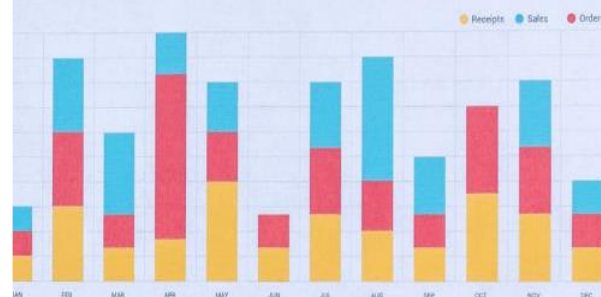


PARTE IV: Cinematografia e Matemática

FAIXA ETÁRIA: 16 – 18



Estadística (Fonte: @Lukas from Pexels.com)

UNIDADE 38: PROBABILIDADE E ESTATÍSTICA NO FILME “MONEYBALL – JOGADA DE RISCO”

SPEL – Sociedade Promotora de Estabelecimentos de Ensino



Co-funded by the Erasmus+ Programme of the European Union

Guia do Professor

Título: Probabilidade e Estatística no filme “Moneyball – Jogada de Risco”

Faixa etária: 16 – 18 anos

Duração: 2 horas

Conceitos matemáticos: Estatística, Probabilidades.

Conceitos artísticos: Sabermétrica

Objetivos gerais: Dotar os alunos do conceito de teoria da probabilidade e estatística

Instruções e Metodologias: Exibir o excerto do filme “Moneyball – Jogada de Risco” no qual o conceito de sabermétrica é mostrado (conferir hiperligação na secção “Informações e Recursos Adicionais”) e sugerir que os alunos assistam ao filme na íntegra em casa.

Recursos: Uma caneta e uma calculadora.

Dicas para o professor: Para um melhor entendimento por parte dos alunos, deverão ser facultadas, de antemão, algumas regras básicas e posições de campo do basebol.

Resultados de aprendizagem e competências: No final desta unidade, o aluno estará apto a:

- Avaliar informação e usá-la para resolver equações envolvendo probabilidades;
- Compreender como as estatísticas poderão ser usadas para prever o resultado de um determinado evento.

Síntese e avaliação:

Indique 3 aspetos que tenha gostado acerca desta atividade:	1. 2. 3.
Indique 2 conceitos que tenha aprendido:	1. 2.
Indique 1 aspeto a melhorar:	1.

Introdução

Por vezes encontram-se aspetos relacionados com a Matemática em séries televisivas ou filmes. Em alguns casos, não é dada muita importância a estes aspetos da Matemática pelo facto de não influenciarem a história em si. Contudo, existem alguns casos em que tal se verifica.

Alguns exemplos incluem: “21” (EUA, 2008), de Robert Luketic; “A Prova” (EUA, 2005), de John Madden; “Uma Mente Brilhante” (EUA, 2001), de Ron Howard; “Enigma” (EUA, 2001), de Michael Apted; “Pi” (EUA, 1998) de Darren Aronofsky; “O Bom Rebelde” (USA, 1997), de Gus Van Sant e “Cubo” (Canadá, 1997), de Vincenzo Natali.

Nesta unidade, o filme “Moneyball – Jogada de Risco” (EUA, 2011), de Bennet Miller, será alvo de estudo e serão analisados os conceitos matemáticos nele abordados tais como a probabilidade e estatística.

Moneyball – Jogada de Risco

O filme norte-americano “Moneyball – Jogada de Risco” (2011) retrata a história verídica de um diretor geral da equipa de basebol de Oakland, da Califórnia, que luta para formar uma equipa competitiva para a temporada desportiva de 2002 com um orçamento muito limitado.

No filme, após ter perdido 3 jogadores-chave, o diretor geral Billy Beane (interpretado por Brad Pitt) e o seu assistente Peter Brand (Jonah Hill) recorrem a um método saber métrico pouco ortodoxo com o fim de detetar jogadores de basebol subestimados por motivos tendenciosos (tais como a idade, aparência e personalidade), longe do radar das grandes equipas, sendo alternativas financeiramente mais acessíveis para uma equipa de orçamento curto como é o caso dos Oakland Athletics.

Enfrentando uma dura resistência por parte dos olheiros veteranos e antiquados de Oakland, que desconsideravam esta estratégia dizendo que a sua experiência e conhecimento no basebol tinham muito mais valor que qualquer estatística, Beane ignora as suas objeções e forma uma equipa seguindo os dados estatísticos saber métricos de Peter Brand.

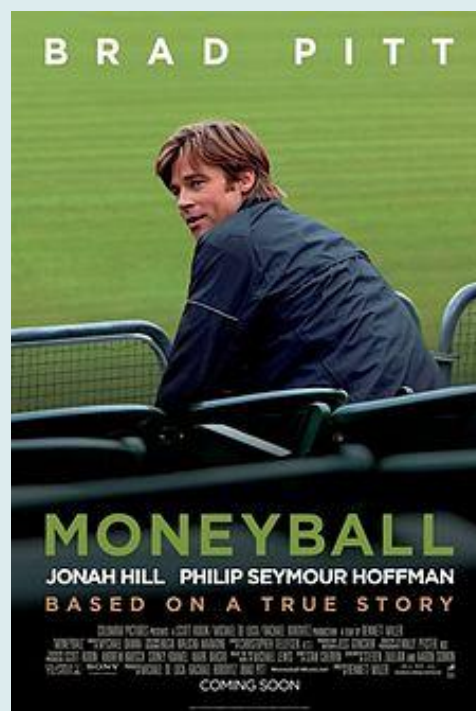


Fig. 1 – Cartaz do filme Moneyball (2011)

(Fonte: <https://pt.wikipedia.org/wiki/Moneyball>)

Glossário

Sabermétrica – um termo inventado pelo escritor e especialista em estatística de basebol Bill James. Deriva da sigla da Sociedade para a Pesquisa de Basebol Americana (SABR) e trata-se de um método que faz a recolha e análise de estatísticas do jogo com o fim de avaliar o desempenho dos jogadores e das equipas em todos os aspetos da partida, tais como:

Batida/Defesa – no basebol, duas equipas opositoras alternam entre turnos, quer na fase de batida, quer na fase de defesa.

→ **Batida** – ato de bater a bola quando esta é arremessada pelo lançador (1) adversário. O jogador que ocupa esta posição é conhecido como recetor (2).

→ **Defesa** - as posições sobre as quais uma equipa está disposta num campo de basebol. Existem 9 posições: o lançador (1) e o recetor (2), que ocupam posições fixas, e o primeira-base (3), segunda-base (4), terceira-base (5), interbases (6), defesa esquerdo (7), defesa central (8) e defesa direito (9), que se podem movimentar livremente pelo campo.

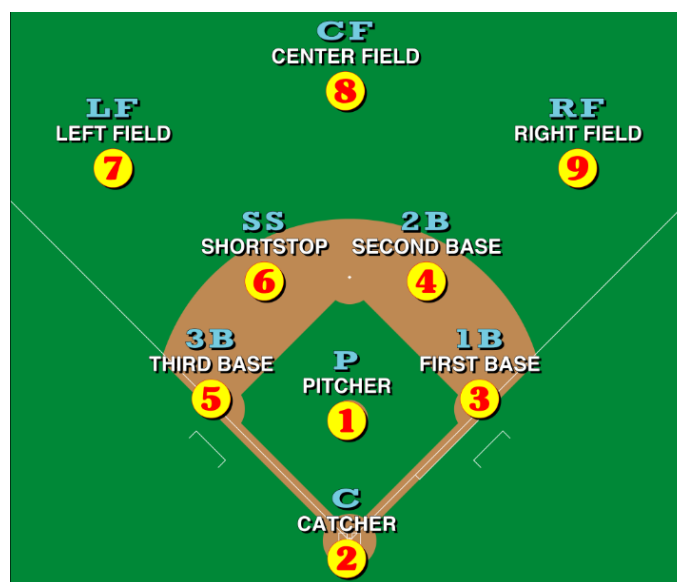


Fig. 2 – Posições do Basebol (Fonte: By Michael J, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=40095322>)

→ **Estatísticas Padrão/Avançadas** – Estatísticas relativas ao desempenho individual/coletivo em qualquer desporto. Abaixo estão representados os termos relativos a basebol abordados nesta unidade:

→ **Assistência (A)** – número de assistências do defesa. O jogador é creditado com uma assistência quando, após a batida, a bola toca o chão, o mesmo defesa captura-a e a arremessa até um defesa situado numa das bases, resultando na eliminação do adversário.

→ **At Bats ou Oportunidades no Bastão (AB)** – número de vezes em que o jogador enfrentou o lançador. O lançador não é creditado com um AB quando:

- É atingido pelo lançador adversário (hit by pitch);
- Recebe um walk;
- Realiza uma batida de sacrifício.

→ **Base on Ball ou Walks (BB)** – Número de walks que o lançador cedeu. Cada vez que o lançador lança quatro bolas contra um mesmo recetor, ele cede um walk.

→ **Caught Stealing ou Apanhado a Roubar (CS)** – Número de vezes em que o jogador foi apanhado pela defesa adversária enquanto tentava roubar a base. Para ser eliminado, o mesmo não pode estar em contacto com a base e um adversário deve tocá-lo com a bola, sem que uma batida tenha ocorrido.

→ **Jogos efetuados (G)** – Total de jogos efetuados por um só jogador. Ao jogador será creditado um jogo caso o mesmo tenha estado presente em qualquer ponto efetuado durante o jogo. Se um jogador totalizar 162 G, significa que se trata de um totalista, tendo efetuado todos os jogos de uma temporada.

→ **Hit ou Batida Válida (H)** – Quando o recetor bate a bola para um território válido e alcança a base. Os hits podem ser singulares, duplos, triplos e home runs;

→ **Plate Appearances (PA)** – Ocorre cada vez que um jogador completa um turno ao bater com um hit, walk, fora ou alcançando a base num erro.

→ **Putout (PO)** – Um jogador é creditado com um Putout quando elimina um adversário, seja ao conseguir um Flyout, pisar numa base para um Groundout ou encostar num adversário para um Tag Out;

→ **Run ou Corrida (R)** – Ocorre quando um jogador cruza a Home Plate (base final que um jogador deve tocar para anotar uma corrida).

→ **Stolen Bases ou Bases Roubadas (SB)** – Número de bases roubadas pelo jogador enquanto corredor. Para roubar uma base, o jogador já deve estar numa base. Assim, enquanto o lançador adversário arremessa para o recetor da sua equipa, o mesmo jogador correrá até à base seguinte mesmo não havendo batida. Se chegar a salvo, o jogador é creditado com um roubo de base.

→ **Total Bases ou Total de Bases (TB)** – Número total das bases somadas que o recetor conseguiu com as suas batidas válidas.

A Matemática por trás de “Moneyball – Jogada de Risco”

No filme “Moneyball – Jogada de Risco”, a certa altura, ao fazer uma antevisão da temporada de 2002, Peter Brand referiu que, para a equipa atingir a fase dos playoffs, teria de ganhar pelo menos 99 das 162 partidas. De forma a alcançar este número, projetou o número mínimo de corridas que a equipa precisaria de completar e o número máximo de corridas que poderia conceder.

Para chegar a este resultado, Brand utilizou uma equação originalmente desenvolvida pelo estatístico desportivo Bill James, conhecida por “Expetativa Pitagórica”, a qual resulta num rácio aproximado de vitórias baseado no número de corridas marcados (a favor) e concedidos (contra). A equação tem a seguinte forma:

$$\text{Rácio de Vitórias} = \frac{\text{Runs a favor}^2}{\text{Runs a favor}^2 + \text{Runs contra}^2}$$

Para a temporada de 2002, Peter projetou que a equipa teria de marcar pelo menos 814 corridas e não conceder mais de 645, o que resulta no seguinte:

$$\text{Rácio de Vitórias} = \frac{814^2}{814^2 + 645^2} = \frac{662596}{1078621} = 0.614299\%$$

O rácio de vitórias é dado em percentagem e, quando multiplicado pelo número de jogos numa época de basebol (162), resulta no número aproximado de jogos que a equipa precisa de ganhar de forma a qualificar-se para a fase dos playoffs.

$$0.614299\% \times 162 = 99.516438 \text{ jogos}$$

Peter Brand revela depois uma base de dados na qual compilou informações individuais de vários jogadores com origem na média de dados estatísticos dos seus jogos (Fig. 3), que será o mote para encontrar os jogadores economicamente mais eficientes.

	OBP	OPS	Runs	% LA
8	0.380	1.038	1246	67%
7	0.419	0.876	1139	53%
5	0.412	0.787	1009	35%
6	0.363	0.819	926	24%
3	0.363	0.806	909	22%
9	0.353	0.812	892	20%
5	0.354	0.799	878	18%
8	0.319	0.797	787	5%

Fig. 3 – Banco de dados de estatísticas saberométricas de jogadores de baseball, de Peter Brand
(Fonte: Filme “Moneyball – Jogada de Risco”)

Apesar de ter apenas 1/3 do orçamento para salários comparativamente aos tubarões do mercado, como os New York Yankees, que se sagraram campeões na Divisão Este da American League, os Oakland A's empataram em número de vitórias com os nova-iorquinos com 103 vitórias na fase regular, conquistando o recorde de 20 vitórias consecutivas na American League e vencendo a Divisão Oeste da competição.

Embora tenham sido eliminados no pós-temporada, esta abordagem saberométrica mudou totalmente a indústria através do uso de Matemática e Estatística.

9

Quem é Bill James?

George William James (nascido em 1949), é um escritor dedicado ao baseball americano, historiador e estatístico, conhecido principalmente por apresentar o método estatístico saberométrico.

Além da já referida “Expectativa Pitagórica”, Bill James apresentou outras inovações no cômputo da Estatística, tais como as Corridas Criadas (RC), Fator de Alcance (RF) e Média Secundária (SecA).



Fig. 4 – Bill James, 2010

(Fonte: https://en.wikipedia.org/wiki/Bill_James)

Corridas Criadas: estatística que estima a contribuição ofensiva de uma equipa/jogador em relação às corridas alcançadas no jogo.

Este método pode também ser utilizado como forma de obter um número aproximado de corridas que uma equipa alcançará quando está a arremessar. A fórmula é a seguinte:

$$RC = \frac{TB * (H + BB)}{PA}$$

Onde:

TB = Total Bases;

H = Hits;

BB = Base on Balls / Walks;

PA = Plate Appearances.

Consideremos as seguintes estatísticas da edição de 2018 da MLB das equipas dos Detroit Tigers (DET) e dos Oakland Athletics (OAK) (baseball-reference.com):

Tm	#Bat	BatAge	R/G	G	PA	AB	R	H	2B	3B	HR	RBI	SB
DET	49	27.9	3.89	162	6029	5494	630	1326	284	35	135	597	70
OAK	53	28.0	5.02	162	6255	5579	813	1407	322	20	227	778	35

Tm	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	LOB
DET	30	428	1341	.241	.300	.380	.680	85	2085	110	52	15	40	18	1071
OAK	21	550	1381	.252	.325	.439	.764	109	2450	136	76	6	44	18	1085

Fig. 5 – Estatísticas dos Detroit Tigers e Oakland Athletics da temporada 2018 da MLB

(Fonte: <https://www.baseball-reference.com/leagues/MLB/2018.shtml>)

Calculemos as Corridas Criadas pelos OAKs:

$$RC (OAK) = \frac{2450 * (1407 + 550)}{6255} = 766.53$$

De acordo com os cálculos, os Oakland Athletics deveriam ter alcançado cerca de 767 corridas. Efetivamente, os Oakland Athletics conseguiram 813 corridas.

Se 813 corridas corresponde a 100%, então 767 corridas corresponde a 94,34%, o que significa que existe um pequeno desvio mínimo de 5,6%.

Apliquemos o mesmo para os DET:

$$\text{RC (DET)} = \frac{2085 * (1326 + 428)}{6029} = 606.58$$

Com base nas estatísticas fornecidas, deveriam ter sido alcançados 607 corridas pelos Tigers na época de 2018. A equipa de Detroit terminou a época com 630 corridas. Uma vez mais, é apresentada uma margem de erro reduzida (3,6%). Este cálculo pode também ser aplicado individualmente a jogadores e revela-se útil para verificar se o recetor cumpriu com a sua principal função: a criação de corridas.

Média Secundária: trata-se de uma versão melhorada da equação da Média de Batidas. Embora opere ainda sob os princípios da Média de Batidas, a Média Secundária abrange também o poder do jogador (bases extra), visão (walks) e velocidade (bases roubadas). A sua fórmula propõe-se a medir a eficácia ofensiva de um jogador/equipa, representando-se da seguinte forma:

$$\text{SecA} = \frac{\text{BB} + (\text{TB} - \text{H}) + (\text{SB} + \text{CS})}{\text{AB}}$$

Na qual:

BB = Base on Balls/Walks;

TB = Total da Base;

H = Hits;

SB = Bases Roubadas;

CS = Apanhado a Roubar;

AB = At Bats.

Observe as estatísticas padrão de batimento dos recetores James McCann, dos Detroit Tigers, e Jonathan Lucroy, dos Oakland A's.

Name	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB
<u>James McCann</u>	118	457	427	31	94	16	0	8	39	0	3	26
<u>Jonathan Lucroy</u>	126	454	415	41	100	21	1	4	51	0	0	29

Name	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB
<u>James McCann</u>	116	.220	.267	.314	.581	58	134	9	2	0	2	0
<u>Jonathan Lucroy</u>	65	.241	.291	.325	.617	71	135	12	3	1	6	1

Fig. 6 – Estatísticas da temporada de 2018

(Fonte: <https://www.baseball-reference.com/players/m/mccanja02.shtml>
e <https://www.baseball-reference.com/players/l/lucrojo01.shtml>)

Usando a equação da Média Secundária, temos:

$$\text{SecA (James McCann)} = \frac{26 + (134 - 94) + (0 + 3)}{427} = 0.161$$

$$\text{SecA (Jonathan Lucroy)} = \frac{29 + (135 - 100) + (0 + 0)}{454} = 0.140$$

12

O resultado final arredondado às milésimas representa a média secundária de um jogador. Neste caso, James McCann apresenta uma eficácia global superior, o que significa que, teoricamente, a longo prazo, é mais eficaz ofensivamente.

Fator de Alcance (RF): estatística que quantifica a contribuição de um jogador a uma determinada posição defensiva. A equação apresenta-se da seguinte forma:

$$\text{RF} = \frac{A + \text{PO}}{G}$$

Onde:

A = Assistências;

PO = Putouts;

G = Jogos efetuados.

Considere as estatísticas defensivas dos mesmos dois jogadores:

Name	Lg	G	GS	CG	Inn	Ch	PO	A	E	DP	Fld%	Rtot	Rdrs	Rtot/yr
James McCann	AL	114	112	111	987.1	902	847	50	5	10	.994	4	-1	5
Jonathan Lucroy	AL	125	119	105	1066.1	950	857	83	10	3	.989	-6	-11	-7

Name	Rdrs/yr	RF/9	RF/G	lgFld%	lgRF9	lgRFG	PB	WP	SB	CS	CS%	lgCS%	PO
James McCann	-1	8.18	7.87	.994	9.07	8.98	5	37	47	27	36%	28%	1
Jonathan Lucroy	-12	7.93	7.52	.994	9.07	8.98	10	63	72	31	30%	28%	0

Fig. 7 – Estatísticas da temporada de 2018

(Fonte: <https://www.baseball-reference.com/teams/DET/2018.shtml>
e <https://www.baseball-reference.com/teams/OAK/2018.shtml>)

Os mesmos jogadores produziram os seguintes resultados ocupando uma posição defensiva:

$$\text{RF (James McCann)} = \frac{50 + 847}{114} = 7,86$$

$$\text{RF (Jonathan Lucroy)} = \frac{83 + 857}{125} = 7,52$$

O Fator de Alcance de James McCann é superior ao de Jonathan Lucroy. Por outras palavras, McCann ocupa um papel significativamente mais relevante ao nível defensivo.

Como em toda a análise de fatores, é importante compreender que, quanto maior for o tamanho da amostra/dados utilizados, mais precisos serão os resultados.

Muitas outras fórmulas foram desenvolvidas por Bill James para tratar outros dados estatísticos padrão e avançados; ao longo do tempo, alguns deles foram refinados e outros foram criados por outros estatísticos. Apesar de terem sido inicialmente concebidos para análise de jogos de basebol, têm vindo a ser desenvolvidos e adaptados de forma a poderem produzir resultados equivalentes noutros desportos.

Em 2006, a revista semanal americana Time enunciou Bill James como uma das 100 pessoas mais influentes do mundo.

EXERCÍCIOS

TAREFA 1

A Divisão Oeste da MLB American League é composta por 5 equipas: Houston Astros (HOU), Los Angeles Angels (LAA), Oakland Athletics (OAK), Seattle Mariners (SEA) e Texas Rangers (TEX).

Observe a tabela abaixo, composta pelas estatísticas da Divisão Oeste da AL relativas à temporada de 2018 e resolva as questões com base nas abordagens mencionadas nesta unidade.

Tm	#Bat	BatAge	R/G	G	PA	AB	R	H	2B	3B	HR	RBI	SB
HOU	41	28.2	4.92	162	6146	5453	797	1390	278	18	205	763	71
LAA	60	29.6	4.45	162	6108	5472	721	1323	249	23	214	690	89
OAK	53	28.0	5.02	162	6255	5579	813	1407	322	20	227	778	35
SEA	53	29.8	4.18	162	6087	5513	677	1402	256	32	176	644	79
TEX	50	27.4	4.55	162	6163	5453	737	1308	266	24	194	696	74

Tm	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	LOB
HOU	26	565	1197	.255	.329	.425	.754	109	2319	156	61	14	45	19	1052
LAA	22	514	1300	.242	.313	.413	.726	100	2260	111	73	7	39	38	1071
OAK	21	550	1381	.252	.325	.439	.764	109	2450	136	76	6	44	18	1085
SEA	37	430	1221	.254	.314	.408	.722	102	2250	128	70	29	41	17	1084
TEX	35	555	1484	.240	.318	.404	.722	88	2204	104	88	33	34	16	1093

Fig. 8 – Estatísticas da Divisão Oeste da American League da edição de 2018
(Fonte: <https://www.baseball-reference.com/leagues/MLB/2018.shtml>)

- 1.1 Calcule o número aproximado de Corridas Criadas pelas cinco equipas.

- 1.2 Compare os resultados obtidos com os números da tabela. Quão grande foi o desvio?

 **TAREFA 2**

Considere o seguinte cenário:

Os Oakland Athletics acabam de ver o seu melhor Primeira Base referenciado para reforçar outra equipa. A fim de encontrar uma alternativa para a sua posição, a equipa fez uma lista dos jogadores da posição Primeira Base disponíveis no mercado que tiveram bons desempenhos na temporada de 2018. Chegaram à conclusão que os jogadores presentes na tabela abaixo encaixavam no perfil, embora só um deles pudesse ser contratado.

Name	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO
Paul Goldschmidt	30	158	690	593	95	172	35	5	33	83	7	4	90	173
Chris Davis	32	128	522	470	40	79	12	0	16	49	2	0	41	192
Joey Votto	34	145	623	503	67	143	28	2	12	67	2	0	108	101
Yuli Gurriel	34	136	573	537	70	156	33	1	13	85	5	1	23	63
Joe Mauer	35	127	543	486	64	137	27	1	6	48	0	1	51	86

Name	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	PO	A
Paul Goldschmidt	.290	.389	.533	.922	139	316	7	6	0	0	11	1323	110
Chris Davis	.168	.243	.296	.539	50	139	5	7	0	4	2	913	67
Joey Votto	.284	.417	.419	.837	125	211	15	9	0	3	6	1047	142
Yuli Gurriel	.291	.323	.428	.751	108	230	22	6	0	7	0	770	48
Joe Mauer*	.282	.351	.379	.729	99	184	9	2	1	3	5	633	61

Fig. 9 – Estatísticas dos jogadores da Divisão Oeste da Liga Americana da Divisão MLB 2018
(Fonte: <https://www.baseball-reference.com/players/>)

2.1 Os Oakland Athletics procuram um Primeira Base eficaz. De acordo com as estatísticas acima, qual deles terá maior propensão a ser mais eficaz globalmente? Descubra através da equação da Média Secundária.

2.2 Calcule o Fator de Alcance do jogador com as melhores estatísticas de Média Secundária.

INFORMAÇÕES E RECURSOS ADICIONAIS

Enredo do filme “Moneyball – Jogada de Risco” (2011)

https://www.imdb.com/title/tt1210166/?ref=mv_sr_1



Sabermetria no filme “Moneyball – Jogada de Risco” (2011)

<https://www.youtube.com/watch?v=phCVVi18aU>

Um olhar sobre as estatísticas que originaram o “Moneyball” (inglês)

<http://www.espn.com/espnw/news-commentary/article/7577771/stats-created-moneyball>

Estatísticas métricas padrão e avançadas (inglês)

<http://m.mlb.com/glossary/standard-stats>

<http://m.mlb.com/glossary/advanced-stats>

Posições do baseball

https://pt.wikipedia.org/wiki/Posi%C3%A7%C3%B5es_no_basebol

Regras do baseball

<http://www.rulesofsport.com/sports/baseball.html> (inglês)

<http://www.blogdobasebol.com/guia-do-iniciante/guia-do-iniciante-regras-do-baseball/>

Base de dados com todos os jogadores, equipas, resultados e líderes de baseball desde o início da modalidade (inglês)

<https://www.baseball-reference.com/>