

Hugo Belisario

Statistics 398: Analysis of Sports Data

Professor Severini

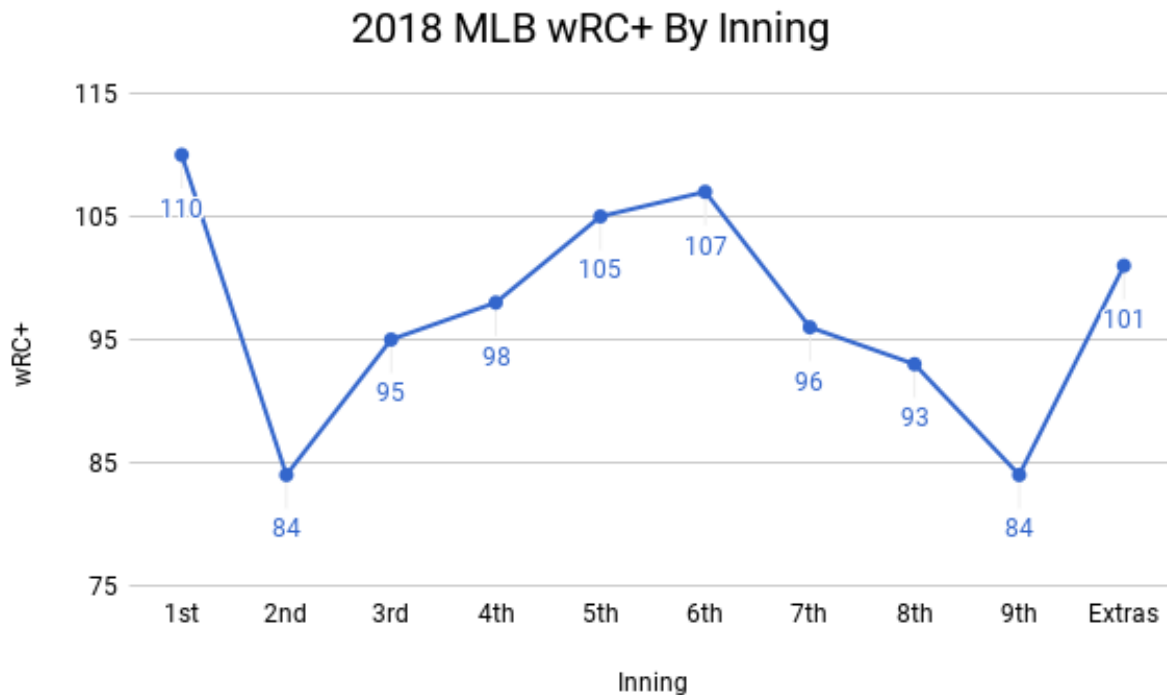
12 June 2019

Openers and Headliners

BACKGROUND

The “opener” strategy in baseball is a fairly new approach to pitching popularized by the Tampa Bay Rays in 2018. The strategy allows the ballclub on defense to set up a good reliever against the top of the offense’s lineup, which are usually three of the best hitters on the team. This explains why the first inning has historically been the highest scoring, in addition to being the only inning where the batting team can guarantee who comes to the plate. It also alleviates the disadvantage of facing the lineup more than twice by having the “bulk guy” or “headliner” pitch after the opener against the middle or bottom of the lineup instead of the top.

According to FanGraphs, $wRC+$ is a rate statistic which attempts to credit a hitter for the value of each outcome (single, double, etc) rather than treating all hits or times on base equally, while also controlling for park effects and the current run environment. The league average $wRC+$ is 100 and every point above or below 100 is equal to one percentage point better or worse than league average. In the line graph below, we can see that in 2018 the first inning had an $wRC+$ of 110, 26 points higher than the second inning where hitters in the middle or bottom of lineup struggle to produce runs.



As mentioned before, the Tampa Bay Rays turned heads when they started reliever Sergio Romo against the Angels twice, throwing two scoreless innings including a perfect inning where he struck out Zack Cozart, Mike Trout, and Justin Upton. However, the Rays weren't the first ones to employ the "opener" according to Jeff Sullivan, formerly from FanGraphs. He tweeted, "everyone thinks the Rays are blazing some new trail but the Orioles have been using one-inning starting pitchers for years" on May 21, 2018. Now, many teams including the A's, Rangers, Angels, Twins, Brewers, Astros, Blue Jays, Dodgers, Mariners, Pirates, and Tigers have followed suit working towards the goal of maximizing back-of-the-rotation starters.

DEFINITIONS

Sabermetrician and Senior Database Architect of Stats for MLB Advanced Media, Tom Tango or “TangoTiger” provided definitions for the “opener” and “headliner” in order to adjust WAR (Wins Above Replacement) for these two types of pitchers. To determine whether we have an “opener” candidate for the starting pitcher of the game, the pitcher has to have either:

- at most two innings (six outs) or
- at most nine batters faced.

To determine whether we have “headliner” candidate for the relief pitcher of the game, the pitcher has to have either:

- at least twelve outs (four innings) or
- at least eighteen batters faced

Also, the relief pitcher is limited to being the:

- first relief pitcher of the game or
- second relief pitcher; if the first relief pitcher entered mid-inning and the second relief pitcher started the next inning.

If both of these pitchers exist, then we have a game with an “opener” and “headliner”.

Otherwise, there is no classification for it.

DATA

The data used in the analysis of the strategy came from The Play Index on the Baseball Reference website. It features game-by-game statistics from the start of the 2018 season to May 30, 2019 using the parameters set by Tom Tango for both types of pitchers. During the process of extracting the “opener” and “headliner” games data, bullpen games (where a team uses multiple relievers for a few outs at a time), games where a regular starter gets injured, or ejected, or hit hard were omitted. Games where a pitcher met the “opener” requirements, but no “headliner” to accompany them following their start were also omitted. By the end, I was left with 85 games where the “opener” and “headliner” were used correctly by Tom Tango standards.

OPENERS

To observe how well these “openers” have performed we look at both their Earned Run Average and Weighted On-Base Average using 2018 linear weights for the outcomes Walks, HBP, Singles, Doubles, Triples, and Home Runs.

ERA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	8.079	9.000	135.000

wOBA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.2200	0.3140	0.3560	0.5078	1.1013

The mean ERA for openers was very high (8.08) during this span with a range of 0.00 to 135.00 ERA. Some openers were used once or twice and had little to no success, hence the inflated maximum ERA. The mean wOBA, on the other hand, was set at an above-average level, however, not too high to indicate any significance that these pitchers were hit hard constantly.

Analysis of Variance

For my linear model, I chose to use ERA as the response variable and Innings Pitched, Batters Faced, Average Leverage Index (aLI), and Win Probability Added (WPA). These parameters stood out as key factors in measuring the performance of pitchers as openers or headliners.

According to Baseball Reference, aLI is the average pressure the pitcher saw in this game with 1 being the average pressure. Openers are susceptible to more pressure since it is a position in which they are not prepared for usually and since they are guaranteed to face the opposing team's best hitters first, there is a certain amount of pressure added to the situation.

According to FanGraphs, WPA captures the change in Win Expectancy from one plate appearance to the next and credits or debits the player based on how much their action increased their team's odd of winning. Clearly, the opener strategy is guided towards giving the team employing it a better chance at winning so including this predictor can show us their value to each game as one inning starters.

```
Linear mixed model fit by REML ['lmerMod']
Formula: Openers$ERA ~ (1 | Openers$Player)
Data: OpenersANOVA
```

```
REML criterion at convergence: 728.6
```

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-2.0778	-0.4048	-0.1393	-0.0076	5.1312

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
Openers\$Player	(Intercept)	1025.2	32.02
Residual		126.2	11.24

Number of obs: 85, groups: Openers\$Player, 30

```
Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	14.275	6.089	2.345

There are 85 observations in this data frame, with 30 openers represented a variety of times.

ERA_{jk} has variance of $\sigma^2_{\alpha} + \sigma^2_{\varepsilon} = 1151.4$ and standard deviation of 33.9323 which is slightly greater than the sample standard deviation (22.99931) of the ERA values in the Openers dataset. $\hat{P} = 0.8904$. About 89.04% of the variation in ERA values is attributable to the opener. The correlation of two ERA values in different years for the same opener is 0.8904. For a given hitter, the correlation of the observed ERA with the opener's true ERA value is 0.9436.

```
Linear mixed model fit by REML ['lmerMod']
Formula: Openers$ERA ~ (1 | Openers$Player) + (1 | Openers$Tm)
Data: OpenersANOVA
```

REML criterion at convergence: 728.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.0875	-0.4075	-0.0881	-0.0320	5.1205

Random effects:

Groups	Name	Variance	Std.Dev.
Openers\$Player	(Intercept)	897.1	29.95
Openers\$Tm	(Intercept)	154.2	12.42
Residual		126.3	11.24

Number of obs: 85, groups: Openers\$Player, 30; Openers\$Tm, 12

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	14.328	7.231	1.981

According to the model, 76.18% of the variation in ERA is attributable to the opener, slightly lower than the previous model, and 13.09% is attributable to the team.

Data: OpenersANOVA

Models:

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
OpenersERA.mod1: Openers\$ERA ~ (1 Openers\$Player)	3	740.02	747.35	-367.01	734.02				
OpenersERA.mod2: Openers\$ERA ~ (1 Openers\$Player) + (1 Openers\$Tm)	4	741.94	751.72	-366.97	733.94	0.0794	1		0.7782

Since the ANOVA table has a small p-value of 0.7782, we conclude that random effect Team is not needed in the model.

Pooling for improved estimates of player-specific parameters

Consider the dataset Openers with 85 observations and 30 pitchers. The variable

OpenersEra.mod1 contains the output for ERA data with a pitcher-specific random term. The function coef when applied to OpenersEra.mod1 provides estimates.

```
$`Openers$Player`
                                (Intercept)
Adrian Houser\\housead01      1.5652101
Alex Claudio\\claudal01      1.5652101
Andrew Kittredge\\kittran01  5.0669864
Brad Peacock\\peacobr01      1.5652101
Cam Bedrosian\\bedroca01     6.3262970
Chad Green\\greench03         6.9026386
Connor Sadzeck\\sadzeco01     0.8279982
Diego Castillo\\castidi01     0.3431411
Gabriel Moya\\moyaga01        0.5628817
Gregory Soto\\sotogr01        9.5784006
Hansel Robles\\robleha01      1.5652101
Hunter Wood\\woodhu01         4.1636033
Jeffrey Springs\\sprinje01     1.5652101
Jesse Chavez\\chaveje01       0.5628817
Jonny Venters\\ventejo01     121.7630667
Jose Leclerc\\leclejo01       0.8279982
Liam Hendriks\\hendrli01      2.7737683
Lou Trivino\\trivilo01        1.5652101
Luis Garcia\\garcilu03        1.5652101
Luke Bard\\bardlu01           0.8279982
Matt Andriese\\andrima01       1.5652101
Michael Feliz\\felizmi01     121.7630667
Montana DuRapau\\durapmo01    37.6245671
Nick Vincent\\vinceni01       1.5652101
Ryne Stanek\\stanery01        4.5479158
Sergio Romo\\romose01        23.3457444
Taylor Cole\\coleta01         1.5652101
Trevor May\\maytr01           33.6179719
Tyler Clippard\\clippty01     17.5915910
Yonny Chirinos\\chiriyo01    13.5849958
```

These estimates are compared to the completely-pooled estimate of $\hat{\mu}$ and the pitcher-specific

sample means, obtained using the aggregate function. Since we want to predict the value of

ERA of opener for an upcoming season, we can find the root-mean-squared-error (RMSE) by summarizing the errors in the predictions. A lower value of the RMSE indicates that the predicted values are more accurate. $RMSE = 10.52653$ which means that numbers are fairly inaccurate predictions. This makes sense because as mentioned before, some of these openers have done it very few times with not much effectiveness. The Pirates tried using Michael Feliz (predicted 121.76 ERA) and Montana DuRapau (predicted 37.62 ERA) as openers once and it clearly backfired, but it does not exactly mean that they will stay at this level if given more opportunities to be openers.

HEADLINERS

ERA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	2.250	3.535	5.400	21.600

wOBA

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.04312	0.21435	0.29148	0.31266	0.39272	0.71794

For Headliners, the mean ERA was at a level below league average at around 3.54 ERA, but not elite either. These headliners were given more innings to pitch and appearances than the openers making the mean and maximum more believable measures of their ability in this situation. The mean wOBA is encouraging to the effectiveness of the strategy since it is fairly low at 0.313.

This could indicate that these pitchers are more comfortable and successful when they are utilized as headliners.

Analysis of Variance

My initial thoughts on the Average Leverage Index for the headliner were that on average the pressure was going to be low since this type of pitcher starts with the middle-bottom of the lineup and faces the whole lineup twice usually. With that in mind, one can also infer that WPA would be higher since the opener set up this ideal situation for the headliner to dominate hitters for about four innings.

```
Linear mixed model fit by REML ['lmerMod']
Formula: Headliners$ERA ~ (1 | Headliners$Player)
Data: HeadlinersANOVA
```

```
REML criterion at convergence: 474
```

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-0.8952	-0.8908	-0.3242	0.4695	4.5642

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
Headliners\$Player	(Intercept)	0.01083	0.1041
Residual		15.65626	3.9568

Number of obs: 85, groups: Headliners\$Player, 31

```
Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	3.5337	0.4304	8.211

There are 85 observations in this data frame, with 31 headliners represented a variety of times.

ERA_{jk} has variance of $\sigma^2_{\sim\alpha} + \sigma^2_{\sim\epsilon} = 15.66709$ and standard deviation of 3.9582 which is slightly greater than the sample standard deviation (3.9581) of the ERA values in the Headliners dataset. $\hat{P} = 0.00069126$. About 0% of the variation in ERA values is attributable to the headliner. The correlation of two ERA values in different years for the same headliner is 0.00069126. For a given hitter, the correlation of the observed ERA with the headliner's true ERA value is 0.0263.

```

Linear mixed model fit by REML ['lmerMod']
Formula: Headliners$ERA ~ (1 | Headliners$Player) + (1 | Headliners$Tm)
Data: HeadlinersANOVA

REML criterion at convergence: 474

Scaled residuals:
    Min       1Q   Median       3Q      Max
-0.8949 -0.8911 -0.3243  0.4698  4.5642

Random effects:
Groups              Name          Variance Std.Dev.
Headliners$Player (Intercept)  0.009184  0.09583
Headliners$Tm      (Intercept)  0.000000  0.00000
Residual                                15.657762  3.95699
Number of obs: 85, groups: Headliners$Player, 31; Headliners$Tm, 12

Fixed effects:
              Estimate Std. Error t value
(Intercept)   3.5338      0.4302   8.214
convergence code: 0

```

According to the model, 0.0005% of the variation in ERA is attributable to the headliner, slightly lower than the model before, and 0% is attributable to the team.

```

Data: HeadlinersANOVA
Models:
HeadlinersERA.mod1: Headliners$ERA ~ (1 | Headliners$Player)
HeadlinersERA.mod2: Headliners$ERA ~ (1 | Headliners$Player) + (1 | Headliners$Tm)
              Df    AIC    BIC  logLik deviance Chisq Chi Df
HeadlinersERA.mod1  3 480.09 487.42 -237.05   474.09
HeadlinersERA.mod2  4 482.09 491.86 -237.05   474.09      0      1
              Pr(>Chisq)
HeadlinersERA.mod1
HeadlinersERA.mod2      1

```

Since the ANOVA table has a small p-value of 1, we conclude that random effect Team is not needed in the model.

Pooling for improved estimates of player-specific parameters

Consider the dataset Headliners with 85 observations and 31 pitchers. The variable HeadlinersEra.mod1 contains the output for ERA data with a pitcher-specific random term. The function `coef` when applied to HeadlinersEra.mod1 provides estimates.

\$`Headliners\$Player`	(Intercept)
Aaron Brooks\\brookaa01	3.532801
Adrian Sampson\\sampsad01	3.532431
Anthony Banda\\bandaan01	3.532227
Ariel Jurado\\juradar01	3.533515
Austin Pruitt\\pruitau01	3.534537
Chance Adams\\adamsch01	3.535912
Chris Bassitt\\bassich01	3.531245
Daniel Mengden\\mengdda01	3.526923
David Hale\\haleda02	3.534356
Drew Smyly\\smylydr01	3.542837
Felix Pena\\penafe01	3.533060
Framber Valdez\\valdefr01	3.531245
Freddy Peralta\\peralfr01	3.531245
Jaime Barria\\barrija01	3.532489
Jalen Beeks\\beeksja02	3.540534
Kohl Stewart\\stewako01	3.526369
Mike Fiers\\fiersmi01	3.542445
Mike Hauschild\\hauscmi01	3.531245
Nestor Cortes Jr.\\cortene01	3.535912
Nick Kingham\\kinghni01	3.540578
Nick Ramirez\\ramirni01	3.532801
Nick Tropeano\\tropeni01	3.534978
Ross Detwiler\\detwiro01	3.534356
Ryan Weber\\weberry01	3.533734
Ryan Yarbrough\\yarbrry01	3.542246
Stephen Gonsalves\\gonsast01	3.531245
Steven Brault\\braultst01	3.533444
Trevor Cahill\\cahiltr01	3.531245
Vidal Nuno III\\nunovi01	3.532579
Yohander Mendez\\mendeyo01	3.531245
Yonny Chirinos\\chiriyo01	3.524553

RMSE = 0.608447 which means that numbers are fairly accurate predictions. These are more reliable predictions since these pitchers on average have a lot more batters faced and innings under their belt than openers. For these starters turned headliners, they are accustomed to adjusting after a rough start or struggling after beginning dominantly, so it is plausible to have this range between 3.52 and 3.54 ERA.

RYAN YARBROUGH

To further look into how effective the “opener” and “headliner” is, I looked at the two pitchers used as such for the most times during the 2018-2019 seasons: Ryan Yarbrough and Ryne Stanek. In his rookie season, Ryan Yarbrough of the Tampa Bay Rays was utilized primarily as a headliner appearing in 38 games while only starting in 6 of those games. He went 16-6 W-L with a 3.91 ERA in 147.1 IP. As show below, Yarbrough performed better as a headliner than a regular starter with wOBA of 0.305 over 0.325, respectively.

wOBA in non-headliner games

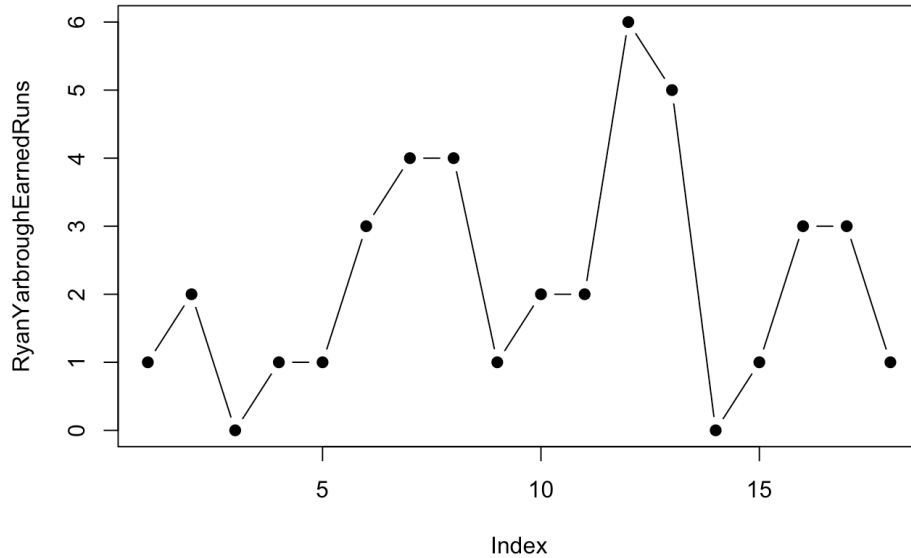
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.2280	0.3250	0.3247	0.3762	0.7850

wOBA in headliner games

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0550	0.2160	0.3150	0.3045	0.4012	0.4946

Runs Test

Let Y_t denote the number of earned runs allowed by Ryan Yarbrough in game t of the 2018-2019 MLB Seasons where he pitched as a headliner.

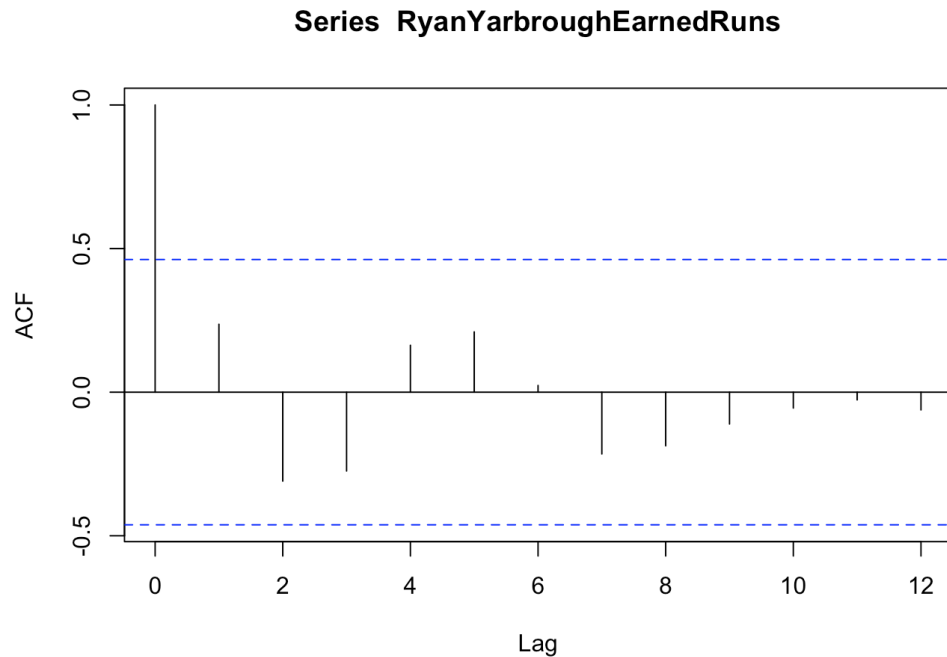


Runs Test

```
data: RyanYarbroughEarnedRuns
statistic = -0.78975, runs = 7, n1 = 7, n2 = 8, n = 15, p-value =
0.4297
alternative hypothesis: nonrandomness
```

The p-value of the test is 0.4297 so there is no evidence to reject the null hypothesis that Yarbrough's game-by-game earned run values is an observed sequence of i.i.d. random variables.

Observations equal to the median of 2 earned runs were dropped from the analysis, leading to $n = 15$. The expected number of earned runs is approx. $15/2 = 7.5$ so that the observed number is slightly less than the expected under the null hypothesis.



Autocorrelations of series 'RyanYarbroughEarnedRuns', by lag

0	1	2	3	4	5	6	7	8	9
1.000	0.237	-0.310	-0.275	0.163	0.210	0.023	-0.215	-0.187	-0.111
10	11	12							
-0.055	-0.027	-0.062							

Yarbrough's earned runs in a given game is slightly positively correlated with his earned runs in the most recent games $\hat{\rho}(1) = 0.237$, earned runs two games earlier or later are negatively correlated with $\hat{\rho}(2) = -0.310$, as well as earned runs three games earlier with $\hat{\rho}(3) = -0.275$.

None of the estimates is statistically significant.

Box-Ljung test

data: RyanYarbroughEarnedRuns

X-squared = 10.742, df = 12, p-value = 0.5512

I chose $m = 12$ which is equivalent to about two weeks of games. The p-value is approx. 0.5512 so we do not reject the null hypothesis of uncorrelated random variables so that Yarbrough's game-by-game earned run values appear to be uncorrelated.

RYNE STANEK

Ryne Stanek, the notable "opener" of the Tampa Bay Rays, made 25 appearances as one in this dataset, however, he started 29 games in 2018 and has already started 16 games this season. His numbers seem to be quite different because of distinct situations with wOBA in non-opener games being significantly lower than in opener games. This could be attributed to the last few starts where he has struggled immensely compared to his superb 2018 campaign.

wOBA in non-opener games

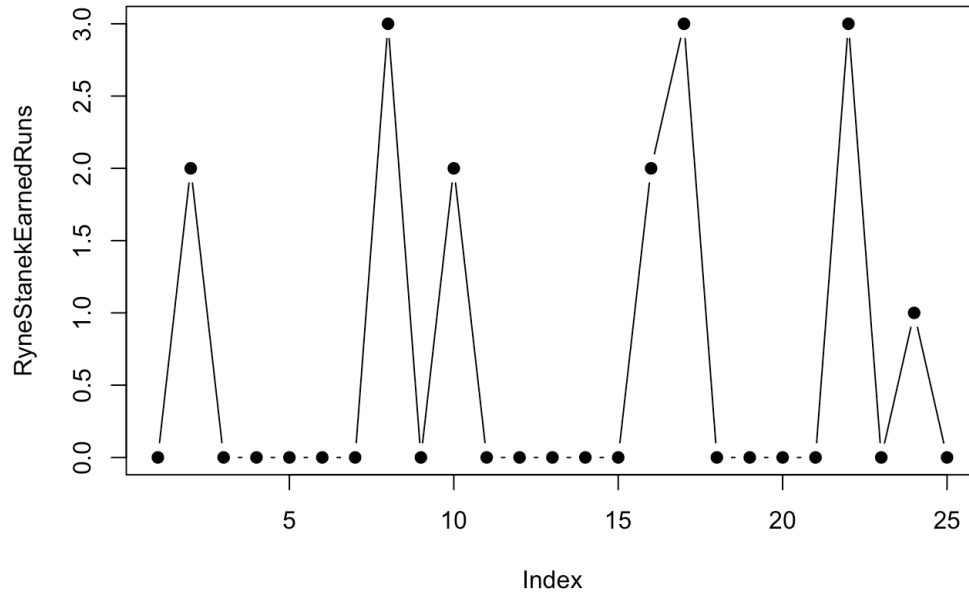
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.1725	0.2438	0.3520	1.4555

wOBA in opener games

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.2200	0.2933	0.3248	0.3767	0.7450

Runs Test

Let Y_t denote the number of earned runs allowed by Ryne Stanek in game t of the 2018-2019 MLB Seasons where he pitched as an opener.

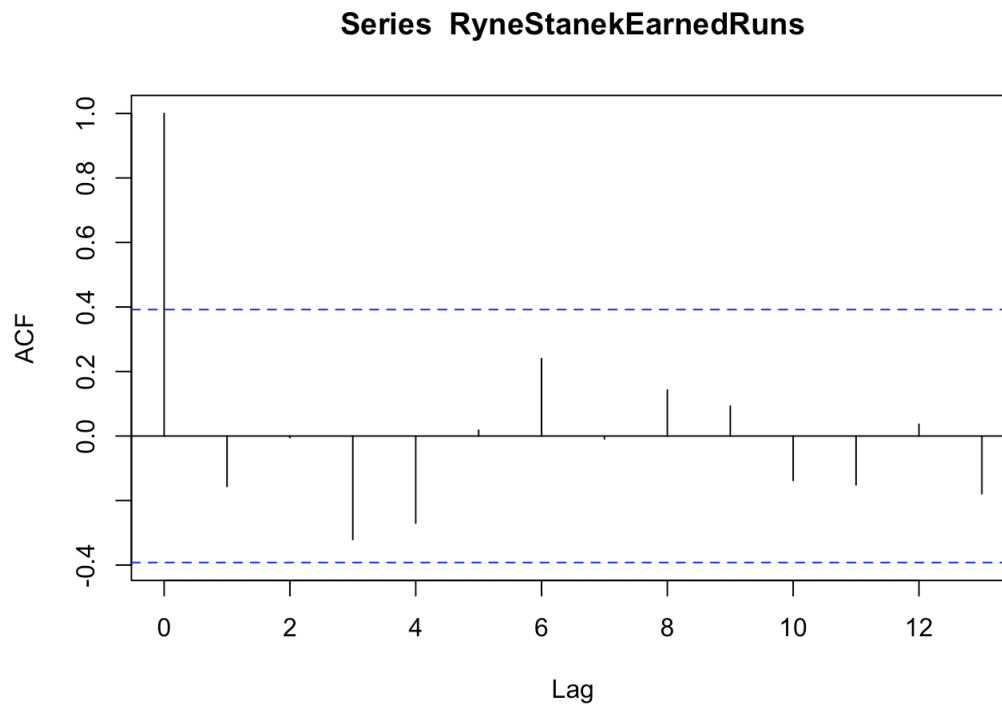


Runs Test

```
data: RyneStanekEarnedRuns
statistic = NaN, runs = 1, n1 = 7, n2 = 0, n = 7, p-value = NA
alternative hypothesis: nonrandomness
```

The p-value of the test is NA so there is no evidence to reject the null hypothesis that Stanek's game-by-game earned run values is an observed sequence of i.i.d. random variables.

Observations equal to the median of 0 earned runs were dropped from the analysis, leading to $n = 7$. The expected number of earned runs is approx. $7/2 = 3.5$ so that the observed number is significantly less than the expected under the null hypothesis.



Autocorrelations of series 'RyneStanekEarnedRuns', by lag

0	1	2	3	4	5	6	7	8	9
1.000	-0.156	-0.005	-0.321	-0.270	0.019	0.240	-0.009	0.143	0.093
10	11	12	13						
-0.138	-0.151	0.036	-0.179						

Stanek's earned runs in a given game is negatively correlated with his earned runs in the most recent games $\hat{\rho}(1) = -0.156$, earned runs two games earlier or later are negatively correlated with $\hat{\rho}(2) = -0.005$, as well as earned runs three games earlier with $\hat{\rho}(3) = -0.321$. None of the estimates is statistically significant.

Box-Ljung test

```
data: RyneStanekEarnedRuns
X-squared = 11.447, df = 12, p-value = 0.4911
```

I chose $m = 12$ which is equivalent to about two weeks of games. The p-value is approx. 0.4911 so we do not reject the null hypothesis of uncorrelated random variables so that Stanek's game-by-game earned run values appear to be uncorrelated.

SUMMARY OF RESULTS

Overall, it is still extremely difficult to encompass all of this information and conclude the effectiveness of the opener and headliner strategy in baseball. The Tampa Bay Rays seemed to have success implementing the strategy by looking at their team ERA recently. However, they had Blake Snell and Tyler Glasnow have elite seasons last year and probably contributing to the low ERA along with other pitchers. There is a great possibility that the headliner is more comfortable in this situation by looking at Yarbrough's success as a headliner over a more traditional starting role. Although Stanek had a great 2018 as a starter/opener, the data shows that he was less effective in official opener/headliner games which could explain the difficulties some pitchers have had when placed in these positions. The predictors used in the linear model for the analysis of variance were indicative measures to view what makes this strategy so unique for the game and its pitchers. There is evidently no proof for streaks in both Yarbrough's and Stanek's performances, despite having stretches where both have consecutive scoreless or low earned run innings. The analysis of pooling for estimates of ERA shows us that it is easier to predict the ERA for headliners than openers since some openers have inflated ERA when they opened once and got absolutely rocked. Many other teams have followed in the Rays' footsteps but have had minor or no success so far, however, once they have multiple opener/headliner games that is when they determine whether or not it is aiding their efforts.

SUGGESTIONS FOR FURTHER ANALYSIS

Suggestions include:

- Waiting until the end of this 2019 season to observe more “opener” and “headliner” games.
- Comparing wOBA, FIP, and other statistics from these openers and headliners to overall team pitching.
- Once WAR (Wins Above Replacement) calculations are finalized for these pitchers, one can analyze their true value or at least close to it.