

# Analysis of Batting Performance in Cricket using Individual and Moving Range (MR) Control Charts

Muhammad Daniyal<sup>1</sup>, Tahir Nawaz<sup>1</sup>, Iqra mubeen<sup>2</sup>, Muhammad Aleem<sup>1</sup>

<sup>1</sup>Department of Statistics, the Islamia University of Bahawalpur.

<sup>2</sup>Department of statistics, the university of Sargodha

*(Received August 24, 2011, accepted September 9, 2011)*

**Abstract.** In cricket the performance of the players whether bowler or batsman is being analyzed with the help of very simple statistical tools. Mostly average scores, strike rate, average runs per wicket are being used. The main idea of this research study is to give the application of quality control charts in evaluating the performance of the players. We have made the comparative study of the two outstanding batsman of this game. We used the individual and moving range control charts for evaluating their performances. We applied basic sensitizing rules to analyze the non random and un-natural variation present in their performance. We will give the opportunity to analysts to check in the light of their knowledge for the assignable causes that have made the scores out of control so that in future these un-natural variations can be controlled and performance can be improved.

**Keywords:** batsman, performance, moving range control charts, individual control charts graphical displays, assignable cause, non-normality.

## 1. Introduction

Cricket is becoming the most popular game of the today's world. It is the game of bat and ball. It includes the team of eleven players. Due to the popularity of this game, media has always given its full coverage to the issues related to it. Most of the analysis is now being conducted on the performance of the players. Most of the studies have been conducted to analyze those factors which have become important in this game. The records and the statistics of each and every moment of this game are available on different sports sites of the world. The experts using these data analyze and give their views. The international cricket council (ICC) is responsible for its rules and regulations.

In cricket; the performance of the players has been analyzed with the help of very basic statistical measures. During the past few years or more lot of work and research papers have been published which measured the performance of the players and their predictions. Most of them paid attention towards the whole matches only. The use of other resources are mentioned in Duckworth/Lewis approach (Duckworth and Lewis, 2002) and its references, (Johnston et al. 1993), (Beaudoin and Swartz, 2003) and (de Silva et al., 2001). Optimal batting orders are discussed in (Swartz et al., 2006) and in (Normanand Clarke, 2010), batting strategies using dynamic programming (Preston and Thomas, 2000) and (Johnston et al. 1993), what will be the effect of winning toss first (de Silva and Swartz, 1997). The methods of the predictions can be found in (Cohen, 2002), (Gilfillan and Nobandla, 2000) and (Swartz et al., 2009). The methods of graphical representation are presented in the study of (Kimber, 1993), (Barr et al., 2008), (Bracewell and Ruggiero, 2009) and (van Staden, 2009). The performance of the batting depends heavily on the average scores of the batting. Various researchers have chosen strike rate and the average scores as the measure of the performance for example (Croucher, 2000), (Barr and Kantor, 2004), 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. Within this two dimensional frame work, they developed the criteria for the selection of the batsman which combines together the average and strike rate. (Basevi and Binoy, 2007) and (Barr et al., 2008). (Barrand van den Honert, 1998) explained a measure which is based on the average and a consistency. (Lemmer, 2008a) showed in his study that the batting average can not be satisfactory in the case of a small amount of scores if

<sup>1</sup> Email: [muhammad.daniyal@iub.edu.pk](mailto:muhammad.daniyal@iub.edu.pk); [tahir.nawaz@iub.edu.pk](mailto:tahir.nawaz@iub.edu.pk); [iqramubeen@yahoo.com](mailto:iqramubeen@yahoo.com); [draleemiub@hotmail.com](mailto:draleemiub@hotmail.com)

the player had a large percentage of not out scores.

The Present study gives the focus on the application of the quality control charts which is different from simple descriptive measures used in previous studies. It will give us the deep insight towards the performance of the players. It will analyze each score of the batsman and will conclude it in control or out of control due to any non random variation.

## 2. Methodology

The main idea of this research paper is to see the application and construction of the control charts in evaluating the batting performance of two world class batsman. Hashim amla, the top order batsman of South African team, has been chosen for this purpose and the scores are taken from ESPN cricinfo website. He has been ranked first in the ICC player rankings of ODI batsman with the rating of 854. Sachin tendulkar has been chosen as a second batsman. He is an excellent and top order batsman of Indian cricket team. He has been ranked 20 in ICC top order ranking of ODI batsman with the rating 644. We are interested in comparing their performances by using quality control charts. The individual and moving average control charts will be utilized to analyze their performance. The main reason of choosing these control charts is that they are applied when the data consist of single observations at a time i.e. sample size is one. Another main reason of choosing these charts is that the assumption of normality is not required in calculating the sigma limits ([http://en.wikipedia.org/wiki/Shewhart\\_individuals\\_control\\_chart](http://en.wikipedia.org/wiki/Shewhart_individuals_control_chart)).

## 3. Objective

The main objective of this study is to compare the performances of two batsman hashim amla and sachin tendulkar by using the statistical control charts. The most suitable control charts which can be used in this case will be individual and moving average control chart. We will apply basic sensitizing rules on the data and will point out the case where there will be non-natural variation. We will compute control limits to analyze that whether the performance is under control or is there any point out of control because of any non random variation. In the end, we will achieve following objectives.

(1) Establish the trial control limits, and check if there is any point out of control. Check for the assignable cause for that out of control signal. If all points are in control, these trial limits will be used as tolerance limits for future analysis

(2) If any out of control point is observed, discard that point and establish the control structure again.

### 3.1. Descriptive Measures

Hashim amla is the top order batsman of the South African team. He made 2705 scores in ODIs with 2945 balls faced. His average scores are 56.35 and SD of the scores is 40.440. The coefficient of variation can be used to check the consistency of his performance. (Barrand van den Honert, 1998) defined a measure which is based on average and consistency measures. The formula to calculate the consistency is

$$\frac{S}{\bar{X}} \times 100 \quad (1)$$

The value of C.V is 71.76 %. On the other hand Sachin tendulkar is the top order batsman of Indian cricket team. He played 458 matches. He made total of 18201 scores and faced 21108 balls. The average scores are 44.83 and the SD is 40.164. The value for the coefficient of variation for Sachin scores is 89.59 %. We can conclude from the value of consistency that Hashim amla is more consistent than Sachin tendulkar as the Value of CV of former is lesser than the latter.

### 3.2. Checking Non-normality of the scores

As already discussed the individual and moving range control charts are good in the situation when the data is non normal. Different tools can be applied to check the non normality of the data. In statistics the Shapiro-wilk test tests the null hypothesis that  $x_1, \dots, x_n$  follows the normal distribution This test was published by Samuel Shapiro and Martin Wilk in 1965. The test statistic is:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

If P-value is less than significance level which is chosen 5% then we will reject the hypothesis that the scores of Hashim Amla follow normal distribution. The SPSS version 15 can be employed to calculate the value of Shapiro-Wilk. The following table is generated for the scores of Hashim Amla.

Table 1: Tests of Normality of the scores of hashim amla

Shapiro-Wilk Value		
Statistic	Degree of freedom	P-value
<b>.927</b>	<b>51</b>	<b>.004</b>

Since the p-value is 0.004 which is less than our significance level, we conclude that the scores of Hashim Amla do not follow the normal distribution.

Similarly the value of Shapiro-wilk can be calculated for the scores of Sachin Tendulkar and the following table is obtained

Table 2: Tests of Normality of the scores of Sachin

Shapiro-Wilk		
Statistic	Degree of freedom	p-value
<b>.871</b>	<b>444</b>	<b>.000</b>

The P-value is less than significance level and therefore can be concluded that the scores of Sachin are non-normal

The normal probability plot is a graphical technique which can be used for checking and assessing whether or not a data set is approximately normally distributed. The scores are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. If the data departs from the straight line, it indicates that the scores are departing from normal distribution. The normal probability plot is a special case of the probability plot, for the case of a normal distribution. The normal P-P plot can be used to check the normality of the scores. The following figures show the normal probability plots for the scores of hashim Amla and Sachin tendulkar. It can be clearly observed that both the scores are showing the non-normal pattern.

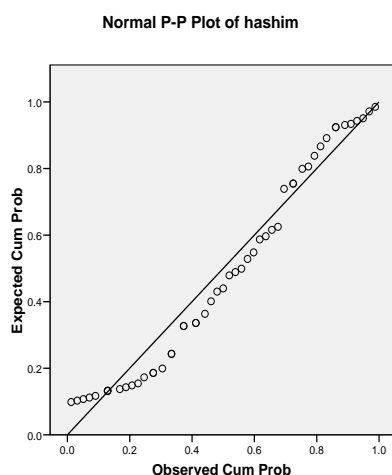


Figure 1: P-P plot for the scores of hashim

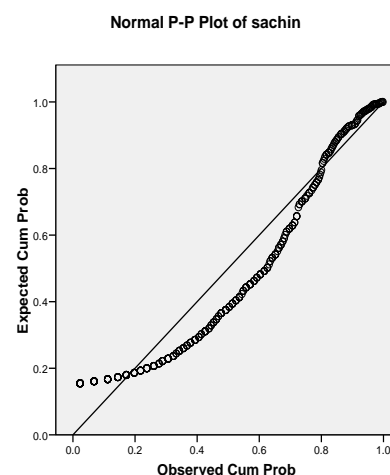


Figure 2: P-P plot for the scores of Sachin

The histogram is a useful and the most commonly device used to explore the shape of the distribution/ scores. If the shape of the distribution is bell shaped, it is the clear indication that the scores/data are following the normal distribution. The following figures are showing the histogram for the scores of the

Hashim and Sachin which are indicating that the scores are not following normal distribution.

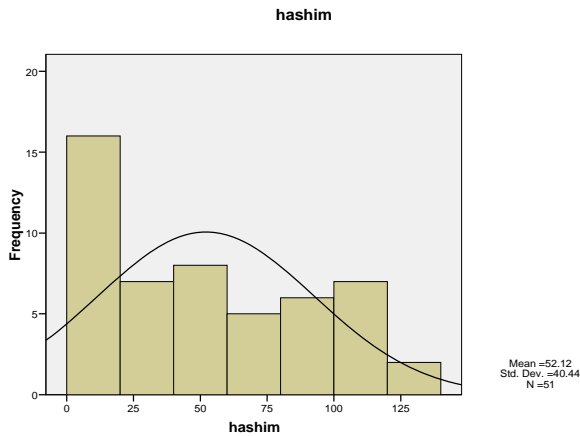


Figure 3: histogram of Hashim scores

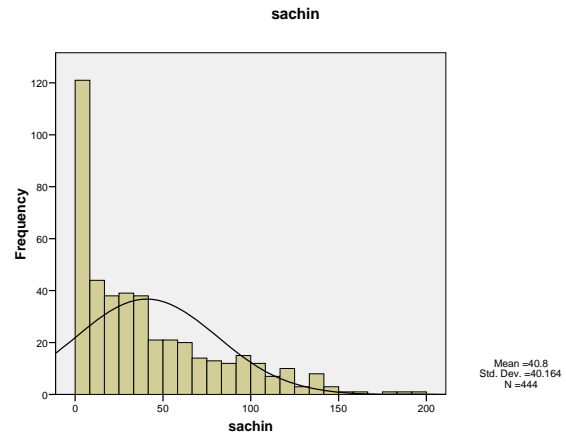


Figure 4: histogram of Sachin scores

### 3.3. Statistical process control chart

The individual and moving range control charts were first proposed by the William shewhart in 1920 and these charts are being used as the graphical way of monitoring the performance of the manufacturing and in services (Montgomery, 2005). The main purpose of the control charts is to analyze the performance and check for the assignable causes which are not random or can be regarded as un-natural variations.

For the purpose of checking assignable causes and variations in the batting performance of the batsman, individual (I) and moving average (MR) control charts are utilized because these control charts are applied when the data is in the form of single observation in each time period. These types of charts are utilized when each unit of measurement is inspected. The first step in the construction of control charts is to establish the trial control limits. Moving range control chart is utilized to control both the location and spread parameter. Once these parameters are controlled, the process is said to be in control. Moving range control limits are constructed with the help of two observations and the statistic can be computed by the following

$$MR = |x_i - x_{i-1}| \quad (3)$$

The first value in the moving range can not be calculated because there is no previous record available of the scores which the player has not played. Therefore the value for the first period will be left blank and the MR will be calculated for the second period and so on.

For the I chart, the upper, lower control limits and the central line are calculated by the following relations

$$UCL = \bar{X} + 3 \frac{\overline{MR}}{d_2} \quad (4)$$

$$CL = \bar{X} \quad (5)$$

$$LCL = \bar{X} - 3 \frac{\overline{MR}}{d_2} \quad (6)$$

Where  $\bar{X}$  denotes the mean of the observations and  $\overline{MR}$  denotes the mean of the moving ranges. The term  $d_2$  denotes the sigma conversation factor for the specific sample size which is taken 2 because we are using two terms  $x_i$  and  $x_{i-1}$  to construct the 3 sigma limits. So the value of  $d_2=1.128$  (Hines et al., 2004).

The control limits for the moving range charts can be calculated by the following

$$UCL = D_4 \overline{MR} \quad (7)$$

$$CL = \overline{MR} \quad (8)$$

$$LCL = D_3 \overline{MR} \quad (9)$$

Where  $D_3$  and  $D_4$  are regarded as the parameters of the control limits and depend upon the sample size which is already discussed that it will be taken two. The values of  $D_3$  and  $D_4$  are chosen 0 and 3.267 respectively (Hines et al., 2004).

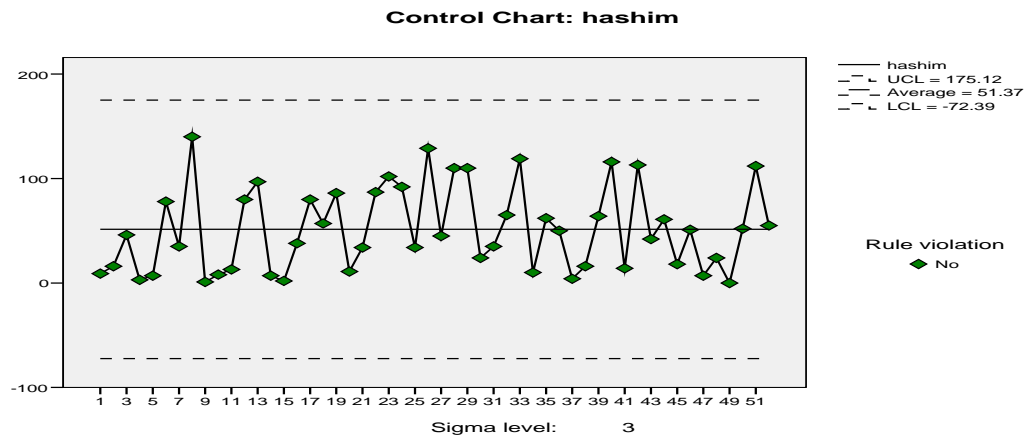


Figure 5: individual control chart for the scores of hashim amla

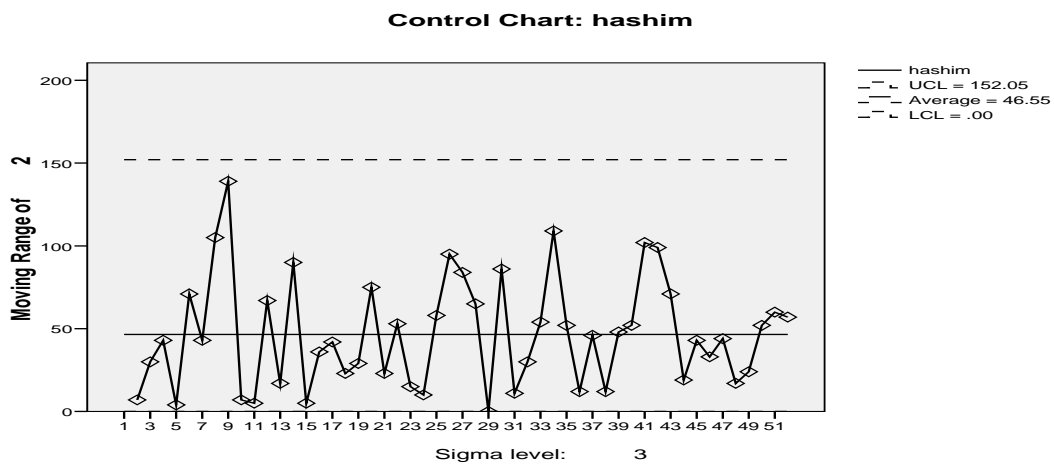


Figure 6: Moving range control chart for scores of hashim

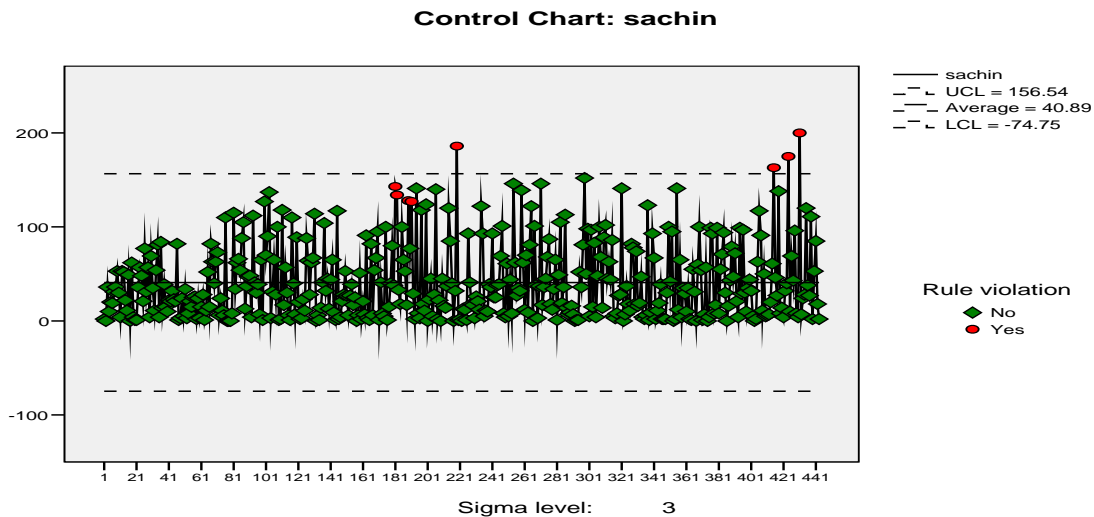


Figure 7: individual control chart for scores of sachin

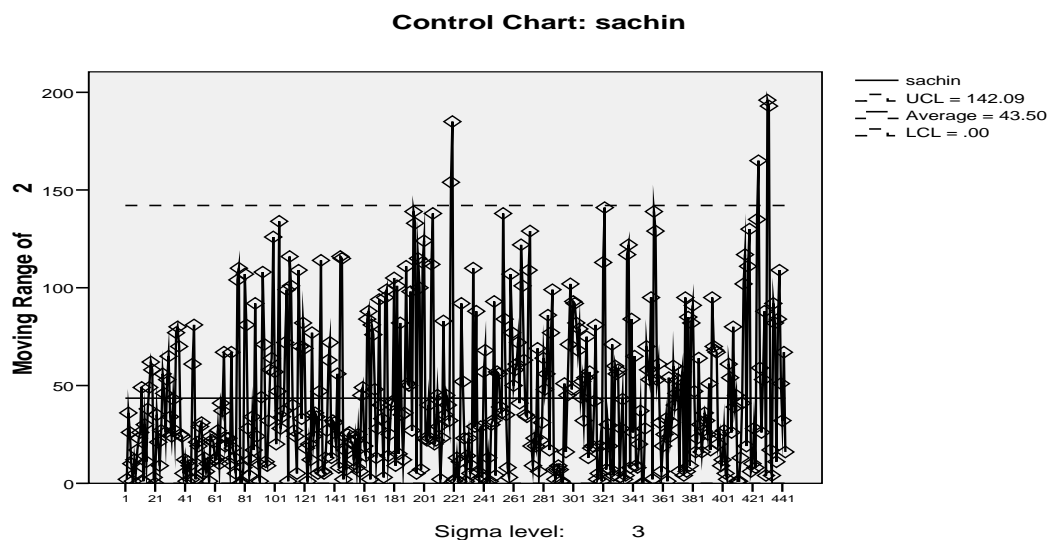


Figure 8: Moving range control chart for scores of sachin

#### 4. Analysis and Conclusions

Having constructed the quality control structure of I and MR chart, we should move towards the analysis of each and every case. We have used two basic sensitizing rules for checking any non random behavior in the performance of players. These rules can be found in Western Electric Company (1956) and Nelson (1984; 1985).

(1) The control chart will be out of control if any point lies above or below the 3 sigma limits.

(2) The control chart will be out of control if 2 points out of 3 above or below the 2 sigma limits.

(3) By considering the above sensitizing rules, the performance of Hashim Amla is totally under control in both charts which can be clearly seen in figure 5 and 6. So we can conclude that there is not any non random or un-natural variation present in the performance of this batsman. These are the trial control limits. Since the structure of the control chart is under control, these trial control limits can be used as tolerance limits which can be used for future analysis.

(4) If we look at the Figure 7 and 8, we will come to the point the control structure of performance of

Sachin tendulkar is under the influence of non-random and un-natural variation. With the help of SPSS version 15, the following table is obtained of the cases which are out of control.

Table 3: Table of sensitizing rules applied on Sachin scores

Case Number	Violations for Points
181	2 points out of the last 3 above +2 sigma
182	2 points out of the last 3 above +2 sigma
189	2 points out of the last 3 above +2 sigma
191	2 points out of the last 3 above +2 sigma
219	Greater than +3 sigma
415	Greater than +3 sigma
424	Greater than +3 sigma
431	Greater than +3 sigma

**8 points violate control rules.**

From the Table 3 it is clear that there are total 8 points which are violating the basic sensitizing rules. So we can make the statement that there is some non random variation present in the performance of the tendulkar which have made the charts out of control. These are trial control limits. If we discard these points and again set the control structure it is possible that charts are under control. These limits will be then tolerance limits and can be used for future analysis.

## 5. References

- [1] 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*. 2004, **55**: 1266-1274.
- [2] Barr, G.D.I. and van den Honert, R. Evaluating batsmen's scores in test cricket. *South African Statistical Journal*. 1998, **32**: 169-183.
- [3] Barr, G.D.I., Holdsworth, G.C. and Kantor, B.S. Evaluating performances at the 2007 cricket world cup. *South African Statistical Journal*. 2008, **42**: 125-142
- [4] Basevi, T. and Binoy, G The world's best Twenty20 players. Available from URL <http://content-rsa.cricinfo.com/columns/content/story/311962.html>. 2007.
- [5] Beaudoin, D. and Swartz, T.B. The best batsmen and bowlers in one day cricket. *South African Statistical Journal*. 2003, **37**: 203-222.
- [6] Bracewell, P.J. and Ruggiero, K. A parametric control chart for monitoring individual batting performances in cricket. *Journal of Quantitative Analysis in Sports*. 2009, **5**: 1-19.
- [7] Cohen, G.L. Cricketing chances. In: *Proceedings of the Sixth Australian Conference on Mathematics and Computers in Sport*. Eds: Cohen, G. and Langtry, T. Sydney University of Technology, NSW. 2002, pp. 1-13.
- [8] 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, pp. 95-106.
- [9] De Silva, M.B. and Swartz, T.B. *Winning the coin toss and the home team advantage in one-day international cricket matches*. The New Zealand Statistician. 1997, **32**: 16-22.
- [10] Duckworth, F. and Lewis, T. Review of the application of the Duckworth/Lewis method of target resetting in one-day cricket. In: *Proceedings of the Sixth Australian Conference on Mathematics and Computers in Sport*. Eds: Cohen, G. and Langtry, T. Sydney University of Technology, NSW. 2002, **1**: 27-140.
- [11] Gilfillan, C. and Nobandla, N. Modelling the performance of the South African national cricket team. *South African Journal for Research in Sport, Physical Education and Recreation*. 2000, **22**: 97-110.
- [12] Hines, W., Montgomery, D., Goldsman, D., and Borror, C. *Probability and Statistics in Engineering* (4th Edition). Wiley & Sons, Inc., Hoboken, NJ. 2004.
- [13] Johnston, M.I., Clarke, S.R. and Noble, D.H. Assessing player performance in one-day cricket using dynamic programming. *Asia-Pacific Journal of Operational Research*. 1993, **10**: 45-55.
- [14] Kimber, A.C. A graphical display for comparing bowlers in cricket. *Teaching Statistics*. 1993, **15**: 84-86.
- [15] Montgomery, D. *Introduction to Statistical Quality Control* (5th Edition). Wiley & Sons, Inc., Hoboken, NJ. 2005.
- [16] Nelson, Lloyd S. The Shewhart Control Chart—Tests for Special Causes. *Journal of Quality Technology*. 1984,

- 16**(4): 237-239.
- [17] Nelson, Lloyd S. Interpreting Shewhart Control Charts. *Journal of Quality Technology*. 1985, **17**(2): 114-116.
- [18] Norman, J.M. and Clarke, S.R. Optimal batting orders in cricket. *Journal of the Operational Research Society*. 2010, **61**: 980-986.
- [19] Preston, I. and Thomas, J. Batting strategy in limited overs cricket. *The Statistician*. 2000, **49**: 95-106.
- [20] Swartz, T.B., Gill, P.S. and Muthukumarana, S. Modelling and simulation of one-day cricket. *Canadian Journal of Statistics*. 2009, **37**: 143-160
- [21] Swartz, T.B., Gill, P.S., Beaudoin, D. and de Silva, B.M. Optimal batting orders in one-day cricket. *Computers and Operations Research*. 2006, **33**: 1939-1950.
- [22] Van Staden, P.J. Comparison of cricketers' bowling and batting performance using graphical displays. *Current Science*. 2009, **96**: 764-766.
- [23] Western Electric Company. Statistical Quality Control Handbook. (1st ed.), Indianapolis, IN: Western Electric Co. 1956.
- [24] Lemmer, H.H. Measures of batting performance in a short series of cricket matches. *South African Statistical Journal*. 2008, **42**: 83-105.
- [25] [http://en.wikipedia.org/wiki/Shewhart\\_individuals\\_control\\_chart](http://en.wikipedia.org/wiki/Shewhart_individuals_control_chart)