

DEL IV: Film & Matematik

ÅLDER: 16 – 18

Statistics (Source: @Lukas from Pexels.com)



UPPGIFT 38: SANNOLIKHET OCH STATISTIK GENOM FILMEN "MONEYBALL"

SPEL – Sociedade Promotora de Estabelecimentos de Ensino



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

Titel: Sannolikhet och statistik genom filmen "Moneyball"

Ålder: 16 – 18 år

Längd: 2 timmar

Matematikinnehåll: Statistik, sannolikhet

Konstinnehåll: Sabermetric

Allmänna mål: Att lära studenterna begreppen sannolikhetsteori och statistik.

Instruktioner: Visa klippet från filmen Moneyball där konceptet sabermetric visas (se länken i avsnittet "Lär dig mer ...") och uppmana eleverna att se hela filmen hemma;

Resurser: Penna och miniräknare.

Tips till läraren: För att eleverna ska förstå uppgiften smidigare, lär dem några grundläggande basebollregler och -positioner i förväg.

Mål: I slutet av denna uppgift ska eleven:

- Kunna bedöma information och använda den för att lösa ekvationer som innehåller sannolikheter;
- Förstå hur statistik kan användas för att förutsäga ett resultat av en händelse.

2

Utvärdering:

Skriv 3 saker du gillar med denna uppgift:	1. 2. 3.
Skriv 2 saker du lärt dig	1. 2.
Skriv en sak som behöver bli bättre	1.

Inledning

Ibland hittar vi matematik i tv-serier eller filmer. Ofta är dessa matematiska begrepp inte så viktiga, eftersom de inte påverkar själva berättelsen. Det finns dock några fall där de gör det.

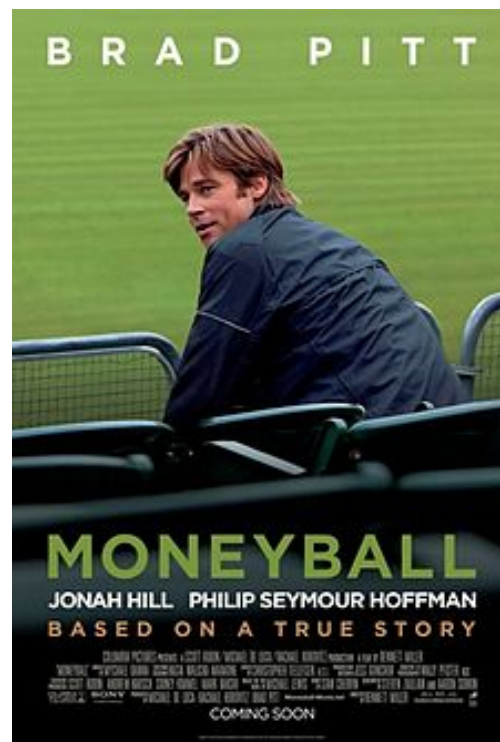
Några exempel är: "21" (USA, 2008), av Robert Luketic; "Proof" (USA, 2005), av John Madden; "A Beautiful Mind" (USA, 2001), av Ron Howard; "Enigma" (USA, 2001), av Michael Apted; "Pi" (USA, 1988), av Darren Aronofsky; "Good Will Hunting" (USA, 1997), av Gus Van Sant och "Cube" (Canada, 1997), av Vincenzo Natali.

I denna uppgift kommer filmen "Moneyball" (USA, 2011), av Bennet Miller, att diskuteras och matematiska begrepp som finns med i filmen, såsom sannolikheter och statistik, kommer att behandlas.

Moneyball

Moneyball (2011) är en amerikansk sportfilm där kassören och sportchefen för Oakland Athletics basebollag försöker sätta ihop ett konkurrenskraftigt lag för säsongen 2002 med en mycket begränsad budget.

Efter att ha förlorat tre nyckelspelare, använder sig Oaklands sportchef Billy Beane (spelat av Brad Pitt) och hans assistent Peter Brand (spelat av Jonah Hill) en udda sabermetrisk metod för att söka upp underskattade basebollspelare. Dessa spelare är underskattade och därmed förbisedda av stora lag på grund av subjektiva skäl (som ålder, utseende och personlighet), vilket gör dem överkomliga för Oakland Athletics lågbudget att investera i.



Bil 1 – Moneyball (2011) filmplansch

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

4

Beane möter tungt motstånd från de ursprungliga, äldre Oakland-agenterna, som nedvärderar denna strategi och hävdar att deras erfarenhet och kunskap inom baseball har mycket mer värde än någon statistik. Beane ignorerar deras invändningar och bildar ett lag utifrån Peter Brands sabermetriska datastatistik.

Ordlista

Sabermetrics – En term som skapades av baseballstatistikern och författaren Bill James. Härstammar från Society for American Baseball Research. Det är en metod som samlar in och analyserar relevant baseballstatistik i spelet för att utvärdera spelarnas och lagens prestation i alla aspekter av spelet, till exempel:

Standard Batting/Fielding – I baseball turas lagen om att slå och att vara utelag.

→ **Batting** - att slå bollen när den kastas av en motståndarlagets pitcher.

Spelaren som har denna position kallas slagman, hitter på engelska (H)

→ **Fielding** - positioner i ett

lag när de sprids ut på planen.

Det finns 9 positioner: Pitcher (P) och fångare (C), som har fasta positioner och den första basmannen (1B), andra basmannen (2B), tredje basmannen (3B), shortstop (SS), vänster fielder (LF), center fielder (CF) och höger fielder (RF), som kan röra sig fritt.



Bild 2 – Baseball-positioner (Källa: Michael J, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=40095322>)

Standard/Avancerad statistik – Statistik från spelarens/lagets prestationer i olika sporter. Nedan hittar du baseballstatistikbegreppen som används i denna uppgift:

→ **Assist (A)** – När en defensiv spelare vidrör bollen innan en putout görs av en annan fielder;

→ **At Bats (AB)** – Inträffar när en slagman måste gå till första basen på grund av ett eget misstag (t ex **BB**);



→ **Base on Ball (BB)** – Kallas även Walks. Händer när en slagman missar fyra försök som domaren godkänt som "Ball";

→ **Caught Stealing (CS)** – En förseelse som inträffar när en baserunner försöker gå framåt från en bas till den andra innan bollen slås och sedan fångas av en fielder medan han gör detta försök;

→ **Games played (G)** – Totalt antal matcher som en spelare deltagit i. En spelare räknas ha spelat om han förekommer när som helst i en match. Om en spelare har 162 G betyder det att han spelade i alla matcher under säsongen;

→ **Hit (H)** – När en hitter (även känd som slagman) slår bollen in i godkänt område och når basen. Hittar kan vara singlar, dubbel, tripplar och homeruns;

→ **Plate Appearances (PA)** – Händer varje gång en spelare genomför en omgång genom att få en hit, Walk, dömas ut eller ta sig till bas på grund av ett eget misstag;

→ **Putout (PO)** – Att besegra motståndaren, antingen genom att trampa på basen, fånga en runner, fånga en utslagen boll eller fånga en tredje strike;

→ **Runda (R)** – När en runner tar sig fram till plattan ;

→ **Stolen Bases (SB)** – Händer när en baserunner framgångsrikt når nästa bas när pitchern kastar;

→ **Total Bases (TB)** – Totala antalet baser en spelare når med en hit.

Matematiken bakom Moneyball

I filmen Moneyball finns det ett tillfälle där Peter Brand studerar lagets prognoser för säsongen 2002 och säger att för att laget ska komma till slutspelet måste det vinna minst 99 av 162 matcher. För att nå detta antal matcher projicerar han det minsta antalet rundor som måste vinnas och det maximala antal rundor som kan tillåtas av motståndare. För att komma fram till dessa resultat använder han en av ekvationerna som ursprungligen utvecklats av sportstatistikern Bill James, känd som "Pythagorisk vinstprocent", vilket resulterar i ett lags ungefärliga vinstförhållande baserat på vunna och förlorade runda. Ekvationen ser ut enligt följande:

$$\text{Vinstförhållane} = \frac{\text{Vunna runs}^2}{\text{Vunna runs}^2 + \text{Förlorade runs}^2}$$

Säsongen 2002 tror Peter att laget ska vinna minst 814 rundor och förlora högst 645, vilket resulterar i följande:

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

7

Vinstförhållandet anges i procent och, multiplicerat med antalet matcher på en basebollsäsong (162), resulterar det i ett ungefärligt antal matcher som laget måste vinna för att komma till slutspelet.

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

Peter visar sedan en databas som han sammanställt med information om enskilda spelare och deras genomsnittliga spelstatistik (fig. 3), där de kommer att träna för att hitta de mest effektiva spelarna.

	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%

Bild 3 – Databas medSabermetric statistik (Källa: filmen "Moneyball")

Trots att de endast hade 1/3 av lönekostnaden jämfört med storlag som New York Yankees, som vann American League East division, nådde The Oakland A med 103 segrar under ordinarie säsong, rekordet med 20 segrar i rad i American League och vann sin division (American League West). Trots att de åkte ut i slutspelet förändrade denna sabermetrisk strategi en hel bransch för alltid genom användningen av matematik och statistik.

Vem är Bill James?

George William James (född 1949), är en amerikansk basebollförfattare, historiker och statistiker mest känd för att införa metoden med sabermetrikstatistik.

Förutom den tidigare nämnda "Pythagorisk vinstprocenten" är andra statistiska innovationer som Bill James introducerade Runs Created (RC), Range Factor (RF) och Secondary Average (SecA):



Bild 4 – Bill James, 2010

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

Runs Created: en statistik som uppskattar den offensiva insatsen från ett lag/spelare jämfört med de poäng som görs i matchen. Denna metod kan också användas som ett sätt för att uppnå ett ungefärligt antal rundor som ett lag kommer att göra när det spelar. Formeln lyder:

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

Där:

TB = Total Bases;

H = Hits;

BB = Base on Balls/Walks;

PA = Plate Appearances.

Studera följande MLB-statistik 2018 från Detroit Tigers (DET) och Oakland Athletics (OAK) (från 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

9

Bild 5 – Statistik från Detroit Tigers och Oakland Athletics från säsongen 2018
(Källa: <https://www.baseball-reference.com/leagues/MLB/2018.shtml>)

Räkna ut Runs Created av OAK:

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

Enligt beräkningarna borde Oakland Athletics ha gjort cirka 767 runs. I själva verket gjorde Oakland Athletics faktiskt 813 runs.

Om 813 rundor motsvarar 100% motsvarar 767 rundor 94,34%, vilket innebär att det finns en minsta avvikelse på 5,6%.

Vi gör samma för DET:

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

Baserat på den statistik som tillhandahålls borde det ha gjorts cirka 607 rundor av laget under den senaste 18-säsongen. Detroit Tigers avslutade säsongen med 630 rundor. Återigen är det bara en minimal avvikelse (3,6%). Denna beräkning kan också tillämpas på enskilda spelare och är användbar för att kontrollera hur väl en slagman har utfört sitt jobb: antalet rundor.

Secondary Average: en förbättrad version av Ekvationen för att räkna ut Slaggenomsnitt. Även om det fortfarande fungerar enligt principen för att räkna ut slaggenomsnitt, täcker Secondary Average också spelarens kraft (extra baser), sikte (Walks) och löphastighet (stulna baser). Dess formel försöker mäta en övergripande offensiv effektivitet hos en spelare/lag och representeras enligt följande:

10

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

Där:

BB = Base on Balls/Walks;

TB = Total Base;

H = Hits;

SB = Stolen Bases;

CS = Caught Stealing;

AB = At Bats.

Här är Standard Batting-statistik från fångarna James McCann, från DET och Jonathan Lucroy, från OAK.

Name	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB
James McCann	118	457	427	31	94	16	0	8	39	0	3	26
Jonathan Lucroy	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
James McCann	116	.220	.267	.314	.581	58	134	9	2	0	2	0
Jonathan Lucroy	65	.241	.291	.325	.617	71	135	12	3	1	6	1

Fig. 6 – Statistics from the 2018 MLB season

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

Om vi använder ekvationen för Secondary average, får vi:

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

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

11

Resultatet avrundat till tusendelar representerar spelarens sekundära medelvärde (Secondary Average). I det här fallet har James McCann en bättre övergripande effektivitet, vilket teoretiskt innebär att han på lång sikt är mer effektiv offensivt.

Range Factor: en statistik som beräknar en spelares bidrag vid en viss defensiv position. Ekvationen är som följer:

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

Där:

A = Assist;

PO = Putouts;

G = Games played.

Studera Standard Fielding-statistik från samma spelare:

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	IgFld%	IgRF9	IgRFG	PB	WP	SB	CS	CS%	IgCS%	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

Bild 7 – Statistik från 2018

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

Samma spelare har följande resultat på en viss defensiv position:

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

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

James McCanns Range Factor är högre än Jonathan Lucroys. Med andra ord är James McCann bättre på defensivt spel.

12

Liksom i varje faktoranalys är det viktigt att förstå att ju större urval/data som används, desto mer exakta och relevanta blir resultaten.

Många andra formler utvecklades av Bill James som användes i andra sammanhang och till annan statistik; med tiden förfinades några av dem och andra gjordes om av andra statistiker. Även om dessa ursprungligen utformades för baseboll, har de sedan dess utvecklats och anpassats för att kunna ge motsvarande resultat i andra sporter.

2006 nominerade den amerikanska veckotidskriften Time Magazine Bill James till utmärkelsen en av de 100 mest inflytelserika i världen..

UPPGIFTER



UPPGIFT 1

American League West i MLB består av 5 lag: Houston Astros (HOU), Los Angeles Angels (LAA), Oakland Athletics (OAK), Seattle Mariners (SEA) och Texas Rangers (TEX).

Studera tabellen från ALW-divisionen från säsong '18 och svara på frågorna med hjälp av sabermetrisk metod som nämnts tidigare i denna text.

American League West Division '18

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

- 1.1 Beräkna antalet Runs Created av samtliga fem lag.
- 1.2 Jämför resultaten som du får med siffrorna från tabellen. Hur stor är avvikelsen?



UPPGIFT 2

Föreställ dig detta:

Oakland Athletics har precis sålt sin bästa First Base-spelare till ett annat lag. För att ersätta honom har de sökt efter tillgängliga First Base-spelare på marknaden som har presterat bra under säsongen 2018. De har kommit fram till att alla spelarna i tabellen nedan är tänkbara, men kan bara anställa en av dem.

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

Bild 9 – Spelarstatistik från American League West Division 2018

(Källa: <https://www.baseball-reference.com/players/>)

2.1 Oakland Athletics vill ha en effektiv First base. Enligt statistiken ovan, vilken är troligen mest effektiv? Ta reda på det med sekundär medels-ekvationen (Secondary Average equation).

2.2 Beräkna Range Factor för spelaren med bäst Secondary Average statistik.

LÄR DIG MER...

Moneyball (2011) filmhandling

https://www.imdb.com/title/tt1210166/?ref=fnv_sr_1

Sabermetrics i Moneyball

<https://www.youtube.com/watch?v=KWPhV6PUr9o>

Statistiken som skapade 'Moneyball'

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

Standard metric statistik

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

Avancerad metrisk statistik

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

Baseboll-positioner

https://en.wikipedia.org/wiki/Baseball_positions

Baseboll-regler

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

Databaser med spelare, lag, resultat och vinnare.

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