A New Handicapping System for Golf

Tim B. Swartz *

Abstract

The official handicapping system of the Royal Canadian Golf Association (RCGA) is very similar to the handicapping system of the United States Golf Association (USGA). Although these handicapping systems are complex and have been carefully studied, the systems do not take statistical theory into account. In 2000, the Handicap Research Committee of the RCGA was formed and challenged with the task of developing a new handicapping system. This paper outlines the proposed system. The proposed system continues to make use of the existing course ratings and slope ratings, but uses statistical theory to drive the methodology. In this paper, we demonstrate that the proposed system has several advantages over existing systems including fairness and improved interpretability. The proposed system is supported by both theory and data analyses. An investigation into the effects of equitable stroke control is also provided.

Keywords : data analysis, golf, handicapping, normal distribution, order statistics.

*Tim Swartz is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. The author has been partially supported by the Natural Sciences and Engineering Research Council of Canada. The author thanks an anonymous referee whose comments lead to an improvement in the manuscript.

1 Introduction

Since 1895, the Royal Canadian Golf Association (RCGA) has served as the governing body for golf in Canada. As part of its mandate, the RCGA oversees the maintenance of a handicapping system whose goal is to enable golfers of differing abilities to compete on an equitable basis (www.rcga.org). At present, the handicapping system of the RCGA is very similar to the handicapping system of the United States Golf Association (USGA).

In 2000, the Handicap Research Committee of the RCGA was formed and challenged with the task of developing a new handicapping system for golf in Canada. Since 2000, the author has served as the technical contributor on the Committee. There were three main goals that the Committee wished to achieve:

1. As golfers are generally resistent to dramatic changes, the Committee sought a system that did not differ too greatly from the current system. In particular, it was hoped that course ratings and slope ratings could continue to be primary ingredients of the handicapping system. Course ratings have existed in various forms since the early 1900's and reflect the difficulty of a golf course from the point of view of an expert golfer. Slope ratings were incrementally adopted during the 1980's and provide a secondary measure of the difficulty of a golf course from the point of view of a non-expert golfer. Details concerning the history of golf handicapping in the United States can be found at www.usga.org.

2. The Committee also sought a handicapping system that is fair. Under the current handicapping system, it is known (Bingham and Swartz, 2000) that the stronger golfer is advantaged in handicapped matches between two golfers whereas weaker golfers are more likely to win in handicapped tournaments involving many golfers.

3. It was hoped that a new handicapping system could be developed where the system is better understood. Currently, the handicapping system of the RCGA is based on "potential ability" (page 73, RCGA 2008), and accordingly, a golfer's net score frequently exceeds the course rating. This may cause some frustration for golfers who expect to "shoot their handicap" 50% of the time.

This paper outlines the proposed system whose adoption is under consideration by the RCGA. An important and unique feature of the proposed system is that statistical theory is used to drive the methodology. Since golf scores exhibit variation, the use of statistical theory is particularly relevant in achieving the three main goals described above. In this paper, supporting theory is also accompanied by extensive data analyses.

In section 2, we review the main elements of the current handicapping system of the RCGA. We also indicate the minor differences between RCGA and USGA handicapping. In section 3, we describe the proposed system which consists of two approaches. In most instances, we advocate an approach for casual play which is implemented in the same way as the current handicapping system, and whose adoption should therefore cause few difficulties. A secondary approach is described for tournament play where inequities involving the current system can be extreme. In addition, we outline the theoretical advantages of the proposed system under both casual and tournament play. In section 4, the controversial topic of equitable stroke control (ESC) is discussed where the intention is to clarify some of the relevant issues in determining a suitable method of ESC. In section 5, data analyses are carried out where we investigate equitable stroke control, the interpretation of handicap and the fairness of matches and tournaments. Some concluding remarks are given in section 6.

2 The Current Handicapping System

As the proposed handicapping system is closely related to the current handicapping system of the RCGA, we provide a brief review of the current handicapping system.

Under the current handicapping system, for every round of golf, a golfer is required to enter an *adjusted score*. An adjusted score is no larger than the corresponding *gross score* (i.e. actual score), and is a consequence of imposing a maximum number of strokes permitted for each hole. The maximum number of strokes is due to ESC and is described more fully in section 4. The implementation of ESC which prescribes the conversion of gross scores to adjusted scores is the only substantive difference between the handicapping systems of the RCGA and the USGA.

Many golfers consider the current RCGA handicapping system to be a black-box procedure where complex calculations are performed on the adjusted scores. The calculations give rise to a quantity called the *handicap factor* or *factor* F which is a measure that quantifies the ability of a golfer. When $F = 0$, a golfer is excellent and is referred to as a "scratch" golfer. It is even possible to have $F < 0$ and these golfers (mostly professionals) are referred to as "plus" golfers. A weak golfer has a factor $F > 30$. The factor is an important quantity as it provides a means for golfers of differing abilities to reasonably compete against one another. In USGA handicapping, the term *handicap index* is used instead of handicap factor. In fact, apart from some differences in terminology and the implementation of ESC, the handicapping systems of the RCGA and the USGA are identical. The handicapping system of the USGA was developed by the USGA, and permission to use the system was given to the RCGA.

The factor F can be calculated for golfers who have completed five or more rounds of golf. We omit discussion of the treatment of 9-hole scores and tournament scores. More information on the intricacies of RCGA handicapping is given in the RCGA Handicap Manual (RCGA 2008).

For ease of exposition, we consider the calculation of the factor for a golfer who has completed 20 or more rounds of golf. For each round of golf, the *differential*

$$D = 113\left(\frac{X-R}{S}\right) \tag{1}$$

is calculated where X is the adjusted score, R is the *course rating*, S is the *slope rating* and D is rounded to the first decimal place. The differential has a long-standing history in golf where small differentials indicate excellent rounds of golf and large differentials indicate poor rounds of golf. The course rating R represents the score that a scratch golfer would be expected to shoot on a given course and the slope rating S is an additional measure of course difficulty. A course with a slope rating greater than (less than) 113 is considered to be a more difficult (less difficult) than average course. Given two courses with the same course rating, the idea behind the slope rating is that weaker golfers struggle more on the course with the higher slope rating. We note that there are different course and slope ratings for men and women, and there are different course and slope ratings for the various sets of tees.

In the next step of the handicapping formula, the best (i.e. lowest) 10 of the 20 most recent differentials are averaged. This quantity is then multiplied by 0.96 and truncated to the first decimal place to give the factor F.

To illustrate the calculation of the factor, consider a golfer whose most recent 20 differentials are given as shown:

21.8	12.4	18.4	15.4	21.2	22.4	13.1	17.4	19.3	15.6
14.9	10.3	17.1	19.9	11.3	11.3	14.2	28.8	10.9	23.4

The 20 differentials are then sorted from smallest (best rounds) to largest (worst rounds). The sorted differentials are given as shown:

10.3	10.9	11.3	11.3	12.4	13.1	14.2	14.9	15.4	15.6
17.1	17.4	18.4	19.3	19.9	21.2	21.8	22.4	23.4	28.8

We then choose the best 10 of the 20 most recent differentials and this corresponds to the first row of the two rows displayed above. We then calculate

$$\begin{aligned} F &= .096(10.3) + .096(10.9) + .096(11.3) + .096(11.3) + .096(12.4) \\ &+ .096(13.1) + .096(14.2) + .096(14.9) + .096(15.4) + .096(15.6) \\ &= 12.4. \end{aligned} \quad (2)$$

We have chosen to write (2) in this slightly unusual way to facilitate a comparison with the proposed handicapping system.

Finally, a golfer determines his *course handicap* by rounding $FS/113$ to the nearest integer where S is the slope rating on the course which the golfer is playing that day. When two golfers are playing one another in a handicapped round of golf, they may choose to play either *medal* or *match* play. In medal play, each golfer obtains their *net score* by subtracting their course handicap from their gross score. The golfer with the lower net score is the winner. In match play, golfers compete hole by hole, and the golfer who wins the most holes wins the match. When applying handicap in match play, the difference in the two course handicaps is the number of strokes allotted the weaker golfer. For example, if the difference in course handicaps is 7, then the weaker golfer reduces his score by one on each of the first 7 handicap holes. When the difference in course handicaps exceeds 18, then strokes are similarly allotted. For example, if the difference in course handicaps is 26, then the weaker golfer reduces his score by two on the first 8 handicap holes and reduces his score by one on the remaining 10 holes. The ranking of handicap holes appears on the scorecard where ranking is done according to hole length and difficulty.

3 The Proposed Handicapping System

The proposed handicapping system has two components; one for casual play and one for tournament play. It is expected that almost all of the time, players will use the approach designed for casual play. However, for large tournaments involving players of differing abilities, it is preferable to use the approach for tournament play.

3.1 Casual Play

Whereas the factor is the main component in the current handicapping system, the proposed approach for casual play introduces the quantity $\hat{\mu}$ which is referred to as the *mean*. Conceptually, this is an extremely simple adjustment for golfers, as everything remains the same, except that the mean $\hat{\mu}$ replaces the factor F.

The calculation of the mean is similar to the calculation of the handicap factor. Referring to the example in section 2, the mean is given by

$$\begin{aligned} \hat{\mu} &= w_1(10.3) + w_2(10.9) + w_3(11.3) + w_4(11.3) \\ &+ w_5(12.4) + w_6(13.1) + w_7(14.2) + w_8(14.9) \\ &+ w_9(15.4) + w_{10}(15.6) + w_{11}(17.1) + w_{12}(17.4) \\ &+ w_{13}(18.4) + w_{14}(19.3) + w_{15}(19.9) + w_{16}(21.2) \\ &= 16.9 \end{aligned} \tag{3}$$

where we have rounded (rather than truncated) to the first decimal place, and the *weights* $w_1, \ldots, w_{16}$ are given in the Appendix.

One of the differences between formula (3) and formula (2) is that we use the best 16 differentials instead of the best 10 differentials. The rationale for using more differentials is that it is wasteful of information to exclude half of the differentials. However, we do not use all 20 differentials since the largest values correspond to a golfer's worst rounds. Sometimes, the worst rounds are not reflective of a golfer's true ability as there is a tendency to lose concentration when playing poorly. Another difference between formula (3) and formula (2) is that the weights $w_1, w_2, \ldots, w_{16}$ have replaced the constant 0.096.

How are the weights chosen? Although golf scores are discrete, the scientific literature suggests that unadjusted (i.e. gross) scores are approximately normally distributed (see Scheid, 1990; Bingham and Swartz, 2000). This implies that for a given round of golf,

$$113\left(\frac{Y-R}{S}\right) \sim \text{Normal}(\mu, \sigma^2) \tag{4}$$

where Y denotes a golfer's gross score, and μ and σ are parameters of the normal distribution which characterize a golfer's ability and consistency respectively. Temporarily ignoring the effect of ESC, the best (i.e. lowest) 16 of the most recent 20 differentials as defined by (1) therefore comprise a right-censored sample from the normal distribution in (4). We estimate μ of the normal distribution by the statistic $\hat{\mu}$ in (3). The statistic $\hat{\mu}$ is an unbiased linear estimator; unbiased in the sense that $E(\hat{\mu}) = \mu$ and linear in the sense that

$\hat{\mu}$ is a linear combination of the best 16 differentials. Moreover, the weights $w_1, \ldots, w_{16}$ are chosen so that the estimator $\hat{\mu}$ is a *blue* (best linear unbiased estimator) where best refers to the property of minimum variance within the class of linear unbiased estimators.

A primary advantage of using the mean $\hat{\mu}$ instead of the factor F is via interpretation. We are interested in the probability that a golfer "shoots his handicap". Shooting one's handicap occurs when a golfer's net score is lower than the course rating, and the corresponding probability is given by

$$\begin{aligned}
\text{Prob}(Y - \hat{\mu}S/113 < R) &\approx \text{Prob}(Y - \mu S/113 < R) \\
&= \text{Prob}(113(Y - R)/S < \mu) \\
&= \text{Prob}(Z < 0) \\
&= 1/2
\end{aligned}$$

according to (4) where Z is a standard normal random variable. Therefore, unlike the factor, the mean provides golfers with an intuitive understanding of handicap. A golfer should shoot his handicap roughly half the time. In contrast, there is no simple interpretation involving net scores using the factor. In fact, with the current handicapping system, some golfers are dismayed and perplexed that they seldom shoot their handicap.

Another advantage of using the mean $\hat{\mu}$ instead of the factor F is that stroke play between two golfers of differing abilities is theoretically fair. Currently, using the factor, the better golfer has an advantage in stroke play. To establish the stated result, suppose that we have two golfers who are characterized by the quantities (Y_1, μ_1, σ_1) and (Y_2, μ_2, σ_2) respectively, and assume that the golfers are playing on a course with course rating R and slope S. The number of handicap strokes in casual play for golfer 1 is therefore $\hat{\mu}_1 S/113$ and the number of handicap strokes in casual play for golfer 2 is $\hat{\mu}_2 S/113$. Assuming statistical independence between the two golfers, the probability that golfer 1 defeats golfer 2 in casual play is therefore

$$\begin{aligned}
\text{Prob} &= \text{Prob}(Y_1 - \hat{\mu}_1 S/113 < Y_2 - \hat{\mu}_2 S/113) \\
&\approx \text{Prob}(Y_1 - \mu_1 S/113 < Y_2 - \mu_2 S/113) \\
&= \text{Prob}(W < 0) \\
&= 1/2
\end{aligned}$$

where $W \sim \text{Normal}(0, S^2(\sigma_1^2 + \sigma_2^2)/113^2)$ follows from (4). Therefore, statistical theory suggests that stroke play competitions between two golfers under casual play are fair. We remark that the theoretical result regarding fair play also applies to the highly unusual situation of two golfers playing different courses.

In this case, the number of handicap strokes need only be modified by the difference in the two course ratings.

We conclude the discussion on the proposed system for casual play with three relevant comments. First, the mean $\hat{\mu}$ always exceeds the factor F for non-scratch golfers. Referring to the Appendix and letting $D_{(i)}$ denote the i-th order statistic of the differentials, the result is established by noting that

$$\begin{aligned}
\hat{\mu} - F &= w_1 D_{(1)} + \cdots + w_{16} D_{(16)} - .096 D_{(1)} - \cdots - .096 D_{(10)} \\
&\geq w_1 D_{(1)} + \cdots + w_9 D_{(9)} + .6134 D_{(10)} - .096 D_{(1)} - \cdots - -.096 D_{(10)} \\
&= .5174 D_{(10)} + (w_1 - .096) D_{(1)} + \cdots + (w_9 - .096) D_{(9)} \\
&\geq .5174 D_{(10)} + (w_1 - .096) D_{(1)} + \cdots + (w_9 - .096) D_{(1)} \\
&= .5174 D_{(10)} - .4773 D_{(1)} \\
&\geq .5174 D_{(1)} - .4773 D_{(1)} \\
&= .0401 D_{(1)} \\
&> 0
\end{aligned}$$

for non-scratch golfers (i.e. $D_{(1)} > 0$). Therefore, it is evident that golfers will require a mental adjustment in having a larger handicap. This may bruise some egos, but the adjustment should not be too difficult as most golfers will experience an increased handicap.

Second, we have noted that the theory developed in this section for interpretation and fairness refers to gross scores whereas the calculation of differentials is based on adjusted scores. We suggest that the discrepancy should not cause too much difficulty as the 16 best rounds are used in the calculation of $\hat{\mu}$. In the best rounds, ESC is invoked less often than in the four worst rounds. We investigate the effect of ESC further in the data analyses of section 5.

Third, although there is no theoretical reason to choose the best 16 rounds from the most recent 20 rounds, in preliminary studies, we have found that 16 seems to be roughly optimal. Our optimality criterion is based on goodness-of-fit tests where we test the normality of the first m order statistics $D_{(1)}, \ldots, D_{(m)}$ where $m \leq 20$. There is some evidence (Scheid, 1990) that golf scores have slightly longer right tails than a normal distribution, and this provides added support for not including all 20 differentials in the handicapping calculation.

3.2 Tournament Play

In large tournaments involving players of differing abilities, it is generally known that the current handicapping system is unfair. In particular, net prizes are rarely won by low handicappers. The explanation for this phenomenon is simply that the scores of high handicappers tend to be more variable than the scores of low handicappers (Bingham and Swartz, 2000).

To focus on the problem, an approach has been proposed for tournament play that takes into account the variability of golfers. This involves the introduction of a second statistic $\hat{\sigma}$ which is referred to as the *spread*. The spread measures the variability of a golfer where smaller values of $\hat{\sigma}$ denote more consistent golfers. Referring again to the example in section 2, the spread is given by

$$\begin{aligned} \hat{\sigma} &= q_1(10.3) + q_2(10.9) + q_3(11.3) + q_4(11.3) \\ &+ q_5(12.4) + q_6(13.1) + q_7(14.2) + q_8(14.9) \\ &+ q_9(15.4) + q_{10}(15.6) + q_{11}(17.1) + q_{12}(17.4) \\ &+ q_{13}(18.4) + q_{14}(19.3) + q_{15}(19.9) + q_{16}(21.2) \\ &= 4.86 \end{aligned} \tag{5}$$

where the weights $q_1, q_2, \ldots, q_{16}$ are given in the Appendix. Note that the calculation of the spread is identical to the calculation of the mean except that the weights $q_1, q_2, \ldots, q_{16}$ replace the weights $w_1, w_2, \ldots, w_{16}$ and the spread is rounded to two decimal places. The spread statistic $\hat{\sigma}$ is an estimator of the standard deviation σ in the normal distribution (4), and again, the estimator is a blue. Previous data analyses have demonstrated that most golfers have a spread statistic lying in the interval $(1.5, 8.0)$. We introduce the additional requirement that $\hat{\sigma}$ is restricted to the interval $(1.5, 8.0)$. Therefore, if the calculation in (5) yields $\hat{\sigma} < 1.5$, then $\hat{\sigma}$ is set equal to 1.5. If the calculation in (5) yields $\hat{\sigma} > 8.0$, then $\hat{\sigma}$ is set equal to 8.0.

Therefore, with the adoption of the proposed handicapping system, the mean $\hat{\mu}$ and the spread $\hat{\sigma}$ will become a standard part of the golfer's lexicon. In summary, the mean measures ability and the spread measures consistency.

For tournament play, the proposed tournament net score is given by

$$N_T = R + \frac{452(Y - R)}{S\hat{\sigma}} - \frac{4\hat{\mu}}{\hat{\sigma}} \tag{6}$$

where Y is the golfer's gross score and N_T can be calculated to as many decimal places as required to break ties. The tournament net score N_T also depends on the course slope S, the course rating R, the mean $\hat{\mu}$ and the spread $\hat{\sigma}$. For example, suppose that the golfer in section 2 shoots a gross score of $Y = 86$ on a course with course rating $R = 69.5$ and slope $S = 120$. Then the golfer's tournament net score is $N_T = 68.38$. We anticipate that tournament net scores will be calculated with the aid of a computer after the rounds are complete.

This is not so different from other post-round handicapping systems such as the Callaway system and the Peoria system.

The formula (6) is based on statistical theory (see the Appendix) and has many appealing properties. Most importantly, tournaments will be much "fairer" than is currently the case. In tournament play, we require a different notion of fairness. In a tournament involving N golfers, we define a fair handicapping system as one in which each golfer wins with probability $1/N$. To investigate the fairness of the proposed tournament net score N_T, consider a golf course with course rating R and slope rating S where we characterize the i-th golfer by (Y_i, μ_i, σ_i), for $i = 1, \ldots, N$. Under this scenario and assuming statistical independence of the golfers, it follows from (4) that the quantities

$$R + \frac{452(Y_i - R)}{S\hat{\sigma}_i} - \frac{4\hat{\mu}_i}{\hat{\sigma}_i} \approx R + \frac{452(Y_i - R)}{S\sigma_i} - \frac{4\mu_i}{\sigma_i}$$

are independent and identically distributed (iid) Normal$(R, 4^2)$ random variables, for $i = 1, \ldots, N$. The fairness criterion follows immediately from the iid property.

In addition, the theory above immediately provides the intuitive result that approximately 50% of tournament net scores will be lower than the course rating and 50% of tournament net scores will be greater than the course rating. Also, tournament net scores will resemble current net scores; in a tournament involving 100 golfers, we expect roughly three N_T values lower than $R - 8$.

In large tournaments, it is typical that golfers are divided into flights of comparable handicap. These golfers then compete for gross and net prizes within their flights. In situations like these, it may seem unnecessary to use the tournament net score (6), and it may be simpler to use the proposed approach for casual play. However, if the tournament awards an overall net prize, then we advocate that the tournament net score provides a much fairer way of determining the winner.

There is one remaining benefit of the proposed approach for tournament play. The approach provides for the atypical case where the N golfers are playing different courses. The added generality of golfer-specific course ratings R_i and golfer-specific slope ratings S_i causes no additional difficulty. Let $R_{\max}$ denote the maximum of the different course ratings. In this case, the tournament net score for the i-th golfer is given by

$$N_T = R_{\max} + \frac{452(Y_i - R_i)}{S_i\hat{\sigma}_i} - \frac{4\hat{\mu}_i}{\hat{\sigma}_i}$$

and the above properties concerning fairness follow in an analogous fashion.

4 Equitable Stroke Control

When a golfer enters a score for handicapping purposes, it is not the gross score which is entered. Rather, it is an adjusted score where the adjustment is due to ESC. ESC places a cap on the number of strokes allowed on each hole, and therefore, an adjusted score is no larger than the corresponding gross score. Since the proposed handicapping system (and the current system) rely on adjusted scores, ESC plays a role in handicapping.

It is well known that a good portion of golfers do not understand ESC and they do not apply ESC as required. Therefore, it would no doubt be simpler for the RCGA to eliminate ESC. However, there are various reasons for retaining ESC including:

- ESC provides a partial safeguard against "sandbagging" (i.e. deliberately inflating scores so as to increase one's course handicap)
- ESC helps maintain pace of play
- ESC provides an unequivocal rule for scoring in cases where a score on a hole is not recorded or is unknown
- ESC provides some conformity with the USGA which has a slightly different rule regarding ESC

As ESC plays a role in handicapping, in section 5, we consider the effect of three methods of ESC on the proposed handicapping system. The three methods are:

1. no ESC
2. restrict each hole to a maximum of 9 strokes
3. the current method of ESC whereby "mean" replaces "factor" and a maximum of bogey/double/triple/quadruple is allowed per hole for golfers with a mean $\leq$ 0/1-18/19-32/33+

Commenting on the three methods, we suggest that method 1 (no ESC) is the simplest to understand but offers no protection against sandbagging and is problematic for pace of play. At the other end of the spectrum, method 3 (the current method) is the most complex but offers some protection against sandbagging. Method 3 also introduces the greatest disparity between gross scores and adjusted scores. Method 2 (maximum 9 strokes) appears to be a compromise between method 1 and method 3.

5 Data Analysis

In this section, we study an extensive data set of full golf rounds recorded by the members of the Coloniale Golf Club in Beaumont, Alberta from 1996 through 1999. We limit our analysis to those golfers who have played a minimum of 20 rounds. We note that the hole by hole data permits the calculation of adjusted gross scores using the three methods of ESC described in section 4. We are left with a dataset consisting of 8000 rounds collected on 178 golfers. The Coloniale data is exceptionally good for testing since it is a large dataset encompassing a wide range of golfing abilities. In figure 1, we provide a histogram of the factor values corresponding to the 8000 rounds at Coloniale.

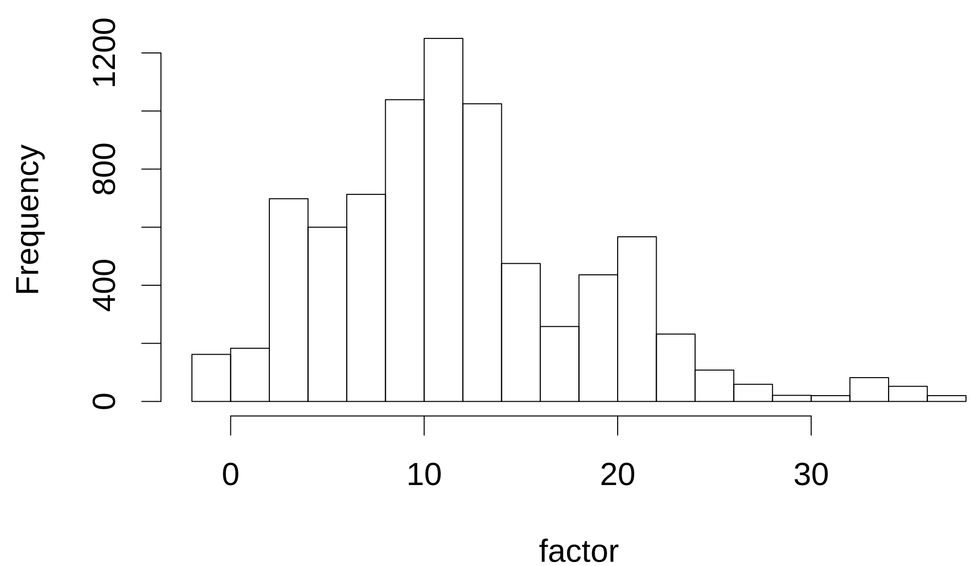

Figure 1: Histogram of factor values from the Coloniale Golf Club.

Since the proposed handicapping system under casual play involves the replacement of the factor with the mean, we investigate the difference between the mean and the factor. In figure 2, we provide three histograms corresponding to the difference between the mean and the factor using the three methods of ESC. In the Coloniale data, the average difference between the mean (with no ESC) and the factor is 4.3 strokes. Similarly, the average difference between the mean (max 9 ESC) and the factor is 4.2 strokes. This suggests that there is no meaningful difference in terms of handicapping between eliminating ESC and introducing ESC with a 9 stroke maximum. The average difference between the mean (current ESC) and the factor is 3.2 strokes. Should any of the three methods of ESC be adopted with the proposed handicapping system, it is evident that most golfers will require a mental adjustment in having a larger handicap.

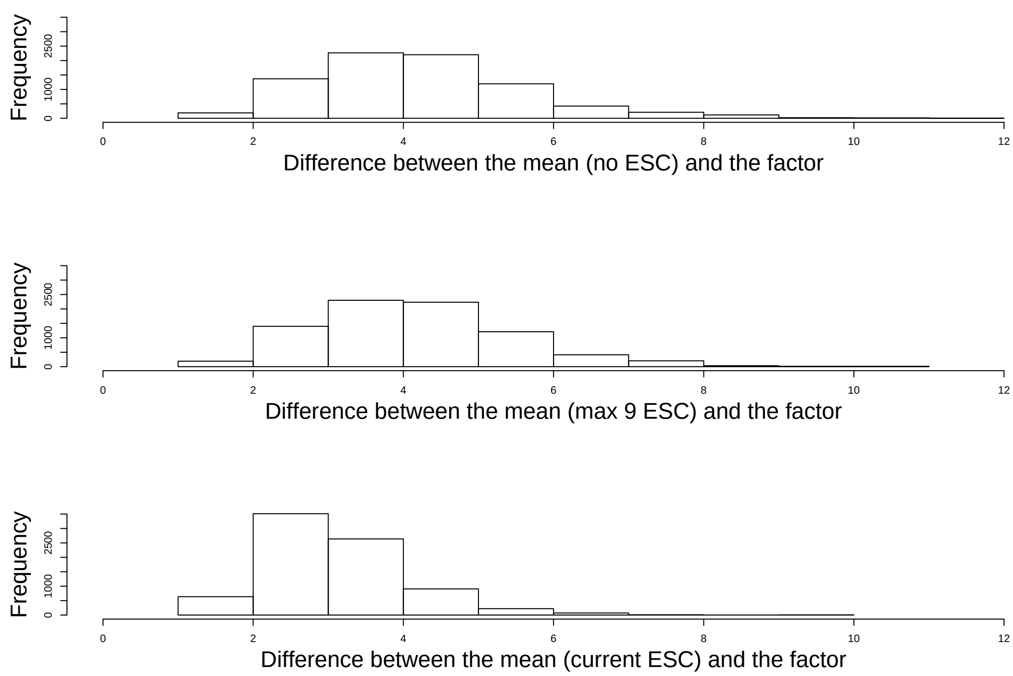

Figure 2: Histograms of the difference between the mean and the factor at the Coloniale Club based on the three methods of ESC.

We next investigate the spread which is a measure of consistency in golf and is used in the proposed handicapping system under tournament play. In figure 3, we provide three histograms of the spread using the three methods of ESC applied to the Coloniale data. The average spread using the three methods of ESC are 3.9, 3.9 and 3.4 strokes respectively. In each of the three histograms we observe that the vast majority of golfers have a spread between 2.0 and 5.0.

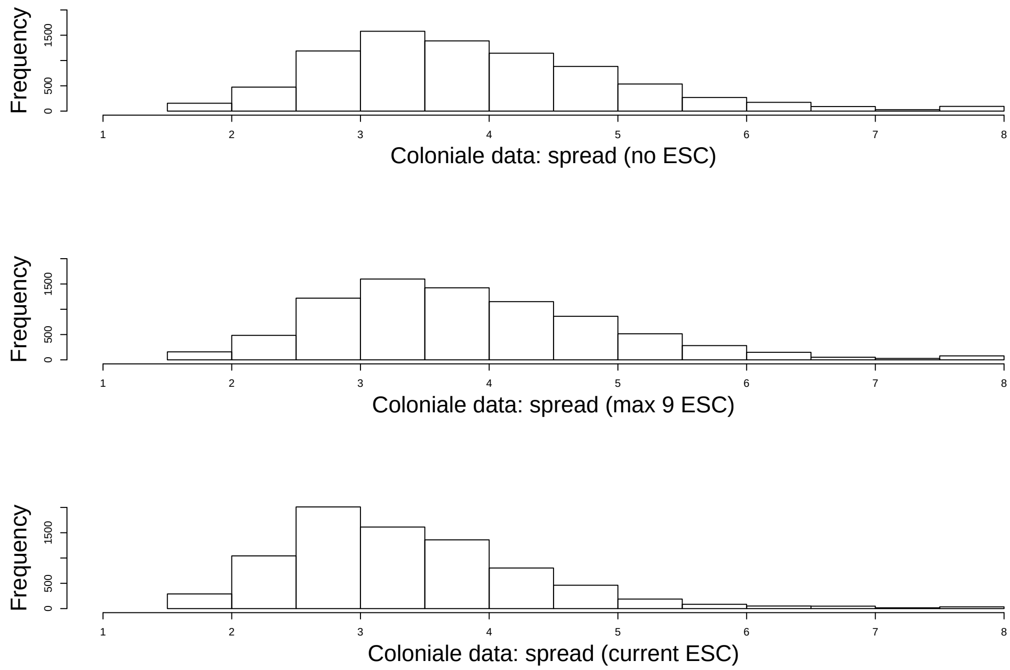

Figure 3: Histograms of the spread at the Coloniale Golf Club based on the three methods of ESC.

It is well-known (eg. Bingham and Swartz, 2000) that low handicap golfers are generally more consistent than high handicap golfers. In figure 4, we provide a scatterplot of the spread versus the mean. The plot is obtained from the Coloniale data using ESC with a 9 stroke maximum. We observe an upward trend in the points and note that similar shapes are observed using the other methods of ESC. Figure 4 confirms the inconsistency of many high handicap golfers. This, in turn, suggests the relevance of the proposed handicapping system under tournament play which takes consistency into account.

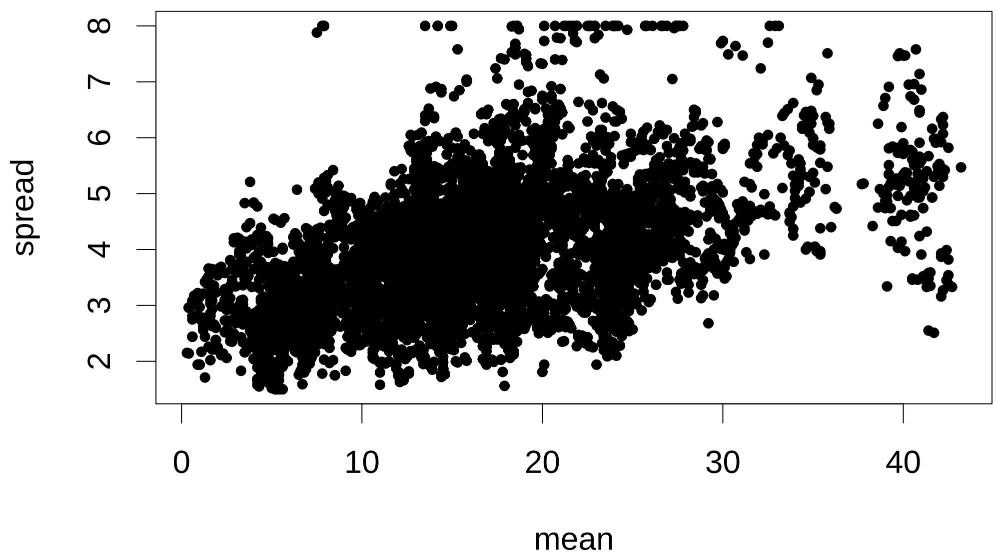

Figure 4: Scatterplot of the spread versus the mean at the Coloniale Golf Club using ESC with a 9 stroke maximum.

5.1 Interpretation of Handicap

Using the Coloniale data, we investigate the interpretation of handicap which is one of the primary motivations for the proposed handicapping system. In particular, we calculate the percentage of time that golfers shoot their handicap in the 8000 rounds. We do this by looking at whether a golfer's net score is lower than or equal to the course rating. This is calculated for the current handicapping system and the proposed handicapping system under casual play. From table 1, we observe that golfers rarely shoot their handicap using the current handicapping system. However, under the proposed system, shooting one's handicap is far more frequent. In fact, without ESC, the percentage of time that Coloniale golfers shoot their handicap is 46% which is close to the theoretical value of 50%. Although close to 50%, a normal test for proportions at level $\alpha = 0.05$ indicates that the percentage 46% is significantly different from 50%. The detection of significance is due to the size of the dataset, random error, the discreteness of scores and departures from normality. These results concerning interpretation suggest that abolishing ESC or introducing a 9 stroke maximum may be preferable to the current method of ESC.

Table 1: Coloniale data: percentage of time golfers shoot their handicap.

Current System	Proposed System under Casual Play		
	no ESC	ESC (max 9)	ESC (current)
11%	46%	45%	36%

5.2 Fairness in Handicapping

We next investigate the issue of fairness which is the main purpose of handicapping. We first compare the current handicapping system with the proposed handicapping system under casual play in a stroke play competition between two golfers. To do this, we randomly generate two scores from the 8000 scores and observe whether or not the lower handicapper wins or ties in stroke play. The random procedure is repeated 40,000 times. From table 2, we observe that the current system significantly favours the lower handicap golfer whereas the proposed system is more fair. The fairness of the proposed system is not greatly affected by the method of ESC.

Table 2: Coloniale data: percentage of time the better golfer wins or ties in stroke play.

	Current System	Proposed System under Casual Play		
		no ESC	ESC (max 9)	ESC (current)
Wins Match	52%	45%	45%	48%
Ties Match	6%	6%	6%	6%
Wins +(1/2)Ties	55%	48%	48%	51%

For investigating fairness under tournament play, we consider a hypothetical tournament involving 99 competitors. From the 8000 scores, we randomly generate 99 scores and determine the winner of the hypothetical tournament involving 99 golfers. We then observe whether the winning golfer comes from the top third (ie. lowest handicappers) or the bottom third (ie. highest handicappers) of the 99 golfers. The random procedure is repeated 40,000 times, and in a fair system, we would expect that the winner would come from each of the two groups 33% of the time. From table 3, we observe that the current system significantly favours the higher handicap golfers as they win 40% of the time. The proposed handicapping system appears to be more fair as the percentages corresponding to the two groups are closer to the ideal value of 33%. We note that the Coloniale Golf Club has fewer higher handicap rounds when compared to a typical golf club (see figure 1). For a typical golf club, we would expect the values in table 3 to be even more extreme in favour of high handicap golfers under the current handicapping system.

Table 3: Coloniale data: percentage of time that the winning golfer in a tournament of 99 players belongs to the top third of handicaps and the bottom third of handicaps.

	Current System	Proposed System under Tournament Play no ESC	ESC (max 9)	ESC (current)
Top Third	27%	29%	29%	29%
Bottom Third	40%	34%	34%	32%

6 Concluding Remarks

The theoretical developments and the data analyses contained in this paper provide evidence that the proposed handicapping system has clear advantages over the current handicapping system. The most important advantages are:

- improved "fairness" in competitions involving golfers of varying ability
- an ease in interpretation of "handicap"

With so many ideas/systems involving handicapping, the proposed system is unique in that its theoretical underpinnings are based on statistical theory. Statistical theory is the study of random systems, and golf scores exhibit randomness to the extent that golfers do not have the same scores from round to round. The important point here is that the underlying statistical theory helps provide confidence in the proposed handicapping system. In summary, the proposed handicapping system is supported by both theory and data analyses.

We suggest that the adoption of the proposed handicapping system should not be a difficult transition. The proposed system continues to use the well-established elements of handicapping, namely (1) course ratings and (2) slope ratings. Furthermore, the proposed handicapping formula is not too different from the current formula where only the weights have changed.

With respect to ESC, the data analyses in section 5 suggest that the proposed handicapping system is not greatly affected by the method of ESC. The choice of ESC method may be more of a "political" decision than a "numbers" decision.

It should be mentioned that under the proposed handicapping system, 9-hole scores are treated the same as in the current system. Therefore 9-hole scores present no additional difficulty for the proposed system. Also, match play does not differ under the proposed handicapping system; again we simply replace the factor with the mean.

It should also be mentioned that although possible, the proposed handicapping system has not introduced tournament scores (T-scores) for special consideration. This may be viewed as a feature of the proposed system as it will simplify the RCGA Handicap Manual. In most clubs, it is noteworthy that tournament scores are rarely considered and that a Handicap Committee has the authority to modify a player's handicap.

Over many years, the proposed handicapping system has been carefully studied and has evolved into an excellent system. We offer the opinion that the proposed system is ready for adoption by the RCGA.

7 Appendix

In (3), the weights $w_1, w_2, \ldots, w_{16}$ correspond to the calculation of the mean $\hat{\mu}$ for a golfer who has completed 20 or more rounds, and in this case, we consider only the most recent 20 rounds. In table 4, we provide the specific values of the weights as taken from Sarhan and Greenberg (1962). Note that when a golfer has completed only n rounds where $5 \leq n \leq 19$, different sets of weights are given. For example, when a golfer has completed only 6 rounds of golf, then only the best four differentials are used in the calculation of the mean, and the weights are 0.185, 0.1226, 0.1761 and 0.6828 respectively.

In (5), the weights $q_1, q_2, \ldots, q_{16}$ correspond to the calculation of the spread $\hat{\sigma}$ for a golfer who has completed 20 or more rounds. In table 5, we provide the specific values of the weights as taken from Sarhan and Greenberg (1962). Note that when a golfer has completed only n rounds where $5 \leq n \leq 19$, different sets of weights are given.

The development of the proposed tournament net score N_T in (6) is based

Table 4: Weights for the calculation of the mean based on n rounds.

	Number of rounds n							
Weights	5	6	7	8	9	10	11	12
w_1	.1252	.0185	.0465	.0569	.0602	.0605	.0320	.0360
w_2	.1830	.1226	.1072	.0962	.0876	.0804	.0609	.0581
w_3	.2147	.1761	.1375	.1153	.1006	.0898	.0741	.0682
w_4	.4771	.6828	.1626	.1309	.1110	.0972	.0845	.0759
w_5			.5462	.1451	.1204	.1037	.0935	.0827
w_6				.4555	.1294	.1099	.1020	.0888
w_7					.3909	.1161	.1101	.0948
w_8						.3424	.4430	.1006
w_9								.3950

	Number of rounds n							
Weights	13	14	15	16	17	18	19	20
w_1	.0380	.0388	.0390	.0257	.0271	.0279	.0283	.0284
w_2	.0555	.0529	.0506	.0408	.0398	.0387	.0376	.0365
w_3	.0633	.0592	.0556	.0474	.0453	.0433	.0415	.0399
w_4	.0693	.0639	.0595	.0524	.0494	.0468	.0445	.0424
w_5	.0745	.0680	.0628	.0566	.0528	.0497	.0469	.0445
w_6	.0792	.0717	.0657	.0604	.0559	.0522	.0491	.0463
w_7	.0836	.0752	.0685	.0638	.0588	.0546	.0510	.0480
w_8	.0880	.0785	.0711	.0671	.0615	.0568	.0529	.0496
w_9	.0924	.0819	.0737	.0704	.0641	.0589	.0547	.0511
w_{10}	.3564	.0852	.0763	.0736	.0666	.0610	.0564	.0526
w_{11}		.3247	.0790	.0768	.0692	.0631	.0582	.0540
w_{12}			.2982	.3649	.0718	.0652	.0599	.0554
w_{13}					.3378	.0674	.0616	.0569
w_{14}						.3144	.0634	.0584
w_{15}							.2940	.0599
w_{16}								.2762

Table 5: Weights for the calculation of the spread based on n rounds.

	Number of rounds n							
Weights	5	6	7	8	9	10	11	12
q_1	-.5117	-.5528	-.4370	-.3638	-.3129	-.2753	-.2852	-.2545
q_2	-.1668	-.2091	-.1943	-.1788	-.1647	-.1523	-.1610	-.1487
q_3	.0274	.-0290	-.0718	-.0881	-.0938	-.0947	-.1038	-.1007
q_4	.6511	.7909	.0321	-.0132	-.0364	-.0488	-.0589	-.0633
q_5			.6709	.0570	.0160	-.0077	-.0194	-.0308
q_6				.5868	.0678	.0319	.0178	-.0007
q_7					.5239	.0722	.0545	.0286
q_8						.4746	.5562	.0582
q_9								.5119

	Number of rounds n							
Weights	13	14	15	16	17	18	19	20
q_1	-.2301	-.2102	-.1937	-.1982	-.1837	-.1713	-.1605	-.1511
q_2	-.1382	-.1292	-.1214	-.1252	-.1179	-.1114	-.1057	-.1006
q_3	-.0970	-.0933	-.0897	-.0935	-.0895	-.0858	-.0823	-.0792
q_4	-.0653	-.0658	-.0655	-.0694	-.0681	-.0665	-.0649	-.0632
q_5	-.0379	-.0423	-.0450	-.0491	-.0501	-.0504	-.0504	-.0500
q_6	-.0128	-.0209	-.0265	-.0310	-.0341	-.0361	-.0375	-.0384
q_7	.0113	-.0006	-.0091	-.0140	-.0192	-.0230	-.0257	-.0277
q_8	.0352	.0192	.0078	.0022	-.0051	-.0105	-.0146	-.0178
q_9	.0598	.0393	.0246	.0182	.0087	.0015	-.0040	-.0082
q_{10}	.4750	.0601	.0417	.0343	.0225	.0135	.0065	.0011
q_{11}		.4438	.0598	.0508	.0364	.0254	.0170	.0103
q_{12}			.4169	.4748	.0509	.0377	.0276	.0196
q_{13}					.4491	.0505	.0385	.0291
q_{14}						.4264	.0499	.0389
q_{15}							.4062	.0492
q_{16}								.3880

on the approximation

$$113\left(\frac{Y-R}{S}\right) \sim \text{Normal}(\hat{\mu}, \hat{\sigma}^2) \qquad (7)$$

which follows from assumption (4) where Y denotes a golfer's gross score, R is the course rating, S is the slope rating and $\hat{\mu}$ and $\hat{\sigma}$ are the mean and spread as given in (3) and (5) respectively. From (7), we have that

$$113\left(\frac{Y-R}{S\hat{\sigma}}\right) - \frac{\hat{\mu}}{\hat{\sigma}} \sim \text{Normal}(0,1)$$

and therefore

$$N_T = R + 452\left(\frac{Y-R}{S\hat{\sigma}}\right) - \frac{4\hat{\mu}}{\hat{\sigma}} \sim \text{Normal}(R,16).$$

This suggests that tournament net scores N_T are approximate realizations from a normal distribution centred about the course rating R with the bulk of the scores (95%) lying in the interval $(R-8, R+8)$. Therefore, on a course with course rating $R = 70$ and 100 competitors, we might expect only a couple of tournament net scores lower than $N_T = 62$. These rates correspond closely to those using existing net scores in tournament settings.

8 References

Bingham, D.R. and Swartz, T.B. (2000). Equitable handicapping in golf. *The American Statistician*, 54(3), 170-177.

RCGA (2008). *The RCGA Handicap Manual*, Royal Canadian Golf Association.

Sarhan, A.E. and Greenberg, B.G. (1962). The best linear estimates for the parameters of the normal distribution. In *Contributions to Order Statistics* (editors A.E. Sarhan and B.G. Greenberg), New York: Wiley, 206-269.

Scheid, F. (1990). On the normality and independence of golf scores, with applications. In *Science and Golf: Proceedings of the First World Scientific Congress of Golf* (editor A.J. Cochran), London: E. and F.N. Spon, 147-152.