Studying Variability in Statistics via Performance Measures in Sport

Clarke, Stephen R
Swinburne University of Technology, Faculty of Life and Social Sciences
John St Hawthorn
Victoria 3122 Australia
E-mail: sclarke@swin.edu.au

1. Introduction

Many students find statistics difficult and are often disinterested. One reason may be that they see it as irrelevant. If the example data are meaningless to the students, they are not interested in the outcome, and confuse the boringness of the data with the technique. If the data analysed are outside their range of experience, they have no basis for questioning how they were obtained, nor any stake in the answer or method of analysis. But show their team has little chance of winning the flag, or that their favourite player does not measure up to past champions, and they will immediately question your basis for the claim. Interest can be heightened if examples are made relevant to the student's present experience. The very students who claim Statistics is boring can often quote you batting averages and such like for many players or teams past and present.

Most statistical concepts can find an application in sport, and its now related area of gambling. In Swinburne's Graduate courses in Applied Statistics, we had two subjects where this approach was taken. "Chance and Gaming" was in reality an introduction to probability, where probability, permutations and combinations, expected value, mean and variance, geometric, binomial, hypergeometric, negative binomial, normal and Chi square distribution, were all taught via examples from gaming. In Sports Performance Modelling, applications in sport were used to introduce analytic and simulation modeling, fitting statistical distributions, linear & logistic regression and Markov chains & stochastic dynamic programming. In this paper we discuss some examples of the use of player performance measures to generate interest or discussion on the topic of variability.

2. Measures of Variability

Most sports have measures of performance which purport to indicate some level of achievement. The most usual is the average – thus cricketers have their batting average, golfers their average score, and basketballers their average points per game. But surprisingly, given the importance many sport followers place on consistency, no measures are quoted which measure the variability in performance. What about the variance or standard deviation of batting, golfing or throwing scores? After all, this would affect the chance of a batsman making a duck or a century, or the likelihood a golfer makes the cut or wins a tournament. The variability of a sportsman can have as much affect of the outcome as the average level of performance, yet it is rarely discussed and usually never measured. In this first example from golf we look not at the golfer's statistics, but the instrument used to measure his performance, the golf course.

During a golf tournament, the media often give an indication of the difficulty of the various holes by showing the average score on the hole. However it is not only the ease or difficulty of a hole that is important. If everybody takes one over par on a hole, that hole is not separating players. The standard deviation gives a measure of the ability of a hole to discriminate between players. It is the holes with the greatest standard deviation on which it is important for players to do well – in so doing they puts more distance between themselves and the other golfers. In a paper investigating the discriminating power of golf courses, Clarke and Rice (1995) obtained hole by hole scores of the 63 qualifiers in the 1992 U.S. Master's tournament. The data obtained from the organisers included the average score on each hole, but the only indication of the variation of scores was the number of eagles, birdies and bogies. These are very difficult to interpret. Table 1 shows the mean and standard deviation of scores on each hole. It is clear that holes 2, 8, 13 and 15 (all par fives) were easier to play than the others. Note that hole 10, clearly the hardest hole over the four rounds, was on only one day in the top half of the holes in order of discrimination (standard deviation). In this regard holes 12 and 13 clearly stand out above the rest, between them sharing the highest and second highest standard deviation in every round. In the first round, Hole 12, a par 3, is one of the easiest holes, but clearly produces the most highly variable scores. In Parsons (1976) the section of the Augusta course from holes 11 to 13 is described thus "The Masters championship has been won or lost so often between the 11th and 13th that this three hole stretch has become known as Amen corner". In particular, Jack Nicklaus describes the 12th as " the most demanding tournament hole in the world". Clearly, in these data, the standard deviation of scores and not the mean score reflects golfers' view of the importance of the holes. Unfortunately this statistic is never quoted. Perhaps it should be.

Table 1. Descriptive Statistics on the Score Relative to Par for the 63 Qualifiers in the 1992 US Masters

		Round One		Round Two		Round Three		Round Four	
Hole	Par	Mean	Std	Mean	Std	Mean	Std	Mean	Std
1	4	-0.05	0.55	-0.02	0.68	0.05	0.55	0.02	0.63
2	5	-0.22	0.73	-0.56	0.56	-0.21	0.72	-0.29	0.79
3	4	0.00	0.48	-0.08	0.55	0.21	0.72	0.10	0.59
4	3	0.11	0.63	0.16	0.60	0.22	0.66	0.14	0.50
5	4	0.17	0.58	0.03	0.44	0.19	0.56	0.13	0.66
6	3	0.08	0.52	-0.03	0.65	0.06	0.54	0.10	0.56
7	4	0.05	0.58	0.05	0.61	-0.24	0.76	0.19	0.72
8	5	-0.25	0.65	-0.29	0.52	-0.17	0.58	-0.40	0.58
9	4	-0.06	0.47	0.10	0.64	0.03	0.59	0.06	0.54
10	4	0.17	0.61	0.11	0.57	0.22	0.58	0.40	0.55
11	4	0.11	0.65	0.11	0.63	0.03	0.44	0.22	0.55
12	3	-0.03	0.78	0.21	0.70	0.17	0.93	0.41	1.10
13	5	-0.41	0.73	-0.35	0.77	-0.46	0.80	-0.44	0.96
14	4	-0.21	0.45	0.11	0.57	0.14	0.64	-0.08	0.52
15	5	-0.24	0.69	-0.54	0.64	-0.57	0.76	-0.65	0.63
16	3	0.00	0.54	-0.02	0.58	0.02	0.55	-0.05	0.68
17	4	-0.02	0.58	-0.19	0.59	0.02	0.55	0.11	0.57
18	4	0.02	0.49	-0.11	0.54	0.19	0.59	0.00	0.54

This concept can be applied in any event in which the total score is made up of a sum of scores on several parts. The triathlon, which consists of a swimming, cycling and running leg, is an interesting example. One view might be that the distances of the three legs should result in equal times. However what is important in determining the winner is how each leg spreads the field. It is the time between athletes that is important, so I would argue the standard deviation of the times should be equal. de Mestre (1992) analyses the energy expended and suggests that the influence of an event in the triathlon in determining the winner is proportional to the variance of the times. Tobin and Clarke (1993) analyse individual times in the 1993 Melbourne Classic triathlon (1.5km swim, 40km ride, 10km run) and show the standard deviation of the individual legs in seconds are 229, 394 and 307. Clearly cyclists have a huge advantage - not surprising as the event was originally designed by cyclists. Using equality of standard deviations of times, the respective legs would be 2, 31 and 10 km. Based on de Mestre's recommendation, a fairer event would consist of legs of length 2, 18 and 7.5 km. Students could investigate other multi-event competitions such as the decathlon.

We now turn to variability in an individual sports person's performance. Athletes and sports followers believe an essential characteristic of excellence is consistency, or low variability of performance. Yet rarely is this aspect of performance measured. Certainly there are few statistics that purport to describe it, and there is plenty of scope for students to investigate. Pollock (1977) in discussing the mean and standard deviation of scores in relation to consistency of golfers says "it seems reasonable that a better player would have low values for both". However, in many sports, it is the exceptional performance, not the average one that is sought after. Thus an increase in variability will often increase the percentage of wins for players competing simultaneously against many opponents. For example, Clarke (1991) shows that a golfer averaging 1 under par will increase the number of tournament wins from 3.3% to 8.5% if the standard deviation increases from 1.5 shots per round to 2 shots per round (assuming 10 under is required to win). Students interested in golf could investigate the statistics of golf handicaps, where early attempts to produce fair handicaps which give all players an equal chance of winning a tournament, just measured the average score of the golfer. Later rules took into account the difficulty of the golfer's home course, and the playing conditions of the day. Some attempt has also been made to account for the variability of a golfer's scores.

If golfers put such a store in 'consistency', and the variance of a player's rounds is important to their chance of winning, why is it not published as a performance statistic? It would be interesting to check if top golfers do vary significantly in their round by round standard deviations. Rotella and Boutcher (1990) use regression analysis on the playing statistics of professional golfers to predict money earned. Over 13 published statistics were used, but standard deviation of scores was not one of them. However birdies divided by greens in regulation was the second most important variable behind scoring average. This statistic may be a surrogate for variance, as it goes up when the number of bogies goes up as well as when the number of birdies increases. Hale & Hale (1990) find that for the leading money winners the performance statistics are not a good predictor of success - perhaps this is further evidence that some others are needed.

Furthermore, in many sports, an increase in variability will actually result in an increase in the average measure of performance. For example, in many field events the athlete's score is the longest out of 3 or even 6 attempts. In such cases, more variable performance will not only increase the chance of winning the event, but will actually produce a greater mean score. Since the average maximum of 3 standard normals is 0.8463 the expected score for an athlete whose individual attempts have a mean μ and standard deviation σ is $\mu+0.8463\sigma$. Clarke (1991) applies this to some actual data for long jumpers, which shows that over 18 cm of the final length of a jumper arises through the variation in the jumps. Of course for world records, or individual athlete's best performances, which are the maximum of many attempts, the effect of variation is even more important.

3. Sources of Variation

Sport can also be used to introduce students to the idea of various sources of variation. A golfer's tournament score is made up of 4 rounds, each made up of scores on 18 holes. The score in a particular round is the sum of the 18 hole by hole scores, Thus if X_i is the score for the ith hole then the score for the round is

$$S = X_1 + X_2 + \quad ..... \quad + X_{18}. \qquad (1)$$

Consider two golfers: one short but straight, a cautious putter who invariably gets par on a hole; the second a long but sometimes wayward hitter and bold putter who has a good chance of a birdie but also a good chance of a bogie. In a round both players could both get par, but the first may get it by having 18 pars, while the other may get 6 pars, 6 birdies and 6 bogies. This difference could be measured by the variance or standard deviation of the hole by hole scores (variance of the Xs) , where the first would show up with a low value and the second with a high value. In a similar way the variance of the Ss would indicate the round by round consistency of a golfer. In general for elite golfers, only the average of S is published – other statistics such as the var(X) and var (S) based on various sources of variation might add to our knowledge of a golfer's makeup.

Similar analysis could be applied to other sports. For example, a cricketer's innings score can be modelled as

$$S = X_1 + X_2 + \quad ..... \quad + X_N. \tag{2}$$

where X_i is the score on the ith ball faced, and N is now a random variable, being the number of balls faced. The relevant formulae for a random sum of random variables is

$$E(S) = E(X).\ E(N) \tag{3}$$

$$Var(S) = E(X)^2.\ Var(N) + E(N).\ Var(X) \tag{4}$$

As before, two batsmen might average 30, but one does it by scoring slower for a longer time (smaller E(X), larger E(N)). This is now covered somewhat, at least in one-day cricket, by run rate, which is essentially a measure of E(X). But again, one batsman might average a run a ball by hitting a single off every ball, while another hits a 6 off every sixth ball. Clearly the 'excitement factor' could be measured by the performance statistic Var(X).

There are many other examples where investigation of sources of variation produces useful statistics. For example, a discussion of variation in performance statistics such as averages over time, could lead to development of moving or exponentially smoothed averages. These would certainly be interesting in cricket for individual batsmen over their careers, and for team run-rates in one day innings. Similar statistics would apply in most other sports ,and could be used to track the development and loss with age of skill levels.

4. Fitting Standard Distributions

Followers of any sport know that performance is variable – a tennis player sometimes gets 70% of his serves in, at other times only 60%; a soccer side sometimes goes scoreless, at other times scores 3 or 4 goals. Sports followers usually put this down to variation in form, but how much is actually due to the inherent variability in the game. Often in sport, outstanding performances such as a large score or a long run of wins, is hailed as evidence the sportsman concerned has played exceptionally well or poorly. However it may equally well be explained by the random occurrences that are expected when players play at a constant level. Pollard et al (1977) has a good discussion. They say

In most sports and games, the winner of a particular contest, be it an individual or a team, is decided both by skill and by luck. There are various ways in which the role played by chance can be investigated. One of these is to examine the frequency distribution of certain events in a game, such as the scoring of a run at baseball. If we are able to demonstrate that this particular event is governed by the laws of a probability distribution, then one can say that within the framework of that distribution, the events are occurring at random. Although the actual event will occur by chance, the rate at which the event occurs depends on the skill of the players involved. The fact that one team may be better than another, in the sense that it has a higher rate of scoring runs, does not alter the concept of runs occurring at random, nor rule out the possibility that the inferior team may win due to a random fluctuation.

This can be investigated and forms a great introduction to the fitting of standard distributions. For example, consider the scores of Jamie Siddons, who batted about number 6 for Victoria in the Australian Sheffield Shield competition in 1985/6. His scores for the year were

33, 17, 76, 5, 74, 7, 7, 107, 1, 45, 17, 2, 36.

Many cricket followers would say that is an inconsistent set of results, since they expect a consistent batsman to have scores with a small standard deviation, like 51, 55, 52, 53, 54. However scores like this mean that a batsman has no chance of going out until he/she reaches 50, and is almost certain to go out soon after. So in terms of probability of dismissal they are very inconsistent. An alternative view of consistency might say a consistent batsman who has a 30% chance of making 50, should turn 30% of those 50s into centuries, and 30% of centuries into 150s etc. Wood (1945), Elderton (1945) and Clarke (1991) discuss this notion of consistency in cricketers, and suggest statistics for its measure.

This assumption of a constant probability of dismissal leads to a geometric distribution (or its continuous counterpart, the negative exponential) for scores. The negative exponential distribution is common as the distribution of waiting times for random events - in this case it is the waiting time (measured by score) until a dismissal. A histogram of Siddons' scores and the geometric distribution with p = 1/33 are shown in Figure 1. The two are virtually identical.

Clearly Siddons' scores follow closely what theory suggests a player with a constant probability and an average of 33 should produce. The standard deviation of Siddons' scores is 34, again agreeing to that predicted by an exponential distribution whose standard deviation is equal to the mean. Followers who judge Siddons to be inconsistent on the basis of his scores would be doing him a great injustice. In this case skill is playing its part in giving Siddons an average of 33. A more skilful player will have a higher average, a less skilful player a lower average. But luck determines on the day whether he will score 100 or go out for a duck.

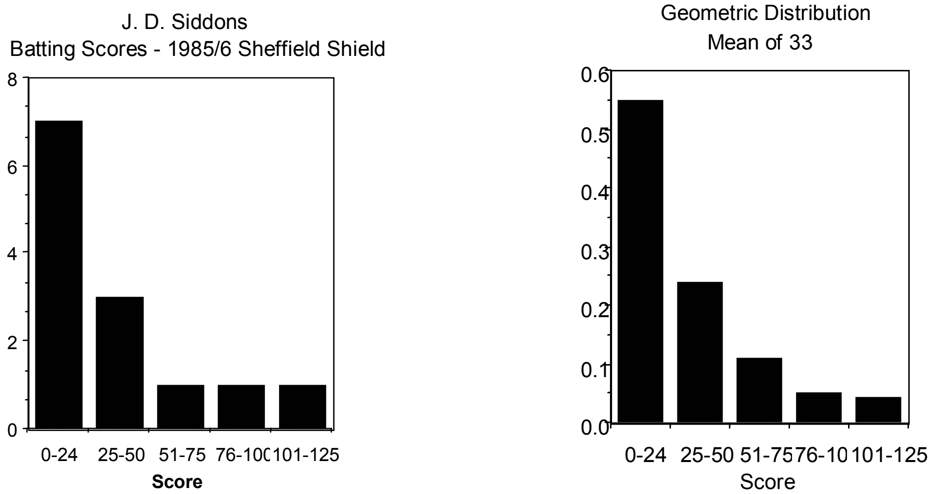

Figure 1. Comparison of Batting Scores with the Geometric Distribution

When fitting distributions, some of the parameters may be obvious from the context, while others have to be estimated from the data. For example, in fitting the binomial distribution to the number of scoring shots in an over of cricket, clearly $n = 6$ while p might be known from previous results or estimated from the current data. Because students have a knowledge of the application area, they will question the assumptions, so in the above case, they might argue that different bowlers or batsman might alter p. The goodness of fit can then be used as a test for the validity of their argument. The failure of a fit often teaches lessons about the distribution, and will usually lead to a modification of the model or the parameters or subsetting the data. So if the geometric distribution fails to fit a player's first class cricket scores, students might suggest splitting the data into test cricket and other first class cricket, as we might expect the former to be more difficult.

Many papers have been written fitting standard distributions to sports scores, and examples can be found from most sports. Some that could be tried with students are given as follows.

The Binomial distribution:

- the number of scoring shots in an over of cricket;
- the number of goals of a particular basketball player in the first 5 attempts from the line;
- the number of first serves faults in the first four points of a tennis game;
- the number of quarters of football a particular team wins each match;
- the number of birdies in a round of golf;
- the number of half-innings in which a run is made by a team in a baseball match.

The Geometric distribution:

- the number of balls faced in a batsman's innings;
- the scores of batsmen in cricket;
- the number of misses or shots until the first goal in soccer;
- the number of sets until a particular tennis player wins the first set;
- the number of holes played until a golfer gets a birdie.

The Poisson distribution:

- The number of goals in a soccer match;
- The number of sixes in a one day cricket innings;
- The number of reports in a game of football;
- The number of dismissals in a session of test cricket.

The Negative Exponential distribution:

- time between goals in a soccer or basketball match;
- time between home runs in a baseball match;
- the number of balls between sixes in a one day cricket innings.

The Normal distribution is a continuous distribution, but it can be applied to many discrete variables which have large means:

- the number of goals in a netball or basketball match;
- the total number of points in an Australian Rules football match;
- the margin in points in an Australian Rules football match;
- the number of runs in a cricket innings.

Once a standard distribution is shown to describe the statistics, probability calculations can be used to answer other questions:

- what is the chance a tennis match will last longer than 100 rallies?
- what is the chance a batsman scores a century?
- what is the chance a team will score more than 3 goals?
- what is the chance a golfer scores less than 60?
- what proportion of Australian rules games are won by more than 60 points.

In some cases, a general statement can be made about all sportsmen or women. For example, it is easily shown that the probability an exponentially distributed random variable exceeds its mean is 1/e, or about 37%. So if a cricketer's batting scores are exponentially distributed, they should exceed their average about 37% of the time. In the small sample of Siddons' scores above, 5 out of 13 innings, or 38%, were above the average of 33. As another example, cricket's greatest batsman, Don Bradman, had a test average of 99.94. Of his 80 innings, 29, or 36% were centuries. Furthermore 15% were double centuries, compared to the 14% predicted by the exponential distribution.

5. Simulation

If standard distributions do not fit, this provides the perfect introduction to simulation. This can be used when the mathematics is too difficult or impossible. As an illustration, I was asked by a journalist to calculate the chance a particular golfer would break 60 in a single round. This is also an example of where historical data is of little use, and modeling needs to be used.

In the 1999 Australian Masters Karrie Webb completed a magnificent 4 days of golf by breaking the record number of shots under par for a women's golf tournament. Her 26 under par included 28 birdies, 43 pars and one double bogey. This one lapse cost her the chance to equal the men's record of 28 under par for a tournament. During her second round Karrie was quoted as having momentarily considered the possibility of scoring a 59. This ranks as one of the most difficult achievements in sport – harder than 300 or a hat-trick in test cricket. Such rare achievements have as much to with luck as skill, and will often be achieved by a good but not necessarily great player having an inordinate amount of good luck, rather than by the greatest player.

This problem could be tackled by using historical data. At the time of the round, three men had achieved a score of 59 in tournament golf in the USA, and as far as I could ascertain no woman had achieved a sub 60 score. But such statistics could only be used to estimate the probability that any random pro golfer playing in any random tournament would break 60. We wanted to determine the chance that the best female golfer in the World, Karrie Webb, playing in this particular tournament, at that particular time, had of achieving that score. The lack of any of the statistics for Webb that measure variability of performance as previously discussed meant an analytic analysis was impossible, and we resorted to simulation.

To estimate the difficulty of the task, we used a resampling technique – assume that the figures actually obtained represent the scores that would be obtained in a retrial. In this case, for each of 18 holes we resampled with replacement from the 28 birdies, 43 pars and 1 double bogey Webb obtained in the tournament. A computer, takes a few seconds to simulate 10,000 tournaments or 40,000 rounds. This analysis showed that she would only equal or better her 26 under 45% of the time, and had a 30% chance of breaking the 28 under record. However the 59 or under in a single round is another matter. Only 136 of the 40000 rounds, or 0.4%, were under 60. Thus the best player in the world, playing the best golf of her life, will break 60 in less than one every 200 rounds. This study was described in Hopkins (1999).

Many other statistics in sport can be calculated using simulations. Examples include chances of teams making finals, chances of two players or teams meeting in a tournament, number of rallies in a tennis, squash or tables tennis match, etc..

6. Conclusion

The use of examples from sport is a great way to involve students in statistics. Here we have discussed the use of performance measures. With the increasing use of computers, the evaluation of statistics which measure the variability of players and team's performance is made simpler. Data at a much more detailed level is now available than in the past. The calculation and publication of alternative statistics could assist players and coaches to a better understanding of their performances, and at least would provide followers with some interesting data. With a suitable choice of names, these could be interpreted by the general public. Of importance here is that they show students how different statistics can be used to describe or summarise a game or a player's makeup.

There are other areas in sport that can be used to great effect to introduce statistical concepts, or to provide interesting examples. One such is forecasting. A large part of media discussion is centred on predicting future results, and this now has important applications in the gaming area. This provides excellent opportunities for the use of time series analysis, and linear, multiple and logistic regression. The general interest of the media in sport also means studies in this area have a fair chance of receiving media coverage. This can be a great motivation for students and reinforces the relevance of their statistical work. For example, Glasson et el (2001) describes an undergraduate student project using logistic regression and simulation to forecast an Olympic games Beach volleyball tournament. Yelas & Clarke (2004) describes a graduate student project using exponential smoothing methods and simulation to successfully forecast the 2003 World Cup of Rugby. Both these studies obtained press coverage.

It is important that you use sports in which students have an interest. American boys may find Australian netball just as big a turn off as examples on the effects of different fertilizer. Teachers wishing to find inspiration for examples could do well to start with Bennet (1998). This forms an excellent summary on research up to 1998, and contains individual chapters on the sports American football, baseball, basketball, cricket, soccer, golf, ice hockey, tennis and track & field, as well as theme chapters on design of tournaments, data graphics, prediction and hierarchical models.

REFERENCES

Bennett, J. (Ed.). (1998). Statistics in Sport, Arnold: London.

Clarke, S.R. (1991). Consistency in Sport - with particular reference to Cricket. 27th NZOR conference, pp 30-35

Clarke, S. R. and J. M. Rice (1995). How well do golf courses measure golf ability? An application of test reliability procedures to golf tournament scores. *ASOR Bulletin* **14**(4): 2-11.

de Mestre, N. (1992). Mathematics applied to sport. In *Mathematics and Computers in Sport*, N. de Mestre Ed., Bond University, Gold Coast, pp137-148

Elderton, W. E. (1945). Cricket Scores and some Skew Correlation Distributions. *Journal of the Royal Statistical Society A*, **108**, 1-11.

Glasson, S., B. Jeremiejczyk and S.R. Clarke (2001). Simulation of Women's Beach Volleyball Tournaments. *ASOR Bulletin* **20**(2): 2-8.

Hale, T. and G. T. Hale, (1990). Lies, Damned Lies and Statistics in Golf. In *Science and Golf*, A.J. Cochran, Ed., Chapman and Hall, London, pp 159-167.

Hopkins, T. (1999). Its official: golfers need lot's of luck. *The Australian Financial Review*, March 6, pp. 69.

Parsons, I., (1976). Augusta: Dream course in a dream setting, *The World Atlas of Golf*, I. Parsons, Ed., Mitchel Beazley, London, pp 116-119.

Pollard, R., B. Benjamin, et al. (1977). Sport and the negative binomial distribution. In *Optimal Strategies in Sports*, S. P. Ladany and R. E. Machol, Eds. North Holland: Amsterdam pp 188-195.

Pollock, S.M., (1977). A Model of the USGA Handicap System and Fairness of Medal and Match Play, *Optimal Strategies in Sport*, S.P. Ladany and R.E. Machol., Ed., North Holland, pp 141-150.

Rotella, R.J. and S.H. Boutcher, (1990). A Closer Look at the Role of the Mind in Golf, In *Science and Golf*, A.J. Cochran, Ed., Chapman and Hall, London, pp 93-97.

Tobin, P. C. and S. R. Clarke (1993). Some statistics of the triathlon. In *Mathematics: Of Primary Importance*, J. Mousley and M. Rice, (Eds.), Mathematics Association of Victoria: Melbourne. pp. 433-440.

Wood, G. H. (1945). Cricket Scores and the Geometrical Progression. *Journal of the Royal Statistical Society (SeriesA)*, **108**, 12-22.

Yelas, S. and S. R. Clarke (2004). Forecasting the 2003 Rugby World Cup. In *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport*, H. Morton, Ed., Massey University: Palmerston Nth. pp. 270-277.

ABSTRACT

Most performance measures in sport estimate mean performance. While consistency in sport is often quoted as an important attribute, there is little effort to measure variability. Tackling this oversight can provide a vehicle for introducing many statistical concepts to students. This paper looks at several examples. Using golf scores at the US masters to investigate the importance of each hole, we show the standard deviation of scores to be a more relevant measure than the mean score. The importance of variability in a golfer's scores can also be measured, and alternative performance measures created. A discussion of the meaning of consistency in cricket (or golf) leads to distributions of scores. There are many other examples in sport where outputs can be fitted with standard probability distributions, which can then be used to estimate probabilities of achieving given scores. We give examples from soccer, Australian rules football, cricket, golf, baseball, basketball and tennis. To allow for a better fit of the data, the concept of modeling can be introduced. In cricket this can be used to investigate various sources of variation and develop new player performance measures. Where standard models do not fit, simulation can be used. An example from golf uses bootstrap sampling to estimate the chance of an elite golfer breaking sixty.