International Journal of Mathematical Education.
ISSN0973-6948 Volume 8, Number 2 (2018), pp. 87-93

http://www.ripublication.com

Cluster Analysis Applied to the Greatest Barbadian Cricketers

Arjun Tan* and Marius Schamschula

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

1. INTRODUCTION

Barbados is a small island having an area of 167 square miles and a population of around a quarter of a million. It appears as a speck on the map. But it can certainly boast the highest per capita production of the ***greatest cricketers*** in the world. In the mid-1960s, Barbadian players formed the majority of the unofficial world champion West Indies team; and, in 1967, the Barbados cricket team played a match against the rest of the world!

Cricket chronicles show that in the space of 34 years between 1924 and 1958 were born a dozen of Barbadians who went on to become the world's greatest cricketers in all positions of the game (Table I). Eight of them were inducted into the ***ICC Hall of Fame***; and six of them received ***knighthood*** – an outstanding achievement for any place in the world. In 2003, an ***all-time Barbados cricket team*** of 12 players was selected by ***Sir Everton Weekes***, which consisted of the dozen players in Table I [1]. And, in 2016, the ***ESPN cricket staff*** selected its own all-time Barbados cricket, which had 11 players in common with those in Table I [2]. In this paper, we perform a ***cluster analysis*** of the 12 players in Table I to justify Everton Weekes' selection of the all-time Barbados cricket team.

2. THE GREATEST CRICKETERS OF BARBADOS

The 12 cricketers in Table I were born within three distinct periods of post-war era. First, in successive years beginning 1924 were born ***Worrell***, ***Weekes*** and ***Walcott*** – ***The Three Terrible Ws*** as they would become known (Table I). They would soon form the greatest middle-order batting in cricket history, at least up to that period and well beyond [3 – 5]. For a decade, the world's greatest batsman would be one of them – Worrell in the late 1940s; Weekes in the early 1950s; and Walcott in the mid-1950s [6]. Worrell was also a tidy medium fast / slow bowler and a great all-rounder amongst the 12; and Walcott was also a wicket-keeper / batsman in the early part of his career. The

Three Ws were the '***Holy Trinity of West Indian Cricket***' [5]. Figure 1 (from [5]) captures them together.

Table I. Greatest Barbadian Cricketers in Chronological Order

Year Born	Cricketer	Position	ICC Hall of Famer	Knight-hood	All-time Barbados Team	
					Weekes	ESPN
1924	Frank Worrell	Middle-order Batsman/ All-rounder/Captain	√	√	√	√
1925	Everton Weekes	Middle order Batsman	√	√	√	√
1926	Clyde Walcott	Middle order Batsman/ Wicket-keeper	√	√	√	√
1932	Conrad Hunte	Opening Batsman	–	√	√	√
1933	Seymour Nurse	Middle order Batsman	–	–	√	–
1936	Garfield Sobers	Middle-order Batsman/ All-rounder/Captain	√	√	√	√
1937	Wesley Hall	Fast Bowler	√	√	√	√
1938	Charlie Griffith	Fast Bowler	–	–	√	√
1951	Gordon Greenidge	Opening Batsman	√	–	√	√
1952	Joel Garner	Fast Bowler	√	–	√	√
1956	Desmond Haynes	Opening Batsman	–	–	√	√
1958	Malcolm Marshall	Fast Bowler	√	–	√	√

Fig. 1. The Three Ws of West Indian Cricket

Following the Three Ws, five more great cricketers were born within a space of six years between 1932 and 1938: ***Hunte***, ***Nurse***, ***Sobers***, ***Hall*** and ***Griffith*** (Table I).

Hunte was an opening batsman; Nurse and Sobers were middle-order batsmen; and Hall and Griffith formed a devastating fast bowling pair. Sobers was an all-rounder in the mould of Worrell, in fact surpassing the latter to become the greatest all-rounder in the history of cricket and selected to be one of five ***Wisden's greatest cricketers of the century*** [7]. However, Worrell reigned supreme as the greatest captain the West Indies had ever produced.

The third batch of great Barbadian cricketers were born between 1951 and 1958 (Table I). Amongst them, ***Greenidge*** and ***Haynes*** became the best West Indian opening pair for a long period; whereas ***Garner*** and ***Marshall*** became two of the finest fast bowlers in the post-war period having the best Test averages.

3. CLUSTER ANALYSIS

Cluster analysis consists of the classification of a group of objects according to certain similarities and criteria [8 − 10]. Cluster analysis has been applied to all fields of empirical science and has led to important discoveries. For example, classification of animals by Darwin resulted in his theory of evolution. Similarly, classification of elements by Mendeleyev gave rise to the periodic table discovery of missing elements. In astrophysics, the classification of stars by the Russell-Hertzsprung diagram led to the theory of stellar evolution. More recently, cluster analysis has been applied to determine the planetary status of Pluto [11].

The basic scheme of cluster analysis consists of the following steps.

Step 1. Suppose there are n objects which are to be grouped according to m properties, which can be ***quantitative*** or ***qualitative*** in nature. The quantitative values for all objects and all properties are assigned and an $n \times m$ data matrix is constructed.

Step 2. The ***similarity*** or ***dissimilarity*** between any two objects is determined by some pre-determined schemes. Numerous different schemes for this purpose have been devised [12]. For instance, the similarity between two objects can be determined by any of the ***resemblance coefficients*** (r) found in the literature. Likewise, the dissimilarity between two objects can be determined by any of the ***distance coefficients*** (d) which have been used [12]. The values of r and d vary between 0 and 1 (100%) subject to the condition that:

$$r + d = 1 \tag{1}$$

One of the most common distance coefficient between two objects denoted by the variables x and y is the ***Euclidean distance*** in m dimensions. Likewise, the most common resemblance coefficient between two objects is the ***Pearson's correlation***

coefficient. For variables having non-negative values, a particularly simple resemblance coefficient has been used [11]:

$$r = \frac{1}{m}\sum_{i=1}^{m}\frac{\min(x_i, y_i)}{\max(x_i, y_i)} \tag{2}$$

The complement of r from 1 (100%) is the distance coefficient known as Wave-Hedges coefficient [12]:

$$d = 1 - \frac{1}{m}\sum_{i=1}^{m}\frac{\min(x_i, y_i)}{\max(x_i, y_i)} \tag{3}$$

Step 3. The resemblance and distance coefficients are calculated between every two objects in the group are calculated and their average values determined. Amongst the averaging schemes, the ***arithmetic mean*** (*AM*) and the ***geometric mean*** (*GM*) are the most common:

$$AM = \frac{x_i + y_i}{2} \tag{4}$$

and

$$GM = \sqrt{x_i y_i} \tag{5}$$

For coefficients having a wide range of values, the geometric mean is preferable [11]. However, for coefficients having comparable values, the arithmetic mean suffices.

The two objects having the greatest value of r or the least values of d are deemed the ***most similar***. They are then joined to form a single new object or ***cluster*** in the group, whereby themselves disappearing in the process. The number of objects in the group is then reduced by one.

Step 4. Steps 2 and 3 are repeated and each time the number of objects is reduced by one. The process is continued until all the objects have been clustered into one. A tree diagram called a ***dendrogram*** is constructed to illustrate the entire clustering process.

4. CLUSTER ANALYSIS OF THE GREAT BARBADIAN CRICKETERS

To begin, we start with two qualitative attributes of the cricketers: (1) the position each cricketer played; and (2) whether the cricketers played during the same period. These two attributes demand that (1) we cluster players of the same playing position; and (2)

amongst the players of the same position, we cluster contemporary players first. The easiest way is to arrange the players in their batting order, viz., the openers; middle-order batsmen (including wicket-keepers); all-rounders; bowlers; and finally 12th man. Table II reflects this order. Next, we record the quantitative attributes for each player. In this study, we choose the ***first class and Test batting averages*** of each batsman; and the ***first class and Test bowling averages*** of each bowler. This means that for all-rounders, there are four quantitative attributes to contend with. The data are readily available in the literature [13].

Table II. Batting & Bowling Statistics of Barbadian Cricketers

Cricketer	Batting Statistics		Bowling Statistics	
	First Class Average	Test Average	First Class Average	Test Average
Greenidge	45.88	44.72		
Haynes	45.90	42.29		
Hunte	43.92	45.06		
Nurse	43.93	47.60		
Weekes	55.34	58.61		
Walcott	56.55	56.68		
Worrell	54.24	49.48	28.98	38.72
Sobers	54.87	57.78	27.74	34.03
Hall			26.14	26.38
Griffith			21.60	28.54
Garner			18.53	20.97
Marshall			19.10	20.94

Next, within the same group, the resemblance coefficients between each player are calculated using Eq. (2) and the average resemblance coefficient r determined using Eq. (4). The results are shown in the dendrogram of Fig. 2. The ***first round*** produces six clusters with the average resemblance coefficients as shown in the figure: (1) Greenidge-Haynes; (2) Hunte-Nurse; (3) Weekes-Walcott; (4) Worrell-Sobers; (5) Hall-Griffith; and (6) Garner-Marshall. Curiously, the first three clusters all had average resemblance coefficients of .973. The sixth cluster had the highest r value of .984 and the fifth cluster had the lowest r value of .843. In the ***second round***: (1) the first two clusters combine to form a mega-cluster of opening batsmen and a middle order batsman; (2) the third and fourth cluster join to form a second mega-cluster of middle-order batsmen and all-rounders (including a wicket-keeper); and (3) the last two clusters

join to form a mega-cluster of fast bowlers. In the ***third round***, the first two mega-clusters join to form a super-mega-cluster consisting of batsmen and all-rounders (including a wicket-keeper). In the ***final round***, the super-mega-cluster and the bowling mega-cluster join to form the All-time Barbados cricket team.

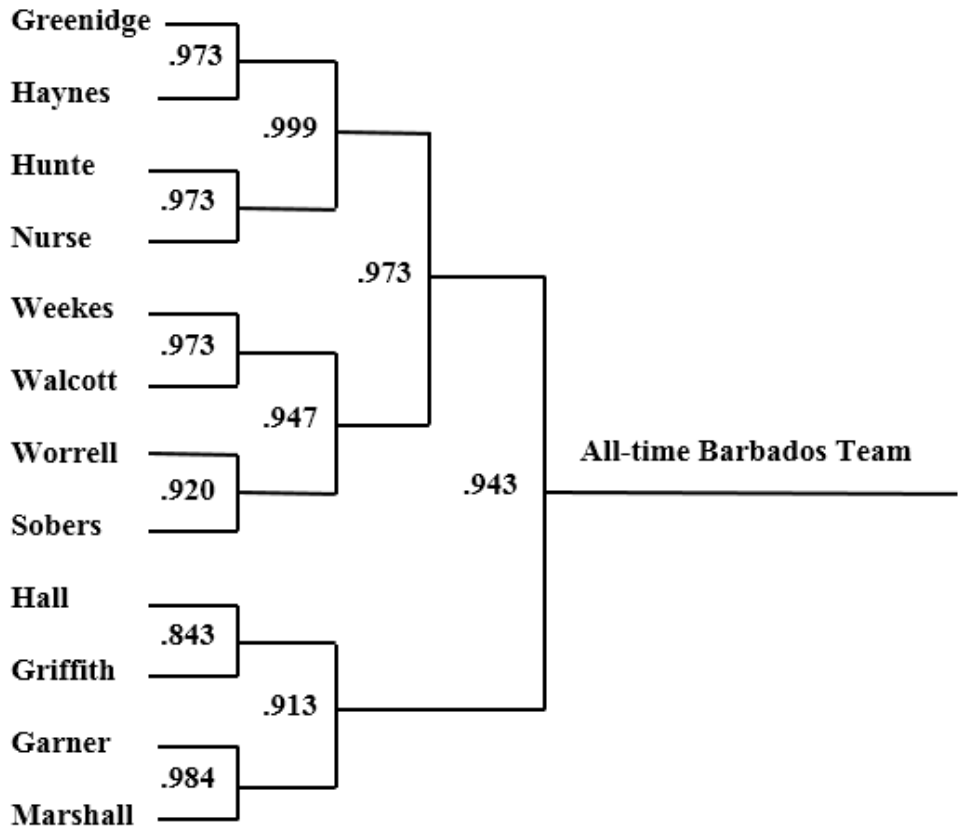

Fig. 2. Cluster Analysis of the Greatest Barbadian Cricketers

5. SELECTING THE ALL-TIME BARBADOS CRICKET TEAM

One can see that the all-time Barbados cricket team almost selects itself from the original 12 players. Cluster analysis only validates this selection. There is still unfinished business at hand. First, one player has to be demoted to be the 12th man. This is both an easy and a difficult task. As we started with three opening batsmen, the easiest thing is to do is to relegate one opener to the 12th man. However, the three openers had nearly identical batting averages. Since Greenidge and Haynes were contemporary players who had many outstanding stands between them, the only rational solution will be to relegate the ***Sir Conrad Hunte*** to that position [1]. Next, the ***captain***, ***vice-captain*** and ***wicket-keeper*** need to be selected. The captaincy, by universal choice, will go to ***Sir Frank Worrell*** – to many, the greatest captain in cricket history [5]. The vice-captaincy goes to ***Sir Everton Weekes***, a close friend of Worrell,

who correctly selected the all-time Barbados team [1]. And since there was only one wicket-keeper in the original 12, that position goes to ***Sir Clyde Walcott***. ***The Three Ws*** – close friends, comrades and called 'inseparable' [5] − remain intact as a unit.

REFERENCES

[1] https://www.tapatalk.com/groups/thevistortrumpercricketboard/all-time-barbados-xi-by-everton-weekes-t6621.html.

[2] www.espncricinfo.com/story/_/id/18170585/1n-all-barbados-xi.

[3] A. Tan, The Three Ws of West Indian Cricket: A Statistical Analysis, *Math. Spectrum,* **33**, 11-14 (2000-1).

[4] Keith A.P. Sandiford & Arjun Tan, *The Three Ws of West Indian Cricket: A Comparative Batting Analysis*, Authorhouse & Univers de Presse (2002).

[5] *Celebrating the 3 W's: A Legacy of West Indian Cricket*, Deborah P. Stoute, ed., The 3W's Celebration Committee, Barbados (2003).

[6] A. Tan, S. Carter, A. Alomari & M. Schamschula, The Three Ws of West Indian cricket – Some mathematical studies, *Intl. J. Math. Educ*., **7**, 19-31 (2017).

[7] http://espncricinfo.com/wisdenalmanac/content/story/153387.html.

[8] B.S. Everitt, *Cluster Analysis*, Heinemann, London (1974).

[9] H. Spaeth, *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*, Ellis Horwood, Chichester (1980).

[10] M.R. Andenberg, *Cluster Analysis for Applications*, Academic Press, New York (1973).

[11] A. Tan, R. Ramachandran & T.X. Zhang, Cluster Analysis and the Planetary Status of Pluto, *Math. Spectrum*, **39**, 102-106 (2006-7).

[12] T.F. Cox and M.A.A. Cox, Multidimensional Scaling, Chapman & Hall, New York (2000).

[13] http://www.stats.espncricinfo.com/guru