

Best Starting Pitcher of the Chinese Professional Baseball League in 2009

Chih-Cheng Chen, Meng-Lung Lin, Yung-Tan Lee, Tien-Tze Chen and Ching-Yu Tseng

Abstract—Baseball is unique among other sports in Taiwan. Baseball has become a “symbol of the Taiwanese spirit and Taiwan’s national sport”. Taiwan’s first professional sports league, the Chinese Professional Baseball League (CPBL), was established in 1989. Starters pitch many more innings over the course of a season and for a century teams have made all their best pitchers starters. In this study, we attempt to determine the on-field performance these pitchers and which won the most CPBL games in 2009. We utilize the discriminate analysis approach to solve the problem, examining winning pitchers and their statistics, to reliably find the best starting pitcher. The data employed in this paper include innings pitched (IP), earned runs allowed (ERA) and walks plus hits per inning pitched (WHIP) provided by the official website of the CPBL. The results show that Aaron Rakers was the best starting pitcher of the CPBL. The top 10 CPBL starting pitchers won 14 games to 8 games in the 2009 season. Though Fisher Discriminant Analysis, predicted to top 10 CPBL starting pitchers probably won 20 games to 9 games, more 1 game to 7 games in actually counts in 2009 season.

Keywords—Chinese Professional Baseball League, starting pitcher, Fisher’s Discriminate analysis

I. INTRODUCTION

TAIWAN is a small place, yet its little League teams have become famous world-wide. How can this possibly have happened? Clearly, baseball is something special in this country. The sport was imported to the colony of Taiwan around 1897. Initially, it was played by colonial bureaucrats, bankers and their sons in Taipei, but spread to southern Taiwan during 1910s. From the 1960s to the 1980s, the Taiwanese teams became very successful in the Little League baseball scene worldwide. Historically, baseball was an important avenue by which the Taiwanese people navigated relationships with the Japanese, the Chinese and their American allies. At the turn of the century, as the search for a uniquely Taiwanese identity received official sanction, baseball became a crucial element of this identity. Professional baseball in Taiwan is a perfect instance of this self-consciousness, ideological combination of the global and the local, the cosmopolitan and the provincial, the international and the Taiwanese [1]. Baseball is unique among other sports as a “symbol of the Taiwanese spirit and

national sport”[1]. Taiwan’s first professional sports league, the Chinese Professional Baseball League (CPBL), was established on October 23rd 1989. The formation committee successfully recruited four business entities to form a four-club league comprised of the Uni-President 7-Eleven Lions, Brother Elephants, La new Bears and Sinon Bulls. The CPBL held its inauguration game on March 17th, 1990 where the Uni-President Lions defeated the Brother Elephants with a score of 4-3. In a regular season (March through early October) each team plays 120 games not including the pre-season and the post-season playoff [2].

Baseball, like other sports, is entertainment. Also like other sports, it can be much more than that. Since 1990, the Wins Award has been presented annually to the best pitcher of the CPBL. Each year, as the season progresses, baseball fans and sportswriters take great interest in predicting the winners of this award. The award is based on how many wins that pitcher receives over the course of the season. However, starting pitchers who pitch very well may still lose if their teammates do not play well. In other words, the award may not go to the best starting pitcher. Over the long history of baseball, starting pitchers have been considered much more important than relief pitchers. Starters pitch many more innings over the course of a season and normally, teams select their best pitchers for starters [3-5]. In this study, we attempt to determine the on-field performance of starting pitchers and which starting pitcher won the most CPBL games in 2009. We choose to solve the problem using Fisher Discriminate analysis. This approach is utilized to examine winning pitchers and their statistics, from which we can reliably find the best starting pitcher.

II. METHOD

A. Data

The data employed in this paper were provided by the official website of the CPBL [2], a website that has collected and posted records of every CPBL baseball game. The most commonly cited statistics for starting pitchers are innings pitched per game, earned run average (ERA), strikeouts per 9 innings pitched (K/9), and walks plus hits per inning pitched (WHIP) [3-5], so we will include them in this study. We want to formulate a simple model, one whose parameters are familiar to all fans, so only those four statistics are used: innings pitched (IP), earned run average (ERA), strikeouts per 9 innings (K/9) and walks plus hits per inning pitched (WHIP). We calculate the ERA, K/9 and WHIP for all starting pitchers using the following formulas:

Chen C. C. and Chen T. T. are with the Department of Sport Management, Aletheia University, No. 32, Chen-Li St., Tamshui, Taipei, 25103 Taiwan (e-mail: ccchen@email.au.edu.tw and au1169@mail.au.edu.tw)

Lin M. L. and Y. T. Lee are with the Department of Tourism, Aletheia University, No. 32, Chen-Li St., Tamshui, Taipei, 25103 Taiwan (e-mail: mllin1976@mail.au.edu.tw and au4300@email.au.edu.tw).

Tseng C. Y. is with the Department of Physical Education, Fu Jen Catholic University, No. 510, Chung Cheng Rd, Hsinchuang, Taipei, 24205 Taiwan (phone: +886-2-29053258; fax: +886-2-29052380; e-mail: 015844@mail.fju.edu.tw).

$$ERA = 9 \times \frac{\text{Earned Run Allowed}}{\text{Innings Pitched}}$$

$$K/9 = 9 \times \frac{\text{Strikeouts}}{\text{Innings Pitched}}$$

$$WHIP = \frac{\text{Walks} + \text{Hits}}{\text{Innings Pitched}}$$

B. Statistical Analysis

Calculated data were analyzed using the mean and standard data descriptive of the performance of all CPBL starting pitchers from IP, ERA, K/9 and WHIP. The association of wins of starting pitchers in 2009 with statistics was examined using t-testing and Fisher Discriminant Analysis.

T-testing followed by multivariate logistic regression analysis determined which statistics were significantly different between those in the win group and those in the non-win group. P-values of less than 0.05 were taken to be statistically significant. The significant statistics were chosen as discriminate variables. Two-group discriminate function analysis (discriminate prediction equation) was used to discriminate between the won group and non-won group, and to determine the discriminate power of the statistics. Discriminate functional analysis was used to determine which variables could be used to discriminate between the win group and the non-win group. Computationally, discriminate functional analysis is very similar to analysis of variance (ANOVA). The basic idea underlying discriminate functional analysis is to determine whether groups differ with regard to the mean of a variable, and then to use that variable to predict group membership. In the two-group case, discriminate functional analysis can also be thought of as (and is analogous to) multiple regression. If we code the two groups in the analysis as 1 and 2, and use that variable as the dependent variable in a multiple regression analysis, then we obtain results that are analogous to those we would obtain via discriminate analysis. In general, in the two-group case, we fit a linear equation of the type: $\text{Group} = a + b_1 \times x_1 + b_2 \times x_2 + \dots + b_m \times x_m$ where a is a constant and b_1 through b_m are regression coefficients. The interpretation of the results of a two-group problem is straightforward and closely follows the logic of multiple regression. We find that those variables with the largest (standardized) regression coefficients are the ones that contributed most to the prediction of group membership. Using these multivariate statistical techniques, it is possible to estimate the effects of each potential critical factor on the winning of professional baseball games in Taiwan. Analyses were conducted using Statistical Packages for the Social Sciences (SPSS) on an HP computer.

III. RESULTS

Of the 475 starters, 173 were wins and 302 non-wins. Win starters pitched 6.1 innings, 2.40 ERA, 5.76 K/9 and 1.57 WHIP per game. Non-win starters on the other hand pitched 4.2 innings, 13.07 ERA, 5.98 K/9 and 3.39 WHIP per game. Starters in the win group had significantly higher levels of IP

count ($p < 0.05$) and lower levels of ERA count ($p < 0.05$) and WHIP count ($p < 0.05$) than starters in the non-win group (Table 1).

TABLE I CHARACTERISTICS OF STARTERS IN THE WIN GROUP AND THE NON-WIN GROUP

Variable	Win (n=173) Mean \pm SD	Non-win (n=302) Mean \pm SD	t-value
IP	6.1 \pm .92	4.2 \pm 1.81	12.235*
ERA	2.40 \pm 2.19	13.07 \pm 51.45	-3.587*
K/9	5.76 \pm 3.15	5.98 \pm 6.61	-0.414
WHIP	1.57 \pm .60	3.39 \pm 7.36	-4.251*

*significant at 0.05

IP: Innings pitched

ERA: Earned run average

K/9: Strikeouts per 9 innings

WHIP: Walks plus hits per inning pitched

Overall Fisher Discriminate analysis revealed that the three statistics (IP, ERA and WHIP) were significantly; the Wilk's $\lambda = .810$, $\chi^2(3) = 99.326$, $p < .05$ (Table 2).

TABLE II INFORMATION FOR DISCRIMINATE ANALYSIS

Wilks' λ value	χ^2	Degree of freedom	p-value
.810	99.326	3	.000

The significant statistics chosen to represent discriminating variables were IP, ERA and WHIP. We then conducted a two-group discriminant model analysis to discriminate between the win group and the non-win group (Table 3). Win function is $-12.988 + 3.484(\text{IP}) - .285(\text{ERA}) + 2.320(\text{WHIP})$. Non-win function is $-9.265 + 2.814(\text{IP}) - 0.279(\text{ERA}) + 2.267(\text{WHIP})$.

TABLE III FUNCTION OF FISHER'S DISCRIMINATE ANALYSIS

	Results	
	Win	Non-win
IP	3.484	2.814
ERA	-.285	-.279
WHIP	2.320	2.267
Constant	-12.988	-9.265

The starters were divided into two groups based on the results of discriminate function analysis showing the following levels: predicted win group (233) and predicted non-win groups (242). The predicted win group had higher statistics levels while the predicted non-win group had lower statistics levels. The discriminate model correctly classified 70.5 % of the starting pitchers of CPBL in 2009 (Table 3 and Table 4). In the predicted win group, 233 starters had the potential to win the game. Among the 233 starters, 133 won; the other 100 did not win. 242 starters were predicted to be in the non-win group. The number of 202 starters was correct, but 40 starters were incorrect.

TABLE IV RESULTS OF FISHER'S DISCRIMINATE ANALYSIS FOR STARTING PITCHERS

Results	Predictions		Total
	Win	Non-win	
Original	133 (76.9%)	40 (23.1%)	173
Non-win	100 (33.1%)	202 (66.9%)	302
Total	233	242	475

70.5% primitive category observations already right distribution.

The number of wins of all starting pitchers in a season is measured. The pitcher who receives the most wins gets a reward in addition to appreciation by fans who think the pitcher is best pitcher in the league. Table 5 shows the top 10 starting pitchers in the 2009 season. Itsuki Shoda was first among starting pitchers of the CPBL, winning 14 games and losing 11. Second place was Aaron Rakers who won 13 games. Third place was Liao, Yu-Cheng who won 11 games. Fourth and fifth were Lin, Ko-Chien and Pan, Wei-Lun. They won 10 games in 2009.

TABLE V CPBL TOP 10 STARTING PITCHERS IN THE 2009 SEASON

Rank	Name	Win	Non-win	Total
1	Itsuki Shoda	14	11	25
2	Aaron Rakers	13	13	26
3	Liao, Yu-Cheng	11	11	22
4	Lin, Ko-Chien	10	14	24
5	Pan, Wei-Lun	10	11	21
6	Kobayashi Ryokan	9	14	23
7	Chang, Chih-Chia	9	10	19
8	Hsu, Wen-Hsiung	9	15	24
9	Tsao, Chin-hui	8	10	18
10	Mike Smith	8	11	19

Table 6 shows the probability for the Top 10 starting pitchers in the 2009 season. Aaron Rakers was predicted to win 20 games, seven more than he actually won. Second place was Itsuki Shoda, who actually won 14 games although predicted to win 19 games. Pan, Wei-Lun was ranked in third place. He was the starting pitcher in 21 games and actually won 10 although predicted to win 18 games. The other top 10 starting pitchers were predicted to win the other 1 to 5 games.

TABLE VI PROBABLE TOP CPBL 10 STARTING PITCHERS IN THE 2009 SEASON

Rank	Name	Win	Non-win	Total
1	Aaron Rakers	20	6	26
2	Itsuki Shoda	19	6	25
3	Pan, Wei-Lun	18	3	21
4	Lin, Ko-Chien	15	9	24
5	Liao, Yu-Cheng	14	8	22
6	Chang, Chih-Chia	14	5	19
7	Kobayashi Ryokan	13	10	23
8	Hsu, Wen-Hsiung	11	13	24
9	Mike Smith	10	9	19
10	Tsao, Chin-hui	9	9	18

IV. DISCUSSION

Throughout the long history of baseball, starting pitchers have been considered much more important than relief pitchers. Great starting pitchers pitch more innings over the course of a season and have the power to dominate the game. Normally teams have made all their best pitchers starters

[3-5]. Starting pitchers have a great effect on game victory or defeat [5-7]. So the purpose of this study is to discover best starting pitcher of the Chinese Professional Baseball League in 2009. Using Fisher's Discriminate model we divided the 475 starters into win and non-win groups. In the win group, 173 starters pitched more innings, having a lower ERA and WHIP than the non-win group (302 starters). Win group starters had significantly higher levels of IP and lower levels of ERA and WHIP than starters in the non-win group. The difference in win count between the actual and probable top 10 CPBL starting pitchers was 1 to 7. The results described above lead to the following conclusions: first, starting pitchers who want to win the game, have to devote more energy to pitching more innings, lose fewer points and have fewer opponent players on base. Second, Aaron Rakers was the best starting pitcher of the CPBL in 2009. According to the information in table 5 and table 6, Aaron Rakers was predicted to win 20 games but won only 13 games in 2009 making him a winning candidate. However when relief pitchers on the pitching mound do not pitch well the game can be lost or tied. Finally Aaron Rakers lost the chance to win the game. The situation is the same for the other 9 starting pitchers. This means that every team, the Uni-President 7-Eleven Lions, Brother Elephants, La new Bears and Sinon Bulls, in the CPBL have to organize a more powerful bullpen. It also means that teams must have powerful relief pitchers in order to lead in the finals.

V. CONCLUSION

This study provides information about the best starting pitcher of the CPBL in 2009. As we can see, the best starting pitcher is Aaron Rakers, who actually won 13 games, and was predicted to probably win 20 games in the 2009 season. Actually, the top 10 CPBL starting pitchers won 14 games to 8 games in the 2009 season. If relief pitchers pitch very well, they can probably win from 1 to 7 more games in the 2009 season. The results also revealed that the bullpens of CPBL teams were too weak to help the teams leading in the finals. Teams that want to be champion of the year, have to organize a powerful bullpen to win more games.

REFERENCES

- [1] A.D. Morris, "Baseball, History, the Local and Global in Taiwan", in *The minor arts of daily life: popular culture in Taiwan*, David K. Jordan, A.D.M., Marc L. Moskowitz. Hawaii: University of Hawaii, 2004, pp. 175-203.
- [2] <http://www.cpbl.com.tw/html/english/cpbl.asp>
- [3] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*. New York: Norton, 2003
- [4] R.L. Sparks, and D.L. Abrahamson, "A Mathematical Model to Predict Award Winners", *Math Horizons*, vol. April, pp. 5-13, 2005.
- [5] C.C. Chen, and T.T. Chen, "Starting Pitchers' Pitch Skills Positioning of CPBL in 2008-Multidimensional Scaling Analysis", *Journal of Physical Education Fu Jen Catholic University*, vol. 8, pp. 109-125, 2009.
- [6] Y.-C. Kao, and C. Liu, "Comparison to Pitching Performance of Chien-Ming Wang in 2005 and 2006 Season of Major League Baseball", *Sports Coaching Science*, vol. 9, pp. 187-196, 2007.
- [7] C. Liu, "Effects of Isokinetic Force and throwing-arm Velocity on Throwing Performance during Pitching Different Weighted Baseballs", *Physical Education Journal*, vol. 42, pp. 1-16, 2009.
- [8] J.C. Bardbury, "Does the Baseball Labor Market Properly Value Pitchers?", *Journal of Sports Economics*, vol. 8, pp. 616-632, 2007.