Segmenting Major League Baseball Teams by Attendance: A Multilevel Analysis of Determinants Across Clusters

Namhun Lim[1], Wonyoung Kim[2], and Mark Vermillion[2]

[1]*Department of Business, Accounting, and Sport Management, Elizabeth City State University*

[2]*Department of Sport and Leadership Studies, Wichita State University*

Abstract

The study examines determinants of game-day attendance in Major League Baseball (MLB) by classifying teams based on their seasonal average attendance and analyzing how established predictors operate within each attendance segment. Using data from all 30 MLB teams across the 2006–2017 regular seasons (*N*=29,056), the analysis proceeded in two stages. First, using cluster analysis, the study identified four attendance-based groups that reflected distinct market and performance profiles. Second, three-level hierarchical linear modeling (HLM) evaluated the effects of economic, demographic, game-attractiveness, and residual-preference factors on single-game attendance within each group. Results revealed meaningful differences across clusters. High-attendance teams were less sensitive to opponent characteristics and more influenced by structural factors such as stadium capacity and home team performance. Lower-attendance teams showed stronger responses to visiting team quality, star players, game uncertainty, and local market conditions. Group-specific patterns also emerged for rivalry games, weekend scheduling, and seasonal progression. Findings demonstrate the value of segmenting MLB teams according to historical attendance patterns and highlight the utility of multilevel modeling for analyzing sport consumption. Results provide practical implications for MLB organizations seeking to tailor marketing strategies, improve scheduling decisions, and enhance fan engagement across diverse market environments.

Keywords: Major League Baseball, sport attendance, cluster analysis, hierarchical linear modeling, market demand

Introduction

Since the 2017 regular season of Major League Baseball (MLB), fans have witnessed markedly different game-day environments. For instance, the familiar ritual of a pitcher throwing four intentional balls to issue a walk was replaced with a manager's signal and an automatic walk to first base. This rule change was part of a broader pace-of-play initiative that also included limiting mound visits, shortening between-inning breaks, and imposing a 30-second decision limit on replay reviews (Posnanski, 2017; McCallister, 2018). Additional adjustments continued in the following years. By the 2024 season, MLB reduced the pitch clock from 20 to 18 seconds with runners on base and expanded the runner's lane to first base (Castrovince, 2023). Ahead of the 2025 season, the league strengthened penalties for infield shift violations and expanded replay capabilities to include abandonment of second or third base (Randhawa, 2025).

MLB Commissioner Rob Manfred explained that these rule adjustments were intended to counter the trend of increasingly longer games, a pattern many analysts have associated with declining attendance. Recent data support this concern. In 2023, the average duration of a nine-inning game fell to just under 2 hours and 40 minutes, which represented the fastest pace recorded in nearly forty years. In the same season, total regular-season attendance rose to 70,747,365, the first time since 2017 that league-wide attendance exceeded 70 million (Adler, 2023). These outcomes contrast with the 2022 season, when league-wide attendance totaled approximately 64.6 million, a figure that remained below the 68.5 million recorded in the last full pre-pandemic season in 2019 (Associated Press, 2022, October 6).

Game duration has demonstrated a meaningful association with spectator demand. Data from the 2004 to 2018 MLB seasons showed a negative correlation ($r = -.57$, $p < .05$) between average game time and next-season attendance, indicating that each additional minute of game time was associated with a decline of approximately 94 spectators per game. Although league-wide patterns provide useful context, they can obscure important differences across individual teams, which operate under varying market, performance, and promotional conditions.

To address this variation, the present study adopts a two-part approach. First, it uses cluster analysis to group MLB teams based on their seasonal average attendance. Second, it applies multilevel modelling to examine how established attendance determinants, including team performance, market characteristics, and opponent attractiveness, influence attendance within each group (DeSchriver & Jensen, 2002; Rivers & DeSchriver, 2002).

Determinants of Sport Attendance

Attendance has long been recognized as a foundational element of sport consumption, and despite the growing financial importance of sponsorship and media, spectator demand remains essential to the sustainability of sport organizations. Early studies examined a wide range of factors influencing attendance (e.g., Baade & Tiehen, 1990). Two widely cited frameworks classify these determinants. Schofield (1983) proposed four categories: economic factors, demographic factors, attractiveness factors, and residual preference factors. Borland and Macdonald (2003) later introduced a similar framework that emphasized viewing quality in addition to consumer preference, economic conditions, contest characteristics, and venue capacity.

Attractiveness refers to game and team characteristics that motivate fans to attend. Indicators such as team performance, star players, playoff contention, and historical success consistently show positive relationships with attendance (Rivers & DeSchriver, 2002; Welki & Zlatoper, 1994). Visiting team quality also plays a significant role, with higher attendance when nationally popular or rival teams appear (Lemke et al., 2010). Promotions and special events provide an additional boost to demand and are controllable tools for sport marketers.

Residual preference factors encompass external elements, including weather and game-day scheduling (weekday vs. weekend), start time, and stadium comfort (Hansen & Gauthier, 1989). Several studies have shown that extreme temperatures, rain, and weekday scheduling reduce attendance (Welki & Zlatoper, 1994; Lemke et al., 2010). Stadium age and capacity influence spectator experience as well, with newer or less crowded venues typically attracting more attendees (McEvoy et al., 2005).

Economic and demographic variables influence the size of the potential market for live sport consumption. Income levels and metropolitan population are consistently identified as significant predictors of attendance (McEvoy et al., 2005; Coates & Humphreys, 2007). Although ticket price is expected to have a direct negative effect on demand, measuring this effect is challenging due to fan loyalty, dynamic pricing systems, and secondary market transactions (Ahn & Lee, 2007; Fort, 2004).

Drawing from these established determinants, the present study classifies all 30 MLB teams into groups according to their seasonal average attendance through cluster analysis. It then examines the extent to which economic, demographic, attractiveness, and residual preference factors influence single game attendance for each group using multilevel modelling. This allows for an understanding of how determinants differ across distinct segments of the league.

Research Questions

Based upon the previous literature, the following research questions guide this study:

RQ 1-1: Into how many groups should the 30 MLB teams be classified based on their seasonal average attendance?

RQ 1-2: Which MLB teams belong to each group?

RQ 2-1: What economic, demographic, attractiveness, and residual preference factors significantly influence single game attendance in each group from the 2006 to 2017 MLB regular seasons?

RQ 2-2: How do these significant determinants differ across the groups?

Methodology

Data

The data set included all regular season games played by the 30 MLB teams from 2006 through 2017. Seasonal average attendance for all teams (N_1=360; "MLB Attendance Report", 2017) was collected from ESPN to classify clubs into attendance-based groups (RQ1). Single game attendance (N_2=29,056; "Arizona Diamondbacks", 2017) served as the dependent variable for examining determinants of demand within each group identified through cluster analysis (RQ2). Independent variables reflected four major categories identified in prior literature: *attractiveness*, *residual preference*, *economic*, and *demographic factors*. These variables included team performance indicators, payroll, rivalry status, game uncertainty, stadium characteristics, ticket prices, income levels, and metropolitan population. Game-level and season-level variables were coded accordingly, and continuous predictors were centered at either the group mean or the grand mean to facilitate interpretation.

Team quality and game uncertainty were calculated following the methods of Tainsky and McEvoy (2012). Stadium age, capacity, ticket prices, household income, and metropolitan population were rescaled to improve model estimation. A complete description of all variables and coding procedures is provided in Table 1 and in the Appendix B.

After the data were collected, several variables required rescaling to facilitate interpretation and model convergence. The final rankings of the home and visiting teams were reverse coded so that higher values indicated better performance. Large-unit variables such as payroll, stadium capacity, household income, and metropolitan population were divided by one thousand. All continuous predictors, except for binary indicators such as rivalry and weekend games, were centered to improve interpretability and model estimation. Game-level variables were centered at their group means, and season-level variables were centered at the grand mean.

Data Analysis

Cluster analysis. A hierarchical cluster analysis was conducted to classify the 30 MLB teams based on their seasonal average attendance across twelve seasons. Similarity among teams was assessed using Euclidean distance, following standard procedures in cluster analysis (Afifi et al., 2003). A dendrogram was used to identify a reasonable number of attendance groups, and analysis of variance was performed to confirm that the resulting groups differed significantly in their average attendance.

Hierarchical linear modeling (HLM). Attendance data in professional sports or recurring sporting events have hierarchical characteristics. In the case of MLB data analyzed in this study, the variance in attendance is influenced by differences across individual games, seasons, and teams. Because hierarchical linear modeling (HLM), also known as multilevel modeling (MLM), is more effective than general linear modeling (GLM) for analyzing multilevel data (Lim & Pedersen, 2022), this study used HLM to examine the relationship between actual attendance and the attendance determinants for each group identified by the cluster analysis. The null model for each group was estimated first to determine the need for HLM by assessing variance components at each level. The full three level HLM for each group was then constructed with twelve game-level variables and thirteen season-level variables. The full model was expressed as:

$$Att_{gijk} = \gamma_{g000} + \gamma_{gn00}X_{gijk} + \gamma_{g0m0}Y_{g\cdot jk} + R_{gijk} + U_{g0jk} + V_{gook}$$

(Where *n*=1, …, 12 and *m*=1, …, 13)

Where, *g*: groups that decided by cluster analysis

X_{gijk}: the twelve game-level independent variables

$Y_{g\cdot jk}$: the thirteen season-level independent variables

σ^2: variance within season (R_{gijk})
τ_0^2: variance between seasons (U_{g0jk})
φ_0^2: variance between teams (V_{gook})

These HLM models were evaluated to ensure that key assumptions were satisfied, including the normality of residuals and the normality of season-level residuals for random coefficients. All statistical analyses in this study were conducted using SAS 9.4.

Results

Cluster Analysis

The mean seasonal average attendance for the 30 MLB teams from the 2006 to 2017 regular seasons was 30,831 (N_1 = 360, *SD* = 8,136). The hierarchical cluster analysis indicated that dividing the teams into four groups provided a reasonable and interpretable structure. The resulting group assignments are presented in Table 2. Although the Philadelphia Phillies recorded a higher twelve-year average attendance than the New York Mets, they were classified into Group 2 due to substantial variation across seasons, including attendance levels below 25,000 in the most recent three years of the dataset. The Toronto Blue Jays exhibited a similar pattern of fluctuation and were therefore grouped in the same category. The results of cluster analysis are provided in Table 2 and in the Appendix B.

A one-way analysis of variance demonstrated that the four groups differed significantly in seasonal average attendance (F = 242.3, p <.001), confirming the appropriateness of the cluster solution. One-way ANOVA tests conducted on the season-level variables, such as team payroll, household income, and metropolitan population, indicated violations of the homogeneity of variance assumption suggesting the variance within each group for these independent variables differed across groups. Although the mean values for some season-level variables appeared to differ across groups, these differences could not be confirmed through statistical testing.

Hierarchical Linear Modeling

Based on the results of the cluster analysis, hierarchical linear modeling was conducted separately for each group to examine the extent to which attendance determinants predicted single game attendance. Descriptive statistics for all variables across the full sample and within each group are presented in Table 3. Because ESPN does not report attendance for the first game of doubleheaders, ninety-nine games were missing from the dataset. In addition, the game between the Baltimore Orioles and the Chicago White Sox on April 29, 2015, which was held without spectators, was excluded. The number of missing games for Groups 1 through 4 was 20, 21, 40, and 19 respectively. The results of descriptive statistics are provided in Table 3 and in the Appendix B.

A three-level null model was estimated for each group following centering and rescaling procedures. Intraclass correlations were calculated to determine the proportion of variance attributable to differences between teams and seasons. As shown in Table 4, variation in Group 1 attendance was distributed across teams (25.2%), seasons (38.5%), and games within seasons (36.2%). In contrast, the proportion of variance between teams in Groups 2, 3, and 4 was relatively small, ranging from 1.2 to 2.8 percent. These groups showed substantially greater variation at the game-level within seasons (Group 2: 65.2%, Group 3: 75.1%, Group 4: 79.8%). Excluding Group 4, which comprised only four teams, these results indicate that single game attendance did not differ meaningfully across teams in Groups 2 and 3. The results of variance and intraclass correlations by level are provided in Table 4 and in the Appendix B.

Given the limited variance at the team level for Groups 2, 3, and 4, Group 1 was estimated with a full three-level model, while the other groups were estimated with simplified three-level models that did not include random team effects. Results of the full models for all groups are displayed in Table 5.

For Group 1, seven game-level variables (i.e., *HTQ*, *VTQ*, *VT_Payroll*, *Rival*, *Weekend*, *Progress*, and $Progress^2$ [p <.001]) and five season-level variables (i.e., *HT_Payroll*, *Capacity*, *Season* [p <.001], *ProTeams*, and $Season^2$ [p <.05]) were significant predictors of single game attendance. Overall, effect sizes were modest relative to the other groups. The influence of the visiting team was limited, as variables such as final rank, team age, star players, and championships were not significant. Although Group 1 attendance generally declined over time ($\gamma_{1\cdot0\cdot12\cdot0}$ = -1099, *p* <.001), the magnitude of decline was smaller compared with Group 3.

Group 2 was the group most strongly influenced by the performance and popularity of the home team Home team quality ($\gamma_{2\cdot1\cdot0\cdot0}$ = 45563, *p* <.001), payroll ($\gamma_{2\cdot0\cdot1\cdot0}$ = 1.87, *p* <.001), and number of star players ($\gamma_{2\cdot0\cdot3\cdot0}$ = 615.83, *p* <.05) were all significant predictors, and several characteristics of the visiting team also influenced attendance, including payroll, championships (*p <.001*), star players, and team age (*p <.05)*. Group 2 was the most sensitive to stadium age ($\gamma_{2\cdot0\cdot6\cdot0}$ = -304.67, *p* <.05) and the presence of other professional teams in the metropolitan area ($\gamma_{2\cdot0\cdot10\cdot0}$ = -1980, *p* <.05). Unlike the overall league trend of declining attendance, Group 2 attendance increased steadily over the twelve-year period ($\gamma_{2\cdot0\cdot12\cdot0}$ = 1128, *p* <.05).

For Group 3, where the average home team winning percentage was relatively low, attendance was heavily influenced by characteristics of the visiting team. Visiting team quality, payroll, previous season final rank, team age, and championships all had positive effects on attendance (*VTQ* [$\gamma_{3\cdot2\cdot0\cdot0}$ = 3664, *p* <.05]; *VT_Payroll* [$\gamma_{3\cdot4\cdot0\cdot0}$= 0.74, *p* <.001]; *VT_FinalRank* [$\gamma_{3\cdot5\cdot0\cdot0}$ = 198.96, *p* <.001]; *VT_Age* [$\gamma_{3\cdot7\cdot0\cdot0}$ = 11.20, *p* <.001]; *VT_Champs* [$\gamma_{3\cdot8\cdot0\cdot0}$ = 129.54, *p* <.001]). The positive effect of the game uncertainty measure ($\gamma_{3\cdot3\cdot0\cdot0}$ = 2670, *p* <.05) further suggests that fans in this group showed heightened interest in matchups involving stronger or more competitive opponents. In contrast to Groups 2 and 4, the presence of star players on the home team did not significantly affect attendance (*p* >.05). Group 3 also experienced the steepest decline in attendance as the season progressed ($\gamma_{3\cdot0\cdot12\cdot0}$ = -1281, *p* <.001).

Although Group 4 had a higher average winning percentage (49.3%) over the 12 seasons compared to Group 3 (47.8%), its average attendance was approximately 6,500 lower. The effects of home team quality ($\gamma_{4\cdot1\cdot0\cdot0}$ = 34604, *p* <.001) and payroll ($\gamma_{4\cdot0\cdot1\cdot0}$ = 1.57, *p* <.001) were smaller in magnitude than in Group 3. Instead, Group 4 showed the strongest response to the presence of star players on the home team ($\gamma_{4\cdot0\cdot3\cdot0}$ = 756.12, *p* <.001). Attendance was also influenced by the strength of the visiting team, including payroll ($\gamma_{4\cdot4\cdot0\cdot0}$ = .78, *p* <.001), star players ($\gamma_{4\cdot6\cdot0\cdot0}$= 221.03, *p* <.05), and championships ($\gamma_{4\cdot7\cdot0\cdot0}$ = 194.04, *p* <.001), as well as by metropolitan income ($\gamma_{4\cdot9\cdot0\cdot0}$ = 301.44, *p* <.001) and population ($\gamma_{4\cdot10\cdot0\cdot0}$ = 16.09, *p* <.05).

R-squared values for the full models indicated that attendance in Group 1 was explained most effectively by the predictors included in the model, with an explained variance of 60.2%. Group 4 showed the lowest explained variance at 41.6%, suggesting that additional variables may be needed to fully account for variation in attendance for this group. The results of full models for each group are provided in Table 5 and in the Appendix B. The next section discusses the implications of these findings for team level and league level marketing strategies.

Discussion

Group-Based Attendance Patterns

The primary aim of this study was to classify MLB teams according to their seasonal average attendance and to examine how established attendance determinants function across these attendance segments. The hierarchical cluster analysis identified four distinct groups that reflected meaningful variation in attendance patterns. Teams with long-standing national reputations, such as the New York Yankees, Los Angeles Dodgers, and St. Louis Cardinals, appeared in the group with the highest average attendance. In contrast, organizations that have consistently faced attendance challenges, including the Tampa Bay Rays, Oakland Athletics, and Miami Marlins, were placed in the group with the lowest average attendance. Analysis of variance confirmed that the four groups were statistically different from one another, supporting the validity of this segmentation approach.

Using this framework, the study applied well-established attendance determinants across economic, demographic, game attractiveness, and residual preference categories (Baade & Tiehen, 1990; Lemke et al., 2010). Given the multilevel structure of MLB attendance data, hierarchical linear modeling provided an appropriate analytical strategy for estimating the influence of these determinants at both the game and season-levels. The null model results confirmed that attendance variation occurred across games, seasons, and teams, validating the use of a multilevel approach. The full models identified significant predictors within each group and revealed that the determinants operated differently depending on each team's attendance profile. These findings provide a structured basis for understanding group-specific attendance dynamics and offer practical direction for differentiated attendance strategies.

Market Demand and Consumer Behavior in Professional Sports

In the past decade, there has been a decline in consumer interest in professional sports, while competition in the sport marketplace has increased. Retaining consumers has become the biggest challenge for the sports industry. To overcome competition, it is essential for marketers to understand market demand, which is related to consumer expectations of the core product's attributes. Market demand is a set of pull factors that an organization can offer to new and returning consumers. Analyzing market demand provides insight into consumer expectations and enables the formulation of an effective marketing mix. This leads to strategic decisions that enhance the success of the business, ultimately satisfying consumer needs and increasing market demands.

Previous studies (e.g., DeSchriver & Jensen, 2002; Rivers & DeSchriver, 2002) have examined market demand primarily in professional and intercollegiate sports. These studies found that game attractiveness, economic consideration, and schedule convenience were essential factors for consumers' decision-making. Game attractiveness was explained by several factors, including individual skills, star players, team records, and stadium quality. Economic consideration included factors such as ticket prices, marketing promotions, substitute forms of entertainment, and competition from other sporting events. Schedule convenience was explained by game time, day of the week, and weather. These factors explained a meaningful portion of the variance in professional sport consumption. The identified market demand variables can be applied to different professional sports.

MLB Attendance Trends

It is worth noting that even before the COVID-19 pandemic in 2020, there had been a gradual decline in MLB attendance over the past several years. Factors such as the rising cost of attending games, the availability of other entertainment options, and changes in viewing habits have all contributed to this trend. Moreover, MLB teams played in empty stadiums or with limited attendance due to health and safety protocols caused by the COVID-19 pandemic in 2020. However, in the 2021 season, some teams were able to have fans back in their venues, though the capacity varied by location and was subject to change depending on local health guidelines.

According to a report published in Forbes (Brown, 2022), MLB attendance for the 2022 season was down nearly 6% from the 2019 season, which was the last full season before the COVID-19 pandemic. Indeed, Forbes indicates that the average attendance per game for the 2022 season was 26,775, compared to 28,044 in 2019. The decline in attendance was not unexpected, given the ongoing impact of the pandemic on live events and public gatherings at the time. There were a variety of factors contributed to the decline in attendance, including but not limited to the rising cost of attending games, changes in viewing habits, and the performance of individual teams. Thus, the analysis of declining MLB attendance is a critical ongoing concern for MLB and the sport industry as understanding the factors that contribute to attendance trends can help teams and leagues make informed decisions about marketing, pricing, and other strategic initiatives.

The study by Zhang et al. (2003) discovered that the market demand factors have the ability to forecast the consumption of live and televised professional sporting events. These results are consistent with previous research conducted by numerous scholars (e.g., Baade & Tiehen, 1990; Becker & Suls, 1983; Hansen & Gauthier, 1989; Whitney, 1988), and highlight the significance of market demand factors, such as game attractiveness, economic considerations, and marketing promotion, in the creation of the marketing mix, including product, price, place, and promotion. By using a 4x3 interactive grid that relates the four marketing mix elements and the three market demand factors, specific marketing strategies can be designed in practice. Different contingency factors that are specific to various sports and competition levels can also be taken into account in different situations. Next, some general marketing implications are briefly discussed.

Implications, Limitations, and Future Research

The group-specific findings in this study offer clear implications for MLB teams seeking to enhance attendance. Recognizing the unique attendance drivers within each group enables organizations to develop more precise strategies related to pricing, scheduling, promotional activities, and brand positioning. For example, teams in lower-attendance clusters may benefit from prioritizing opponent-based attractiveness, while high-attendance teams may focus on strengthening brand equity and enriching in-game experiences. From a theoretical standpoint, the study contributes to the sport management literature by illustrating the value of clustering approaches in segmenting sport markets and by demonstrating the advantages of multilevel modeling for examining attendance determinants.

A key limitation concerns the temporal scope of the dataset, as the analyses rely on pre-pandemic attendance data. COVID-19 pandemic and its aftermath may have altered attendance segments and the salience of key determinants through shifts in live event consumption, media habits, and pricing dynamics. Future research should replicate the analysis with post-pandemic data and compare pre- and post-COVID pandemic to evaluate whether clusters and group-specific effects remain consistent. Additional extensions may incorporate evolving media consumption habits, dynamic and secondary ticket pricing mechanisms, and fan-level psychological motivations. These extensions would provide a more comprehensive understanding of attendance decision-making in today's MLB environment.

Appendix A

Computation of Team Quality and Game Uncertainty

A1. Team Quality Calculation

Team quality for each game was calculated following Tainsky and McEvoy (2012).

The value reflects a weighted combination of the previous season winning percentage and the current season winning percentage prior to game *i*.

$$TQ_{ijk} = \left[Win\%_{\cdot(j-1)k} \times \left(162 - Progress_{(i-1)jk}\right) + \left(Win\%_{ijk} \times Progress_{(i-1)jk}\right)\right]/162$$

Where,

i: i^{th} home game of team *k* in season *j* (*i* ≅ 1, …, 81)

j: season *j* (*j* = 1, …, 12)

k: team k (k = 1, …, 30)

$Win\%_{\cdot(j-1)k}$: team k's winning percentage in the previous season of season j

$Win\%_{ijk}$: team k's winning percentage prior to game i in season j

$Progress_{ijk}$: team k's number of games that have been played in season j, including game i

A2. Game Uncertainty Calculation

Game uncertainty was computed using the home and visiting team qualities, following the uncertainty measure based on Bill James' approach.

$$Uncertainty_{ijk} = |0.5 - C_{ijk}|$$
$$C_{ijk} = (HTQ_{ijk} - HTQ_{ijk}VTQ_{ijk})/[(HTQ_{ijk} + VTQ_{ijk}) - 2HTQ_{ijk}VTQ_{ijk}]$$

Where,

HTQ_{ijk}: home team quality for game i of team k in season j

VTQ_{ijk}: visiting team quality for the same game

Appendix B

Table 1

Variable Descriptions of MLB Attendance Determinants

Symbol	Description
Dependent Variable	
Att_{ijk}	Team k's attendance of game i in the season j
(Game-level) Independent Variables	
HTQ_{ijk}	Home team's winning percentage prior to game i
VTQ_{ijk}	Visiting team's winning percentage prior to game i
$Uncertainty_{ijk}$	Bill James's baseball game uncertainty
$VT\_Payroll_{ijk}$	Total payroll of visiting team
$VT\_FinalRank_{ijk}$	Visiting team's final rank in the season j
$VT\_StarPlayers_{ijk}$	Visiting team's number of star players
VT_Age_{ijk}	Visiting team's age
VT_Champs_{ijk}	Visiting team's number of previous Championships
$Rival_{ijk}$	Dummy of local or divisional rivalry games
$Weekend_{ijk}$	Dummy of weekend games (from Friday to Sunday)
$Progress_{ijk}$	Number of games that have been played, including game i

$Progress_{ijk}^2$	Quadratic term of *Progress*
(Season-level) Independent Variables	
$HT\_Payroll_{\cdot jk}$	Total payroll of home team *k* in the season *j*
$HT\_FinalRank_{\cdot jk}$	Home team *k*'s final rank in the season *j*
$HT\_StarPlayers_{\cdot jk}$	Home team *k*'s number of star players
$HT\_Age_{\cdot jk}$	Home team *k*'s age
$HT\_Champs_{\cdot jk}$	Home team *k*'s number of previous championships
$STD\_Age_{\cdot jk}$	Arena age of home team *k*
$Capacity_{\cdot jk}$	Arena capacity of home team *k*
$Ticket_{\cdot jk}$	Fan Cost Index of home team *k*
$Income_{\cdot jk}$	Median household income for home city
$ProTeams_{\cdot jk}$	Number of other professional teams in same area
$Population_{\cdot jk}$	Population of home city
$Season_{\cdot jk}$	Season progress of home team k from 2006 season
$Season_{\cdot jk}^2$	Quadratic term of *Season*

Note. *i*: home game *i* ($i \cong 1, \ldots, 81$); *j*=season *j* (j=1, …, 12); *k*: team *k* in MLB (k=1, …, 30)

Table 2

Results of Cluster Analysis

Group 1 (n_1=8)		**Group 2** (n_2=7)		**Group 3** (n_3=11)		**Group 4** (n_4=4)	
NY Yankees	(44,956)	Philadelphia	(36,010)	Toronto	(29,890)	Cleveland	(21,960)
LA Dodgers	(44,948)	Detroit	(34,071)	Houston	(29,614)	Oakland	(20,749)
St. Louis	(41,930)	Milwaukee	(33,677)	San Diego	(28,448)	Miami	(19,709)
San Francisco	(39,775)	Colorado	(32,736)	Washington	(28,302)	Tampa Bay	(18,650)
LA Angels	(39,086)	Texas	(32,399)	Arizona	(28,226)		
Chicago Cubs	(37,627)	Minnesota	(30,275)	Seattle	(26,600)		
Boston	(36,728)	Atlanta	(29,890)	Cincinnati	(26,550)		
NY Mets	(35,205)			Baltimore	(26,525)		
				White Sox	(26,129)		
				Pittsburgh	(26,016)		
				Kansas City	(24,695)		
Mean	40,032	Mean	32,723	Mean	26,777	Mean	20,267
SD	3,644	*SD*	2,146	*SD*	1,784	*SD*	1,417

Note. The average of seasonal average attendance during over 12 seasons in parentheses

Table 3

Descriptive Statistics of Overall and each Group

Variables	**Overall** (*N*=29155)		**Group 1** (*N*=7774)		**Group 2** (*N*=6806)		**Group 3** (*N*=10688)		**Group 4** (*N*=3886)	
	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
Dependent Variable										
Att	30848	(10376)	40038	(6382)	32736	(8484)	26792	(8788)	20279	(7467)
(Game-level) Independent Variables										
HTQ	0.500	(0.061)	0.531	(0.054)	0.502	(0.057)	0.478	(0.061)	0.493	(0.055)
VTQ	0.500	(0.061)	0.498	(0.060)	0.497	(0.061)	0.503	(0.061)	0.503	(0.062)
Uncertainty	0.072	(0.051)	0.072	(0.050)	0.069	(0.050)	0.075	(0.053)	0.070	(0.049)
VT_Payroll	3619108	(1493494)	3487997	(1460296)	3555978	(1455710)	3677144	(1517856)	3832326	(1525020)
VT_FinalRank	3.00	(1.44)	3.06	(1.44)	3.03	(1.44)	2.96	(1.46)	2.94	(1.41)
VT_StarPlayers	2.00	(1.37)	1.97	(1.34)	1.93	(1.34)	2.03	(1.38)	2.10	(1.43)
VT_Age	80.56	(43.5)	77.86	(45.2)	80.06	(43.56)	83.03	(43.31)	80.01	(39.97)
VT_Champs	3.55	(5.1)	3.27	(4.76)	3.31	(4.49)	3.81	(5.36)	3.82	(5.92)
Rival	0.12	(0.33)	0.17	(0.37)	0.14	(0.35)	0.08	(0.28)	0.12	(0.32)
Weekend	0.48	(0.50)	0.48	(0.50)	0.48	(0.50)	0.48	(0.50)	0.48	(0.50)
(Season-level) Independent Variables										
HT_Payroll	3619142	(1493643)	4928661	(1401240)	3606724	(1251562)	3161349	(1145882)	2280406	(810076)
HT_FinalRank	3.00	(1.44)	2.42	(1.37)	2.84	(1.4)	3.49	(1.37)	3.10	(1.36)

HT_StarPlayers	2.00	(1.37)	2.38	(1.5)	2.23	(1.29)	1.77	(1.26)	1.50	(1.22)
HT_Age	80.56	(43.5)	104.00	(32.71)	85.20	(43.51)	66.51	(40.31)	63.25	(47.42)
HT_Champs	3.55	(5.10)	7.70	(7.78)	1.68	(1.58)	1.83	(1.85)	3.25	(3.42)
STD_Age	22.68	(24.32)	41.34	(37.43)	12.45	(6.20)	15.08	(10.99)	24.13	(14.00)
Capacity	43670	(5288)	45515	(6057)	45490	(3853)	43503	(3918)	37252	(3779)
Ticket	27.14	(9.54)	35.81	(11.71)	24.52	(6.59)	24.28	(6.32)	22.21	(3.58)
Income	47859	(13160)	53832	(13585)	42657	(9639)	50095	(12135)	38873	(12214)
ProTeams	2.70	(1.39)	4.00	(1.73)	2.57	(0.73)	2.00	(0.95)	2.25	(0.43)
Population	1523285	(2042557)	3190825	(3225563)	714617	(354119)	1214125	(883478)	454054	(90001)

Table 4

The results of Variance and Intraclass Correlations by Level

Parameter	**Group 1**		**Group 2**		**Group 3**		**Group 4**	
	Estimate	(ICC)	Estimate	(ICC)	Estimate	(ICC)	Estimate	(ICC)
Within Season (σ^2)	14,773,113	(0.362)	46,939,214	(0.652)	57,969,572	(0.751)	45,738,924	(0.821)
Between Seasons (τ_0^2)	15,748,719	(0.386)	23,066,066	(0.320)	17,904,909	(0.232)	9,314,891	(0.167)
Between Teams (φ_0^2)	10,280,387	(0.252)	1,980,803	(0.028)	1,335,942	(0.017)	679,817	(0.012)
Total	40,802,219		71,986,083		77,210,423		55,733,632	

Table 5

Results of Full Models for each Group

	Group 1 (ML)		Group 2 (ML)		Group 3 (ML)		Group 4 (ML)	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Fixed Effect								
Intercept	** 41301	878	** 25332	1613	** 22906	894	** 18158	895
(Game-level) Independent Variables								
HTQ	** 15821	1768	** 45563	2835	** 34604	2410	** 28228	3000
VTQ	** 4502	1084	2797	1809	* 3664	1451	2964	2304
Uncertainty	1455	988	-1248	1533	* 2670	1311	3532	2020
VT_Payroll	** 0.29	0.04	** 0.60	0.06	** 0.74	0.05	** 0.78	0.09
VT_FinalRank	28.31	38.38	106.45	69.43	** 198.96	54.64	-20.09	88.60
VT_StarPlayers	-38.31	36.37	* 189.43	65.93	75.48	52.29	* 221.03	80.84
VT_Age	0.92	1.04	* 4.04	1.89	** 11.20	1.58	* 7.63	2.63
VT_Champs	0.50	11.04	** 126.31	19.32	** 129.54	14.18	** 194.04	20.98
Rival	** 1203	116	** 991	205	** 2525	233	** 1571	292
Weekend	** 2427	79	** 6210	139	** 7595	116	** 5846	183
Progress	** 51.93	3.44	** 92.82	6.05	** 98.84	5.01	** 43.66	7.90
Progress^2	** -0.28	0.02	** -0.49	0.04	** -0.55	0.03	** -0.20	0.05
(Season-level) Independent Variables								
HT_Payroll	** 1.24	0.25	** 1.87	0.35	** 1.57	0.32	** 1.52	0.20
HT_FinalRank	67.80	168.86	222.70	244.19	245.91	218.43	141.44	122.7
HT_StarPlayers	9.96	153.58	* 615.83	256.24	319.01	250.82	** 756.12	110.81

HT_Age	-14.07	21.73	** -70.49	18.92	* -32.38	16.47	** 106.22	17.54
HT_Champs	67.15	81.81	762.01	543.18	-294.92	378.56	-372.20	227.87
STD_Age	17.10	15.40	** -304.67	67.29	** -160.97	24.84	** -195.86	36.34
Capacity	** 740.29	71.16	* 441.97	136.32	-121.12	77.34	68.45	66.46
Ticket	16.57	38.45	85.85	123.49	-2.33	60.93	* 152.21	75.25
Income	2.59	31.15	76.84	62.17	* -98.63	34.79	** 301.44	41.04
ProTeams	* -1533	594	* -1980	712	* -680	323	** 8084	1409
Population	-0.15	0.25	* 6.00	1.91	0.56	0.36	* 16.09	6.01
Season	** -1099	265	* 1128	472	** -1281	304	193	283
Season^2	* 50.15	19.74	** -117.75	32.01	** 95.83	23.66	** -79.29	21.98
Random Effect								
Residual (σ^2)	11,966,598	193,388	32,586,742	562,974	35,795,248	493625	32091766	731122
τ_0^2	3,381,069	557,345	6,301,612	1,034,718	6,960,686	911783	450757	239440
φ_0^2	872,514	846,856						
-2 Log Likelihood	148692.1		136867.6		215813.6		77832.2	
***R*-Square**								
R^2	0.602		0.459		0.446		0.416	

Note. * $p<.05$, ** $p<.001$, and ML means using maximum likelihood estimation.

References

Adler, D. (2023, October 2). *Attendance up, game time down in record-setting '23 season*. MLB.com. https://www.mlb.com/news/mlb-attendance-mlb-tv-post-big-numbers-in-2023

Afifi, A., Clark, V. A., & May, S. (2003). *Computer-aided Multivariate Analysis* (4th ed.). CRC Press.

Ahn, S. C., & Lee, Y. H. (2007). Lifecycle demand for Major League Baseball. *International Journal of Sport Finance, 2*(2), 79-93. https://doi.org/10.1177/155862350700200203

Arizona Diamondbacks Schedule. (2017). *ESPN*. http://www.espn.com/mlb/team/schedule/_/name/ari/season/2017

Associated Press. (2022, October 6). *MLB crowds up from 2021, but still below pre-pandemic levels*. *ESPN*. https://www.espn.com/mlb/story/_/id/34741056/mlb-crowds-2021-pre-pandemic-levels

Baade, R. A., & Tiehen, L. J. (1990). An analysis of Major League Baseball attendance, 1969–1987. *Journal of Sport and Social Issues*, *14*(1), 14–32. https://doi.org/10.1177/019372359001400102

Becker, M. A., & Suls, J. (1983). Take me out to the ballgame: The effects of objective, social, and temporal performance information on attendance at major league baseball games. *Journal of Sport and Exercise Psychology*, *5*(3), 302-313. https://doi.org/10.1123/jsp.5.3.302

Borland, J., & Macdonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, *19*(4), 478–502. https://doi.org/10.1093/oxrep/19.4.478

Brown, M. (2022, October 6). *MLB attendance for 2022 down nearly 6% from 2019, last year before the pandemic*. Forbes. https://www.forbes.com/sites/maurybrown/2022/10/06/mlb-attendance-for-2022-down-nearly-5-from-2019-last-year-before-the-pandemic/

Castrovince, A. (2023, December 21). *Widening of runner's lane to 1B among rule changes for '24*. MLB.com. https://www.mlb.com/news/mlb-2024-rule-changes-runners-lane-to-first-base

Coates, D., & Humphreys, B. R. (2007). Ticket prices, concessions and attendance at professional sporting events. *International Journal of Sport Finance*, *2*(3), 161-170. https://doi.org/10.1177/155862350700200305

DeSchriver, T. D., & Jensen, P. E. (2002). Determinants of spectator attendance at NCAA Division II football contests. *Journal of Sport Management*, *16*(4), 311–330.

Fort, R. (2004). Inelastic sports pricing. *Managerial and Decision Economics*, *25*(2), 87-94. https://www.jstor.org/stable/25151279

Hansen, H., & Gauthier, R. (1989). Factors affecting attendance at professional sporting events. *Journal of Sport Management*, *3*(1), 15-32.

Lemke, R. J., Leonard, M., & Tlhokwane, K. (2010). Estimating attendance at Major League Baseball games for the 2007 season. *Journal of Sports Economics*, *11*(3), 316-348. https://doi.org/10.1177/1527002509337212

Lim, N., & Pedersen, P. M. (2022). Examining determinants of sport event attendance: A multilevel analysis of a Major League Baseball season. *Journal of Global Sport Management*. *7*(1), 181-198. https://doi.org/10.1080/24704067.2018.1537675

McCallister, D. (2018, February 20). MLB sets new rules to speed up the game. *National Public Radio*. https://www.npr.org/sections/thetwo-way/2018/02/20/587181780/rules-for-2018-aim-to-speed-up-how-long-it-takes-to-play-a-mlb-game

McEvoy, C. D., Nagel, M. S., DeSchriver, T. D., & Brown, M. T. (2005). Facility age and attendance in Major League Baseball. *Sport Management Review*, *8*(1), 19-41. https://doi.org/10.1016/S1441-3523(05)70031-0

MLB Attendance Report. (2017). *ESPN*. http://www.espn.com/mlb/attendance/_/year/2017

Posnanski, J. (2017, February 22). Anticipated rule could continue decline of IBB. *MLB.com*. https://www.mlb.com/news/2018-19-mlb-offseason-analysis-for-each-move/c-302712514

Randhawa, M. (2025, January 23). MLB implementing 2 minor rules changes starting this season. *MLB.com*. https://www.mlb.com/news/mlb-shift-baserunning-rules-changes-2025

Rivers, D. H., & DeSchriver, T. D. (2002). Star players, payroll distribution, and Major League Baseball attendance. *Sport Marketing Quarterly*, *11*(3), 164–173. https://doi.org/10.1177/106169340201100304

Schofield, J. A. (1983). Performance and attendance at professional team sports. *Journal of Sport Behavior*, *6*(4), 196–206.

Tainsky, S., & McEvoy, C. D. (2012). Television broadcast demand in markets without local teams. *Journal of Sports Economics*, *13*(3), 250-265. https://doi.org/10.1177/1527002511406129

Welki, A. M., & Zlatoper, T. J. (1994). US professional football: The demand for game-day attendance in 1991. *Managerial and Decision Economics*, *15*(5), 489-495. https://doi.org/10.1002/mde.4090150510

Whitney, J. D. (1988). Winning games versus winning championships: The economics of fan interest and team performance. *Economic Inquiry*, *26*(4), 703-724. https://doi.org/10.1111/j.1465-7295.1988.tb01524.x

Zhang, J. J., Lam, E. T., & Connaughton, D. P. (2003). General market demand variables associated with professional sport consumption. *International Journal of Sports Marketing and Sponsorship*, *5*(1), 24-46. https://doi.org/10.1108/IJSMS-05-01-2003-B003