Test Match Batting Form Slumps

This project uses test match data aquired from https://cricsheet.org/downloads/. The raw .csv files are processed and compiled into three datasets: 'batting_stats', 'bowling_stats' and 'match_information'.

The goal of this project is to test whether batsmen experience periods of poor performance or form. Commentators and pundits alike often comment when a batsman fails to produce good scores over a number of matches. The inspiration for this project was taken from the batting results of the 2019 Ashes test series between Australia and England. In this series David Warner scored a total of 95 runs in 10 innings (9.5 average)- an extremely low average. By contrast Steve Smith scored 774 runs in 7 innings (110.57 average)- a remarkably good performance.

The null hypothesis can be stated as: "there is no time dependent correlation of the runs scored by a batsman"

To reject this hypothesis we must show that periods of poor form cannot be accounted for by pure statistical sampling.

The most subjective part of this project is defining what constitutes poor form. One consideration is the number of innings across which we measure form. Choosing too large a window may pick up on longer term trends in batting performance, such as a new player improving over time, or an old players performance declining. Likewise choosing too small a window is likely to be meaningless since scores vary wildly on an innings by innings basis. Motivated by the Ashses test series I choose a window size of 10 innings, though I also look at a smaller 6 innings window as well.

Another consideration is the temporal aspect of each innings and not just the ordinal aspect. Two low scores may be seperated by a year or more. Should this be considered a continuation of poor form? We may want to restrict our analysis to poor batting seasons, where the average number of runs scored across the season is significantly below average. For simplicity I choose not to follow this line of investigation but the results may prove interesting.

Finally the grounds on which the batsman plays, or the opposition that they face, may also influence their average over a period of time. For example a batsman facing a very strong bowling lineup is likely to score lower than against a weaker opposition. Again for simplicity I choose not to follow this line of investigation.

This project will be split into 6 parts.

```
1. Data Exploration and Visualisation
2. Resampling and Bootstrapping
3. Smith and Warner Batting Comparison
4. Ensemble Batting Performances
5. Possible Longer Term Trends
6. Summary
```

Test Match Batting Form Slumps

This project uses test match data aquired from https://cricsheet.org/downloads/. The raw .csv files are processed and compiled into three datasets: 'batting_stats', 'bowling_stats' and 'match_information'.

The goal of this project is to test whether batsmen experience periods of poor performance or form. Commentators and pundits alike often comment when a batsman fails to produce good scores over a number of matches. The inspiration for this project was taken from the batting results of the 2019 Ashes test series between Australia and England. In this series David Warner scored a total of 95

runs in 10 innings (9.5 average)- an extremely low average. By contrast Steve Smith scored 774 runs in 7 innings (110.57 average)- a remarkably good performance.

The null hypothesis can be stated as: "there is no time dependent correlation of the runs scored by a batsman"

To reject this hypothesis we must show that periods of poor form cannot be accounted for by pure statistical sampling.

The most subjective part of this project is defining what constitutes poor form. One consideration is the number of innings across which we measure form. Choosing too large a window may pick up on longer term trends in batting performance, such as a new player improving over time, or an old players performance declining. Likewise choosing too small a window is likely to be meaningless since scores vary wildly on an innings by innings basis. Motivated by the Ashses test series I choose a window size of 10 innings, though I also look at a smaller 6 innings window as well.

Another consideration is the temporal aspect of each innings and not just the ordinal aspect. Two low scores may be seperated by a year or more. Should this be considered a continuation of poor form? We may want to restrict our analysis to poor batting seasons, where the average number of runs scored across the season is significantly below average. For simplicity I choose not to follow this line of investigation but the results may prove interesting.

Finally the grounds on which the batsman plays, or the opposition that they face, may also influence their average over a period of time. For example a batsman facing a very strong bowling lineup is likely to score lower than against a weaker opposition. Again for simplicity I choose not to follow this line of investigation.

This project will be split into 6 parts.

```
1. Data Exploration and Visualisation
2. Resampling and Bootstrapping
3. Smith and Warner Batting Comparison
4. Ensemble Batting Performances
5. Possible Longer Term Trends
6. Summary
```

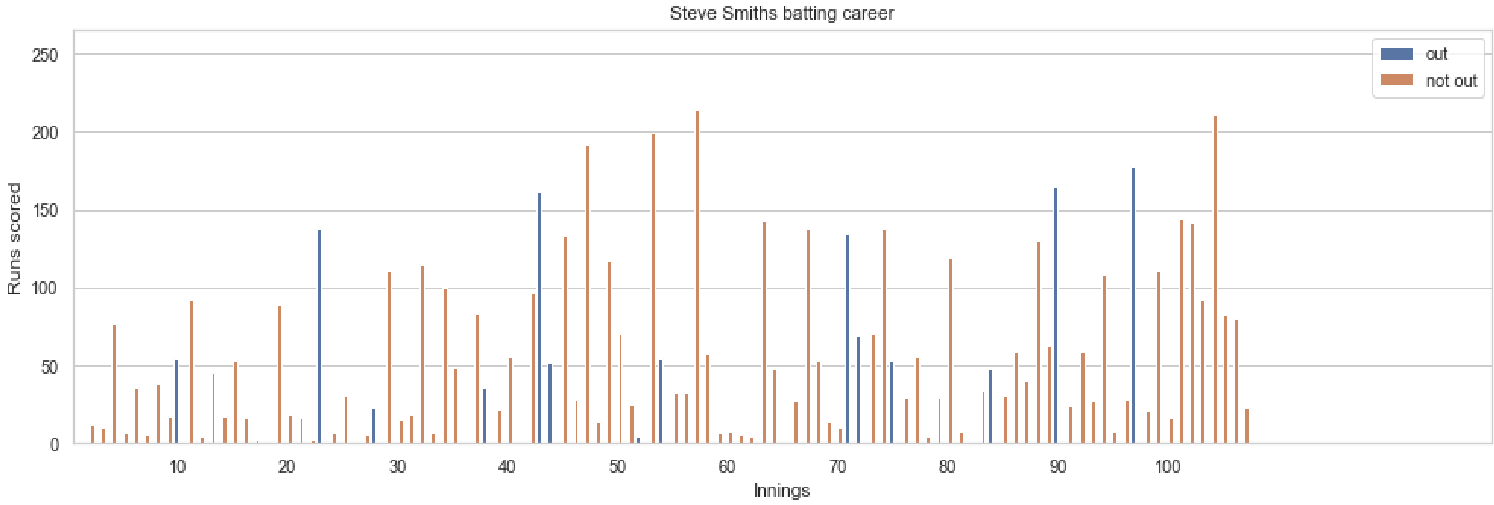
1. Data Exploration and Visualistation

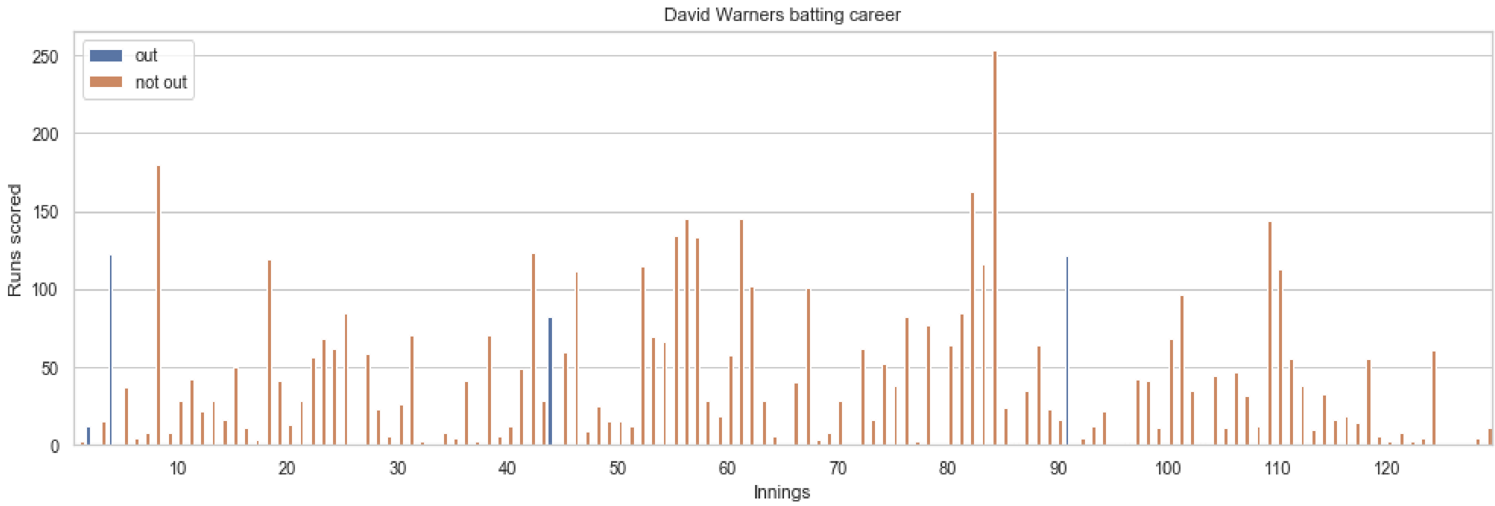
Here I plot a bar plot of each score for David Warner and Steve Smith placed in chronological order. This gives a sense of the volitility of the data. It appears that a large score is as likely to be followed by a low score as another large score.

Note also that I have colored each score by the 'out' column of the batting dataframe. When considering an average over a period of time the total number of runs scored is divided by the number of times they have been given out.

It is possible to append career averages as well as rolling averages to these plots, and this is done in section 3.

For now these batting charts show the variability in scoring. Note that there appear to be a large number of relatively lowe scores (<20) in both batsmens careers, punctuated by a number of higher scores (> 50 or 100) which may be pulling the average up.





2. Resampling and Bootstrapping

To find how significant a peak or dip in performance is I will perform bootstrapping.

The process first involves resampling a batting career. The resampling chooses (with replacement) scores at random from a career until the total number of innings played is matched by the players actual career. A rolling average with a certain window size is then performed. Note that the average incorporates whether the player is given out or now.

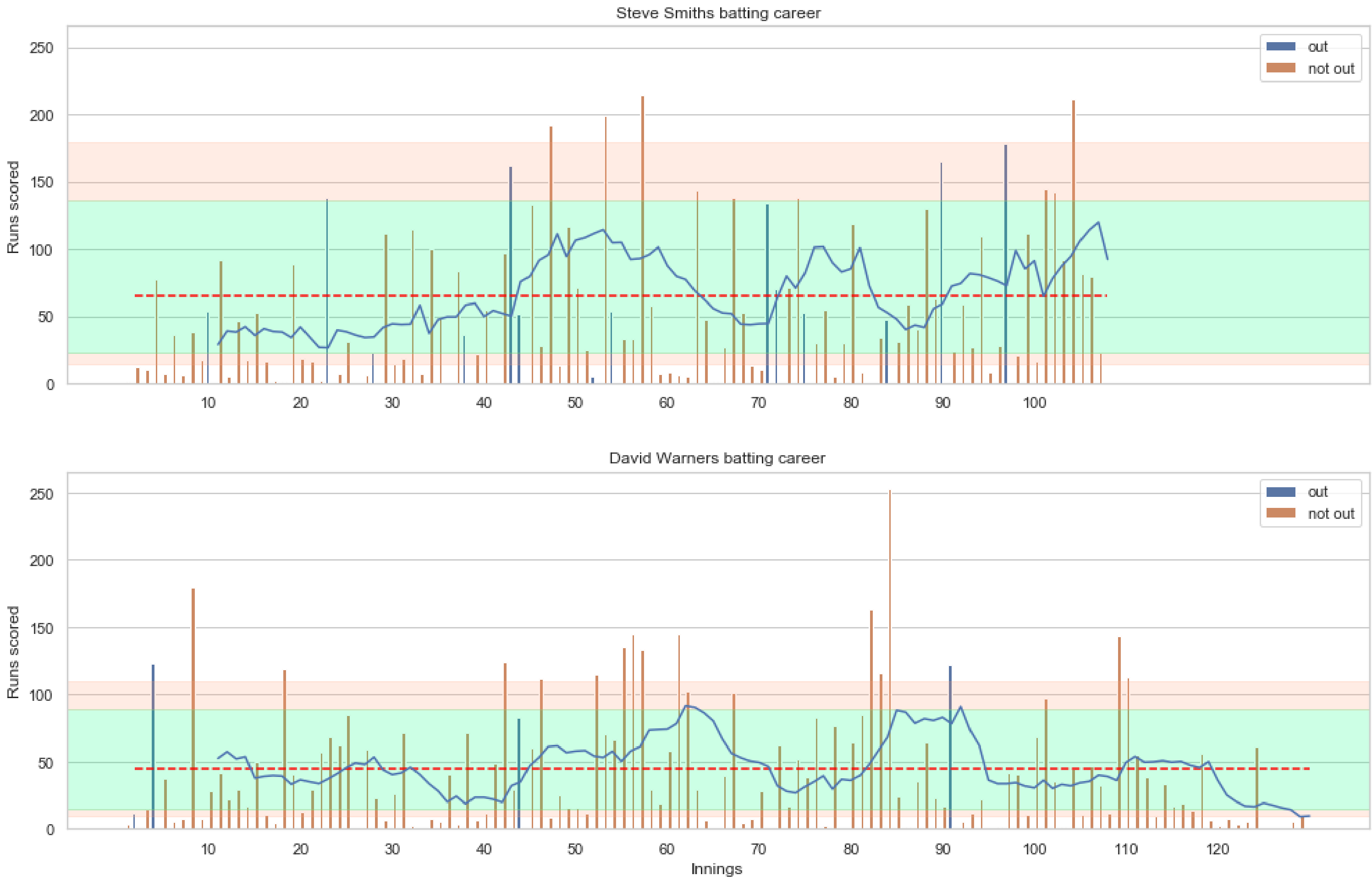
The complete bootstrapping process repeats the resampling step thousands of times (10000 default in this project). Comparisons can then be made between the bootstrap samples and the original career of the player. For example the proportion of bootstrap samples contains a rolling average equal to or less than the players actual lowest rolling average will give a confidence interval for that score.

Experimental Cumulative Distribution Functions (ECDFs) are a way of visualising the confidence intervals corresponding scoring below a certain number or runs. Functions for bootstrapping and plotting confidence intervals are shown in the code below. By default the ECDF function returns confidence intervals for a rolling average below 5%, below 32%, which are appoximately 1 and 2 standard deviatios respectively. The code also calculates the likelyhood of good performance spikes, and confidence intervals of 5% and 32% are given for obtaining a rolling average above a certain number of runs throughout a career.

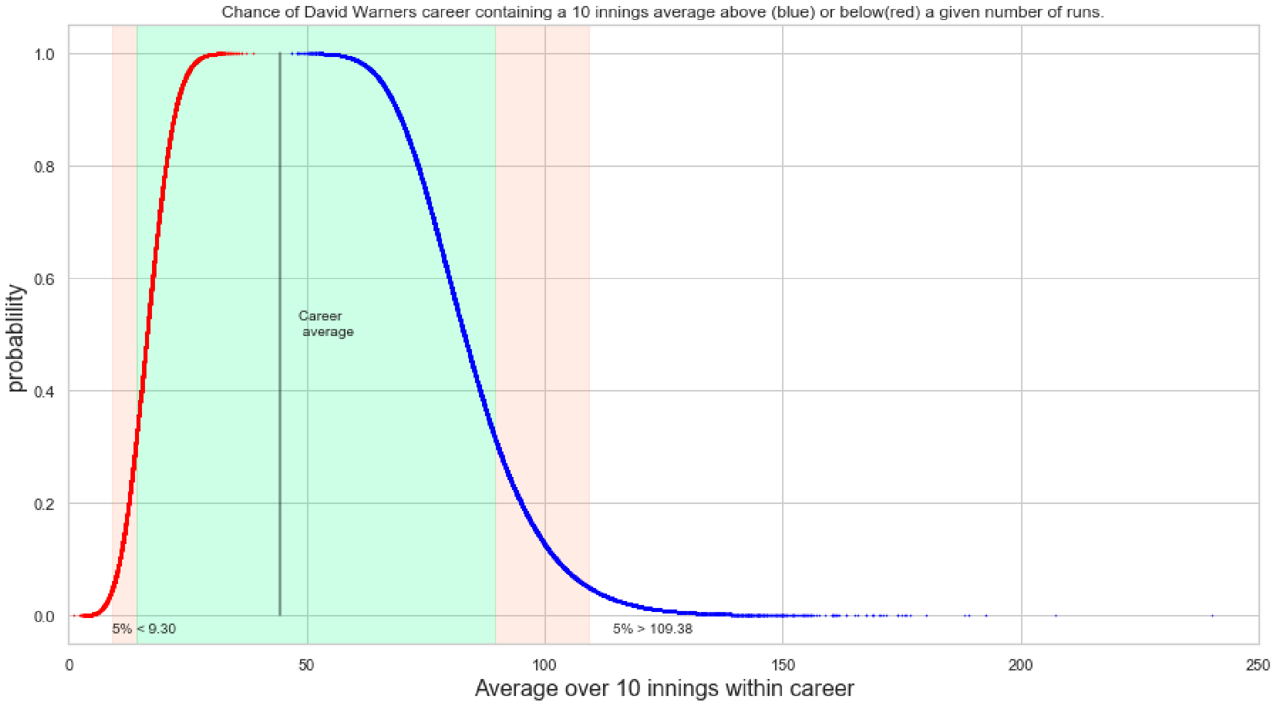
3. Smith and Warner Batting Comparisons

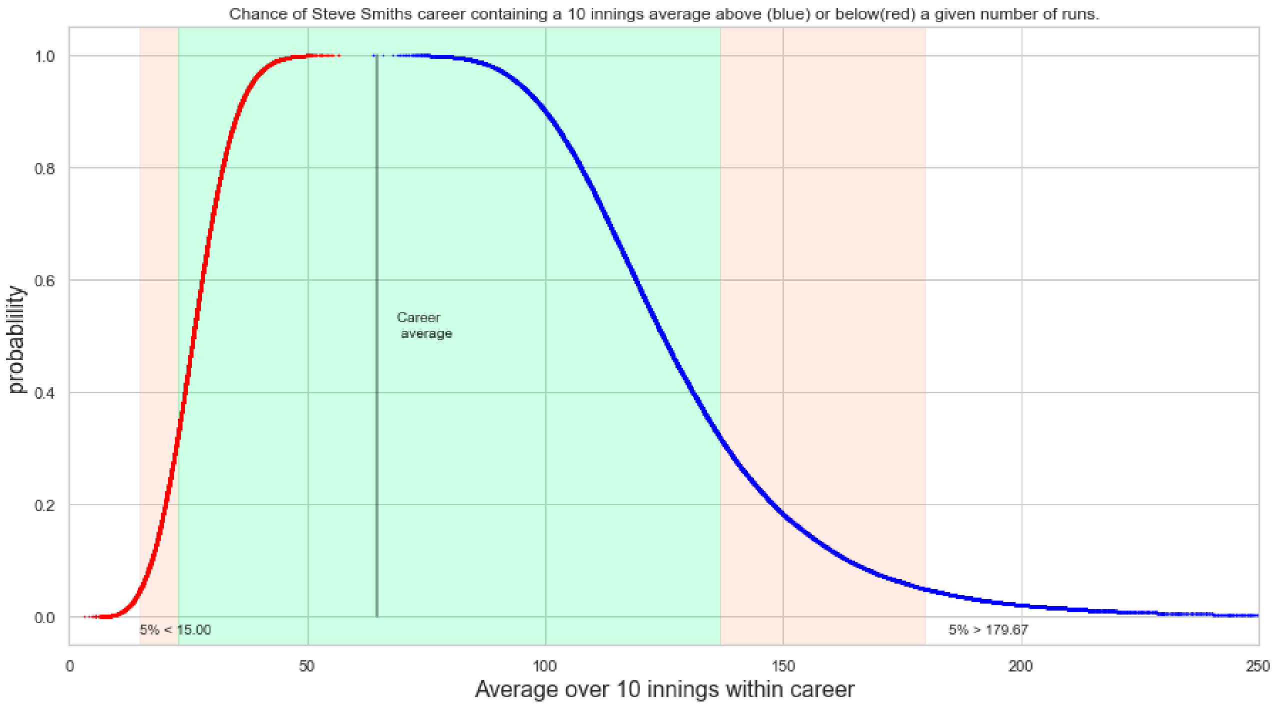
In this section I compare Steve Smiths and David Warners batting performances. I reproduce the graphs in section 1 as well as adding confidence bounds and displaying rolling averages over 10 innings. To do this I use the functions defined in section 2.

Shown below are the careers for each batsman with a rolling average in blue and an overall average as a dashed red lines. For an average over 10 innings confidence intervals of 68% (1 standard deviation) and 95% (approximately 2 standard deviations) are colored in green and salmon respectively.



I have also generated cumulative distribution functions for each player using bootstrapping. The confidence intervals are calculated from these functions.





3.1 Results

Comparing the two batsmen we see that Smith has a higher average and has played slightly fewer innings.

Smiths rolling average over 10 innings does not exceed the 1 standard deviation confidence interval (we would expect the careers of 68% of batsmen to do the same). Even Smiths strong performance during the 2019 Ashes did not exceed this interval. I note that Smiths performance appears to improve after around 30 or 40 innings.

Warners rolling averages to not exceeds the 1 standard deviation confidence interval around his 60th innings (on the high end) and again around 130 innings (on the low side). Even the very poor performance of Warner in the 2019 Ashes series does not quite push his rolling average outside the 95% confidence interval, though it comes very close.

An interesting note is that while Smiths average is around 18 runs per innings higher than Warners, this has a larger effect on their comparative confidence intervals. The 95% high confidence intervals for Smtih and Warner are 109 and 180 respectively, a difference of 71 runs.

In order to discern whether Warners performance is caused by poor form I look at a group of batsmen and compare their performance compared to statistical confidence intervals. This is covered in the next section.

4. Ensemble Batting Performances

In the section above I looked at the batting performance of Steve Smith and David Warner. The condifence interval of David Warners recent streak of poor performance during the 2019 ashes was 5.4% - placing it just above the usual bounds of statistical significance. However this does not mean that no effect is present (and similarly if the ci was below 5% it does not necessarily mean that there is an actuall effect present).

To further test the hypothesis that batsmen go through periods of poor performance we can generate 5% confidence intervals for a large group of batsmen. We can then generate confidence intervals of performance for each of these batsmen. Take the 5% confidence interval as an example. If 5% of these batsmen have careers with periods of poor performance (as measured by the 5% ci) then we can say that it is unlikely that players undergo actual periods of poor performance. Conversely if the rate at which players perform poorly is significantly greater than 5% we can say that players do indeed experience periods of poor performance.

In either case we can not definitively say whether David Warner experienced actual poor form during the Ashes, but we may be able to say whether he was unlucky or not.

The selection criteria for batsmen is to try and get a fairly large sample. Since I intend to look at the 5% confidence interval hundreds of samples are needed to accurately find the proportion of batsmen under this interval. The number of batsmen in the data set proved to be too few to achieve this, so instead i will consider anything below the 10% confidence interval to be poor form.

Selecting batsmen with a minimum number of innings as 40 and a minimum average of 20 returns 93 eligible batsmen. This is fewer than I would like, and a more complete dataset with matches extending back beyond 2019 would be useful to increase the accuracy of this analysis. Even so I will continue the analysis.

Results

Of 94 batsmen tested only 8 (8.5%) have poor performing periods which fall outside the 90% confidence interval. This result falls within a range of values which would be expected from statistical sampling. Note that the executed code finds only 7 batsmen who have poor periods, and I have added David Warner to the tally because his final four innnings push him below the confidence interval threshold.

In terms of statistical significance the number of batsmen used (94) is quite low. For this reason a confidence interval of 90% was chosen rather than 95% in order to aquire a more representitive sampling of positive and negative results. To improve the analysis a larger selection of batsmen is required. Extending the data set to matches before 2009, looking at first class matches (not just internationals) or even lookinng at one day formats would all be possible avenues of further analysis.

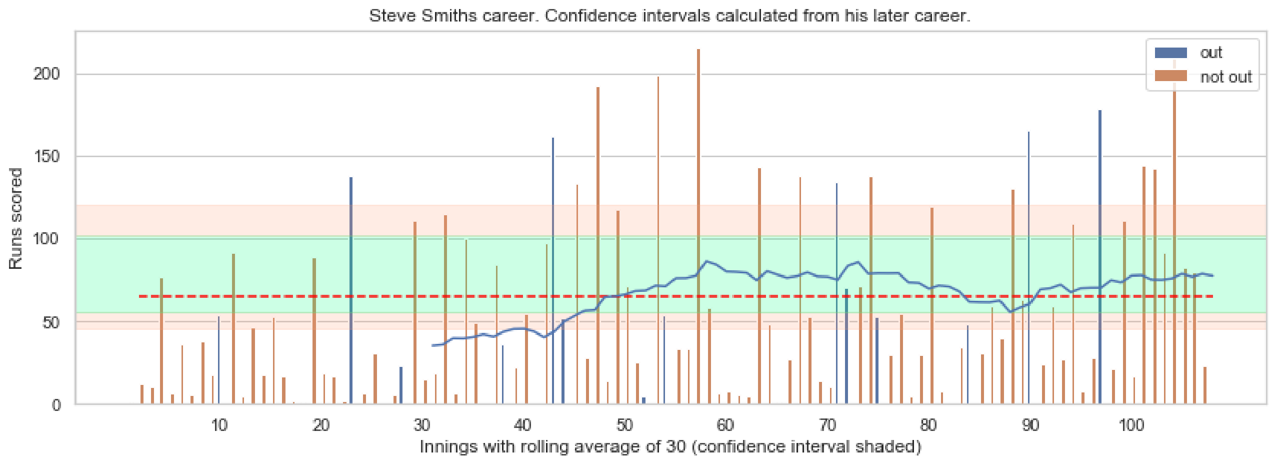
I also note that what constitutes poor form is fairly arbitrary. I set the benchmark at poor results over a window of 10 innings becuase this corresponds to 5 matches. 5 matches are the largest number of matches which are frequently played between two test match playing sides. Additionally 10 innings was motivated by the poor performance of David Warner in the 2019 Ashes series, which is widely regarded as a very dismal showing.

Looking at longer timescale trends does produce statistically significant results, as is the case of Steve Smiths early career vs overall career. The increase in performance is likley due to a real increase in his performance as he developed as a batsman.

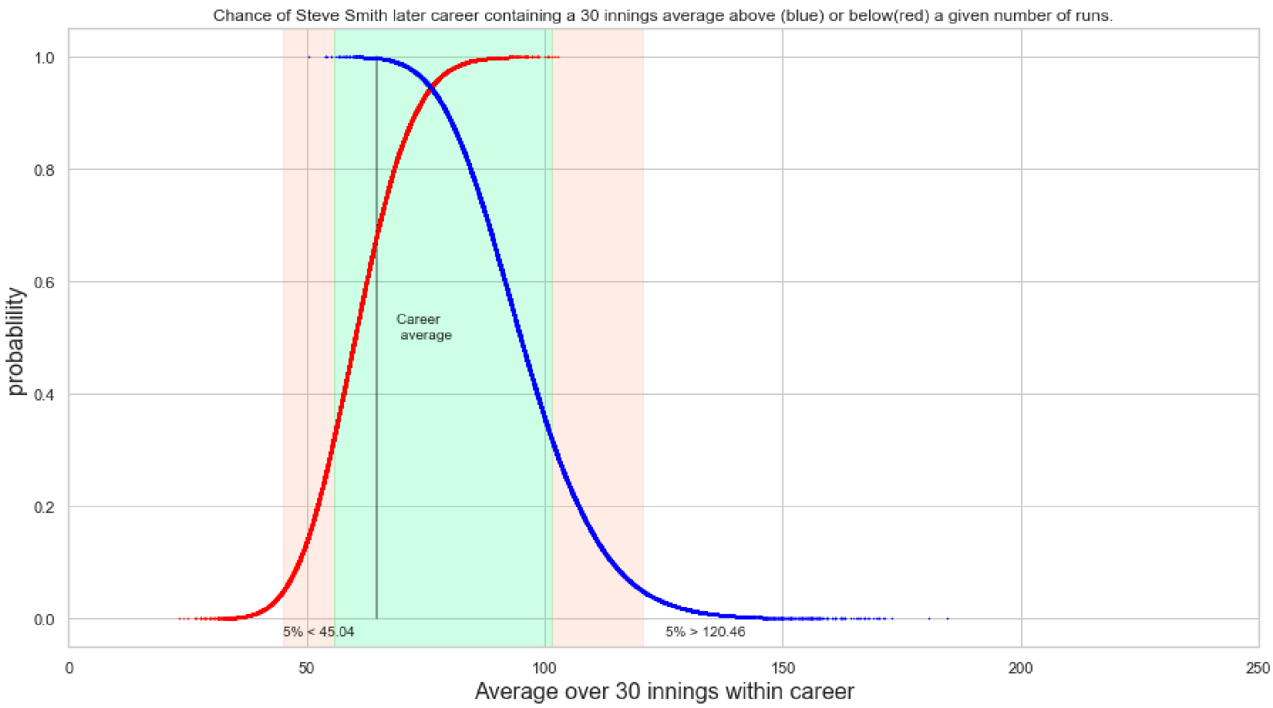
A further consideration when considering whether 10 innings is a fair benchmark for poor form is whether a player is dropped from the team for poor form prior to 10 innings, and then returned when they are in better form. This could artificially prevent the number of actual periods of low scoring in real careers. For this reason I reran the code in the previous cell with a rolling window of 6. There were 8 results (including Warner) which fell below the confidence interval. Again this is in line with those results being a result of statistical chance rather than actual poor form.

5. Possible longer term trends

As a quick appendix to the main results of this project I want to investigate whether players performance increases or decreases significantly over a longer period of time. These periods (30 innings) would not be considered a reflection of good or bad form, but rather reflect long term improvement or decline in a batsmans ability. While this coud be calculated for a large pool of batsmen, as above, for this section I will look specifically into Steve Smith.



The confidence intervals were created using a cumulative distribution function and bootstrapping. The CDF function is shown below.



5.1 Results

Adjusting the window size from 20 innings to 40 innings there were no periods where Smiths early career exceeds a 95% confidence interval. That said Smith did consistently fall close to the confidence interval over varying window sizes.

However if we assume that Smiths performance increase is due to a temporal change (an absolute improvement over time) then we can restrict sampling for the ECDF graph to innings recorded after 'window_size'. When this is performed the rolling averages of Smiths early career are significantly lower than the 95% confidence interval. For a window size of 30 the 95% ci is 45.04 runs and Smiths first 4 rolling averages are below 40 runs. This indicates that there is a strong possibility that Smiths performance has increased from his early career to his late career.

Not all batsmen has a consistent improvement in their average from their early career, and any improvement does not necessarily occur over the same timeframe. For this reason I do not provide a comparitive analysis using multiple batsmen, as I did in the case of the poor form analysis. With that being said a more sophisticated look into improvement over a career may be of interest.

6. Summary

I have shown that there does not appear to be any periods of poor form over fewer than 10 innings which could not be reasonably attributed to statistical chance. There does appear (at least in the case of Steve Smith and possibly others) to be longer term trends in performance (greater than 20 innings) are likely to be real phenomena as a batsman improves or declines throughout their career.

Selectors and players are under tremendous pressure by pundits and fans to perform consistently well. It is not uncommon for a few poor results from a batsman to cause people to question their ability. While it is hard to argue that a batsman who has recently produced poor results should not be dropped in favour of a batsman who is producing good recent results, this project shows that pundits and fans may be too quick to jump to conclusions about trends that may not exist.