

Factors Affecting Attendance of Major League Baseball: II. A Within-Season Analysis

John P. Marcum
University of Mississippi

Theodore N. Greenstein
University of Texas at Arlington

This study examines game-by-game attendance data for one National League team (St. Louis Cardinals) and one American League team (Texas Rangers) for the 1982 season to explore factors related to attendance at professional sporting events. Multiple regression analyses indicate that the major factors affecting attendance are day of the week, opponent, and type of promotion. Recent and season-long performance measures for both the home and visiting teams have relatively little effect on daily attendance.

Every year millions of Americans attend professional and amateur sporting events. A recent nationwide survey (Miller Brewing Company, 1983) found that 79% of those polled had attended at least one sporting event during the previous year, and a majority of men (as well as 48% of women) reported attending six or more events during that period. Despite the pervasiveness of this spectator role in American society, the forces that affect sports attendance are not completely understood. Why do some events attract more fans than others?

A variety of factors have been proposed to explain variation in sports attendance, from team quality (Demmert, 1973; Canes, 1974) to population size of the host community (Demmert, 1973; Noll, 1974) to the vagaries of weather (Siegfried & Hinshaw, 1977). In a previous paper (Greenstein & Marcum, 1981), we suggested the factors that affect attendance at sports events fall into three broad categories. The first category consists of *sociodemographic* characteristics such as population composition (age-sex structure, size, distribution), social status, and economic conditions. Also included would be the presence of competing recreational and cultural activities, such as another professional sport team (Demmert, 1973). A second group of factors concerns the *accessibility* of the event, to include relatively permanent factors such as seating availability, location of the event, and

An early version of this paper was presented at the annual meetings of the American Sociological Association, San Antonio, August 1984. The authors wish to acknowledge the assistance of Dave Fendrick of the Texas Rangers and Marty Hendin of the St. Louis Cardinals for providing some of the data reported herein, and Kofi Benefo (now of the University of Michigan) for his help in constructing the database.

Direct all correspondence to John Marcum, Dept. of Sociology-Anthropology, The University of Mississippi, University, MS 38677.

transportation facilities as well as more transient factors such as day of the week (Jones, 1969) and weather (Siegfried & Hinshaw, 1977). The third category, *performance*, includes indicators of team quality such as won-lost record and relative position in league standings, as well as individual achievement such as the presence of a star player (Davenport, 1969; Scully, 1974).

Studies of attendance at sporting events have examined a variety of professional sports: basketball (Noll & Okner, 1972), cricket (Schofield, 1983), football (Federal Communications Commission, 1978), and hockey (Jones, 1969). However, disproportionate attention has been directed at baseball (e.g., Demmert, 1973; Scully, 1974; Davenport, 1969; Becker & Suls, 1983; Hunt & Lewis, 1976; Greenstein & Marcum, 1981) for a variety of reasons. First, its form has remained largely unchanged for decades, in contrast to other popular American professional team sports (Demmert, 1973:55). Combined with the fact that the baseball season has many more contests than any other professional sport, the result is an extremely large data base for analysis. Second, there is, in Guttman's words, "the tendency of baseball toward extremes of quantification" (1978:100). Baseball officials and fans both seem to have a mania for record-keeping; statistics are readily available (e.g., Reichler, 1985) on every team and individual ever to play in the major leagues. Third, baseball tickets are typically lower than those of other professional sports, making attendance financially practical for virtually anyone. Fourth, season ticket sales in Major League Baseball comprise a relatively small proportion of total ticket sales, thus allowing greater variability in attendance levels.¹ Finally, a relatively large share of the American population is fascinated with baseball. Even though football seems to have surpassed baseball as the favorite sport of Americans, 28% of those polled in a national survey reported that they "always or usually are interested" in baseball (Miller Brewing Company, 1983).

The emphasis in previous research has been on the effect of performance on attendance. Despite a variety of measures, studies of Major League Baseball uniformly reveal a strong, positive relation between success on the field and attendance when the entire season is the unit of analysis. Demmert (1973:64-66), using data for 16 clubs from 1951-1969, found that the number of games a team finished behind the divisional leader was negatively associated with home attendance even when a variety of sociodemographic factors had been controlled. On average, teams had 25,000 fewer paying fans over the course of a season for each game they finished behind the division leader. Similarly, in an examination of major league records for 1920-1968, Canes (1974) found that a one-place rise in the standings was accompanied by an average increase of 84,000 fans in combined road and home attendance. Noll (1974) found attendance for the 1970 and 1971 seasons positively associated with the number of recent pennants won, number of star players on the team, and closeness of the pennant race (cf. Scully, 1974; Becker & Suls, 1983; Davenport, 1969).

Our own previous research (Greenstein & Marcum, 1981) documented the importance of won-lost record for attendance using season-long performance and home attendance measures for all teams in the National League of Professional Baseball Clubs from 1946-1975 (a dataset of 282 cases in which each team-season constituted an observation). Multiple regression analyses predicting proportion of seats filled indicated that linear composites representing offense (team doubles, triples, home runs, batting average, slugging average, and stolen bases), defense (team errors, fielding average, and double plays), and pitching (team complete games, walks, strikeouts, shutouts, saves, and earned-run average)—along with won-lost percentage—accounted for about 27% of the variation in attendance. Surprisingly, however, *only* won-lost percentage had a significant direct effect on attendance. It alone accounted for just over one-fourth of the variation in attendance.

Fans, of course, do not attend "seasons," they attend individual games, and season-long attendance is a sum of the drawing power of individual games. As we noted previously (Greenstein & Marcum, 1981), many variables (e.g., weather, day of the week, promotions) likely to affect attendance fluctuate throughout the season, and to study their effects on attendance requires game-by-game information. Yet, except for the occasional study of a star pitcher's effect on attendance at those games he starts,² game-by-game analyses of baseball attendance have been lacking. This omission has been attributed both to the difficulty of acquiring data for each game (Davenport, 1969) and to the need, once acquired, of "laboriously analyzing attendance for every team for every game during the season" (Scully, 1974:918). Although we are unable to overcome these problems entirely in the present research, we undertake a beginning step toward narrowing the focus (and unit of analysis) from the *season* to the *individual game*.

We analyze game-by-game records for two teams for the 1982 season—the Texas Rangers of the American League and the St. Louis Cardinals of the National League. The research is exploratory, and we make no formal hypotheses. While it would be unwise to generalize from this limited dataset, the Cardinals and Rangers had very different seasons in 1982, and a comparison of the relative effects of different factors on these two teams' home attendances should tell us a great deal. The Cardinals won their division by three games and defeated Milwaukee in seven games to win the World Series,³ while Texas lost 98 games and finished 29 games out of first place. Perhaps even more important in an analysis of attendance, the Cardinals were in first place for over three-fourths of their home openings in a hotly contested race whereas the Rangers were never competitive.

The data for this study comes from a variety of sources. Some of the data—attendance figures and information on team promotions—was provided by the club offices. Team performance information was gathered from published box scores and league standings. Weather information was obtained from the official NOAA publication *Local Climatological Data* (1983).

We have selected a variety of independent variables based on previous research (see especially Demmert, 1973; Noll, 1974; Jones, 1969; Greenstein & Marcum, 1981). Some variables that might otherwise have been included were omitted due to practical considerations, mainly difficulty of measurement. A good example would be other events taking place in the metropolitan area that might siphon off potential baseball fans. However, Noll (1974) notes that population provides a proxy for competing entertainments. Because the populations of Dallas-Fort Worth and St. Louis are so similar in size—3,821,622 and 3,684,727, respectively, based on 1980 census data (U.S. Bureau of the Census, 1981) for counties within a 100-mile radius of Arlington and Busch Memorial stadiums—this factor is unlikely to differ significantly for the two clubs. Further, local television coverage of road games, a practice found by Demmert (1973:69) to be associated with lower home attendance, was similar for the two clubs during the 1982 season in that both clubs televised about half of their away games via commercial television.

Method

The data for this research consists of one observation for each home opening for the Texas Rangers (78 openings) and the St. Louis Cardinals (80 openings) for the 1982 regular season.⁴ Measures of 10 factors that could reasonably be expected to affect attendance were gathered: day of the week,⁵ type of team promotion, opposing team, oppo-

ment's won-lost percentage, opponent's games behind first place, home team's won-lost percentage, home team's games behind first place, home team's record over the previous 10 decisions, weather conditions, and whether the opening involved a doubleheader.

The dependent variable in these analyses is attendance, which, by necessity, is operationalized in slightly different ways for the two teams. American League attendance data is reported in terms of total ticket sales (i.e., those who purchase tickets but are not present in the stadium are included in the attendance total), while National League figures are for turnstile count only (no-shows are not included in the attendance total even though they may have purchased a reserved seat ticket). This difference should not present a major problem for interpretation, however; inspection of data available from the Rangers organization (not reported here) indicates that the American League attendance figures overstate turnstile count data only slightly.

To allow for direct comparisons between the two ballclubs' data, we standardized the attendance figures by dividing by stadium capacity. This is especially important because Busch Stadium in St. Louis officially seats 50,222 in its baseball configuration while Arlington Stadium seats 41,284.

Dummy-variable coding was used for categorical variables such as day of the week, opponent, and so forth. Weather conditions were coded as 0 = bad (temperature 5 degrees lower than normal in April and September, 5 degrees higher than normal in other months; rain or other inclement weather), and 1 = other. Promotions were classified into two categories: major promotions involved major giveaways such as Cap Night or Bat Night; minor promotions were all others. The Cardinals held 8 major and 36 minor promotions; the Rangers, 7 major and 34 minor promotions.

Results

Table 1 presents data on the distribution of the attendance variable for both ballclubs. The median proportion of seats filled was .499 for St. Louis and .310 for Texas, with means of .526 and .358, respectively.

Table 2 presents results of the multiple regression analyses for both clubs. These are unstandardized multiple regression coefficients; because the dependent variable has been rescaled as the *proportion* of seats filled, the coefficients are interpretable as the increase in proportion of seats filled attributable to a specific factor, controlling for the ef-

Table 1
Distribution of Proportion of Seats Filled, 1982 Season

Percentile	Cardinals	Rangers
90	.773	.705
75	.630	.419
50 (median)	.499	.310
25	.396	.220
10	.314	.205

Table 2
Unstandardized Regression Coefficients Predicting
Proportion of Seats Filled, 1982 Season

Factor	Cardinals	Rangers
Intercept	.458*	.320*
Sunday	.086	.106**
Monday	.024	.051
Tuesday	.003	—
Wednesday	—	.036
Thursday	.066	.033
Friday	.083	.093*
Saturday	.220***	.157***
Games behind	.008	-.001
Percentage	-.611*	-.111
Last 10	.316*	.114
Opp. record	-.013	-.429*
Opp. behind	-.002	-.007
Weather	.008	.052*
Twinbills	.325*	.076
Minor promotions	.049	.060*
Major promotions	.277***	.471***
R ²	.66	.88

* $p < .05$, ** $p < .01$, *** $p < .001$ for test of significance from zero.

fects of all other factors in the model. For example, Table 2 shows that the unstandardized regression coefficient for the weather factor had a value of .052 for the Rangers' data. This means that there was an increase of 5.2% in seats filled on days with good weather as opposed to days with poor weather. The asterisk next to this value indicates that the significance test for this particular coefficient shows that it is significantly different from zero at the .05 probability level.

It is interesting to note that only two of the factors (in addition to the intercept value) are significantly different from zero for both ballclubs: Saturday attendance and major promotions. If the opponent variable is coded into a single factor, it also shows a significant coefficient although it is not directly interpretable and so is not included here. The net increase in attendance attributable to opponent is about 3% for the Cardinals and just over 6% for the Rangers.

The overall fit of the model to the data is quite good for both ballclubs;⁶ the 10 factors account for 66% of the variation in Cardinal attendance and 88% of the variation in the Rangers' data. Both of these values are significantly different from zero. To get a better idea of the relative effects of these factors, Table 3 presents the results of the multiple regression analyses in a different way. The entries in this table represent the proportion of variation in attendance attributable to a particular factor. For example, a little over 14% of the variation in Cardinals' attendance is accounted for by the opponent factor (after controlling for the effects of other factors in the model), while opponent accounts

Table 3
Proportion of Variation in Attendance Explained
by 11 Major Factors, 1982 Season

Factor	Cardinals	Rangers
Opponent	.143	.287
Day of the week	.289	.211
Games behind	.000	.015
Percentage	.004	.004
Last 10	.045	.004
Opp. record	.001	.010
Opp. behind	.008	.005
Weather	.000	.011
Twinbills	.030	.003
Minor promotions	.005	.005
Major promotions	.147	.335
Total variation explained (R^2)	.66	.88

for about twice that amount—28.7%—of the variation in Rangers' attendance. Table 3 shows that three factors—day of the week, opponent, and type of promotion—account for the bulk of the explained variation in attendance for both clubs. Nearly 60% of the variation in Cardinals' home attendance and over 80% of the variation in Rangers' home attendance is attributable to just these three factors.

Discussion

Perhaps the most interesting finding in these data is that most of the variation in attendance can be attributed to three factors: opponent, day of the week, and promotions. While the order of effects is different—promotions are the single most important factor in the Rangers' data whereas day of the week contributes the most in the Cardinals' analysis—nearly all the explained variation in both datasets can be attributed to these three variables. None of the other factors accounts for as much as 5% of the variation in either dataset.

In comparing the differences in effects for the two clubs, we find that promotions and opposing team have far more effect on attendance for the Rangers. Considering the poor season that the Rangers had, this is not unexpected. As Canes (1974) has observed, quality of the visiting team likely also relates positively to attendance, and for a poor home team the visitors' won-lost records may take on greater importance in fans' decisions about whether to attend particular games. Similarly, performance over the previous 10 decisions had a significant effect on Cardinal attendance, but not for the Rangers, suggesting that nonperformance measures have much more effect on attendance for losing teams than for winners. When a team is doing well, attendance will be good regardless of which club

is in town or what the management is giving away at the ballpark. When a team is doing poorly, however, fans need an extra incentive to show up; thus, giveaways and first-place opponents increase attendance.

Another interesting point is the difference in total explained variation for the two ballclubs. The 10 factors account for nearly 90% of the variation in Ranger attendance but less than 70% for the Cardinals. Again, we believe the difference in team performance may explain this finding. In particular, some writers (e.g., Davenport, 1969:6-7; Demmert, 1973:14-15) have argued that the existence of a close pennant race, independent of won-lost record, increases attendance. While prior empirical evidence is mixed (Demmert, 1973:67, fails to find support for this hypothesis while Davenport, 1969:6-7, does), our results seem consistent with this argument: Notice how much larger the intercept term is for the Cardinals' data than for the Rangers'. At the same time, past performance may be more important for Cardinals' as opposed to Rangers' fans. With its longer and more successful history, St. Louis may attract fans based on memories of past Cardinal championships, a memory Ranger fans do not have (Becker & Suls, 1983).

We advise caution in extrapolating these findings to other teams because there are major differences between the Cardinals and the Rangers beyond won-lost percentage. For example, differences in importance of day of the week are probably better understood in terms of the Cardinals' greater regional following. For many fans, distance from the ballpark precludes attending any games except on weekends, and such long excursions to St. Louis are often planned well in advance of the actual game and hence are relatively unaffected by current won-lost record or weather at game time.

One other issue is raised by the rather high values of explained variation for these data. We are already accounting for well over 60% of the variation in the Cardinals' attendance and nearly 90% for the Rangers'. Such high values lead us to suspect there are probably few, if any, other variables we could add to this model that would produce an appreciable increase in explained variation. Adding such esoteric factors as whether a star player is nearing a record or who the starting pitchers are will not likely increase the amount of explained variation appreciably. Rather, our future analyses will focus on interactions among these variables. Such analyses must await larger datasets, however, because even a limited study of interaction effects will require a model with a large number of terms and a corresponding decrease in the degrees of freedom term.

In our future research on this topic, we would like to study the effects of socio-demographic factors on attendance. How have population shifts affected attendance of Major League Baseball? What about changes in ethnic and racial composition, both in the stands and on the playing field? What is the role of economic factors such as ticket prices and unemployment? How does accessibility to the stadium affect attendance? Finally, while research suggests that Major League Baseball attendance is not unique (e.g., Jones, 1969; Noll & Okner, 1972), amateur and minor league professional sports have yet to be subjected to the same scrutiny. It would be useful to extend this research to all levels of sport.

References

Becker, M.A. and J. Suls

- 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 Psychology*, 5:302-313.

- Canes, M.E.
 1974 "The social benefits of restrictions on team quality." Chapter 3 in R.G. Noll (ed.), *Government and the Sports Business*. Washington: The Brookings Institution.
- Davenport, D.S.
 1969 "Collusive competition in Major League Baseball: its theory and institutional development." *The American Economist*, 13:6-30.
- Demmert, H.G.
 1973 *The Economics of Professional Team Sports*. Lexington, MA: D.C. Heath.
- Federal Communications Commission
 1978 *The Effect of the Sports Anti-Blackout Law. Fifth Annual Report*. Washington: Government Printing Office.
- Greenstein, T.N. and J.P. Marcum
 1981 "Factors affecting attendance of Major League Baseball: I. team performance." *Review of Sport and Leisure*, 6:21-33.
- Guttman, A.
 1978 *From Ritual to Record*. New York: Columbia University Press.
- Hunt, J.W. Jr. and K.A. Lewis
 1976 "Dominance, recontracting, and the reverse clause: Major League Baseball." *The American Economic Review*, 66:936-943.
- Jones, J.C.H.
 1969 "The economics of the National Hockey League." *Canadian Journal of Economics*, 2:1-20.
- Miller Brewing Company
 1983 *The Miller Lite Report on American Attitudes Toward Sport*. Milwaukee: Miller Brewing Company.
- National Oceanic and Atmospheric Administration
 1983 *Local Climatological Data*. Washington, DC: Government Printing Office.
- Noll, R.G.
 1974 "Attendance and price setting." Chapter 4 in R.G. Noll (ed.), *Government and the Sports Business*. Washington: The Brookings Institute.
- Noll, R.G. and B.A. Okner
 1972 *Professional Basketball. Testimony before the subcommittee on Antitrust and Monopoly of the Committee on the Judiciary, United States Senate. 5:2373, A Bill to Allow the Merger of Two or More Professional Basketball Leagues. September 21, 1971*. Washington: Government Printing Office.
- Reichler, J.L. (ed.)
 1985 *The Baseball Encyclopedia* (6th edition). New York: Macmillan.
- Schofield, J.A.
 1983 "The demand for cricket: the case of the John Player League." *Applied Economics*, 15:283-296.
- Scully, G.W.
 1974 "Pay and performance in Major League Baseball." *The American Economic Review*, 64:915-930.
- Siegfried, J.J. and C.E. Hinshaw
 1977 "Professional football and the anti-blackout law." *Journal of Communication*, 27:169-174.
- U.S. Bureau of the Census
 1981 *Census of Population and Housing, 1980: Summary File 1C* (microfiche). Washington: U.S. Bureau of the Census.

Notes

¹Jones (1969) reports that the Toronto and Montreal National Hockey League teams sell out their respective arenas year after year, and attendance at professional football games is so high

that studies have focused on "no-shows" rather than on attendance per se (Siegfried & Hinshaw, 1977).

²Davenport (1969), for example, found that during the 1966 season Sandy Koufax (with a 27-9 record) increased attendance by an average of 6,000 at home and 7,000 on the road per game.

³Our analysis is limited to regular season play.

⁴An opening is a game or pair of games for which a separate admission is charged. Most openings are single games, and most doubleheaders are a single opening.

⁵Time of game (day or night) was excluded because of potential multicollinearity problems with day of the week. Texas plays few day games and, for both clubs, day games were concentrated on weekends.

⁶The opponent variable does not appear in this table due to the noncomparability of opponents for the two teams. Opponent was coded as a series of dummy variables for each team and included in the analysis, however, so the effects reported in Table 2 control for the effects of opponent. See Table 3 for an estimate of the amount of variation in attendance attributable to the opponent variable.