan Hable

Gliederungspunkt

2. Problem der großen Vielfalt von statistischen Möglichkeiten im Basketball

Gliederungspunkt

3. Prinzip der gleichen Anzahl an Possessions pro Spiel

Ausführliche Definition
Allgemeine Schätzformel
Spezialisierungen
Vergleich der beiden Box Score Varianten
Vergleich der verschiedenen Formeln

Ausführliche Defintion

Eine Possession bedeutet, dass eine Mannschaft gerade im Ballbesitz ist.

Es gibt drei Wege, wie eine Mannschaft ihr Recht auf den Ball an den Gegner abgibt:

- 1) Punkten
- 2) Defensivrebound des Gegners
- 3) Turnover

Aber: Ein Offensivrebound bedeutet keinen neuen Ballbesitz

→ Die Anzahl der Possessions pro Mannschaft in einem Spiel sind demnach fast gleich (maximale Differenz von zwei Possessions)

Allgemeine Schätzformel

$$POSS_t = (FGM_t + \lambda FTM_t) + \alpha [(FGA_t - FGM_t) + \lambda (FTA_t - FTM_t) - OREB_t] + (1 - \alpha) DREB_o + TO_t$$

FGA = Wurfversuche

FGM = erfolgreicher Wurf

OREB = Offensivrebound

TO = Turnover

λ = Anteil der Freiwürfe, die zu einem Ballbesitzwechsel führen

α = Parameter zwischen 0 und 1

FTA = Freiwurfversuche

FTM = erfolgreicher Freiwurf

DREB = Defensivrebound

Spezialisierungen

Fall 1: "verlorene Ballbesitze" ($\lambda = 0, 44$ und $\alpha = 1$)

$$POSS_t = FGA_t + 0,44 * FTA_t - OREB_t + TO_t$$

Fall 2: "gewonnene Ballbesitze" ($\lambda = 0, 44$ und $\alpha = 0$)

$$POSS_t = FGM_t + 0,44 * FTM_t + DREB_o + TO_t$$

Vergleich der beiden Box Score Varianten

Variable	(1)		(2)	
	Coeff	Std Error	Coeff	Std Error
Field goals attempted	0.9640	0.0039	0.9492	0.0036
Field goals missed	-0.3452	0.0086	--	--
Free throws attempted	0.4637	0.0034	0.4437	0.0030
Free throws missed	-0.2073	0.0098	--	--
Offensive rebounds	-0.6227	0.0100	-0.9599	0.0074
Defensive rebounds (for opponent)	0.3643	0.0086	--	--
Turnovers	0.9767	0.0053	0.9550	0.0061
Constant	3.2258	0.5200	3.8810	0.6070
R ²	0.9615		0.9473	
Number of observations	5,178			

Data are from the 2002-03 through 2005-06 seasons.

Fehlwurf = $0,9640 + (- 0,3452) = 0,6188$

Fehlwurf an der Freiwurflinie = $0,4637 + (- 0,2073) = 0,2564$

Fehlwurf + Defensivrebound = $0,6188 + 0,3643 = 0,9831$

4. Basisstatistiken im Basketball

- Offensiv- und Defensivrating
- Average League Ratings
- Plays
- Per-Minute Statistiken
- Trefferquoten
- Rebound Rate
- Four Factors
- Plus/Minus Statistiken
- Individuelle Ballbesitzbehandlungs- und Effektivitätsstatistiken
- Lineare Gewichtung
- Pythagoreische Siegesquote
- Methode der glockenförmigen Kurve

Rating = Bewertung im jeweiligen Bereich

$$\text{Offensive Rating (ORtg}_t) = \text{PTS}_t / \text{POSS}_t * 100$$

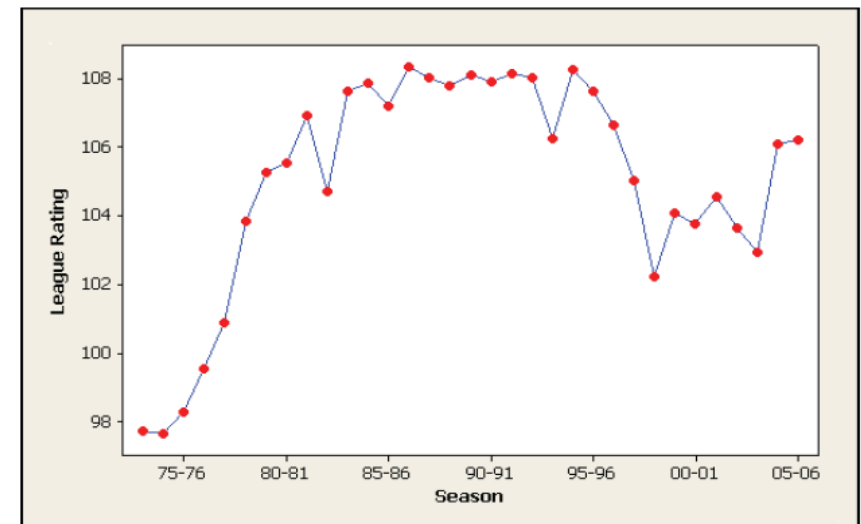
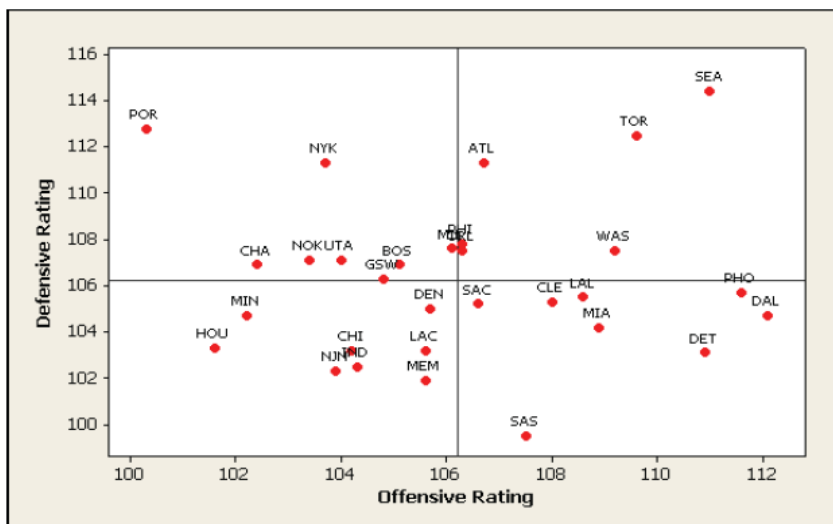
► erzielte Punkte pro 100 Possessions

$$\text{Defensive Rating (DRtg}_t) = \text{PTS}_o / \text{POSS}_o * 100$$

► kassierte Punkte pro 100 Possessions

Beispiel: Saison 2005-2006

Average League Ratings



Plays

Play = Spielzug
Possession = Ballbesitz

$$\text{PLAYS}_t = \text{FGA}_t + 0,44 * \text{FTA}_t + \text{TO}_t$$

Im Zeitraum von 2002-2003 bis 2005-2006 (insgesamt 5178 Spiele):

- durchschnittlich 105,0 plays (sowohl geschätzt als auch exakt)
- durchschnittlich 91,7 possessions
- ▶ 1,145 plays pro possession

Per-Minute Statistiken

scoring rate = Erzielte Punkte eines Spielers in 40 Minuten:

$$\text{PTS}_{40p} = \text{PTS}_p / \text{MIN}_p * 40$$

NBA Spiele dauern 48 Minuten, aber

- einzelne Spieler selten länger als 40 Minuten auf dem Feld
- Spieldauer im College: zweimal 20 Minuten
- Europäische Ligen sowie Europa- und Weltmeisterschaften: vier mal 10 Minuten

Anpassung des Spieltempos:

$$\text{adjPTS}_{40p} = \text{PTS}_{40p} * (\text{POSS}_i / \text{POSS}_t)$$

Wichtig: Minimum von 500 bzw 1000 gespielten Minuten pro NBA Saison

Trefferquoten

Man unterscheidet 3 verschiedene Trefferquoten:

- 1) field goal percentage ▶ $\text{FG}\% = \text{FGM} / \text{FGA}$
- 2) effective field goal percentage ▶ $\text{eFG}\% = (\text{FGM} + 0,5 * 3\text{PM}) / \text{FGA}$
- 3) true shooting ▶ $\text{TS}\% = (\text{PTS} / 2) / (\text{FGA} + 0,44 * \text{FTA})$

Durchschnittswerte eines NBA Spielers im Intervall von 1996-97 bis 2005-06:

- $\text{FG}\%$ = 44,6 %
- $\text{eFG}\%$ = 47,9 %
- $\text{TS}\%$ = 52,3 %

Rebound Rate - Player

= Wahrscheinlichkeit eines Spielers einen Rebound zu holen, wenn er im Spiel ist

$$\text{REB}\%_p = \left(\frac{\text{REB}_p}{\text{REB}_i + \text{REB}_o} \right) \left/ \left(\frac{\text{MIN}_p}{\text{MIN}_i} \right) \right.$$

Lässt sich wiederum in Angriff und Verteidigung unterteilen:

$$\text{OREB}\%_p = \left(\frac{\text{OREB}_p}{\text{OREB}_i + \text{DREB}_o} \right) \left/ \left(\frac{\text{MIN}_p}{\text{MIN}_i} \right) \right.$$

$$\text{DREB}\%_p = \left(\frac{\text{DREB}_p}{\text{OREB}_o + \text{DREB}_i} \right) \left/ \left(\frac{\text{MIN}_p}{\text{MIN}_i} \right) \right.$$

Rebound Rate – Team

- = Wahrscheinlichkeit der Mannschaft den Rebound zu holen
- Rebounds werden verstärkt von der Verteidigung eingesammelt
- ähnlich den Rebounds per Game
- isoliert aber besser die Fähigkeiten eines Teams zu Rebunden

$$OREB\%_t = \frac{OREB_t}{OREB_t + DREB_o}$$

$$DREB\%_t = \frac{DREB_t}{OREB_o + DREB_t}$$

$$REB\%_t = \frac{OREB\%_t + DREB\%_t}{2}$$

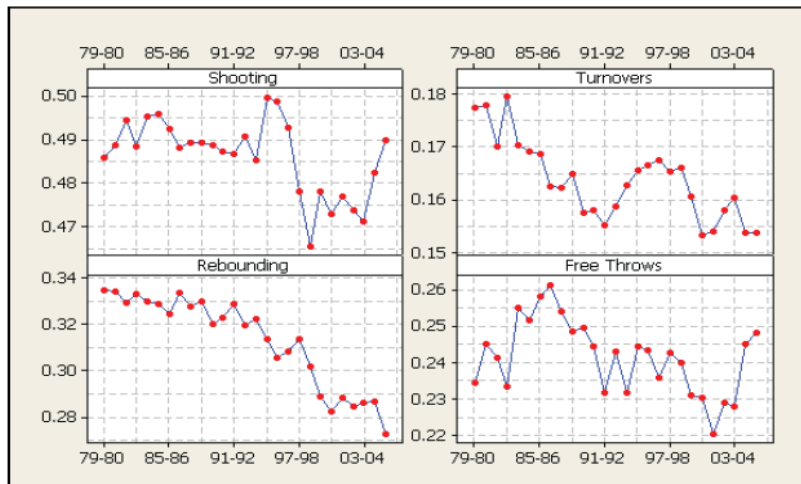
ABER: negativer Zusammenhang zwischen OREB% und DREB%

Four Factors

- Methode zum Abschätzen des Siegers der zwei Mannschaften
 - einzelne Faktoren müssen dazu "gewonnen" werden
 - unterschiedliche Gewichtungen
- 1) effective field goal percentage (eFG%) mit Gewichtung 10
 - 2) Turnovers pro Possession ($TO_t / POSS_t$) mit Gewichtung 6
 - 3) Offensivreboundquote (OREB%) mit Gewichtung 3
 - 4) Free throw rate (FTM_t / FGA_t) mit Gewichtung 3

Four Factors - Entwicklung

Entwicklung der Vier Faktoren anhand des Ligadurchschnitts



Plus/Minus Statistiken

- 1) Plus/Minus Statistik:
 - Punktedifferenz (erzielte Punkte minus kassierte Punkte), wenn ein bestimmter Spieler auf dem Parkett ist
 - oft wird eine Basis pro Minute oder pro Possession verwendet
- 2) net Plus/Minus Statistik:
 - Plus/Minus Statistik für einen bestimmten Spieler im Vergleich zu einer relativen Plus/Minus Statistik des Teams, wenn der bestimmte Spieler auf der Bank ist
 - wurde früher auch als "Roland Rating" bezeichnet
- 3) adjusted Plus/Minus Statistik:
 - gibt Auskunft über die Leistung eines Spielers in Zusammenspiel mit seinem Team
 - gibt Auskunft über die Leistung eines Spielers gegen seine Gegenspieler

Größter Vorteil:

+ Anerkennung individueller Leistungen, welche nicht im Box Score notiert werden

Größter Nachteil:

- Leistung der Mit- und vor allem der Gegenspieler hat immer großen Einfluss

Individuelle Ballbesitzbehandlungs- und Effektivitätsstatistiken

J. Hollinger:

Individuel possession rate = wie intensiv nutzen Spieler ihre Possessions für Wurfversuche, Freiwurfversuche, Assists und Turnovers

D. Oliver:

Individuel possession rate = wie intensiv nutzen Spieler ihre Possessions für Wurfversuche, Freiwurfversuche, Assists, Turnovers und Offensivrebounds

→ ein Spieler nutzt ein Fünftel des Ballbesitzes

Individuel offensive rating = wie effizient nutzen Spieler ihre Possessions zum Punkten

→individuel offensive rating ist in der Summe das Offensivrating des Teams

Pythagoreische Siegesquote

$$PYTH_t = \frac{PTS_t^x}{PTS_t^x + PTS_o^x}$$

- ursprüngliche Entdeckung im Baseball
- Exponent x zwischen 13 und 17 in der NBA
- Exponent x wurde auf 16,8 festgelegt
- kleiner Werte sind aufgrund weniger Possessions für andere Ligen besser geeignet
- Punkte können auch ohne Verluste durch das Offensiv- bzw Defensivrating ersetzt werden

Lineare Gewichtung

$$NBA_EFF_p = PTS_p + REB_p + AST_p + STL_p + BLK_p - TO_p - Missed_FG_p - Missed_FT_p$$

Nachteile:

- Nur Stats aus dem Box Score; keine Help-Verteidigung oder Wurfsschirme
- Kein Zusammenhang über Sieg oder Niederlage der Mannschaft

Andere bekannte Versionen der Linearen Gewichtung:

- > Player Efficiency Rating (PER) von John Hollinger
- > TENDEX von Dave Heeren
- > Points Created von Bob Bellotti

Methode der glockenförmigen Kurve

$$Win\% = NORMSDIST \left[\frac{PPG_t - PPG_o}{StDev(PPG_t - PPG_o)} \right]$$

- net points: Differenz der erzielten minus der kassierten Punkte pro Spiel
- Normalisierung der net points durch die Dividierung ihrer Standardabweichung
- Z-Statistik
- Einsetzen in die Dichtefunktion der Normalverteilung liefert das Schätzergebnis

Vorteile gegenüber pythagoreischen Siegesquote:

- + keine Modifikation für andere Ligen notwendig
- + minimal bessere Genauigkeit

► meistens wird dennoch die pythagoreische Siegesquote aufgrund seiner Einfachheit benutzt

Gliederungspunkt

- 6. Zusammenfassung und Ausblick**
- Fazit der statistischen Analyse**
- Diskussion des Papers**
- Ausblick auf weitere statistische Möglichkeiten**

Diskussion des Papers

- Hintergrundwissen muss vorhanden sein
- oft unnötige Schätzungen
- nicht jede Leistung wird im Box Score notiert
- Aussagekraft der Statistiken immer kontrollieren
- Trefferquoten könnten logischer sein

Fazit der statistischen Analysen

Basis: Prinzip der gleichen Anzahl an Possessions pro Mannschaft pro Spiel

→ Schätzungen fast identisch mit der Realität

Davon ausgehend werden einige Grundstatistiken im Basketball veranschaulicht

Somit lassen sich die Mannschaften und deren Spiele sehr gut in den verschiedenen Bereichen (Scoring, Rebounding, Effektivität, Trefferquoten, etc) miteinander vergleichen.

Ausblick auf weitere Möglichkeiten

- Punkte in Zusammenhang mit Trefferquoten
- Verhältnis von Assists gegenüber Turnover
- Einfluss von Heim- und Auswärtsspielen
- Einfluss der Spielerrotation
- Statistiken zu Trainern und Coaches
- Unterschiede der Statistiken in der Regular Season und den Playoffs

ENDE

ENDE