ANOVA FOR TESTING THE CONSISTENCY AMONGST THE TOP TEN T20 ICC RANKED BATSMEN

[1]Dr. Nazimuddin Ahmed, [2]Prerona Sen, [3]Namrata Bhattacharjee
Department of Statistics, D.H.S.K. College, Dibrugarh- 786001, Assam

ABSTRACT

In cricket, the batsman performances are compared with their averages (Arithmetic mean) runs they have scored over their careers. However, the mean of a set of data is very far from telling the whole story. The obvious next thing to do is calculate its standard deviation. This will describe how 'spread out' a batsman's scores are. For better batsmen (i.e. those with higher averages) have more 'space' to spread into, which will result in larger standard deviations. On this measure alone the 'most consistent' performers are likely to be the lowest average scores. In this paper an attempt has been made to test the significance of the consistency of the batsman performances in T20 among the top 10 in T20, ICC ranked cricketer as on February, 2018. To meet the stated objective, secondary data was collected, and ANOVA has been carried out.

KEY WORDS

Players Runs, Mean, Standard Deviation, ANOVA.

1 Introduction

Cricket is a bat-and-ball game played between two teams of eleven players on a cricket field, at the centre of which is a rectangular 22-yard-long pitch with a wicket (a set of three wooden stumps) at each end. One team bats, attempting to score as many runs as possible, whilst their opponents field. Each phase of play is called an innings. After either ten batsmen have been dismissed or a fixed number of overs have been completed, the innings ends and the two teams then swap roles. The winning team is the one that scores the most runs, including any extras gained, during their innings. The international cricket council (ICC), the game's governing body is responsible for its rules and regulations.

The ICC Player Rankings are a sophisticated moving average. Players are rated on a scale of 0 to 1000 points. If a player's performance is improving on his past record, his point's increase; if his performance is declining his points will go down. The value of each player's performance within a match is calculated using an algorithm, a series of calculations (all pre-programmed) based on various circumstances in the match.

The factors for a batsman in the ratings are: runs scored ratings of the opposing bowling attack; the higher the combined ratings of the attack, the more value is given to the batsman's innings (in proportion) the level of run-scoring in the match. Batsmen who score highly in victories receive a bonus. That bonus will be higher for highly rated opposition teams. In T20 (Twenty20 cricket) batsmen gain significant credit for rapid scoring. They only get a small amount of credit for being not out.

T20 cricket, often abbreviated to T20, is a short form of cricket recognized by the International Cricket Council (ICC). It was originally introduced by the England and Wales Cricket Board (ECB) in 2003 for the inter-county competition in England and Wales. In a Twenty20 game the two teams have a single innings each, which is restricted to a maximum of 20 overs and a break between the innings.

2 A brief review of literature

Many research papers have been published which measured the performance of the players and their predictions. Duckworth and Lewis developed the D/L method, which is a mathematical formulation designed to calculate the target score for the batting second in a limited overs match interrupted by weather or other circumstances [7]. Croucher has chosen strike rate and the average scores as the measure of the performance [4]. Barr and Kantor used a new graphical representation together with the strike rate on one axis and the probability of getting out on the other [2]. Barr and Van Den Honert, explained a measure which is based on the average and a consistency [3]. Lemmer showed in his study that the batting average cannot be satisfied in the case of a small number of scores if the player had a large percentage of not out scores [6]. In a comparative study with four key players of India (Test match), Ahmed found Sachin Tendulkar as a wall in Indian cricket team (test matches) [1]. Dey et. al proposed a statistical based multi-criteria decision making analysis which provides a comparison between the batsmen in three IPL and evaluate the overall performances of batsmen [5]. Ghosh, A. et. al. analyzing the variance of the batting performance of Sachin Tendulkar in all venue by "*t*" test and ANOVA [6].

3 Different measures used in the present study

The measure, which is used in the present study is as follows:

The mean (or average) is the most popular and well known measure of central tendency. So, if we have n values in a data set having values x_1, x_2, Λ, x_n, then the mean and standard deviation are respectively given by:

$$\bar{x} = \frac{x_1 + x_2 + \Lambda + x_n}{n}, \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \quad .$$

3.1 Analysis of variance (ANOVA)

Analysis of Variance (ANOVA) is a hypothesis-testing technique used to test the equality of two or more populations (or treatment) means by examining the variances of samples that are taken. ANOVA allows one to determine whether the differences between the samples are simply due to random error (sampling errors) or whether there are systematic treatment effects that cause the mean in one group differ from the mean in another. The basic purpose of ANOVA is to test the homogeneity of several means. Most of the time ANOVA is used to compare the equality of three or more means, however, when the means of two samples are compared the ANOVA is equivalent to *t-test*.

Since ANOVA assumes the populations involved follow a normal distribution, ANOVA falls into a category of hypothesis tests known as parametric tests. If the populations involved did not follow a normal distribution, an ANOVA test could not be used to examine the equality of the sample means.

The assumptions of ANOVA are as follows:

- The samples are randomly selected and independent of one another.
- Parent population from which observations are taken is the normal distribution.
- Various treatments and environmental effects are additive in nature.
- Variances of populations are homogeneous (equal).
- The errors are independently, identically and normally distributed for fixed effects models.

3.1.1 ANOVA for one-way classified data

A one-way analysis of variance considers one factor/treatment (a single independent) with two or more levels. For example, if the independent variable (factor) is the price of a certain commodity, the classes may correspond to different seasons or different districts, etc. (called as levels) or if the variable (factor) is the crop of a variety of cereals, the classes may correspond to different manures (levels).

In one-way ANOVA, the total sum of squares (*SS*); can be partitioned into two parts- *SS* due to factor effect and *SS* due to error effect. It is one-way ANOVA, since there is only one independent variable (factor). The *F* value in one way ANOVA is a tool, help us to find whether the variances between the means of two populations are significantly different or not. The aim of the analysis is to test for differences among the means of the levels and to quantify these differences. If there are two factor levels, this analysis is equivalent to a *t* test comparing two group means. The one-way ANOVA table is as follows:

One-way ANOVA table

Source of variation	*df*	*SS*	*MSS*	*F*
Between	$k-1$	$S_A^2=\sum_{i=1}^{k} n_i(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2$	$s_A^2=\frac{S_A^2}{k-1}$	$F=\frac{s_A^2}{s_E^2}$
Within	$n-k$	$S_E^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})^2$	$s_E^2=\frac{S_E^2}{n-k}$	-
Total	$n-1$	$S_T^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\bullet\bullet})^2$	-	-

If the observed (calculated) value of *F* is greater than the tabulated value of *F* for $(k-1,\ n-k)$ *df* at a specified level of significance (usually 5%), then H_0 is rejected at that level of significance; otherwise it may be retained (or accepted).

4. Pre-requisite tables related to the present study

The names, ranks and the points of top ten Cricketer of T20 ICC ranking are shown in the following table:

Table No. 2: List of top ten T20 ICC Cricketers as on February, 2018

Rank	Name	Country	Innings played	Average	Points
1	Colin Munro	New Zealand	45	33.51	801
2	Glenn Maxwell	Australia	43	34.58	799
3	Babar Azam	Pakistan	17	48.08	786
4	Aaron Finch	Australia	36	40.20	763
5	Martin Guptill	New Zealand	75	34.40	747
6	Virat Kohli	India	57	50.84	741
7	Evin Lewis	West Indies	14	36.00	734
8	Lokesh Rahul	India	12	50.88	683
9	Alex Hales	England	52	31.65	679
10	Mohammad Shahzad	Afghanistan	60	31.85	653

Table No. 3: Players' Match by Match Runs (Partial)

Sl. No.	Player	Match by match Runs Scored						Total				Avg
								Runs	innings	Not Out	DNB	
1	C Munro	23	8*	3	…	57	29	1173	45	7	3	33.51
2	G Maxwell	4	27	DNB	…	31	20*	1072	43	7	5	34.58
3	B Azam	15*	55*	19	…	50*	18	577	17	5	0	48.08
4	A Finch	15*	53*	36	…	36*	18*	1206	36	6	0	40.20
5	M Guptill	0	41	10	…	62	21	2271	75	7	2	34.40
6	V Kohli	26*	DNB	28	…	89	4	1983	57	14	4	50.84

9	A Hales	0	62*	2	…	47	1	1456	52	6	0	31.65
10	M Shahzad	10	18	12	…	20	17	1816	60	3	0	31.85

on February, 2018

Notes: DNB means Did Not Bat and * means Not Out.

5 Objectives of the paper

The objective of the study is to test the consistency by using ANOVA of batting performances in T20 of the top 10 ranked cricketers, February, 2018. Present study analyses batsmen's performance in terms of two criteria, namely batting average and consistency or dependability.

6 Material for the study

As my objective is to analysis the batting performance in T20 of the top 10 ranked, ICC cricketers, so I have collected the secondary data from the website [9, 10] of information regarding the T20 runs of the top 10 T20 ranked cricketers on February, 2018.

7 Methodology

In this paper an attempt has been made to study the most consistent batsman in T20 among the top 10 T20 ranked cricketer. To meet the stated objective, the authors have collected the secondary data, i.e., downloaded the information of each score of top 10 ICC ranked batsman in T20 career since from his debut match from the website [9, 10].

For calculation of an average of the player runs, the number of times not-out and DNB are not considered, therefore, the authors have prepared a modified table of player's runs by ignoring the DNB and innings where the players are not-out is being added to the preceding innings or otherwise succeeding innings. By considering zero in place of DNB affect the results in the computations of the sum of squares. Also, we consider a player who scored an average run of more than 30 is considered as a good player. The ANOVA test was used to test the significance of batsman performances. We also considered the mean; standard deviation and graphical representation were considered in concluding the performance.

8 Null Hypothesis

$H_0: \mu_1 = \mu_2 = \Lambda = \mu_{10}$ i.e., the players' performances are homogeneous.

9 Calculation

We prepare a modified table of player's runs by ignoring the DNB and innings where the players are not-out is being added to the preceding innings or otherwise succeeding innings. The modified table is as follows:

Table No. 4: Modified Players' Match by Match Runs (Partial)

Sl. No.	Player	Match by match Runs Scored (y_{ij})								Total Runs ($\sum_{j=1}^{n_i} y_{ij}$)	Total Innings (Adjusted) (n_i)	Average $\bar{y}_{i\bullet} = \sum_{j=1}^{n_i}\left(\frac{y_{ij}}{n_i}\right)$
1	C Munro	31	3	28	7	74	…	57	29	1173	35	33.51
2	G Maxwell	4	31	12	1	27	…	39	51	1072	31	34.58
3	B Azam	116	29	27	43	38	…	91	18	577	12	48.08
4	A Finch	104	1	7	4	156	…	42	82	1206	30	40.20
5	M Guptill	0	41	10	51	8	…	62	21	2271	66	34.40
6	V Kohli	54	14	4	15	22	…	26	1	1983	39	50.84
7	E Lewis	0	100	7	1	3	…	144	51	468	13	36.00
8	L Rahul	47	132	8	71	22	…	89	4	458	09	50.88
9	A Hales	62	2	11	99	22	…	47	1	1456	46	31.65
10	M Shahzad	10	18	12	30	76	…	20	17	1816	57	31.85

9.1 Calculation of mean and standard deviation (SD)

The mean, and SD are respectively calculated by the following formulae:

$$\bar{y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}\ ,\ \bar{y}_{\bullet\bullet} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} \quad \text{and} \quad \sigma_i = \sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2}$$

From Table No. 3, the average runs, the sum of the square of deviation from average, and the standard deviation of each individual are calculated as follows:

Table No. 5: Calculation of SD and SS of the top 10, T20 cricketers

Sl. No.	Top T20 ICC Cricketer	Total runs	Adjusted Innings (n_i)	Average Runs ($\bar{y}_{i\bullet}$)	$\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2$	SD (σ_i)	Rank (Consistency)
1	C Munro	1173	35	33.51	48506.74	37.23	6th
2	G Maxwell	1072	31	34.58	53457.55	41.53	8th
3	B Azam	577	12	48.08	12130.92	31.79	3rd
4	A Finch	1206	30	40.20	38348.80	35.75	4th
5	M Guptill	2271	66	34.40	86577.96	36.22	5th

6	V Kohli	1983	39	50.84	76747.08	44.36	9th
7	E Lewis	468	13	36.00	26288.00	44.97	10th
8	L Rahul	458	09	50.88	14148.89	39.65	7th
9	**A Hales**	1456	46	**31.65**	36047.44	**27.99**	**1st**
10	M Shahzad	1816	57	31.85	55180.88	31.11	2nd

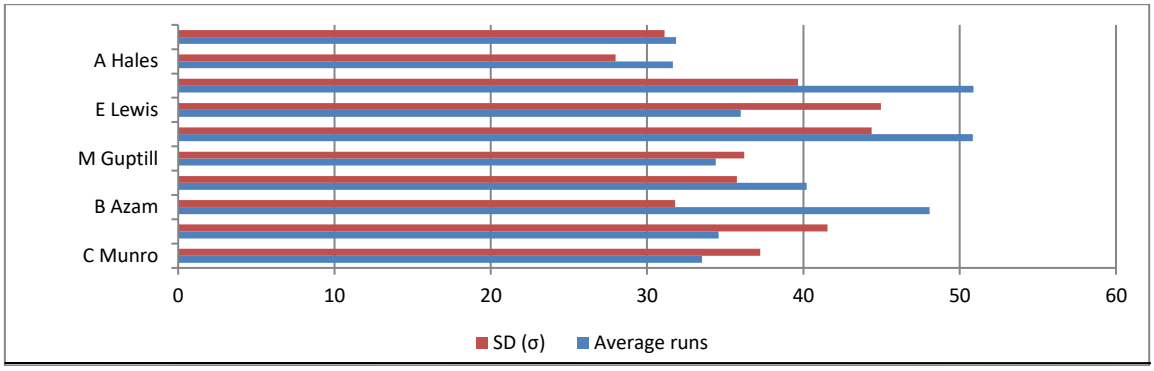

Figure 1: Mean and Standard deviations of the runs scored by the respective player

9.2 Calculation of *SSE*, *SST* and *TSS*

Table No. 6: Calculation of SSE, SST and TSS of the top 10, T20 cricketers

Sl. No.	Top T20 ICC Cricketer	Total runs scored	Adjusted Innings (n_i)	Average Runs ($\bar{y}_{i\bullet}$)	$n_i(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2$ Where, $\bar{y}_{\bullet\bullet}=36.92$	$\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})^2$	$\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\bullet\bullet})^2$
1	C Munro	1173	35	33.51	406.98	48506.74	48912.70
2	G Maxwell	1072	31	34.58	169.74	53457.55	53627.20
3	B Azam	577	12	48.08	1494.55	12130.92	13626.36
4	A Finch	1206	30	40.20	322.75	38348.80	38671.55
5	M Guptill	2271	66	34.40	419.13	86577.96	86994.06
6	V Kohli	1983	39	50.84	7556.89	76747.08	84310.65
7	E Lewis	468	13	36.00	11.00	26288.00	26299.00
8	L Rahul	458	09	50.88	1753.93	14148.89	15905.09
9	**A Hales**	1456	46	**31.65**	1277.55	36047.44	37640.13
10	M Shahzad	1816	57	31.85	1465.18	55180.88	56640.48
Total		**12480**	**338**	-	$SST=\sum_{i=1}^{k}n_i(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2$ **=14877.71**	$SSE=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\bullet})^2$ **=447434.26**	$TSS=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\bullet\bullet})^2$ **=462627.20**

9.3 Calculation of variance ratio (*F*)

Table No. 7: ANOVA table

Source of variation	*df*	*SS*	*MSS*	*F*	$F_{9,\ 328}$ at 5%
Between	9	14877.71	1653.08	1.21	1.91
Within	328	447434.26	1364.13		
Total	337	462627.20	-	-	-

Since, the calculated value of F at (9, 328) d.f. is smaller than the tabulated value of F at 5% specified [11] level of significance, therefore, H_0 is accepted at that level of significance and conclude that the consistency of the batsman performances are homogeneous.

10 Conclusions:

From table 5, if we compare the standard deviations of all the ten players, the standard deviation of Alex Hales (σ = 27.99) of England is less as compared to the other cricketers, therefore we conclude that Alex Hales is the most consistent or uniform and successful performer among the top 10 ICC T20 cricketer as in February, 2018 in the world. Also, we may expect from Hales that he can score his average runs (i.e., 31.65) more often.

Next followed by Mohammad Shahzad and Babar Azamof Afganistan & Pakistan are the next most consistent player with their respective averages and standard deviations. Two Indian Batsman Lokesh Rahul and Virat Kohli are stood at 7th and 9th position respectively.

From ANOVA, table 7, batsman performances are homogeneous.

REFERENCES

[1] Ahmed, N. "The 'WALL' of the Indian Cricket Team (Test Match)". *VAK*. 2 (2011), 44-54, Print.

[2] Barr, G.D.I. and Kantor, B.S. "A criterion for comparing and selecting batsmen in limited overs cricket". *Journal of the Operational Research Society*. 55 (2004), 1266-1274, Print.

[3] Barr, G.D.I. and Van Den Honert, "R. Evaluating batsmen's scores in test cricket". *South African Statistical Journal*. 32 (1998), 169-183.

[4] Croucher, J.S. "Player ratings in one-day cricket. In: Proceedings of the Fifth Australian Conference on Mathematics and Computers in Sport". *Eds: Cohen, G. and Langtry, T. Sydney University of Technology*, *NSW*. (2000), 95-106, Print.

[5] Dey, P.K., Ghosh, D.N., Mondal, A.C. (2012): Statistical Based Multi-Criteria Decision Making Analysis for Performance Measurement of Batsmen in Indian Premier League, International Journal of Advanced Research in Computer Science, Volume 3, No. 4, 51 - 57.

[6] Duckworth, F. and Lewis, T. "Review of the application of the Duckworth/Lewis method of target resetting in one-day cricket". *Proceedings of the Sixth Australian Conference on Mathematics and Computers* in Sport. Eds: Cohen, G. and Langtry, T.Sydney University of Technology, *NSW*. 1(2002), 27-140, Print.

[7] Ghosh, A., Mondal, P. and Das, A. (2012): An analysis of Sachin Tendulkar's ODI performance, International Journal of Behavioral Social and Movement Sciences, Vol.01, Issue04, 10 – 17.

[8] Lemmer, H.H. "Measures of batting performance in a short series of cricket matches. *South African Statistical Journal*. 42(2008), 83-105, Print

[9] www.stats.cricinfo.com

[10] www.howstat.com/cricket/statistics/Players/PlayerMenu.asp

[11] https://www.danielsoper.com/statcalc/calculator.aspx?id=4

*****