Middle-East Journal of Scientific Research 19 (3): 373-377, 2014
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2014.19.3.13609

Analysis of Cricket Scores Using Statistical Control Charts

Muhammad Daniyal, Tahir Nawaz and Muhammad Rajab

Department of Statistics,
The Islamia University of Bahawalpur, Pakistan

Abstract: Today's in cricket, the performance of the teams is being analyzed by using very simple statistical tools and graphics. Most of the times, the average scores, strike rates, coefficient of variation and graphical measures are being utilized to measure the performance and to make comparisons between different teams. The main focus of this research paper is to provide the application of statistical quality control charts in analyzing the performance of the team by using its batting scores. For this purpose, we analyzed the scores of Pakistan cricket team. We used individual and moving range control charts because it may be applied in normal conditions otherwise false special cause signals can be generated. We then used the basic sensitizing rules to check whether the un-natural or non random variations are present in the batting scores or not. We will provide the specialists of this game an opportunity to check for the reasons for the un-natural and non random behavior in the batting scores so that in future the performance can be improved.

Key words: Scores · Performance · Moving range control charts · Individual control charts · Graphical displays · Assignable cause and normality

INTRODUCTION

Cricket has become one of the world class famous and popular games. It is the game of bat and ball which includes eleven players. Most of analysis is being conducted to measure the performance of the teams and individual players from all perspectives. Different sites are available which include the complete information and time to time record of every event of the cricket. The experts of this game give their opinions in the light of this information. As far as the ODI career of Pakistan team is concerned, it started its one day career in 1972/1973. It is ranked at 6th position according to ICC criteria in one day with 2620 points and 105 rating [1].

The performance of the players and teams in cricket has been analyzed with the help of very simple statistical tools. During the few years back, a huge study and work has been conducted to measure the performance of the teams and the prediction of their performance in the next series. The utilization of other sources are given in Duckworth/Lewis approach [2] its references, [3-5]. The batting orders with optimality are mentioned in [6] and in (Normanand Clarke, 2010), different batting strategies using the dynamic programming [7-3], what kind of impact will have winning the toss first are discussed in the study of (de Silva and Swartz, 1997). The methods of prediction and forecasting can be found in the study of [8-10]. Graphical representation of the results are mentioned in the research study of [11-14]. The performance of batting is keenly measured by the method of batting average scores and various scholars and researchers have used strike rate and the average score of batsman as a measure of performance e.g [15, 16]. Bar and Kantor used a new idea of graphical representation of the results with the strike rate on one axis and the probability of getting out on the other axis. In this two dimensional structure, they created the criteria for the selection of the batsman which combines together the average and strike rate [17-12]. [18] gave a measure which is based on both averages and consistencies. [19] concluded in his research study that the batting average can not be reasonable or satisfactory in the case of a small amount of scores if the player had a large percentage of not out scores.

The present research gives the application of statistical quality control charts as compared to the simple statistical measures for measuring the batting performance of Pakistan team. It will help us to see the non random pattern and un-natural variation in the scores of Pakistan team which will help the experts to make some decisions.

Corresponding Author: Muhammad Daniyal, Department of Statistics,
The Islamia University of Bahawalpur, Pakistan.

These decisions will help us in future to improve the batting performance of Pakistan team and avoid un-natural situations due to assignable causes.

MATERIALS AND METHODS

The main theme of this research study is to see the application of statistical control charts in evaluating the batting performance of Pakistan cricket team by using its ODI scores from 11 Feb 1973 to 22 Mar 2012. Pakistan is ranked at 6th position according to ICC criteria in one day with 2620 points and 105 rating. We are interested in evaluating the performance of Pakistan cricket team by applying quality control charts on its batting scores. The individual and moving average control charts were used for this purpose. The individual and moving range control charts can be utilized when the data follows the normal distribution otherwise the false special cause signals will be generated.

(http://www.smartersolutions.com/blog/forrestbreyfogle/2009/12/08/non-normal-data-and-the-creation-of-their-control-charts-and-predictive-statements/).

Objective: The basic purpose of this research study is to evaluate the performance of Pakistan cricket team using the statistical control charts. The appropriate control charts which will be used for this analysis is the individual and moving range control charts. The sensitizing rules will be used to check which score is indicating an out of control signal or un-natural variation. In the end, we will achieve following objectives.

- We will develop the trial control limits and analyze if there is any out of control point. We will look for the assignable cause for that out of control signal. If all the scores are found in control, the trial limits will be utilized as tolerance limits for the future analysis.
- If any out of control signal is observed, we will remove that score and analyze the scores again using control structure.

Checking Normality of the Scores: As it has already been mentioned that individual and moving range control charts can be utilized when there is normality of data otherwise false special cause signals will be generated. Different techniques exist to check the normality of the scores. In statistics the Shapiro-Wilk (1965) test checks the null hypothesis that $x_1 \ldots x_n$ follows the normal distribution. The test statistic used is:

Table 1: Tests of Normality of the scores of Pakistan Team

	Shapiro-Wilk value	
Statistic	Degree of Freedom	P-value
0.996	409	0.303
0.997	330	0.883

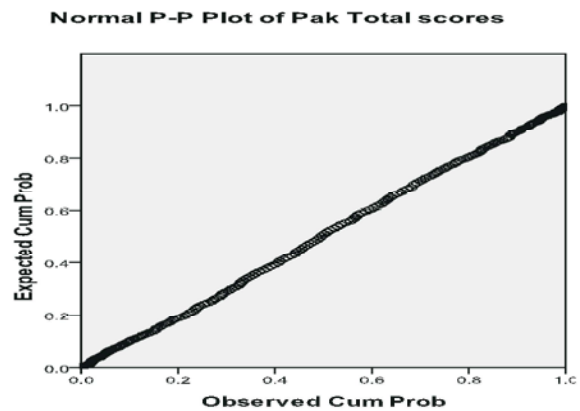

Fig. 1: P-P plot for the scores of Pakistan Cricket Team

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

If P-value is less than significance level which is normally taken 5%, then we will reject the hypothesis that the scores follow normal distribution. The SPSS version 17 was utilized for this purpose and the following table was obtained.

Since the p value is greater than the significance level 0.05, therefore we can conclude that the scores follow the normal distribution. The graphical technique which is being utilized to measure the normality of the scores is the normal probability plot. The scores are plotted against a theoretical normal distribution in such a manner that the points plotted show the straight line approximately. If the points deviate from the straight line, this is the indication that the scores are non-normal. The normal probability plot is a special case of the probability plot in the case of a normal distribution. The normal P-P plot can be utilized to check the normality of the scores. The following figure is showing the P-P plot of the batting scores. It can be clearly observed that the scores are following the normal distribution.

The histogram is the most common and the most useful graphical measure used to explore the shape of the distribution of the scores. If distribution is bell shaped then it clearly means that they follow the normal

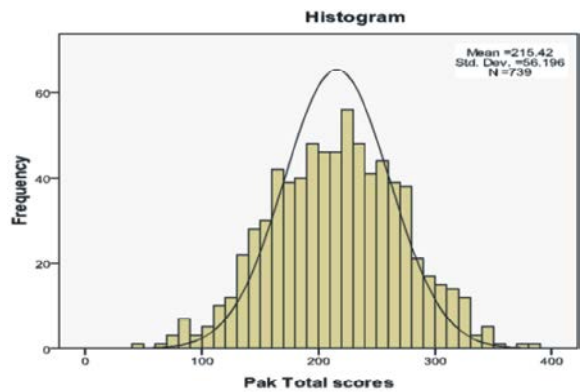

Fig. 2: Histogram of Pakistan cricket team scores

distribution. The following figure is showing the histogram for the scores of Pakistan cricket team made against different teams in its whole ODI career.

Statistical Process Control Chart: The individual and moving range control charts were first introduced by the William Shewhart in 1920. These control charts are utilized as the graphical measure to evaluate the performance of the manufacturing units and in services [20]. The main use of these charts is to evaluate the performance and look for any non random and un-natural variations present due to any assignable cause.

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 analyzed. The initial step in the creation of control charts is to set up the trial control limits. Moving range control chart is used to control the spread and location parameter. The process will be considered under control once these parameters are controlled. Moving range control limits are made with the help of two observations and its statistic can be computed as:

$$MR = |x_i - x_{i-1}| \tag{2}$$

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

For the individual control chart, the upper, lower control limits and the central line are calculated by the following statistic

$$UCL = \bar{X} + 3\frac{\overline{MR}}{d_2} \tag{3}$$

$$CL = \bar{X} \tag{4}$$

$$LCL = \bar{X} - 3\frac{\overline{MR}}{d_2} \tag{5}$$

Where $\bar{X}$ denotes the mean of the observations and $\overline{MR}$ denotes the mean of the moving ranges. The term d_2 shows the sigma conversation factor for the specific sample size which is taken 2 because we are utilizing two terms x_i and x_{i-1} to construct the 3 sigma limits. So the value of d_2=1.128 [20].

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

$$UCL = D_4\overline{MR} \tag{6}$$

$$CL = \overline{MR} \tag{7}$$

$$LCL = D_3\overline{MR} \tag{8}$$

D_3 and D_4 are considered as the parameters of the control limits. They depend upon the sample size which has been mentioned that it will be taken two. The values of D_3 and D_4 are chosen 0 and 3.267 respectively [20]. The charts were constructed using MINITAB 15.

CONCLUSIONS

The next step after the construction of the I and MR charts is to analyze each score or case. For this purpose, we selected two basic sensitizing rules to check which score is out of control and which is showing non random pattern. These rules can be found in [21-22].

- The control chart will be out of control if any point lies above or below the 3 sigma limits.
- The control chart will be out of control if 2 points out of 3 above or below the 2 sigma limits.

By considering these rules, we can see that the individual control chart of the scores in figure 3 is showing seven out of control signals. On the other hand if we look at the figure 4, the moving range control chart is showing nine points out of control. These out of control points can be generated due to any assignable cause. We leave this to the specialists and the experts of

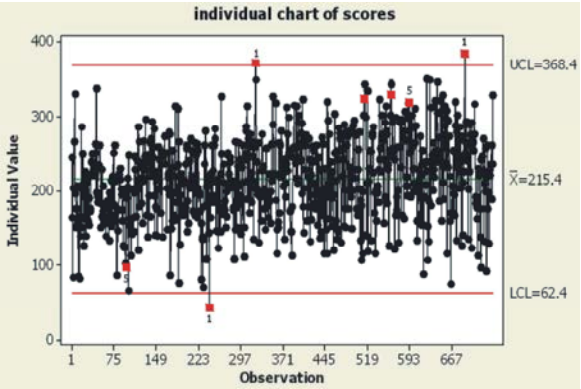

Fig. 3: Individual control chart for the batting scores of Pakistan Team

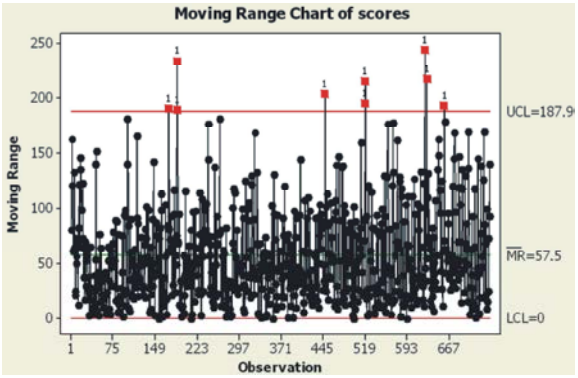

Fig. 4: Moving Range control chart for the batting scores of Pakistan Team

Table 2: Table of sensitizing rules applied on batting scores

Rule Violations	
Case Number	Violations for Points
96	2 points out of the last 3 below -2 sigma
242	Less than -3 sigma
324	Greater than +3 sigma
324	2 points out of the last 3 above +2 sigma
515	2 points out of the last 3 above +2 sigma
561	2 points out of the last 3 above +2 sigma
592	2 points out of the last 3 above +2 sigma
690	Greater than +3 sigma
	7 points violate control rules.

the cricket to look for these assignable causes which have made the scores out of control. The sensitizing rules applied were formulated in the table using SPSS 17 as follow.

The above charts indicate that the process which is being monitored is not in control. If we go into the deep analysis of the chart, we will find different the sources of variation. This deep analysis is the work of the specialists and the experts of the game who deeply analyze the game and its techniques. They would be in the better position to look for the assignable causes which are making these control structures out of control. They would look for the non random and un-natural pattern which is showing the out of control signal and disturbing the batting performance of the team. The limits which are constructed are the trial control limits. If we remove the out of control points again calculate the control structure, it may be possible that charts are then under control. These limits will be then tolerance limits and can be used for future analysis.

REFERENCES

1. http://www.smartersolutions.com/blog/forrestbreyfogle/2009/12/08/non-normal-data-and-the-creation-of-their-control-charts-and-predictive-statements/).
2. Duckworth, F. and T. Lewis, 2002. 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, pp: 127-140.
3. Johnston, M.I., S.R. Clarke and D.H. Noble, 1993. Assessing player performance in one-day cricket using dynamic programming. Asia-Pacific Journal of Operational Research, 10: 45-55.
4. Beaudoin, D. and T.B. Swartz, 2003. The best batsmen and bowlers in one day cricket. South African Statistical Journal, 37: 203-222.
5. De Silva, M.B. and T.B. Swartz, 1997. Winning the coin toss and the home team advantage in one-day international cricket matches. The New Zealand Statistician, 32: 16-22.
6. Swartz, T.B., P.S. Gill, D. Beaudoin and B.M. de Silva, 2006. Optimal batting orders in one-day cricket. Computers and Operations Research, 33: 1939-1950.
7. Preston, I. and J. Thomas, 2000. Batting strategy in limited overs cricket. The Statistician, 49: 95-106.
8. Cohen, G.L., 2002. 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, pp: 1-13.
9. Gilfillan, C. and N. Nobandla, 2000. Modelling the performance of the South African national cricket team. South African Journal for Research in Sport, Physical Education and Recreation, 22: 97-110.

10. Swartz, T.B., P.S. Gill and S. Muthukumarana, 2009. Modelling and simulation of one-day cricket. Canadian Journal of Statistics, 37: 143-160.
11. Kimber, A.C., 1993. A graphical display for comparing bowlers in cricket. Teaching Statistics, 15: 84-86.
12. Barr, G.D.I., G.C. Holdsworth and B.S. Kantor, 2008. Evaluating performances at the 2007 cricket world cup. South African Statistical Journal, 42: 125-142.
13. Bracewell, P.J. and K. Ruggiero, 2009. A parametric control chart for monitoring individual batting performances in cricket. Journal of Quantitative Analysis in Sports, 5: 1-19.
14. Van Staden, P.J., 2009. Comparison of cricketers' bowling and batting performance using graphical displays. Current Science, 96: 764-766.
15. Croucher, J.S., 2000. 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, pp: 95-106.
16. Barr, G.D.I. and B.S. Kantor, 2004. A criterion for comparing and selecting batsmen in limited overs cricket. Journal of the Operational Research Society, 55: 1266-1274.
17. Basevi, T. and G. Binoy, 2007. The world's best Twenty20 players. Available from URL http://content-rsa.cricinfo.com/columns/content/story/311962.html.
18. Barr, G.D.I. and R. van den Honert, 1998. Evaluating batsmen's scores in test cricket. South African Statistical Journal, 32: 169-183.
19. Lemmer, H.H., 2008a. Measures of batting performance in a short series of cricket matches. South African Statistical Journal, 42: 83-105.
20. Montgomery, D., 2005. Introduction to Statistical Quality Control, 5th Edition, Wiley & Sons, Inc., Hoboken, NJ.
21. Western Electric Company, 1956. Statistical Quality Control Handbook. (1st ed.), Indianapolis, IN: Western Electric Co.
22. Norman, J.M. and S.R. Clarke, 2010. Optimal batting orders in cricket. Journal of the Operational Research Society, 61: 980-986.
23. Nelson and S. Lloyd, 1984. The Shewhart Control Chart-Tests for Special Causes, Journal of Quality Technology, 16(4): 237-239.
24. Nelson and S. Lloyd, 1985. Interpreting Shewhart Control Charts, Journal of Quality Technology, 17(2): 114-116.
24. Hines, W., D. Montgomery, D. Goldsman and C. Borror, 2004. Probability and Statistics in Engineering, 4th Edition, Wiley & Sons, Inc., Hoboken, NJ.
22. Nelson and S. Lloyd, 1984. The Shewhart Control Chart-Tests for Special Causes, Journal of Quality Technology, 16(4): 237-239.
23. Nelson and S. Lloyd, 1985. Interpreting Shewhart Control Charts, Journal of Quality Technology, 17(2): 114-116.