

Selecting the Best Hitter in a Professional Baseball League Using the Technique for Order of Preference by Similarity to Ideal Solution

Selecting Best Hitter in Pro-Baseball League

Chih-Cheng Chen^{1,*}, Yen-Whea Kuo², Chih-Jen Chuang¹, Chun-Hsiao Chu³, Tian-Shaing Kuo⁴ and Kuang-Tsan Hung⁵

¹Department of Sport Management, Aletheia University, New Taipei City, Taiwan

²Center for Physical Education, Aletheia University, New Taipei City, Taiwan

³Department & Graduate Institute of Tourism, Aletheia University, New Taipei City, Taiwan

⁴Bachelors Degree Program of Dynamic Recreation and Health Management, Tainan

⁵Department of Physical Education, De Lin Institute of Technology, New Taipei City, Taiwan

*Corresponding author

Abstract—Selecting the best hitter in a professional baseball league is considered a multi-attribute decision-making problem. We developed an evaluation model based on literature review and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to select the best hitter in the Chinese Professional Baseball League. We determined attribute weights and used the TOPSIS to calculate the individual performance ranking. We conducted an empirical analysis to demonstrate the applicability of the proposed model for addressing the problem associated with hitter performance. The results demonstrate the effectiveness and feasibility of the proposed model.

Keywords—hitter; Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS); Chinese Professional Baseball League; plate appearances (PAs); on-base percentage (OBP); slugging percentage (SLG)

I. INTRODUCTION

A baseball game involves two teams: offense and defense. The offense attempts to score by using a bat to hit a ball thrown by the pitcher, thus putting the ball in play, and the hitter scores a run by moving counter-clockwise around a series of four bases: first, second, third, and, finally, home plate. The defense attempts to prevent the scoring of runs by getting the hitter out. To win a baseball game, a team must score more runs than the opponent. Pitchers, whose aim is to prevent the offense from hitting and subsequently scoring, are particularly crucial in a baseball game [1-2]. Therefore, pitchers and hitters are equally valuable to a baseball team. Each year, as the playing season progresses, baseball fans and sportswriters worldwide take great interest in discussing which pitcher or hitter is the best in a professional baseball league. In addition, selecting a pitcher or hitter is a decision-making problem for a complex system that includes many quantitative attributes [1]. Most baseball-related decision-making studies have focused on selecting pitchers for particular themes [1, 3-4], and rarely on selecting hitters. Thus, in this study, we focused on selecting hitters. As mentioned, in a professional baseball game, scoring runs is the

main objective of the batting team, and this is accomplished through a combination of skills, particularly the skills of hitting the ball and avoiding getting out [5-6]. However, in the world of sports, this message often appears lost. The best baseball players are often valued if they demonstrate the five attributes of hitting for average, hitting for power, running, throwing, and fielding. In other words, a nonspecialist hitter is preferred in baseball [7]. Thus, selecting the best baseball hitter is a type of multi-attribute decision-making (MADM) problem. In this study, we applied methods used in previous studies on professional baseball [5-6] to determine the weight of each attribute. Specifically, we used the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), which is a multi-criteria decision analysis method for calculating the distance of each alternative from positive and negative ideal solutions, instead of ranking the alternatives [3]. The remainder of this article is organized as follows. Section 2 presents the methodology for evaluation. Section 3 focuses on empirical analysis for determining the best hitter in the Chinese Professional Baseball League (CPBL). Finally, Section 4 draws conclusions and offers recommendations for future study.

II. MATERIAL AND METHODS

A. TOPSIS Method

The TOPSIS, which was initially proposed by Hwang and Yoon, is based on the concept of distance measures [3, 8]. An ideal solution (a positive ideal solution) maximizes the benefit criteria or attributes and minimizes the cost criteria or attributes. Conversely, an antiideal ideal solution (a negative solution) maximizes the cost criteria or attributes and minimizes the benefit criteria or attributes [3, 8].

The TOPSIS is explained as follows [8]:

Step1. Obtain performance data for m alternatives over n attributes. Raw measurements are generally standardized by converting raw measures X_{ij} into standardized measures S_{ij} .

Step2. Develop a set of importance weights W_k for each criterion. These weights are typically ad hoc reflective of relative importance. Scale is not a concern if the weights are standardized in Step 1.

Step3. Identify an ideal alternative (extreme performance in each criterion) S^+ .

Step4. Identify a nadir alternative (reverse extreme performance in each criterion) S^- .

Step5. Develop a distance measure over each criterion to include both ideal (D^+) and nadir (D^-).

Step6. For each alternative, determine a ratio R equal to the distance to the nadir divided by the sum of the distance to the nadir and the distance to the ideal as follows:

$$R = D^- / (D^+ + D^-).$$

Step7. Rank order alternatives by maximizing R obtained in Step 6.

Thus, the TOPSIS minimizes the distance to the ideal alternative while maximizing the distance to the nadir. Several procedures can be used for Step 2 (developing weights) and Step 5 (distance measures). Additionally, different conventions can be used for defining the best performance (Step 3) and the worst performance (Step 4).

B. Data

We obtained data from the official CPBL website (<http://www.cpbl.com.tw>), which collected and posted records of every CPBL game in 2013. We selected alternative players for the teams in 2013 from the website. Every alternative player chosen had at least 372 plate appearances (PAs) in the 2013 season. The commonly cited statistics for hitters found in the empirical analysis were on-base percentage (OBP) and slugging percentage (SLG) [5-6], both of which significantly contribute to the team winning percentage [5]. We included these attributes and performed calculation by using the TOPSIS, in which parameters used as attributes are familiar to many baseball fans.

III. RESULTS AND DISCUSSION

A. Information on the Performance of Alternative Hitters

To safeguard CPBL player interests, we did not use the actual player names in the paper. A brief description and the code names of 22 CPBL hitters is provided in Table 1. In the 2013 season, all alternative hitters had an OBP > .288 (Table 1). Player A3 had the best OBP (.419). There was a tie between Players A4 and A5 for second place. The SLG of all of the hitters was > .319. Player A18 was the best performer; Players A4 and A5 ranked second and third, respectively. However, based on these results, we could not clearly determine which player was the best hitter in the 2013 season.

TABLE I. ALTERNATIVE HITTERS' PERFORMANCE INFORMATION

Player	G	PAs	OBP	SLG
A1	119	505(1)	0.329(15)	0.417(9)
A2	117	486(2)	0.298(21)	0.319(22)
A3	116	477(3)	0.419(1)	0.454(6)
A4	113	467(4)	0.415(2)	0.549(2)
A5	112	460(5)	0.415(2)	0.498(3)
A6	113	449(6)	0.308(18)	0.330(20)
A7	112	447(7)	0.331(13)	0.383(10)
A8	105	445(8)	0.373(6)	0.451(7)
A9	103	438(9)	0.336(12)	0.369(12)
A10	113	436(10)	0.306(19)	0.342(16)
A11	100	436(10)	0.322(16)	0.364(13)
A12	98	434(12)	0.351(10)	0.361(15)
A13	101	433(13)	0.376(5)	0.439(8)
A14	99	414(14)	0.368(7)	0.466(5)
A15	114	412(15)	0.288(22)	0.374(11)
A16	114	407(16)	0.310(17)	0.331(19)
A17	101	401(17)	0.331(14)	0.323(21)
A18	102	398(18)	0.411(4)	0.566(1)
A19	106	398(18)	0.357(8)	0.478(4)
A20	99	388(20)	0.352(9)	0.362(14)
A21	111	384(21)	0.303(20)	0.338(17)
A22	101	382(22)	0.347(11)	0.338(17)

B. Using the TOPSIS to determine the best hitter

Based on the data in Table I, we performed the TOPSIS as follows. We used Step 1 to calculate the normalized decision matrix depending on whether the objective of the selection attribute is minimization or maximization. Table II shows the normalized decision matrices of each hitter. The objective of all attributes was maximization.

TABLE II. NORMALIZED DECISION MATRIXES

Player	OBP	SLG
A1	0.043	0.047
A2	0.039	0.036
A3	0.055	0.051
A4	0.054	0.062
A5	0.054	0.056
A6	0.040	0.037
A7	0.043	0.043
A8	0.049	0.051
A9	0.044	0.042
A10	0.040	0.039
A11	0.042	0.041
A12	0.046	0.041
A13	0.049	0.050
A14	0.048	0.053
A15	0.038	0.042
A16	0.041	0.037
A17	0.043	0.036
A18	0.054	0.064
A19	0.047	0.054
A20	0.046	0.041
A21	0.040	0.038
A22	0.045	0.038

We reviewed previous studies [5-6] to determine the weights of OBP and SLG. The coefficient of OBP was approximately 2.6 times more than that of SLG [5-6]. Thus, we calculated the OBP and SLG coefficients as .722 and .278, respectively. We calculated the weighted normalized decision

matrix by using Step 2. The weighted normalized decision matrix for each selection criterion is shown in Table III.

TABLE III. WEIGHTED NORMALIZED DECISION MATRIXES

Player	OBP	SLG
A1	0.031	0.013
A2	0.028	0.010
A3	0.040	0.014
A4	0.039	0.017
A5	0.039	0.016
A6	0.029	0.010
A7	0.031	0.012
A8	0.035	0.014
A9	0.032	0.012
A10	0.029	0.011
A11	0.030	0.011
A12	0.033	0.011
A13	0.036	0.014
A14	0.035	0.015
A15	0.027	0.012
A16	0.029	0.010
A17	0.031	0.010
A18	0.039	0.018
A19	0.034	0.015
A20	0.033	0.011
A21	0.029	0.011
A22	0.033	0.011
PIS	0.040	0.018
NIS	0.027	0.010

We calculated the positive (S^+) and negative (S^-) ideal solutions by using Steps 3 and 4, respectively; the values are shown in Table 3. We calculated the distance of each alternative by using Step 5; the values are listed in Table 4. We calculated the closeness coefficient CC_i by using Step 6. The all-river tracing route' CC_i value and its rank are shown in Table IV.

TABLE IV. POSITIVE AND NEGATIVE IDEAL SOLUTIONS AND DISTANCE FOR EACH ALTERNATIVE, CLOSENESS COEFFICIENT AND RANK

Player	d^+	d^-	CC_i	Rank
A1	4.71E-05	1.22E-05	0.206	12
A2	9.54E-05	4.46E-07	0.005	22
A3	6.18E-06	8.55E-05	0.933	4
A4	2.14E-07	9.79E-05	0.998	1
A5	2.35E-06	8.77E-05	0.974	3
A6	8.24E-05	1.84E-06	0.022	18
A7	5.10E-05	1.03E-05	0.167	14
A8	1.60E-05	4.08E-05	0.719	6
A9	4.98E-05	1.15E-05	0.188	13
A10	8.17E-05	1.70E-06	0.020	19
A11	6.21E-05	6.15E-06	0.090	16
A12	4.13E-05	1.85E-05	0.310	10
A13	1.62E-05	4.16E-05	0.720	5
A14	1.65E-05	3.91E-05	0.703	7
A15	9.47E-05	1.49E-06	0.015	20
A16	8.02E-05	2.23E-06	0.027	17
A17	6.36E-05	8.26E-06	0.115	15
A18	2.86E-07	9.75E-05	0.997	2
A19	2.10E-05	3.37E-05	0.616	8

A20	4.05E-05	1.92E-05	0.321	9
A21	8.56E-05	1.18E-06	0.014	21
A22	4.87E-05	1.57E-05	0.244	11

After employing the TOPSIS, we identified Player A4 as the best hitter in the 2013 season. This finding was expected because Player A4 ranked second in OBP and SLG. By contrast, Player A18 was the best performer in SLG, but ranked fourth in OBP; therefore, Player A18 ranked second among all candidates as the best hitter. Player A3 was the best performer in OBP; however, according to the TOPSIS, he ranked fourth among all candidates as the best hitter. Because of his performance in SLG, Player A3 ranked sixth among all players.

IV. CONCLUSION

Evaluating the best hitter in a professional baseball league is difficult, and can be classified as a type of MADM problem. Most baseball fans or commentators judge the best hitter based on the number of home runs or hits scored by players. To evaluate the performance of the hitters in the 2013 CPBL season more objectively, we employed the TOPSIS to determine the best hitter in the Chinese professional baseball league. By using this method, we found that the best hitter was an outstanding performer in OBP and SLG. From a methodological perspective, the present study results demonstrated that the TOPSIS is an adequate tool for selecting the best alternative.

ACKNOWLEDGMENT

This research is partially supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2410-H-156-013.

REFERENCES

- [1] C. C. Chen, M. L. Lin, Y. T. Lee, T. T. Chen and C. L. Huang, "Selection best starting pitcher of the Chinese professional baseball league in 2010 using AHP and TOPSIS methods," *Advances in Intelligent and Soft Computing*, 2002, vol. 133, pp.643-649.
- [2] C. C. Chen, M. L. Lin, Y. T. Lee, T. T. Chen and C. Y. Tseng, Best Starting Pitcher of the Chinese Professional Baseball League in 2009, *Proceeding of World Academy of Science, Engineering and Technology*. 2010, vol. 72, pp. 210-212.
- [3] C. C. Chen, Y. T. Lee and C. M. Tsai, Professional Baseball Team Starting Pitcher Selection using AHP and TOPSIS Methods, *International Journal of Performance Analysis in Sport*. 2014, vol.14, pp. 545-563.
- [4] W. B. Lin, I. w. Tung, M. J. Chen and M. Y. Chen, An analysis of an optimal selection process for characteristics and technical performance of baseball pitchers, *Perceptual and Motor Skills*, 2011, vol. 133, pp. 300-310.
- [5] J. K. Hakes and R. D. Sauer, An Economic Evaluation of the Moneyball Hypothesis, *Journal of Economic Perspectives*, 2006, vol. 20, pp. 173-185.
- [6] J. K. Hakes and R. D. Sauer, The Moneyball Anomaly and Payroll Efficiency: A Further Investigation, *International Journal of Sport Finance*, 2007, vol. 2, pp. 177-189.
- [7] D. J. Berri and M. B. Schmidt, On the Evaluation of Kickers in the National Football League, *International Journal of Sport Finance*, 2013, vol.8, pp. 263-278.
- [8] D. L. Olson, Comparison of Weights in TOPSIS Models, *Mathematical and Computer Modelling*, 2004, vol. 48, pp. 721-727.