SAS® GLOBAL FORUM 2018

USERS PROGRAM

Using the LIFETEST Procedure to Calculate an Alterative to Strike Rate in Limited Overs Cricket

April 8 – 11 | Denver, CO

#SASGF

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Using the LIFETEST Procedure to Calculate an Alterative to Strike Rate in Limited Overs Cricket

Keith Curtis

USAA

ABSTRACT

In limited overs cricket, batsmen are measured by their ability to score quickly. The most common statistic is the strike rate, which is the average number of runs scored per 100 balls faced. One drawback to the strike rate is that two batsmen with similar values may differ in their ability to sustain power. For example, both a batsman who scored 25 runs on 20 balls and one who scored 5 runs on 4 balls yielded a strike rate of 125. Strike rates also do not distinguish batsmen on their ability to avoid outs. For example, two batsmen who score 7 runs on 7 balls will both have a strike rate of 100, even though one batsman was out on the seventh ball, whereas the inning ended after the seventh ball for the other.

This E-Poster proposes an alternative measure, namely expected runs at x balls, where x is a positive integer. Different batsmen's ability to sustain power will be compared by plotting their expected run values against the corresponding ball count. Calculations will be performed by the Kaplan-Meier Product-Limit Estimator within the SAS/STAT® LIFETEST procedure. This procedure allows run totals in matches where the batsman was not out to be treated as censored observations.

Results from 2016-17 Season of the KFC Big Bash League (BBL) will be analyzed. The BBL is an Australia-based Twenty20 league.

METHODS

80 batsmen who faced at least 20 balls during the 2016-17 BBL season were first identified. At each ball count from 1 – 60, these batsmen's expected run values were calculated. This was performed by the Kaplan-Meier Product-Limit Estimator within the SAS/STAT® LIFETEST procedure. For example, a batsmen's expected run count at 12 balls is determined in the following manner:

1. In matches where the batsman exceeded 12 balls, record their run total at 12 balls.
2. In matches where the batsman did not exceed 12 balls, record their final run total.
3. In matches where the batsman was not out and faced 12 or fewer balls, record their final run total as a censored observation.

A detailed calculation example is shown in Appendix A for Michael Klinger of the Perth Scorchers.

After the Kaplan-Meier estimates were obtained, the selection of ball counts to represent short-term and long-term power were made. These choices were facilitated by Ordinary Least Squares (OLS) regression models that were fit by the SAS/STAT® REGRESSION procedure. Each model contained strike rate and average runs as explanatory variables. The response variable is an expected run value. Ten models were fit, or one model for each expected run value between 1 and 10 overs (6 – 60 balls). The results are summarized in the table below:

RESULTS

Expected runs at 6 and 30 balls were chosen to represent a batsman's short-term and long-term performance, respectively. These two rows are highlighted in grey. At 6 balls, strike rate is the main driver of expected runs, since the average runs coefficient yields a high p-value of 0.1556. At 30 balls, expected runs is a composite of power and endurance, as both strike rate and average runs are significant at $\alpha = 0.05$. Above 30 balls, strike rates have a negligible impact on expected run values. Since the regression coefficients for average runs (β) approach one as the number of balls increases, a 1-1 relationship between average runs and expected runs is approached.

		Strike Rate		Average Runs	
# Balls	R^2	β	p-value	β	p-value
6	0.378	0.059	0.0000	-0.026	0.1556
12	0.552	0.081	0.0000	0.104	0.0004
18	0.660	0.080	0.0001	0.272	< 0.0001
24	0.704	0.071	0.0031	0.420	< 0.0001
30	0.815	0.044	0.0487	0.607	< 0.0001
36	0.858	0.023	0.3004	0.743	< 0.0001
42	0.900	0.014	0.4903	0.843	< 0.0001
48	0.923	0.008	0.6782	0.916	< 0.0001
54	0.933	0.0009	0.9631	0.972	< 0.0001
60	0.933	0.0001	0.9957	0.973	< 0.0001

Based on their short-term (6 ball) and long-term (30 ball) expected run values, each batsmen was assigned to one of 6 clusters. This clustering was performed by the SAS/STAT® FASTCLUS procedure, and it is displayed in the scatterplot below. The x-axis contains the expected run count at 6 runs, while the 30 run count is represented by the y-axis. The six batsman that are closest to their respective cluster's centroid are denoted by the filled circle, and they will represent their cluster in the subsequent expected run graph:

The six clusters are also summarized in the table below. Based on this table, clustering of short-term and long-term expected run values creates batsmen groups with distinct power and endurance characteristics. Batsmen that yield similar strike rates, such as those in Clusters B and C, were segmented based on their ability to sustain power:

				Cluster Means			
Cluster	# Batsmen	Power	Endurance	Strike Rate	Avg. Runs	Exp. At 6 Balls	Exp. At 30 Balls
A	6	Poor	Poor	102.31	9.40	3.10	10.59
B	28	Average	Poor	124.02	11.68	6.39	13.08
C	27	Average	Average	130.03	25.43	4.92	22.23
D	9	Strong	Average	151.49	23.84	8.49	25.44
E	6	Strong	Strong	147.10	40.43	5.24	34.35
F	4	Elite	Elite	164.70	44.19	10.61	39.82
Overall	80			131.3	21.30	6.01	20.3

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

Using the LIFETEST Procedure to Calculate an Alternative to Strike Rate in Limited Overs Cricket

Keith Curtis

USAA

RESULTS CONTINUED

Each cluster is described below. Separate Tukey multiple comparison tests were performed to compare strike rate and average run means by cluster. These tests were performed by the SAS/STAT® GLM procedure, and they facilitated determining the power and endurance ratings for each cluster.

Cluster A: This cluster performs poorly on both power and endurance. It yields the lowest strike rate and average run values, and is statistically worse than every cluster except Cluster B.

Cluster B: This cluster contains 35% of the batsmen. It yields strike rates that are close to the average. However, the average run output is statistically worse than every cluster except for Cluster A.

Cluster C: This cluster contains 33.8% of the batsmen. It yields comparable power to Cluster B, and comparable endurance to Cluster D. For average runs, it is statistically different from every cluster except for Cluster D.

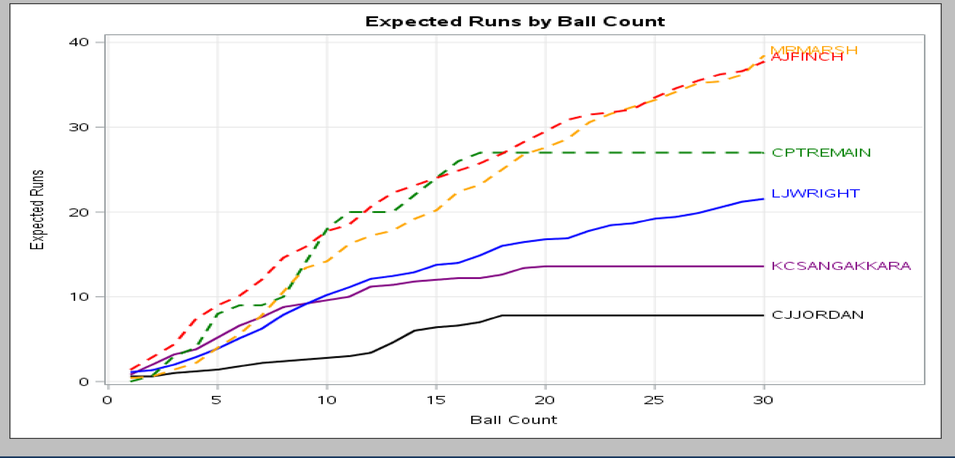
Cluster D: This cluster yields similar endurance to Cluster C, and for average runs is statistically different from every cluster except for Cluster C. However, this cluster yields more power than Cluster C, and its average strike rate of 151.49 is only below the best cluster (Cluster F).

Cluster E: This cluster produces similar power to Cluster D, but far better endurance. Only Cluster F yields a higher average run value. The average run values for Cluster E are statistically higher than Clusters A – D.

Cluster F: This cluster contains four elite batsmen, or the top 5%. They produce the best average results for both power and endurance. Its strike rates are statistically better than Clusters A – C, and only Cluster E has statistically equivalent average run values.

In the graph below, expected runs by ball count are compared for each cluster representative for five overs (30 balls). The line for CJ Jordan is markedly lower than the others, and that is consistent with Cluster A being poor for both power and endurance. The Cluster B and C representatives, namely KC Sangakkara and LJ Wright, show consistent expected run values for about two overs (12 balls). However, Wright is better at sustaining power, and his long-term expected run values approach the output from the Cluster D representative (CP Tremain). Tremain is among the best at achieving power through about 15 balls. However, his ability to sustain power beyond that is below the level of the Cluster E and F representatives. The Cluster E and F representatives are MR Marsh and AJ Finch, respectively. They both yield exceptional long-term expected run values. However, Marsh distinguished himself by achieving more power in the short term, as evidenced by his curve being the highest during the first over.

RESULTS CONTINUED



CONCLUSIONS

Strike rates do not distinguish batsmen in their ability to sustain power. By instead calculating expected run values at different ball counts, a batsman's short-term and long-term ability to hit for power can be assessed. Performing cluster analysis on these values allows batsmen with similar power and endurance characteristics to be grouped together.

Strike rates are also sensitive to outlier matches, and matches where batsmen are out quickly have minimal impact (See Appendix B).

A potential next step is to stratify batsmen by their typical position in the batting order (i.e. opening, upper, middle or lower). Other possible refinements include accounting for game situations, such as whether the batsmen's team hit or bowled first, and whether the match situation dictated more attempts at boundary scores (4s or 6s).

If you would like more information, you can reach the author at keithc71@yahoo.com.

REFERENCES

The data source for this E-Poster is www.cricsheet.org.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

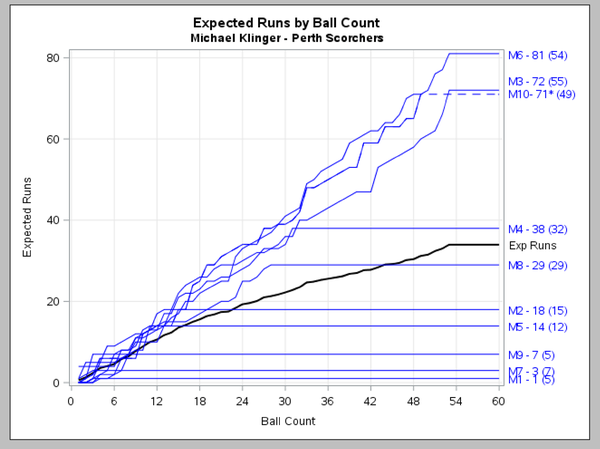
Using the LIFETEST Procedure to Calculate an Alternative to Strike Rate in Limited Overs Cricket

Keith Curtis

USAA

APPENDIX A: EXPECTED RUNS CALCULATION EXAMPLE

This section illustrates the expected run calculations for Michael Klinger of the Perth Scorchers. He scored 334 runs on 263 balls, which corresponds to a strike rate of 127.00. Klinger played in 10 matches, and was not out in one. The graph below plots Michael Klinger's run totals by number of balls. Each blue line represents a single match, while the thicker black line shows the expected number of runs at each ball count. In the match where Michael Klinger was not out at 49 balls, the line is dashed from 50 – 60 balls:



The table below contains a portion of the SAS/STAT® LIFETEST procedure that calculated Michael Klinger's expected number of runs at 50 balls. For eight of the ten matches, he faced fewer than 50 balls, and so his final run counts are used for these matches. His run count of 71 is treated as a right censored observation, however, since he was not out when this was achieved. This not out occurred on his 49th ball in the final, and was a 6 that clinched the KFC Big Bash League championship for the Perth Scorchers.

In Klinger's remaining two matches, namely Match 3 and 6, his run totals at 50 balls were obtained. These values are 61 and 72, respectively. Klinger increased his run total after 50 balls in both matches, as he yielded 72(55) and 81(54) in Match 3 and 6, respectively.

Based on the Kaplan-Meier Product-Limit Estimator, Micheal Klinger's expected run total at 50 balls is 31.5. If no censoring was performed, and a simple average of the 10 values is calculated, the expected run total is 31.4.

Product-Limit Survival Estimates

runs		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0		1	0	0	0	10
1		0.9	0.1	0.0949	1	9
3		0.8	0.2	0.1265	2	8
7		0.7	0.3	0.1449	3	7
14		0.6	0.4	0.1549	4	6
18		0.5	0.5	0.1581	5	5
29		0.4	0.6	0.1549	6	4
38		0.3	0.7	0.1449	7	3
61		0.2	0.8	0.1265	8	2
71	*	.	.	.	8	1
72		0	1	.	9	0

APPENDIX B: BEN DUNK V. D'ARCY SHORT

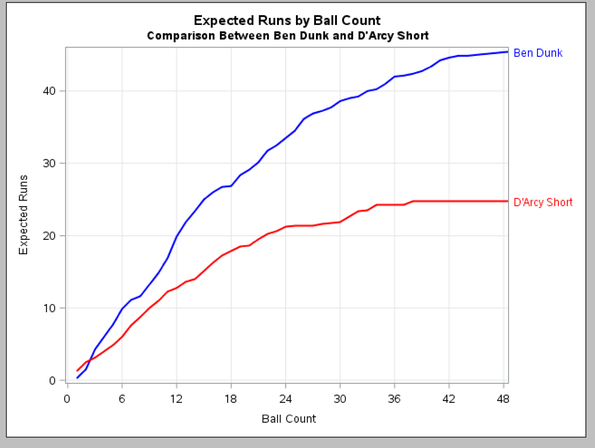
One challenge with interpreting a batsman's strike rate for a season is that matches are weighted by the number of balls faced. Therefore, a season strike rate can be driven by a strong outlier performance, and poor strike rate matches where a batsman is out quickly are not overly detrimental to the season strike rate.

Consider two batsmen, namely Ben Dunk of the Adelaide Strikers and D'Arcy Short of the Hobart Hurricanes. Dunk and Short's season strike rates nearly identical at 163.96 and 163.64, respectively. Their matches are tabulated below, and are sorted in descending order by strike rate:

Ben Dunk provides power on a more consistent basis. He yielded a strike rate higher than 160 in five matches, as compared to just two matches for D'Arcy Short. Short also had three matches where he faced three or fewer balls, as compared to only one for Dunk. However, these three matches are not overly detrimental to Short's strike rate, since the six total balls represent only 4.96% of his total ball count. In Short's highest strike rate match (210.34), he scored 61 runs from 29 balls, and that represents 24% of his season run count. When this match is removed, his season strike rate drops from 163.64 to 148.91. Dunk's best strike rate match (197.67) is less of an outlier. The 43 balls represented only 19.4% of his season count, and removing this match only drops Dunk's strike rate from 163.96 to 155.87.

When expected runs by ball count are examined instead of strike rates, Ben Dunk's better ability to sustain power is apparent. In the graph on the right, Dunk is already showing higher expected run values by one over (6 balls). At five overs (30 balls), his expected run value is 38.63. This compares to a value of only 21.88 for D'Arcy Short.

Ben Dunk					D'Arcy Short				
Match	Runs	Balls	Cumul % Total	Strike Rate	Match	Runs	Balls	Cumul % Total	Strike Rate
1	85	43	19.4%	197.67	1	61	29	24.0%	210.34
3	37	20	28.4%	185.00	7	18	11	33.1%	163.64
4	31	17	36.0%	182.35	4	60	39	65.3%	153.85
8	65	40	54.1%	162.50	8	36	24	85.1%	150.00
5	79*	49	76.1%	161.22	3	18	12	95.0%	150.00
6	35	25	87.4%	140.00	5	4	3	97.5%	133.33
7	32	25	98.6%	128.00	2	1	2	99.2%	50.00
2	0	3	100.0%	0.00	6	0	1	100.0%	0.00
Total	**364**	**222**		**163.96**	**Total**	**198**	**121**		**163.64**



SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

SAS® GLOBAL FORUM 2018

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.