Three Effects of a Shortened Baseball Season

More than any other sport, baseball is a game of probabilities. Players go on hot streaks, cold streaks, and everything in between. Over the course of a full 162 game season, though, these streaks tend to cancel out. Because of the length of the season, short-term anomalies won’t have as much of an impact on your season long stats.

However, in the shortened 2020 season we can expect many unusual things to happen. There isn’t as much time for a player that has a hot month to cool off. There isn’t as much chance for a pitcher who gives up 9 runs one night to bring his ERA down. In this article, I want to point out three interesting things about the 2020 season. The three topics I’ll discuss are:

* There will be extremes in rate-based stats (BA, OBP, ERA, etc.)
* The playoff teams might be more surprising than normal and, because of the importance of every game, teams might use pitchers more aggressively.
* The lack of present fans will let us learn more about home-field advantage.

**Effect 1: Extremes in Rate Stats due to the Shortened Season**

I am going to focus on perhaps the simplest stat here to illustrate what I mean: Batting Average. First, I need to talk about how I interpret batting average. There is a difference between a ‘true .320 hitter’ and someone who happens to hit .320 for a season.

A true .320 hitter is someone who, against league average pitching, will record a hit 32% of the times they get up to bat. This is a measure of a player’s true hitting ability, a description of expected future performance. You can think of it as a long-term average.

Hitting .320 for a season is a description of past performance. Someone who hits .320 for a season might actually be a true .290, .300, or .330 hitter. But, due to the natural variability in a single season, someone’s true batting average will rarely equal their season batting average.

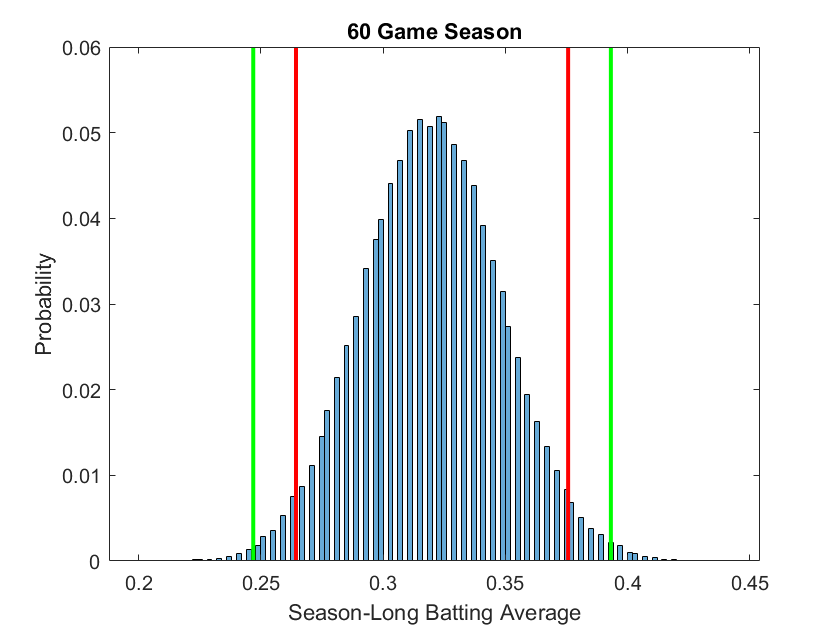
In a shortened season, this effect is exaggerated because of the smaller number of at bats. What would have last year been ‘a hot month’ is now closer to ‘a hot first half’. What kinds of averages can we expect to see this year?

**Variability in One Player’s Stats**

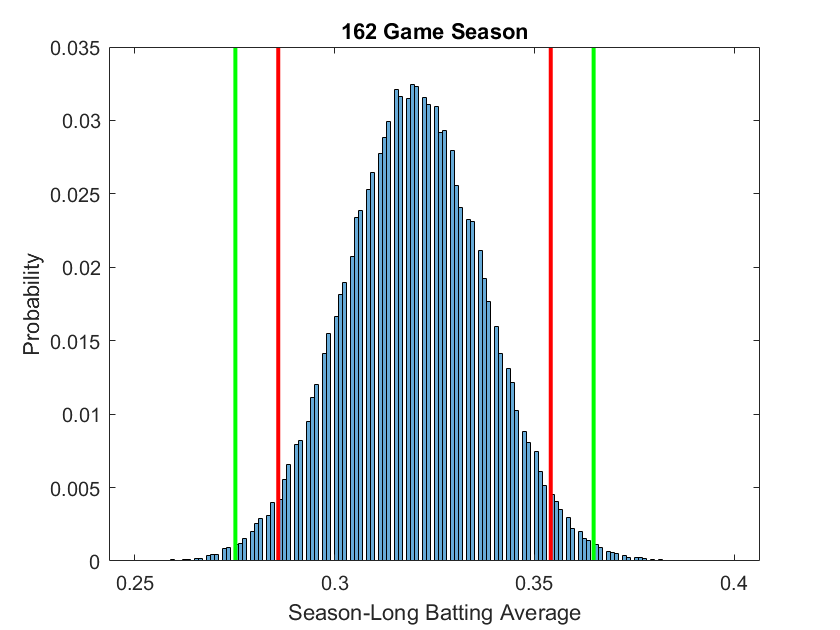
I’ve heard more than once that ‘somebody is going to hit .400 this year’. As you may remember from my [most recent post](https://thedatajocks.com/quick-hit-will-someone-hit-400-this-year/), I argued it was unlikely anyone would do this. The technique I used is as follows.

I simulated 270 at bats (roughly 60 games) for a .320 hitter by flipping a coin that says ‘hit’ 32% of the time and ‘out’ the other 68%. I flipped this coin 270 times to see what their season long average would be in a particular season. Then I repeated this a few thousand times to get a feel for what the distribution of season-long batting averages look like for a true .320 hitter.

Between the red lines below is the middle 95% of simulated seasons and between the green lines the middle 99% of simulated seasons. We notice that most often the player hits close to .320 but it is not uncommon for the player to do much better or much worse than .320 for the season. The key takeaway, and we’ll see this after looking at the second graph below, is that the shortened season increases the variance in season long batting average. It is more likely to have large anomalies in a shortened season.

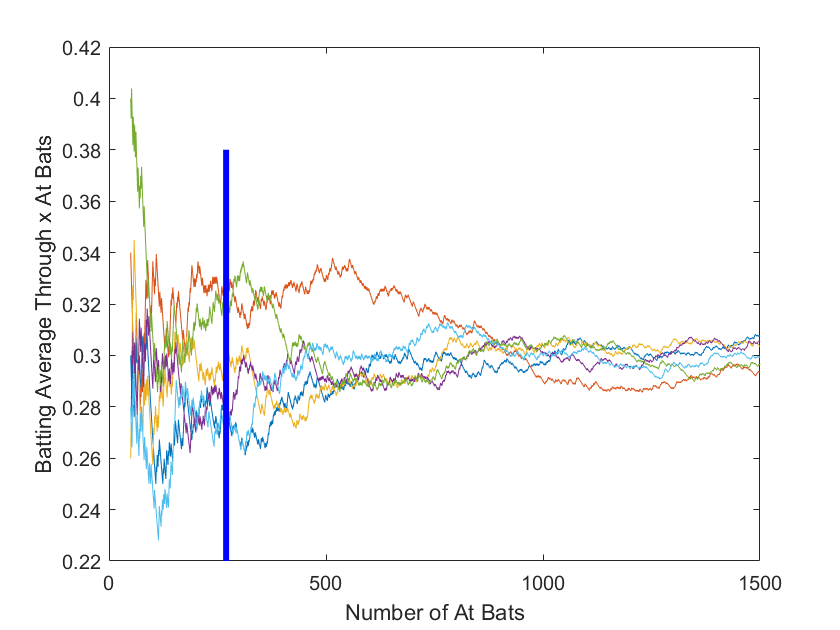


You can interpret the above chart like: There is a 95% chance that a true .320 hitter will have their season-long batting average fall between (roughly) .260 and .380. If we compare this to the same simulation for a 162 game season, we see that the range of viable season-long batting averages shrinks.



For a 162 game season (with 4.5 at-bats per game), we can be 95% certain that a true .320 hitter will have a season-long batting average between about .285 and .335.

In fact, to show this effect graphically one more time, the following chart simulated 6 seasons of a player who is a true .300 hitter. Notice that as the number of at-bats (or, the length of the season) increases, the season-long averages approach the player’s true ability. However, at the beginning of the season, their averages are free to vary quite a bit. The thick blue line below is approximately placed at the 60-game mark.



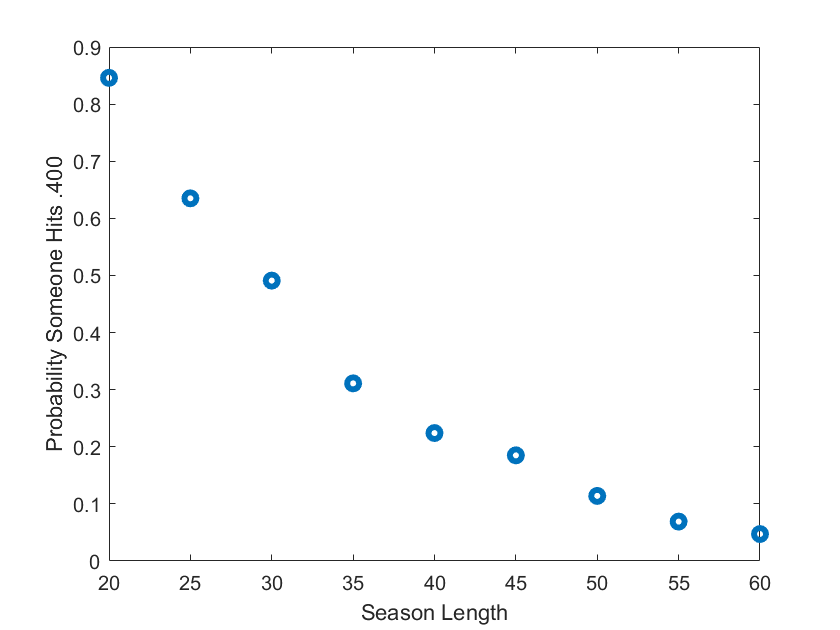
**How Good is the Average *League-Best* Batting Average in a Shortened Season?**

Above, we showed two things. First, that having more games compresses the range of viable batting averages for a hitter. That is, as the season goes on, their season average regresses towards their true ability. Second, we showed that it was quite unlikely (<1%) that a true .320 hitter would break 400 this year.

The thing is, though, it makes sense that out of the top 100 hitters, at least one of them will have a ‘1%’ type season. I mean, when we expand our scope from a single player to the entire league, the odds of a .400 season will increase.

We performed a few experiments to show this effect. First, we took the top 100 hitters from last season and assumed their season long averages from last year reflected their true average (Homework assignment: What might be wrong with this assumption?) Then, we simulated all 100 of these hitters’ shortened seasons simultaneously and recorded the best batting average from among them. Then we repeated these simulations thousands of times to see what would happen.

**In only 5%** of our simulated 60-game seasons did anyone hit .400. Moreover, the average league leading BA was about .372 over the shortened season. This is higher than normal, but still not the .400 number being claimed. After this I got curious and asked: What happens if the season is shortened to 50, 40, or less games? Is it likely someone will hit .400 then?



This is the probability that any individual hits .400 versus season length. We see that the probability drops off rapidly. The season length would need to be about 30 games for it to be reasonable to expect someone to break .400 for the season. It could happen this season, but it probably won’t.

**Other Rate Stats**

While we just discussed batting average in the prior discussion, the exact same analysis could be done for other statistics which measure some rate (as opposed to *counting* stats like HR). Just like the league batting average leader will probably have a pretty high number this year, the OBP leader will too. The league leader in ERA will probably be quite a bit lower than average. The same is true for OPS, WHIP, and anything else that measures performance ‘per’ something.

Counting stats like HR, RBI, strikeouts, wins, etc., however, will probably see very low numbers for the season leader simply because there is not enough time to accumulate significant statistics in the shortened season.

**Effect 2: Each Game Matters More in a Shortened Season**

I was watching the opening day Cubs-Brewers game and Alex Rodriguez commented that he did the math, each game this season is worth 2.7 regular season games. He claimed a 4 game losing streak was kind of like an 11 game losing streak in a full season. While I don’t totally agree with that logic, the sentiment he is trying to convey is undeniably important.

Because of the shortened season, the difference between a division leader and 2nd and 3rd place is going to be proportionally much smaller. For example, let’s look at last year’s NL Central Race. The Cardinals won the division with a .562 winning percentage. Next came the Brewers, then the Cubs, Reds, and Pirates at 2, 7, 16, and 22 games back, respectively.

If we scale those numbers to the shortened season, then the Brewers, Cubs, Reds, and Pirates would have only been 0.7, 2.6, 6, and 8 games back of first place. While the Reds and Pirates certainly would have had a tough time making up that deficit, less than 3 games would have separated first place from third! Every single game is going to be crucially important this year. One salvaged win or one blown save is going to have a much larger effect on the standings.

**So What Should We Look For?**

Because every game means so much more, we may see some strange coaching decisions this year. In fact, it is probable that the regular season will feel more like playoff baseball. There are two things I think managers might do to take advantage of the shortened season.

First, I think it is possible or even probable that managers may pitch their ace starters more often. That is, the rotations may resemble four man rotations especially towards the end of the season. This is an especially viable strategy because the shortened season also reduces season-long fatigue that might otherwise be a deterrent to this strategy.

Second, we may see teams be much more liberal with their bullpen usage. In the playoffs, we get a lot more ‘bring this pitcher in for a single inning or a single at-bat’ than in the regular season. We might even see one of my favorite things: a starter brought in to throw in a relief inning in a tight game. [Scherzer did it in the playoffs last year](https://www.mlb.com/news/max-scherzer-pitches-in-relief-vs-dodgers). I think we may see a lot more of this in the regular season as teams fight for every last win.

**Effect 3: The Lack of Fans Yields Home-Field Advantage Insights**

My friends and I have had an unreasonable number of debates about what home-field advantage is most attributable to. There are three possibilities that explain why the home-team wins more often:

* Familiarity with the specifics of the field/court/stadium etc. give an advantage
* The home-town fans actually give a huge psychological boost.
* The effects of travelling give the home team a physical (exhaustion) boost.

While I think the first bullet does not contribute much to home field advantage (except, maybe the pre-2016 Astros with their disastrous centerfield warning track hill). Most of the home field advantage is probably due to travel effects and fan effects. But which is more important? To determine the relative importance we either need to remove the influence of fans (which has never been done) or remove the influence of travelling.

We could remove the influence of travelling if we look at past data for Lakers-Clippers, Yankees-Mets, or White Sox-Cubs, etc. However, the sample sizes are small here and the conclusions we make won’t necessarily generalize to the entire league. For instance, even at Clippers home games, many of the fans are still Lakers fans, tainting the data.

Luckily this year with fans being not present or minimally present we finally have a control group! This year, fans aren’t going to be in the stadium for at least the beginning of the season. Though the MLB is pumping in fake noise, complete with cheering (but, interestingly, [no booing](https://www.cbssports.com/mlb/news/mlb-fake-crowd-noise-will-not-include-booing-per-report/)), there is certainly a difference between playing noise out of a speaker and the energy associated with in-person fans.

To put that another way, the component of home field advantage associated with hometown fans is mostly gone this year. Because of this, the data from this year will help us determine the amount of home field advantage that is attributable to the effects from travelling.

After this season is complete, I will compute the home field advantage this year and compare it to years past (using the [home-field advantage metric described here](https://thedatajocks.com/home-court-advantage-with-maximum-likelihood-estimation/)) to determine precisely the amount of the home field advantage that is fan based and the amount that is travelling based.

**Conclusions**

We are all undoubtedly glad to have baseball (and soon basketball and hockey) back for the entertainment value. In addition to just the presence of sports, I am particularly interested to see what small curiosities come out of the next few seasons. Certainly, any discussion about sports in the future will have an asterisk and many footnotes after the 2020 season.