

Joey Eovaldi  
12/9/20  
Syracuse University

## MiLB Attendance – Eugene Emeralds:



### **Introduction**

The Eugene Emeralds play their home games at PK Park in Eugene, Oregon. They are a Class A Short Season Affiliate of the Chicago Cubs and have been playing at PK Park since 2010. The capacity of PK Park is approximately 4,000, with standing room. The Emeralds mascot is Sluggo, a version of *Sasquatch*, because of the widely-known belief that *Sasquatch* lives in the mountains very close to Eugene. The Emeralds are in the Northwest League and have won six championships, dating back to 1955 (when they won the first championship in League history). Recently, the Emeralds won the Northwest League in 2016 and in 2018 and appeared in the championship series in 2016, 2017, and 2018. There was no season this past summer (2020) due to the worldwide COVID-19 pandemic, so I collected data from the 2019 season, which took place from June until September. The Emeralds home opener in 2019 was on Friday, June 14 vs. the Hillsboro Hops. They played 38 home games spanning from their home opener to their home closer on Saturday, August 31 vs. the Salem-Keizer Volcanoes.

### **Summary**

I looked at each home game the Emeralds the summer of 2019 and collected data on multiple variables to measure the biggest factors of attendance in Eugene. My model is below.

```
Model <- lm(Attendance ~ DivisionGame + Sunday + Monday + Tuesday + Wednesday +  
            Thursday + Friday + HomeWinPercentage + HomeWinTotal +  
            Temperature + Humidity + WindSpeed + Precipitation +  
            TwoPromotions + ThreePromotions + Giveaway + HomeOpener + June +  
            July, data = MiLB)
```

There were 24 division games, 5 games played on Sunday, 4 on Monday and Tuesday, 6 on Wednesday, 7 on Thursday and Friday, and 5 on Saturday. The Emeralds hosted 7 home games in June, 14 in July, and 17 in August. They won 18 of the 38 games at home in 2019. On game days, the average temperature in Eugene was 67.4 degrees Fahrenheit. The average humidity was 61.53% and the average wind speed was 6.45 miles per hour (MPH). There was a total of half an inch of rain all summer (on game days) – sounds like a great summer for baseball. Of the 38 home games last summer, there were 33 games with two promotions, 3 games with three promotions, and two games with one promotion. Of those promotions, there were 12 games with giveaways to early arriving fans. So while the sample size is relatively small, the data is pretty decently distributed across the variables being used.

### **Model Overview**

The model I originally ran had 19 variables, which is a lot for a model. But with the small sample size, I wanted to include all possible factors (that I had data on) to get the most accurate results on what impacts attendance in Eugene the most. As I had data on every Emeralds game, I had their opponents and was able to create the variable, “DivisionGame.” I used this variable to see how many more fans come to divisional games, as those games are more important throughout a season. I didn’t expect these games to bring that many more fans though, as they are minor league games, which are much more about entertainment than competition (especially

compared to MLB). Although, the Emeralds have had recent successful seasons, so I definitely expected some raise in attendance for divisional games. I included each day of the week as a variable, to look more specifically at how the attendance on each of the week fares. This had its pros and cons. As there were only 38 home games, each day of the week did not have a lot of games played on them. But we have more specific data looking at each day of the week. I also included variables on both the Emeralds home win totals and home win percentage to see how their recent success at home impacted attendance. I included four variables about weather, each representing something different to see which variable fans cared about the most when attending games. I included variables representing if there were two or three promotions and if there was a giveaway promotion, to see what drew fans to games. I included a variable for the home opener to see how many more fans go to that game and I had variables for each month played in. This helped show me which month drew the most fans over the summer. After I ran the model and saw some problems, I removed a couple variables to increase its effectiveness.

## Results

My first model has the following regression and coefficients.

Minimum	1Q	Median	3Q	Maximum
-1011.9	-304.29	-44.51	367.09	1209.04

Coefficients	Estimate	Standard Error	T Value	P Value
(intercept)	11320.36	8916.92	1.27	0.2204
DivisionGame	497.13	426.06	1.167	0.2585
Sunday	70.84	648.14	0.109	0.9142
Monday	-2023.9	1410.73	-1.435	0.1685
Tuesday	-1447.62	677.28	-2.137	.0465*
Wednesday	-618.27	571.36	-1.082	0.2035
Thursday	108.05	590.99	0.193	0.857
Friday	253.82	519.67	0.488	0.6311
HomeWinPercentage	-5850.95	3994.35	-1.465	0.1602

HomeWinTotal	-71.21	85.76	-0.83	0.4172
Temperature	-40.56	85.49	-0.474	0.6409
Humidity	-9.31	33.26	-0.28	0.7827
WindSpeed	22.19	97.4	0.228	0.8223
Precipitation	1691.89	2409.86	0.702	0.4916
TwoPromotions	-799.38	992.5	-0.805	0.4311
ThreePromotions	393.94	1083.61	0.364	0.7204
Giveaway	-537	467.33	-1.149	0.2656
HomeOpener	2803.79	2374.02	1.181	0.253
June	-631.3	1344.65	-0.469	0.6444
July	-115.34	1000.48	-0.115	0.9095

While most variables showed at least semi-accurate trends for how they impact attendance, the standard error for each variable is relatively high. Looking at the days of the week, Monday, Tuesday, and Wednesday all negatively impact attendance for the Emeralds. This makes sense, as the Emeralds had 7 Thursday games and those lead into the weekend. Three promotions had a positive impact on attendance while two promotions had a negative impact on attendance (most of the games had two promotions, so there may have been other factors at play though). The home opener had a very positive impact on attendance as well. Games in June and July had a negative impact on attendance compared to games in August. The multiple r-squared for this model was .7075.

### Diagnostics

My first model had no heteroscedasticity or autocorrelation, which was nice.

Heteroscedasticity	<i>Breusch-Pagan test</i>
BP Statistic	17.724
Degrees of Freedom	19
p-value	0.5409

Autocorrelation	<i>Durbin-Watson Test</i>
DW Statistic	2.3569
p-value	0.6391

The p-values of both are larger than .05, so we can accept the null hypothesis, which is that there is no heteroscedasticity (same thing for autocorrelation). There was multicollinearity in 9 variables (the VIF score of a variable was larger than 5).

Multicollinearity	VIF Table
Variable	Score
Division Game	2.65407
Sunday	3.016148
Monday	11.777624
Tuesday	2.714592
Wednesday	2.727445
Thursday	3.298002
Friday	2.550038
HomeWinPercentage	9.086071
HomeWinTotal	16.465832
Temperature	6.147283
Humidity	3.108435
WindSpeed	2.479515
Precipitation	1.552015
TwoPromotions	7.072547
ThreePromotions	5.364923
Giveaway	2.965034
HomeOpener	9.074024
June	17.072863
July	14.634647

With a decent amount of multicollinearity, I took out two variables and re-ran my model. This lowered the multiple r-squared to .689, but it vastly improved multicollinearity issues.

```
> vif(Model)
  DivisionGame      Sunday      Monday      Tuesday
  2.160330      2.863745      6.313355      2.687603
  Wednesday      Thursday      Friday HomeWinPercentage
  2.597409      3.268143      2.514594      9.066605
  Temperature      Humidity      WindSpeed      Precipitation
  5.046565      2.632434      2.333592      1.516110
  TwoPromotions      Giveaway      HomeOpener      June
  6.476757      2.646539      8.979244      5.122624
  July
  4.276349
```

No variables were highly correlated with

each other and only 5 had multicollinearity.

## Conclusions

My final model was very similar to my first model, with a few smaller changes in coefficients of variables and in the VIF table. As I mentioned, the multiple r-squared of the model below is .689.

```
Model <- lm(Attendance ~ DivisionGame + Sunday + Monday + Tuesday + Wednesday +  
Thursday + Friday + HomeWinPercentage +  
Temperature + Humidity + WindSpeed + Precipitation +  
TwoPromotions + Giveaway + HomeOpener + June +  
July, data = MiLB) |
```

Most things did remain similar though, and my conclusions remain as they were for my first model. The factors that negatively influenced attendance were: Monday, Tuesday, and Wednesday games, the teams home performance going into the game, higher temperature, more humidity, and when there were two promotions vs. three promotions. The variable “Giveaways” had a negative coefficient, but I believe that has to do with other variables on the same game days. The other eight variables all positively impacted attendance for the Eugene Emeralds over the 2019 season. Besides the home opener, the biggest draw of fans to Emeralds games is if it is a division game. I think this shows that the fans do care about the success and competition of the team (along with the entertainment, of course). The Emeralds go a great job of selling tickets, as they averaged 3,469 people per game, which is only about 500 less than their capacity. This very well may be due to 24 of the 38 home games being divisional games, as Emeralds fan enjoy going to rivalry games in Eugene.

## Data Sources

“Eugene Emeralds.” *MiLB.com*, [www.milb.com/eugene](http://www.milb.com/eugene).

“Eugene Emeralds Schedule: Schedule.” *MiLB.com*,  
[www.milb.com/eugene/schedule/2019/fullseason](http://www.milb.com/eugene/schedule/2019/fullseason).

“Eugene, OR Weather Historystar\_ratehome.” *Weather Underground*,  
[www.wunderground.com/history/monthly/us/or/eugene/KEUG/date/2019-6](http://www.wunderground.com/history/monthly/us/or/eugene/KEUG/date/2019-6).