

PARTIE IV : Cinématographie & Mathématiques

ÂGE : 16 – 18 ans

Statistics (Source: @Lukas from Pexels.com)



OUTIL 38 : PROBABILITÉ ET STATISTIQUES DANS LE FILM "LE STRATÈGE"

SPEL – Sociedade Promotora de Estabelecimentos de Ensino



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

Guide de l'éducateur

Titre : Probabilités et statistiques dans le film "Le Stratège"

Âge : 16 – 18 ans

Durée : 2 heures

Concepts Mathématiques : Statistiques, Probabilités

Concepts Artistiques : Sabermétrie

Objectifs Généraux : Présenter aux étudiants la théorie des probabilités et des statistiques.

Instructions et Méthodologies : Montrez l'extrait du film Moneyball dans lequel le concept de sabermétrie est présenté (cf. lien de la section "Pour en savoir plus...") et suggérez aux élèves de regarder le film complet à la maison ;

Ressources : Un stylo et une calculatrice.

Conseils pour l'éducateur : Pour une meilleure compréhension, faites-leur comprendre au préalable quelques règles et positions de base du base-ball.

Compétences et Résultats ciblés : À la fin de cet outil, l'élève sera capable de :

- Évaluer les informations et les utiliser pour résoudre des équations impliquant des probabilités ;
- Comprendre comment les statistiques peuvent être utilisées pour prédire le résultat d'un événement.

Compte-rendu et Évaluation :

Écrivez trois aspects que vous avez aimé dans cette activité	1. 2. 3.
Écrivez deux choses que vous avez apprises :	1. 2.
Écrivez un aspect à améliorer	1.

Introduction

On trouve parfois des aspects liés aux mathématiques dans des séries télévisées ou des films. Dans ces cas-là, il arrive que ces concepts mathématiques n'aient pas beaucoup d'importance, car ils n'influencent pas l'histoire elle-même. Cependant, il y a quelques films dans lesquels c'est le cas.

En voici quelques exemples : "Las Vegas 21" (USA, 2008), de Robert Luketic ; "Proof" (USA, 2005), de John Madden ; "Un homme d'exception" (USA, 2001), de Ron Howard ; "Enigma" (USA, 2001), de Michael Apted ; "Pi" (USA, 1988), de Darren Aronofsky ; "Will Hunting" (USA, 1997), de Gus Van Sant et "Cube" (Canada, 1997), de Vincenzo Natali.

Dans cet outil, le film "Le Stratège" (USA, 2011), de Bennet Miller, sera abordé et ses concepts mathématiques, tels que les probabilités et les statistiques, seront traités.

Le Stratège

Le Stratège (2011) est un film sportif américain dans lequel un responsable et directeur général de l'équipe de baseball Oakland Athletics tente de monter une équipe compétitive pour la saison 2002 avec un budget très limité.

In the film, after losing 3 key players, Oakland's general manager Billy Beane (played by Brad Pitt) and his assistant Peter Brand (played by Jonah Hill) resort to an unorthodox sabermetric approach in order to scout underrated baseball players. Underestimated due to biased reasons (such as age, appearance and personality), these players are overlooked by big teams, which makes them affordable for the low-budget Oakland Athletics to invest on.

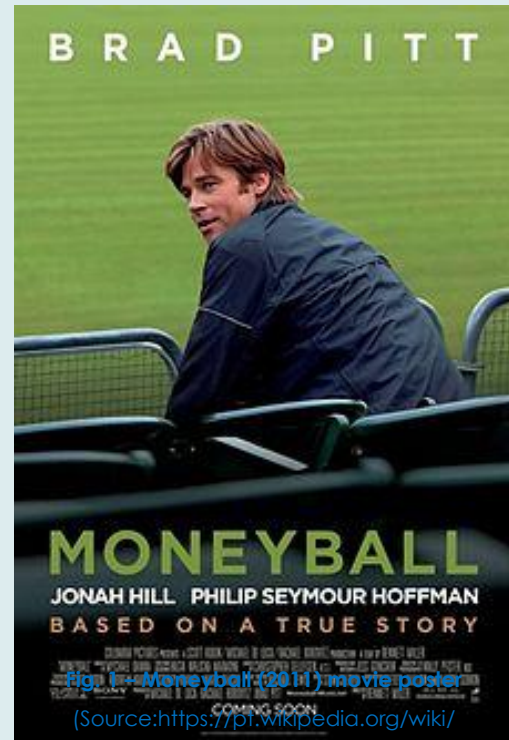


Fig. 1- Moneyball (2011) movie poster
 (Source: <https://pt.wikipedia.org/wiki/Moneyball>)

Facing heavy resistance by the original, old-fashioned Oakland scouts, who lessen this approach arguing that their experience and knowledge in baseball has far more value than any statistics, Beane ignores their objections and forms a team following Peter Brand's sabermetric data statistics.

Glossaire

Sabermétrie – Un terme inventé par le statisticien de base-ball et écrivain Bill James. Dérivé de la Society for American Baseball Research. Il s'agit d'une méthode qui permet de recueillir et d'analyser des statistiques pertinentes sur le base-ball en cours de jeu afin d'évaluer les performances des joueurs et des équipes dans tous les aspects du jeu, tels que :

Frappe et champs – Au baseball, deux équipes opposées se relaient à la frappe et au champ.

→ **Frappe** - l'acte de frapper la balle lorsqu'elle est lancée par le lanceur de l'adversaire. Le joueur qui occupe cette position est connu sous le nom de " batteur "

→ **Champ** - les positions dans lesquelles un L'équipe est répartie autour d'un terrain de base-ball.

Il y a 9 positions : le lanceur et le receveur, qui occupe une place fixe et le joueur de première base, joueur de deuxième base , joueur de troisième base, arrêt-court, joueur de champ gauche, joueur de champ central et joueur de champ droit, qui peuvent se déplacer librement.

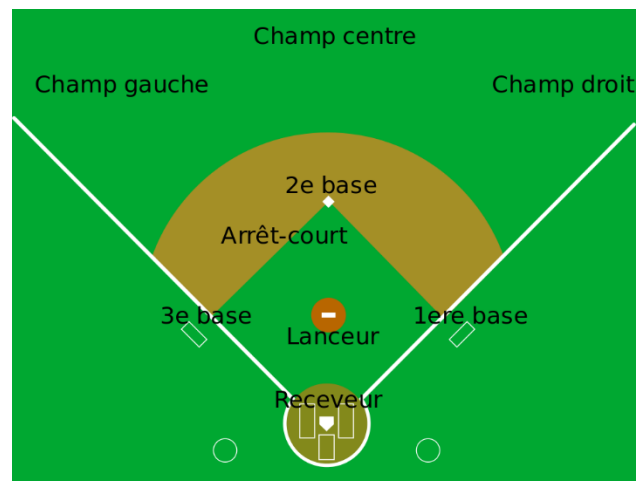


Fig. 2 – Positions de Baseball (Source : par Ziit ; CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:Baseball_positions_fr_ench_svg.svg?uselang=fr)

Statistiques standard/avancées – Statistiques sur les performances des joueurs/équipes dans n'importe quel sport. Voici une liste des statistiques de baseball abordées dans cet outil :

→ **Assistance (A)** – Se produit lorsqu'un joueur défensif touche le ballon avant qu'un coup roulé ne soit enregistré par un autre joueur de champ ;



- **Présence officielle au bâton** – Se produit lorsqu'un batteur atteint la base par choix, frappe ou erreur du joueur de champ ;

- **But sur balles** – Se produit lorsqu'un frappeur reçoit quatre lancers jugés hors de la zone des prises, donc étant déclarées balles par l'arbitre ;

- **Retrait sur tentative de vol** – Faute qui se produit lorsqu'un coureur de base tente d'avancer d'une base à l'autre avant que la balle ne soit frappée et est ensuite touché par un joueur de champ lors de cette tentative ;

- **Matches joués** – Total des matchs joués par un seul joueur. Un joueur est crédité d'un match joué s'il apparaît à un moment quelconque dans un match. Si un joueur a 162 matchs, cela signifie qu'il a joué pendant toute la saison;

- **Frappe** – Lorsqu'un batteur frappe la balle dans un territoire de jeu et atteint la base. Les frappes peuvent être des simples, des doubles, des triples et des Coup de circuit ;

- **Passage au bâton** – Se produit à chaque fois qu'un joueur termine son tour de batte avec un coup, qu'il marche, qu'il sort ou qu'il atteint la base suite à une erreur ;

- **Retrait** – L'acte de compléter physiquement un retrait, que ce soit en marchant sur la base, en touchant un coureur, en attrapant une balle frappée ou en attrapant une troisième prise ;

- **Point** – Se produit lorsqu'un joueur franchit le marbre (le quatrième but) ;

- **But volé** – Se produit lorsqu'un coureur réussit à atteindre la base suivante alors que le lanceur effectue un lancer ;

- **Total de buts** – Le nombre total de buts qu'un joueur a marqué en un coup.

Les Maths dans “Le Stratège”

Dans le film Le Stratège, à un certain moment, en faisant les pronostics de l'équipe pour la saison 2002, Peter Brand déclare que pour que l'équipe se rende aux éliminatoires, elle doit gagner au moins 99 des 162 matchs. Pour atteindre cette valeur, il projette le nombre minimum de points à marquer et le nombre maximum de points qui peuvent être concédés.

Pour obtenir ces résultats, il utilise une des équations développées à l'origine par le statisticien sportif Bill James, connue sous le nom d'"Estimation pythagoricienne", qui donne le taux de victoire approximatif d'une équipe en fonction des points marqués et des points concédés. L'équation se présente comme suit :

$$\% \text{ de Victoire} = \frac{\text{Points marqués}^2}{\text{Points marqués}^2 + \text{Points concédés}^2}$$

Pour la saison de 2002, Peter prévoit que l'équipe devrait marquer au moins 814 points et en concéder au maximum 645, ce qui donne les résultats suivants :

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

Le taux de victoire est indiqué en pourcentage et, multiplié par le nombre de matchs d'une saison de base-ball (162), donne un nombre approximatif de matchs que l'équipe devrait gagner pour se rendre aux éliminatoires.

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

Peter montre ensuite une base de données qu'il a compilée avec des informations sur les joueurs individuels sur leurs statistiques moyennes de jeu (Fig. 3), avec laquelle ils vont travailler afin de trouver les joueurs les plus rentables.

	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 – Base de données des statistiques sabermétriques des joueurs de base-ball présentée par Peter (Source : Film "Le Stratège")

Bien qu'ils n'aient qu'un tiers du salaire des équipes des grands marchés comme les New York Yankees, champions de la division Est de la Ligue américaine, les Oakland A ont égalé le record de 103 victoires en saison régulière, ont remporté 20 victoires consécutives dans la Ligue américaine et ont été champions dans leur division (Ligue américaine Ouest). Même s'ils ont été éliminés dans les séries éliminatoires, cette approche sabermétrique a changé à jamais toute une industrie en utilisant les mathématiques et les statistiques.

Qui est Bill James?

George William James (né en 1949), est un écrivain, historien et statisticien américain spécialisé dans le baseball, surtout connu pour avoir introduit la méthode statistique sabermétrique.

Outre l'"Estimation pythagoricienne" déjà mentionnée, Bill James a introduit d'autres innovations statistiques, telles que les "Runs Created" (RC), le "Range Factor" (RF) et la "Second Average" (SecA) :



Fig. 4 – Bill James, en 2010
(Source : https://en.wikipedia.org/wiki/Bill_James)

« **Runs Created** » : une statistique qui estime la contribution offensive d'une équipe/joueur aux points marqués dans le jeu. Cette méthode peut également être utilisée pour obtenir un nombre approximatif de points qu'une équipe marquera lorsqu'elle est au bâton. La formule est la suivante :

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

Où :

TB = Total des buts ;

H = Frappes ;

BB = Buts sur balle ;

PA = Passages au bâton.

Voici les statistiques du MLB 2018 des Detroit Tigers (DET) et d'Oakland Athletics (OAK) (de 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 – Statistiques des Tigres de Detroit et de l'Oakland Athletics pour la saison 2018 de la MLB
(Source: <https://www.baseball-reference.com/leagues/MLB/2018.shtml>)

Calcul des points créés (RC) par les OAKs :

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

Selon les calculs, l'OAK aurait dû créer environ 767 points. En réalité, l'OAK a marqué 813 points.

Si 813 runs correspondent à 100%, alors 767 points correspondent à 94,34%, ce qui signifie qu'il y a un écart minimal de 5,6%..

Faisons de même pour les DET :

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

D'après les statistiques fournies, l'équipe aurait dû marquer environ 607 point au cours de la saison de 2018. Les Detroit Tigers ont terminé la saison avec 630 points. Une fois de plus, le pourcentage d'écart est minime (3,6 %).

Ce calcul peut également être appliqué à des joueurs individuels et est utile pour vérifier si un batteur a bien fait son travail : marquer des points.

« **Secondary Average** » : une version améliorée de l'équation de la moyenne au bâton. Tout en continuant à fonctionner selon les principes de la moyenne au bâton, la moyenne secondaire couvre également la puissance (bases supplémentaires), l'oeil (buts sur balle) et la vitesse (buts volés) d'un joueur. Sa formule tente de mesurer l'efficacité offensive globale d'un joueur/équipe et est représentée comme suit :

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

Où :

BB = Buts sur Balle ;

TB = Total de butts ;

H = Frappes ;

SB = Buts volés ;

CS = Retraits sur tentative de vol ;

AB = Présence officielle au bâton.

Observe les statistiques Standard de frappe des receveurs James McCann, du DET, et Jonathan Lucroy, des OAKs.

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 – Statistiques de la saison 2018 du MLB

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

En utilisant l'équation de la moyenne secondaire, nous avons :

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

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

11

Le nombre résultant, arrondi au millième, représente la moyenne secondaire d'un joueur. Dans ce cas, James McCann a une meilleure efficacité globale, ce qui, théoriquement, signifie qu'à long terme, il est plus efficace sur le plan offensif.

« **Range Factor** » : une statistique qui quantifie la contribution d'un joueur à une position défensive donnée. L'équation est la suivante:

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

Où :

A =Assistances ;

PO = Retraits ;

G = Matchs joués.

Examine les statistiques des deux mêmes joueurs lorsqu'ils jouent dans le champ :

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

Fig. 7 – Statistiques de la saison 2018 du MLB

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

Les mêmes joueurs ont produit les résultats suivants lorsqu'ils étaient en position dans le champ :

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

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

Le "Range Factor" de James McCann est plus élevé que celui de Jonathan Lucroy. En d'autres termes, James McCann a un jeu défensif nettement plus efficace.

Comme dans toute analyse factorielle, il est important de comprendre que plus la taille de l'échantillon/les données utilisées sont importantes, plus les résultats seront précis et justes.

Beaucoup d'autres formules ont été développées par Bill James et représentent de nombreuses autres statistiques standard et avancées ; au fil du temps, certaines d'entre elles ont été affinées et d'autres ont été créées par différents statisticiens. Alors que celles-ci ont été initialement conçues pour les matchs de baseball, elles ont depuis été développées et adaptées afin de pouvoir produire des résultats équivalents dans d'autres sports.

En 2006, l'hebdomadaire américain Time a désigné Bill James comme l'une des 100 personnes les plus influentes au monde.

TÂCHES



TÂCHE 1

La division Ouest de la Ligue américaine du MLF est composée de 5 équipes : Houston Astros (HOU), Los Angeles Angels (LAA), Oakland Athletics (OAK), Seattle Mariners (SEA) et les Texas Rangers (TEX).

Observe le tableau ci-dessous avec la division de la saison de 2018 de l'ALW et résous les problèmes en utilisant les méthodes sabermétriques mentionnées dans cet outil.

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 – Statistiques de la Division Ouest de la Ligue américaine de la MLB 2018
(Source: <https://www.baseball-reference.com/leagues/MLB/2018.shtml>)

1.1 1.1 Calcule le nombre approximatif de "Runs Created" (RC) par les 5 équipes.

1.2 Compare les résultats obtenus avec les chiffres du tableau. Quelle a été l'importance de cet écart ?

 **TÂCHE 2**

Considérons le scénario suivant :

L'Oakland Athletics vient d'envoyer son meilleur joueur de première base dans une autre équipe. Afin de remplacer son poste sur le terrain, ils ont cherché sur le marché les joueurs de première base disponibles qui avaient réalisé de bonnes performances lors de la saison de 2018. Ils ont conclu que les joueurs du tableau ci-dessous sont aptes à occuper le poste, mais ils ne peuvent en engager qu'un.

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 – Statistiques des joueurs de la division Ouest de la Ligue américaine de la MLB 2018
(Source : <https://www.baseball-reference.com/players/>)

2.1 L'Oakland Athletics veut un joueur de première base efficace. D'après les statistiques ci-dessus, lequel est susceptible d'être le plus efficace ? Découvrez-le à l'aide de l'équation de la moyenne secondaire.

2.2 Calcule le "Range Factor" du joueur ayant les meilleures statistiques de moyenne secondaire.

POUR EN SAVOIR PLUS...

L'intrigue du film Le Stratège (2011)

https://fr.wikipedia.org/wiki/Le_Strat%C3%A8ge

Sabermétrie dans le film "Le Stratège" (en anglais) :

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

Les statistiques qui ont créé "Le Stratège" (en anglais) :

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

Statistiques métriques standard (en anglais) :

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

Statistiques métriques avancées (en anglais) :

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

Positions au baseball

https://fr.wikipedia.org/wiki/Positions_au_baseball

Règles du baseball

<http://www.baseball-essentiel.fr/regles-baseball.htm>

Base de données avec les joueurs de baseball, les équipes, les scores et les leaders de tous les temps (en anglais) :

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