International Journal of Mathematical Education.
ISSN 0973-6948 Volume 13, Number 1 (2023), pp. 15-24

http://www.ripublication.com

Some Quirks of Batting Statistics in Cricket: A Quantitative Examination

Arjun Tan

Department of Physics, Alabama A & M University,
Normal, AL 35762, U. S. A.

Abstract

In the game of cricket, the statistical categories recorded are abundant. In the batting department, the meaning of some of the statistics are unclear and require further examination. We analyze three specific cases for ambiguities and provide alternative suggestions. In this context, we examine the batting statistics of the 25 greatest batsmen who have averaged at least 50 in Test cricket while batting at least 100 times and ended their careers by 2022. First, by half centuries (50s) is meant innings score of between 50 and 99. Any innings which progresses to 100 is discounted. In order to count all innings of 50 or greater, we propose to introduce a quantity called total half centuries (Σ50s) by adding the centuries (100s) to the half centuries: Σ50s = 50s + 100s. This quantity is highly correlated to the aggregate Test runs, with a correlation coefficient of .98456, and therefore best represents the half century statistics. Second, the centuries (100s) include the multiple centuries (double, triple and quadruple centuries) by counting the latter as single centuries, even though double centuries are two centuries, triple centuries are three centuries, and so on. This can severely diminish the total centuries made by the batsmen. To represent the last quantity, we propose the statistics: Σ100s = 100s + 200s + 300s + 400s. The correlation coefficient of Σ100s with the batting average increases only marginally in this process. Third, Test cricket statistics are included in the First Class cricket statistics. Statistics for First Class matches which are not Test matches are not kept separately. It is proposed to name the latter matches as Other First Class (OFC) matches. One would expect the Test batting averages to be smaller than the OFC batting averages owing to the stronger opposition in Test cricket. However, it is found that a slight majority of the 25 great batsmen actually recorded their Test averages higher than their OFC averages. This shows that many great batsmen are able to raise their batting performance in the Test cricket arena.

INTRODUCTION

The game of cricket is unsurpassed in terms of statistical categories recorded. Amongst the individual batting statistics are: innings batted, unbeaten innings (not outs), runs scored, batting average, centuries made (100s), double centuries (200s), triple centuries (300s), quadruple centuries (400s), etc., half centuries (50s), boundaries made (4s), over boundaries (6s), modes of dismissals (caught, bowled, leg before wicket, stumped, hit wicket, run outs, etc.), partnerships records, and so on. Also, before the advent of one-day cricket, cricket matches ratified and recorded by MCC (Marylebone Cricket Club) and later by ICC (International Cricket Council) were of two categories: First Class Cricket (matches played over three days or longer) and Test Cricket (matches played over 5 days or longer, usually between two national cricket teams). We find that there are ambiguities in some of these batting statistics as to what they actually mean and what they could or should have meant. In this paper, we examine three such cases: (1) Half centuries; (2) Total centuries; and (3) First Class records. We accomplish this by examining the batting careers of the 25 greatest batsmen who have batted in at least 100 Test innings averaging at least 50 and ended their careers by 2022 [1]. The statistics are readily available in the internet [2 − 4] and are compiles in Tables for the relevant cases.

HALF CENTURIES IN CRICKET

We first investigate the ***half centuries*** made in Test cricket by the 25 great batsmen in history [1]. These are listed under **50s** in **Table I** along with other relevant batting statistics. We find that these are actually innings which reached 50s but came short of 100s since the latter are included in **100s** and are excluded from 50s. Thus, the half centuries named in this category actually refer to ***additional half centuries***. These are scores between 50 and 99 and exclude those which reach 100s. This scheme is therefore ***exclusionary*** in nature. In order to include all half centuries (including those which progress to centuries), we define a statistical category of ***total half centuries*** denoted by ***Σ50s***. Mathematically, this can be written as:

$$\Sigma 50s = 50s + 100s \tag{1}$$

It has been pointed out that most batting statistics depend upon the number of innings batted and are therefore ***extensive*** in nature [1]. These include the ***aggregate runs scored***, 50s, 100s, Σ00s, etc. The ***batting average***, on the other hand, is an ***intensive*** quantity, which is fairly independent on the number of innings batted [1]. According to Eq. (1), Σ50s, the sum of two extensive quantities, is also an extensive quantity.

Since the batting averages of the 25 batsmen fall in a fairly narrow range (50 - 58), it is expected that the extensive quantities would exhibit some ***positive correlations*** between them. As such, 50s, 100s and Σ50s, should all show positive correlations with the aggregate runs. **Figure 1** shows the scatterplot of these three quantities against the total runs scored by the batsmen. In all three cases, ***linear trend lines*** sufficiently describe the inter-relationship between the abscissas and ordinates two variables *x* and *y*:

$$y = mx + c \tag{2}$$

The ***slope*** *m* and ***y-intercept*** *c* are given, respectively, by:

$$m = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{3}$$

and

$$c = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{4}$$

where the summation Σ runs from $n = 1$ to 25. The scatter of the data points about the trend lines is clearly the least for the Σ50s line, suggesting that this variable perhaps best represents the half centuries made by a batsman. The ***Pearson's correlation coefficient*** *r* is calculated as:

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \tag{5}$$

where the averages of the variables are given by: $\bar{x} = \Sigma x_i / n$ and $\bar{y} = \Sigma y_i / n$. The results show that the addition of 100s increases the *r* value from .91696 for 50s to .98456 for Σ50s. This confirms that ***Σ50s best represents the half centuries made by a batsman in Test cricket***.

CENTURIES & MULTIPLE CENTURIES IN CRICKET

Centuries and ***multiple centuries*** are another quirky statistics in cricket. Centuries (**100s**) are defined as all individual innings of 100 or more runs. Thus, they include ***double centuries*** (**200s**), ***triple centuries*** (**300s**), ***quadruple centuries*** (**400s**), and so on. Likewise, double centuries include the triple and quadruple centuries; the triple centuries include the quadruple centuries and quintuple centuries (if any), and so on. Technically, however, double centuries consist of two centuries, triple centuries comprise three centuries, and so on. Hence counting multiple centuries as single centuries deprives the batsmen of their dues. This is an example of yet another exclusionary scheme in batting statistics in cricket. In order to correct this shortcoming and account for all centuries, one can define a statistical category called ***total centuries*** (**Σ100s**) as follows (cf. [5]):

$$\Sigma 100s = 100s + 200s + 300s + 400s + \cdots \tag{6}$$

Table II lists all the centuries (100s) and multiple centuries (200s, 300s & 400s) registered by the 25 great batsmen in Test cricket, from which the total centuries (Σ100s) are calculated. Note that the lone quadruple century recorded in Test cricket

was made by Brian Lara, who also made the lone quintuple century in First Class cricket.

Scatterplots of 100s and Σ100s were drawn against the aggregate runs and correlation coefficients were calculated. It was found that the *r* values were high as usual but the value actually decreased from .94663 for 100s to .91893 for Σ100s. This is probably due to some factors as yet undetermined. Examining the multiple centuries data, we find that the enhancements of 5 or more centuries from 100s to Σ100s were dependent on the batting averages and were the highest for the upper and middle ranks of the batmen (highlighted in Table II). For this reason, scatterplots of 100s and Σ100s were drawn against the batting averages instead in **Fig. 2**. The scatters were quite random and correlations were small. Nonetheless, the *r* value increased from −.04459 for 100s to .03819 for Σ100s. This only marginally supports the concept of using the total centuries Σ100s. Nevertheless, ***counting all centuries Σ100s instead of just 100s is logically justifiable***.

FIRST CLASS MATCHES & TEST MATCHES

As stated earlier, ***First Class cricket*** matches are MCC/ICC sanctioned cricket matches of three days or more duration. These are generally matches between two local/regional teams but may also involve a touring foreign team. This is where promising young players prove their skills and progress to play for their national teams. ***Test cricket*** matches are matches between two national teams of five days or more duration. They are the venues where the worth of the Test players are ultimately judged. By definition, Test matches are also First Class matches, but the converse is not always true. The majority of First Class matches are not Test matches. The last category of matches can be termed ***Other First Class*** (OFC) matches [5]. However, the statistics of these matches are not kept separately and have to be calculated from those of First Class and Test match statistics [5]. For example, the OFC batting average of a player A_{OFC} can be calculated as follows. If R_{FC} = Total First Class runs; R_T = Total Test runs; D_{FC} = First Class outs; and D_T = Test outs, then (cf. [5]):

$$A_{OFC} = \frac{R_{FC} - R_T}{D_{FC} - D_T} \tag{6}$$

Contrary to the cases of half centuries and centuries, this scheme of lumping Test matches into the First Class category is ***inclusionary*** in nature. Here, a mathematical operation has to be conducted to separate the Other First Class statistics from the Test statistics. In this section, the OFC batting averages of the 25 great batsmen are determined and entered in **Table III**. The mean values of the First Class, Test and OFC batting averages are shown in the table. Normally, one would expect the Test averages to be inferior to the OFC averages since the Test opponents are generally stronger than the OFC counterparts. However, the table shows that the Test averages of a slight majority of the great batsmen (13 out of 25), highlighted in the table, were actually higher than their OFC averages. The mean Test average of 53.60 was also decidedly higher than the mean OFC average of 52.34. Interestingly, Ken Barrington

recorded the greatest enhancement in batting average of 12.66 from the lowest OFC average of 43.01 to the top of the Test batting average list of 58.67. The formidable Sachin Tendulkar, scorer of the most Test runs and Test centuries registered the largest decline of 12.47, but from the highest OFC average of 66.25 to a still very high Test average of 53.78.

Figure 3 is a scatterplot of the Test average against the OFC average. There is a great scatter of data points showing the randomness of the variables. The solid line representing the least squares error regression line, determined from Eqs. (2−4), shows a small negative correlation with a correlation coefficient of −.09381 obtained from Eq. (5). The dashed line represents the locations on which the Test average equals the OFC average. The data points northward and leftward of this line represent the batsmen who had raised their Test averages above their OFC averages in spite of playing generally against stronger opposition. This merely confirms the fact that ***many great batsmen are able to raise their batting performance in the greater arena of Test cricket***.

Table I. 50s, 100s & Total 50s made by Great Batsmen in Test Cricket

		Team	Ave	Runs	50s	100s	Σ50s
1	Barrington	E	58.67	6,806	35	20	55
2	Hammond	E	58.45	7,249	24	22	46
3	Sobers	WI	57.78	8,032	30	26	56
4	Sangakkara	SL	57.40	12,400	52	38	90
5	Hobbs	E	56.94	5,410	28	15	43
6	Hutton	E	56.67	6,971	33	19	52
7	Kallis	SA	55.37	13,289	58	45	103
8	G Chappell	A	53.86	7,110	31	24	55
9	Tendulkar	I	53.78	15,921	68	51	119
10	Lara	WI	52.88	11,953	48	34	82
11	Miandad	P	52.57	8,832	43	23	66
12	Dravid	I	52.31	13,288	63	36	99
13	Md Yousuf	P	52.29	7,530	33	24	57
14	Younis Khan	P	52.05	10,099	44	34	78
15	Ponting	A	51.85	13,378	62	41	103
16	Flower	Z	51.54	4,794	27	12	39
17	Hussey	A	51.52	6,235	29	19	48
18	Chanderpaul	WI	51.37	11,867	66	30	96
19	Gavaskar	I	51.12	10,122	45	34	79
20	S Waugh	A	51.06	10,927	50	32	82
21	Hayden	A	50.73	8,625	29	30	59
22	De Villars	SA	50.66	8,765	46	22	68
23	Border	A	50.56	11,174	63	27	90
24	Richards	WI	50.23	8,540	45	24	69
25	Compton	E	50.06	5,807	28	17	45

A = Australia; E = England; I = India; P = Pakistan; SA = South Africa; SL = Sri Lanka; WI = West Indies; Z = Zimbabwe

Table II. 100s, 200s, 300s, 400s and Total 100s made by Great Batsmen in Test Cricket

Rank	Batsman	Ave	Runs	100s	200s	300s	400s	Σ100s
1	Barrington	58.67	6,806	20	1	-	-	21
2	Hammond	58.45	7,249	22	7	1	-	30
3	Sobers	57.78	8,032	26	2	1	-	29
4	Sangakkara	57.40	12,400	38	11	1	-	50
5	Hobbs	56.94	5,410	15	1	-	-	16
6	Hutton	56.67	6,971	19	4	1	-	24
7	Kallis	55.37	13,289	45	2	-	-	47
8	G Chappell	53.86	7,110	24	4	-	-	28
9	Tendulkar	53.78	15,921	51	5	-	-	56
10	Lara	52.88	11,953	34	9	2	1	46
11	Miandad	52.57	8,832	23	6	-	-	29
12	Dravid	52.31	13,288	36	5	-	-	41
13	Md Yousuf	52.29	7,530	24	4	-	-	28
14	Younis Khan	52.05	10,099	34	6	1	-	41
15	Ponting	51.85	13,378	41	6	-	-	47
16	Flower	51.54	4,794	12	1	-	-	13
17	Hussey	51.52	6,235	19	-	-	-	19
18	Chanderpaul	51.37	11,867	30	2	-	-	32
19	Gavaskar	51.12	10,122	34	4	-	-	38
20	S Waugh	51.06	10,927	32	1	-	-	33
21	Hayden	50.73	8,625	30	2	1	-	33
22	De Villars	50.66	8,765	22	2	-	-	24
23	Border	50.56	11,174	27	2	-	-	29
24	Richards	50.23	8,540	24	2	-	-	26
25	Compton	50.06	5,807	17	2	-	-	19

Table III. Batting Averages of Great Batsmen in First Class, Test & Other First Class Cricket

Rank	Batsman	First Class Cricket			Test Cricket			OFC
		Runs	Outs	Ave	Runs	Outs	Ave	Ave
1	Barrington	31,714	695	45.63	6,806	116	58.67	43.01
2	Hammond	50,551	901	56.10	7,249	124	58.45	55.72
3	Sobers	28,314	516	54.87	8,032	139	57.78	53.79
4	Sangakkara	20,911	399	52.40	12,400	216	57.40	46.50
5	Hobbs	61,760	1218	50.70	5,410	95	56.94	50.17
6	Hutton	40,140	723	55.51	6,971	123	56.67	55.28
7	Kallis	19,695	364	54.10	13,289	140	55.37	51.66
8	G Chappell	24,535	470	52.20	7,110	132	53.86	51.55
9	Tendulkar	25,396	439	57.84	15,921	296	53.78	66.25
10	Lara	22,156	427	51.88	11,953	226	52.88	50.76
11	Miandad	28,663	537	53.37	8,832	168	52.57	53.74
12	Dravid	23,794	430	55.33	13,288	254	52.31	59.69
13	Md Yousuf	10,152	206	49.28	7,530	144	52.29	42.29
14	Younis Khan	17,116	343	49.90	10,099	194	52.05	47,09
15	Ponting	24,150	432	55.90	13,378	258	51.85	61.90
16	Flower	16,379	303	54.05	4,794	93	51.54	55.16
17	Hussey	22,783	437	52.13	6,235	121	51.52	52.36
18	Chanderpaul	27,545	518	53.17	11,867	231	51.37	54.62
19	Gavaskar	25,834	502	51.46	10,122	198	51.12	51.68
20	S Waugh	24,052	463	51.94	10,927	214	51.06	52.71
21	Hayden	24,603	468	52.57	8,625	170	50.73	53.61
22	De Villars	10,689	215	49.71	8,765	173	50.66	45.80
23	Border	27,131	528	51.38	11,174	221	50.56	51.97
24	Richards	36,212	733	49.40	8,540	170	50.23	49.15
25	Compton	38,942	751	51.85	5,807	116	50.06	52.18
	Mean	27,328	521	**52.50**	9,404	177	**53.60**	**52.34**

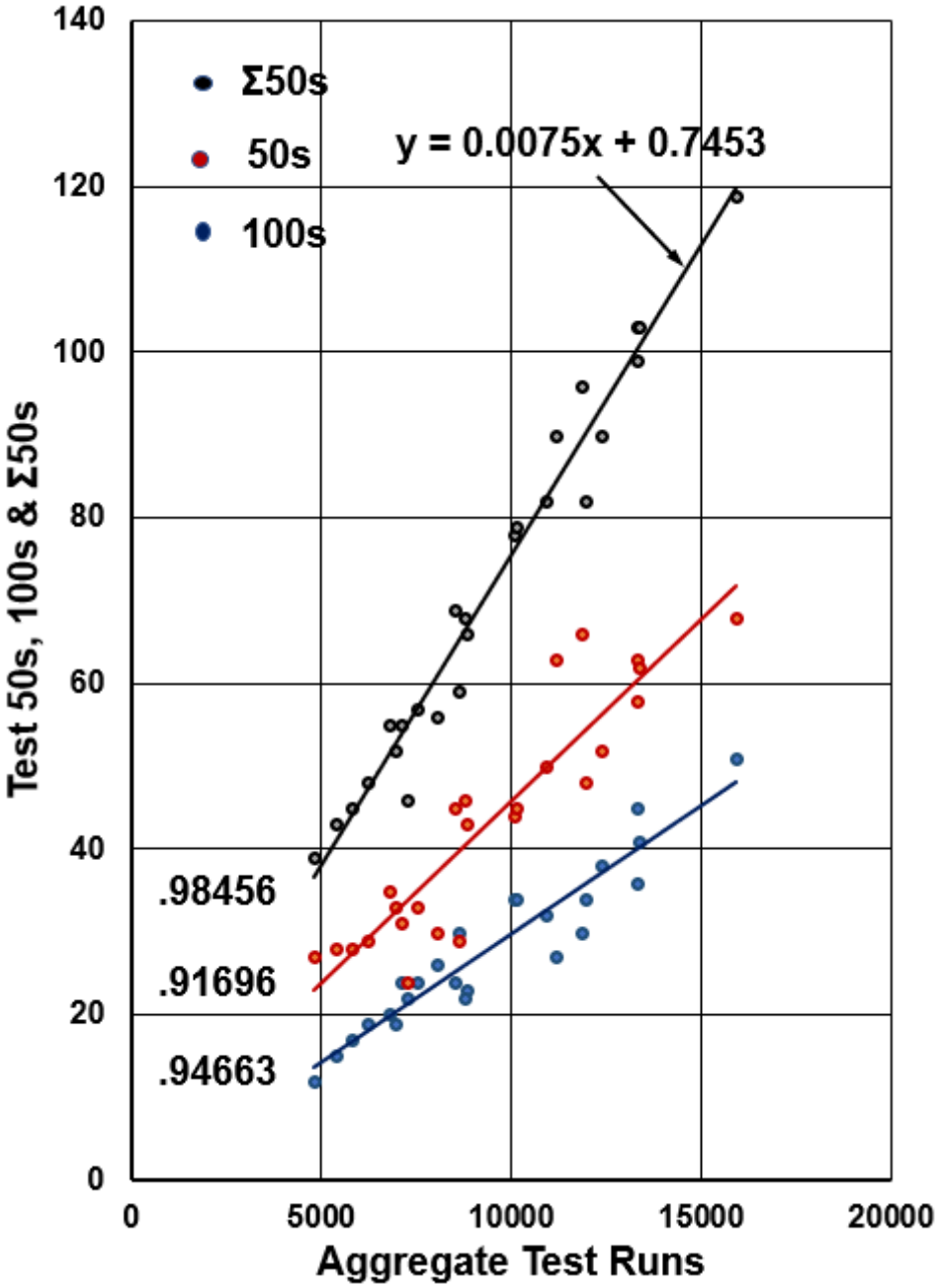
Σ50s
50s
100s
y = 0.0075x + 0.7453
.98456
.91696
.94663
Test 50s, 100s & Σ50s
Aggregate Test Runs
0
20
40
60
80
100
120
140
0
5000
10000
15000
20000

Figure 1

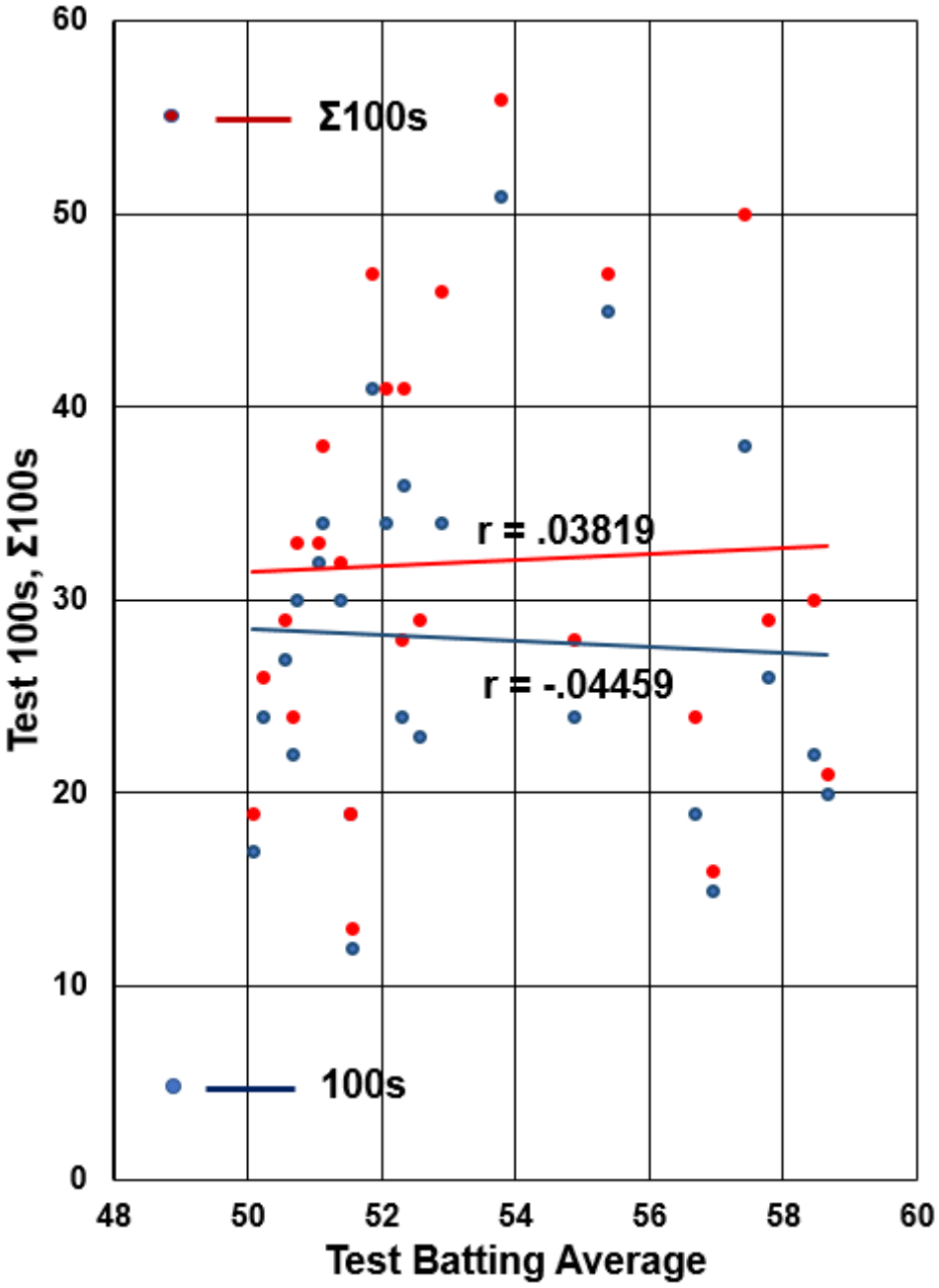
Σ100s
r = .03819
r = -.04459
100s
Test 100s, Σ100s
Test Batting Average
0
10
20
30
40
50
60
48
50
52
54
56
58
60

Figure 2

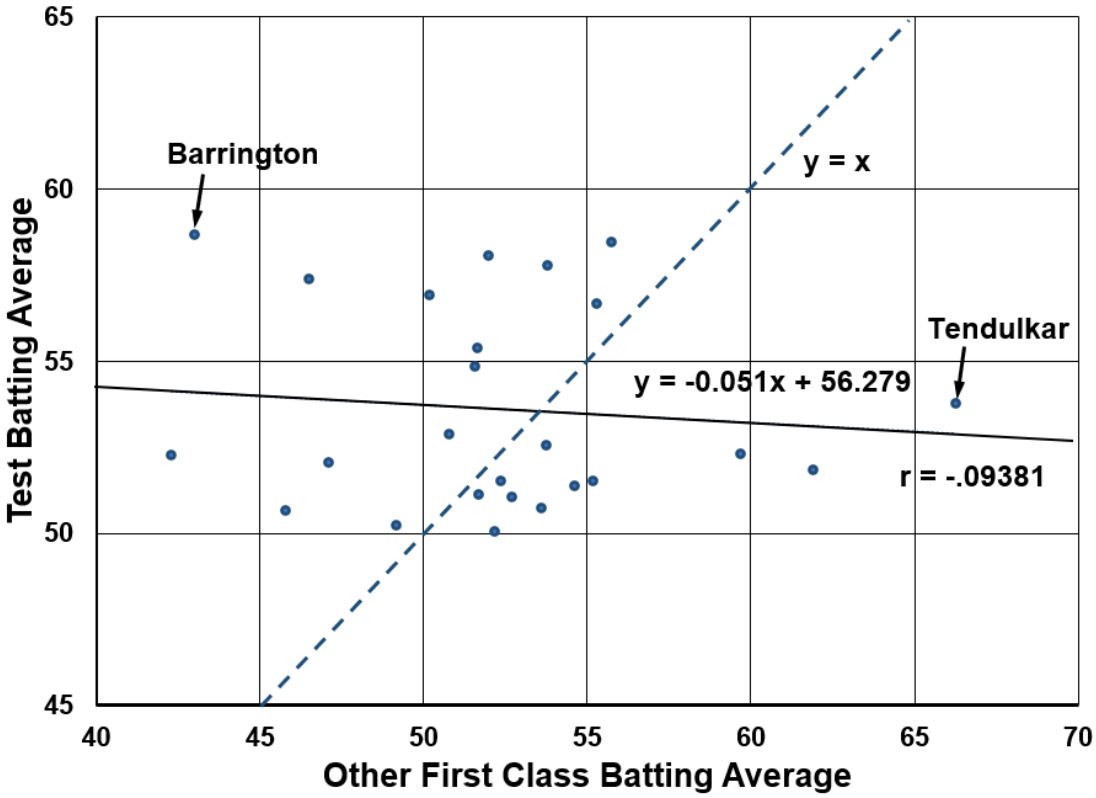

Figure 3

REFERENCES

[1] Arjun Tan, Categorizing the best batsmen in Test cricket, *Int. J. Math. Educ.*, **13**, 1 (2023).
[2] https://cricketarchive.com.
[3] https://espncricinfo.com.
[4] https://en.wikipedia.org>wiki>Cricket_statistics.
[5] A.Tan, The Three Ws of West Indian Cricket – A statistical analysis, *Math. Spectr.,* **33**, 11 (2000-1).