**Final report**

**KBO Statistics Analysis**

**Little Café**

Hyunjae Lee / Hosoo Lee / Duyoung Choi

**Introduction**

What our team try to analyze is correlation between teams and spectators. We decide to analyze professional baseball data because Korean professional baseball welcomes 34th anniversary in this year and it has the largest number of fan in Korean professional sports increasing the popularity. We were wondering about what part can be associated with the number of spectators. So, we tried to analyze whether each team performance can affect on the attendance.

However, there are lots of similar researches in the US Major League, it is hardly to find the studies and outcomes in Korean Professional baseball. So, we take the analysis about KBO and not just rank, but also attack and defense ability such as batting rate and earned run average, because those can attract spectators’ interest and concentration.

**Body**

To assess whether the correlation between team performance and the number of spectators is exist, we collect the data about attendance, rank, batting rate and earned run average as teams and as years from 2002 based on the KBO records. Now, Korean professional baseball operates 10 clubs, but KT(join to first division in 2015) and NC(join to first division in 2013) are excluded because they have less samples. Nexen is also joined to first division in 2008, so have comparatively less samples. But it was formerly Hyundai that was joined to first division until 2007, so for convenience, we combine two teams’ records.

Analyzing the correlation between attendance as teams and performances(Rank, AVG, ERA), we mark the plot and it is tested through cor, cor.test function and p-value.

Because our purpose is analyzing whether the correlation between performance and attendance is exist, we select attendance to core features. And although a number of factors can affect to attendance, we select rank, batting rate and earned run average to sub data because these are most important mark representing attack and defense ability. Depending on the situation, though rank is high, batting rate or earned run average can be low. So, it is the characteristic to try to analyze approaching with diversified factors.

Before the regular analysis, it should be noted the limitations in our analysis as follows.

1) Baseball international contest grade

There are WBC, Olympic and Asian game and so on. But relative evaluation of various levels of contests is impossible. And it can’t be reflect the impact the contest has to the number of spectators.

2) Social event

For example, in 2014, there is a social issue so called “Ferry Sewol sunk”. At that time, Korean professional baseball did not conduct group cheering event such as cheer leading, using amplifier, kiss time running between innings. The impact those have to the number of spectators can’t be measured.

So, in the assumption that the limitation about two kinds of contents above is excluded, our analysis is started.

Correlation between attendance and Rank

|  |  |  |  |
| --- | --- | --- | --- |
| Plot | | | |
| Samsung | C:\Users\내문서\Desktop\Samsung-Rank.jpg | Doosan | C:\Users\내문서\Desktop\Doosan-Rank.jpg |
| Kia | C:\Users\내문서\Desktop\Kia-Rank.jpg | Hanhwa | C:\Users\내문서\Desktop\Hanhwa-Rank.jpg |
| Lotte | C:\Users\내문서\Desktop\Lotte-Rank.jpg | SK | C:\Users\내문서\Desktop\SK-Rank.jpg |
| LG | C:\Users\내문서\Desktop\LG-Rank.jpg | Nexen | C:\Users\내문서\Desktop\Nexen-Rank.jpg |

H0 : In Samsung’s attendance and rank, there is no correlation.

Cor=-0.1559318, p-value=0.611 H0, fail to reject H0

H0 : In Kia’s attendance and rank, there is no correlation.

Cor=-0.1144, p-value = 0.7097 H0, fail to reject H0

H0 : In Lotte’s attendance and rank, there is no correlation.

Cor=-0.9396, p-value = 1.869e-06, reject H0

H0 : In LG’s attendance and rank, there is no correlation.

Cor=-0.1142, p-value=0.7103 H0, fail to reject H0

H0 : In Doosan’s attendance and rank, there is no correlation.

Cor=-0.2684, p-value=0.3752 H0, fail to reject H0

H0 : In Hanhwa’s attendance and rank, there is no correlation.

Cor=0.4112, p-value=0.1628 H0, fail to reject H0

H0 : In SK’s attendance and rank, there is no correlation.

Cor=-0.4770, p-value=0.0993 H0, fail to reject H0

H0 : In Nexen’s attendance and rank, there is no correlation.

Cor=-0.0181, p-value=0.9532 H0, fail to reject H0

Between attendance and rank, we concluded that the correlation is not exist except Lotte.

Correlation between attendance and AVG

|  |  |  |  |
| --- | --- | --- | --- |
| Plot | | | |
| Samsung | C:\Users\내문서\Desktop\Samsung-AVG.jpg | Doosan | C:\Users\내문서\Desktop\Doosan-AVG.jpg |
| Kia | C:\Users\내문서\Desktop\Kia-AVG.jpg | Hanhwa | C:\Users\내문서\Desktop\Hanhwa-AVG.jpg |
| Lotte | C:\Users\내문서\Desktop\Lotte-AVG.jpg | SK | C:\Users\내문서\Desktop\SK-AVG.jpg |
| LG | C:\Users\내문서\Desktop\LG-AVG.jpg | Nexen | C:\Users\내문서\Desktop\Nexen-AVG.jpg |

H0 : In Samsung’s attendance and AVG, there is no correlation.

Cor=0.2488, p-value = 0.4124 H0, fail to reject H0

H0 : In KIA’s attendance and AVG, there is no correlation.

Cor=0.4286, p-value = 0.1439 H0, fail to reject H0

H0 : In Lotte’s attendance and AVG, there is no correlation.

Cor=0.7779, p-value=0.0017 H0, reject H0

H0 : In LG’s attendance and AVG, there is no correlation.

Cor=0.6650, p-value=0.0131 H0, reject H0

H0 : In Doosan’s attendance and AVG, there is no correlation.

Cor=0.3825, p-value=0.1971 H0, fail to reject H0

H0 : In Hanhwa’s attendance and AVG, there is no correlation.

Cor=-0.0521, p-value=0.866 H0, fail to reject H0

H0 : In SK’s attendance and AVG, there is no correlation.

Cor=0.1244, p-value=0.6855 H0, fail to reject H0

H0 : In Nexen’s attendance and AVG, there is no correlation.

Cor=0.5747 p-value=0.04 H0, reject H0

Between attendance and AVG, we concluded that the correlation is not exist except Lotte, LG, Nexen..

Correlation between attendance and ERA

|  |  |  |  |
| --- | --- | --- | --- |
| Plot | | | |
| Samsung | C:\Users\내문서\Desktop\Samsung-ERA.jpg | Doosan | C:\Users\내문서\Desktop\Doosan-ERA.jpg |
| Kia | C:\Users\내문서\Desktop\Kia-ERA.jpg | Hanhwa | C:\Users\내문서\Desktop\Hanhwa-ERA.jpg |
| Lotte | C:\Users\내문서\Desktop\Lotte-ERA.jpg | SK | C:\Users\내문서\Desktop\SK-ERA.jpg |
| LG | C:\Users\내문서\Desktop\LG-ERA.jpg | Nexen | C:\Users\내문서\Desktop\Nexen-ERA.jpg |

H0 : In Samsung’s attendance and ERA, there is no correlation.

Cor=0.058, p-value=0.8506 H0, fail to reject H0

H0 : In KIA’s attendance and ERA, there is no correlation.

Cor= 0.4157, p-value=0.157 H0, fail to reject H0

H0 : In Lotte’s attendance and ERA, there is no correlation.

Cor=-0.3141, p-value=0.296 H0, fail to reject H0

H0 : In LG’s attendance and ERA, there is no correlation

Cor=-0.0936, p-value=0.761 H0, fail to reject H0

H0 : In Doosan’s attendance and ERA, there is no correlation

Cor=0.4118, p-value=0.1621 H0, fail to reject H0

H0 : In Hanhwa’s attendance and ERA, there is no correlation

Cor=0.3704, p-value=0.2128 H0, reject H0

H0 : In SK’s attendance and ERA, there is no correlation

Cor=-0.1382, p-value=0.6526 H0, fail to reject H0

H0 : In Nexen s attendance and ERA, there is no correlation

Cor=-0.0207, p-value=0.9466 H0, fail to reject H0

Between attendance and ERA, we concluded that the correlation is not exist except Hanhwa

We could obtain spectator record from 1982 when the first korean professional baseball started, but there were no record about each team's performance. There are days SK performs well and Hanhwa also do well and attract attendance at the same time after the new manager came. If there were more records available, we could have drawn more accurate correlation of each team's winning rate with spectator.

And in korean professional baseball, not only a team's winning rate, but which hometown the team's based on influences the number of spectator. For example, it is normal that high ranking attracts large spectators and low ranking does not. But our analysis on hanhwa's spectator and ranking shows the opposite result.

Although all teams' correlation coefficient between spectator and ranking were not big enough to be meaningful, they were all positive, except for hanhwa. only hanhwa's case shows unusual negative coefficient. hanhwa's hometown fan, daejeon citizen is faithful to genuinely enjoy the match, not cling to the ranking

**Conclusion**

Professional baseball gaining huge popularity in Korean professional sports, sets a goal of 800 million attendance. From 2008 Olympic gold medal and 2009 WBC second place, the number of baseball audience is soaring and it continues the tendency of increase. Also, because we thought that KBO has lots of data easy on the analysis, we took the analysis. Through the analysis, we know that the reason spectators come to baseball stadium is very complex and various. Although we can’t judge the variation of attendance only by the simple numerical performance, it is clear that team performance can affect to the attendance. We speculate that It will show clearer correlation after the materials are stacked as time goes by.