1.

Does Travel affect athletic performance in (badminton) athletes

2.

* Travelling can be very taxing!
  + Long hours on the road or in a plane
  + Crossing time-zones can ruin a player’s energy levels
  + Sleeping is much harder in general in new environments
  + Additional everyday tasks require more energy to accomplish

In the context of professional sports leagues like the NBA, NHL where athletes have to play a scheduled number of games, this has historically shown to have some minor affects. However, what about sports where the schedule is more or less self-made and athletes may need to travel longer distances? What I’m really describing is badminton, but I think this may be generalizable for other individual sports where the barrier to entry is not as high as professional leagues (martial arts, squash, cycling, track sports). But this also adds another element – what kind of athletes in these sports are travelling long distances?

* When do athletes choose to travel?
  + Travelling further means more sacrifice/money
  + Athletes will travel to tournaments they feel are appropriate to their skill level/they can win
  + Usually when we look at tournaments, there are a lot of home country players, and the top players are from elsewhere

3.

Despite the difference between funded professional sports, I believe there is still going to be some effect of travel distance and results.

4.

There will be two primary measures for this experiment. We will look to see if players who travel less have a greater chance of winning a match, AND we will see if players play better at home/reach further rounds on average at home.

5.

To answer these questions, I needed to collect a lot of data. This is actually where the bulk of my time was spent. I decided to look at tournaments that were sanctioned by BWF and their data was all available at tournmentsoftware.com. Unfortunately there was no easy way to download data from that site, so I had to manually build a scraper in python to do it for me. The packages used were: pandas, requests, regex, beautiful soup. The main pipeline I hatched together was designed to scrape the tournament links, determine the match page links, and using beautiful soup/regex I was able to parse the important data and then export them as a pandas dataframe in a csv format. I managed to scrape data for all the tournaments in 2017 and half of the tournaments in 2018.

6.

Once the data was gathered, I needed to extract win/loss information, distance information, rounds won and format everything into a usable form. I had to match the country codes/clean up some non-traditional country encodings and used ArcGIS to calculate the distance travelled between the player/team and the host country. I also looked to see if the host country was also the team/player country to indicate if the players are playing at home or not. Determining the number of rounds a player reached in a tournament is simply cutting to the player and grouping by the match location (Although we had to be careful with multiply tournaments in the same location), and counting the number of rounds.

7.

So now that we have our data, let’s answer our first question: Do winners travel less than losers of a match?

Really simply we can first look at the difference of means in the population. We have the mean of Winner distances as 4131 KM and Losers as 3412 KM. This is definitely not what I expected as I seems like Winners travel further on average. But let’s look at the distributions.

The first two images are the winner distances and loser distances for each match. The distributions are kind of interesting as it seems most players are playing closer to home regardless of winning or losing. These two distributions are clearly not normal. I actually did a quick test to see if they were folded normal by doing normality tests doubling the sample, and also Anderson/KS tests for exponential but they all failed to match those distributions. We do note that we DO have close variances and quite a large sample size.

8.

Luckily, we still have non-parametric tests that can determine if these two populations are actually different. Especially since both sample are independent – The opponents distances do not depend on each other, we have close variance values, and our values are ordinal the Mann-Whitney U test is appropriate.

The Mann-Whitney U test is specifically designed to test two sets, where the null hypothesis is:

**Equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample.**

P(A>B) = P(B>A) where A and B are the random variables from group A and group B

In this case, we feel like this null hypothesis is not true since one group (winners) have a much higher mean, and we believe the winner distribution is stochastically greater than the loser distribution.

Very briefly, the Mann-Whitney U test assigns rankings to each sample point in both samples, and then for the values of set A, we count how many values of set B each value from set A beats. The sum will give a test statistic U and the test statistic is a known normally distribution for large enough sample size so we can arrive at a p-value for the each test!

Luckily python did this test for me and we arrived at a very large statistic with a p values less than 0.01 -> Winner distances and Loser distances do not come from the same distribution. This does go against what I expected, but I did wonder if this was due to the number of local amateurs entering big tournaments that were conveniently local. Maybe we need to take a look at this from another angle.

9.

So I decided to cut the data in a different way. For each player, I recorded the mean distance for wins and losses. We then ran the same Mann-Whitney U tests to see if the distributions were different. The results are much closer, but it does still seem like losers travelled less distance. However, if I force the dataset to have players with more than 6 matches, we see the p-value indicates they are drawn from the same set. Maybe these scrubs were ruining my hypothesis!

10.

This actually lead me to think about home court advantage as the calculations and results are a simple corollary of the previous player specific win/loss means. However, in this case I gathered all the matches played by a team/player and calculated the number of rounds they made and took the average of the tournaments that were at home, against the rounds away from home. I limited the set of players to only players with home and away wins (removing the scrubs) and

11.

Valided learning and next steps:

In general this study made an effort to isolate the effects of distance on performance directly.

It seems that my initial hypothesis that travel negatively affects badminton player performance does not appear to be widespread over the competing international athletes. However in the process of the investigation, we noticed that players that enter a tournament in general are closers to the location –based on the distributions - And did also note that a disproportionate amount of scrub players play in convenient home tournaments. The final note is that home court advantage appears to exist, although our sample size is small.

These results can be extended by scraping more data from previous years as well as investigating related effects of distance. Distance does not appear to have a direct relationship with performance according to this study. I tried my best to isolate the effects of distance here, but there are many other factors that may have impact in conjunction with distance: time of match, how many timezones crossed, hotel ratings on average.