NEWS
MULTI-DAY EVENTS - A FAIR SCORING SYSTEM (full version of the article published in O-Sport 1/2004)
2004-04-4
Article by David Rosen

At some multi-day events, the overall results list is something of a lottery. The order of competitors for the overall multi-day competition is strongly influenced by chance factors. This article is an attempt to explain what goes wrong, and to encourage organisers to think more carefully about their scoring system.

WHAT DO WE MEAN BY A FAIR EVENT?

Mappers, planners, controllers and organisers all work hard to prepare a good orienteering event. And one of the key criteria of a good event is that it is fair. A fair event is a good test of orienteering so that the results list reflects the relative orienteering skills of the competitors.

Conversely, an unfair event is one where the order of competitors in the results list is affected by chance factors. For example, if the map is particularly poor around one control, then competitors won't know that it will be much better to approach the control from one particular direction. Some fortunate people will by chance come in that way, others will waste time as they battle through an unmarked thicket.

All aspects of an event have to be as fair as possible. Unfairness in any one part makes the whole competition unfair.

DIFFERENT TYPES OF MULTI-DAY EVENT

In a multi-day event, the results of the individual competitions have to be amalgamated in some way to give the overall result. The "classic" way to do this is simply to add up the times of all the competitions - as in the Swedish O-Ringen. This also makes it easy to stage an exciting chasing start on the last day.

However, some multi-day events, for good reasons, want to base the results on only some of the events - say best four out of six. That allows people to miss out one or two events. If you run every day, then a bad run or two can be discarded.

However hard we try, it is impossible to make the courses for the six days so similar in length and difficulty that you can just add up each person's fastest times. Inevitably, one or two days will have longer times than the others, so those days will be the ones most people will discard, even if they had their best runs on those days. Therefore, it is necessary to turn the times into points - a process statisticians call standardisation.

TURNING TIMES INTO POINTS – THE WRONG WAYS

How we convert the times into points is absolutely critical. Far too often, an organiser spends only a few minutes making this choice. They choose a “simple” method. Yet a wrong decision can undo the hundreds of hours of work that the event officials spend in trying to ensure that other aspects of the event are fair.

No conversion method will ever be completely fair, but we should choose a method that is as fair as possible.

There are several common methods which are variations of one of the following:
1. The winner gets 100, the second gets 99, the third get 98 etc.
2. The winner gets 1000 points and every extra 10 seconds loses a point.
3. The winner gets 1000 and other runners get points according to the formula 1000 x Winner's Time/Runner's Time

All three of these have fatal flaws.

Ad 1) The winner gets 100, the second gets 99, the third get 98 etc.
This takes no account of times, so if first and second are separated by only a few seconds, but then third trails by 10 minutes, the scores don't reflect that. For example, if the top results are:

Day 1
1. Andrew 60.03 (100) 2. Brian 65.13 (99) 3. Colin 75.33 (98)
Day 2
1. Andrew 68.24 (100) 2. Colin 69.04 (99) 3. Brian 69.14 (98)
then it is unfair that Brian's reasonably good run on Day 2 is only worth the same as Colin's poor run on Day 1.

Ad 2) The winner gets 1000 points and every extra 10 seconds loses a point.
This is highly dependant on the winner's performance. Often there is an outstanding runner in the class who wins by a long way on most days, and even when having a poor run, wins by a small margin. Andrew is that outstanding runner but he had a relatively poor run on Day 2. The results are:

Day 1
1. Andrew 60.03 (1000) 2. Brian 65.13 (969) 3. Colin 75.33 (907)
Day 2
1. Andrew 68.24 (1000) 2. Colin 69.04 (996) 3. Brian 69.14 (995)
Brian ran equally well both days and it is unfair that his Day 1 run scores rather poorly.

Ad 3) The winner gets 1000 and other runners get points according to the formula 1000 x Winner's Time/Runner's Time
This method is also highly dependant on the winner's performance. The results are:

Day 1
1. Andrew 60.03 (1000) 2. Brian 65.13 (921) 3. Colin 75.33 (795)
Day 2
1. Andrew 68.24 (1000) 2. Colin 69.04 (990) 3. Brian 69.14 (988)

Again, if Brian ran equally well both days, it is unfair that his Day 1 run scores rather poorly. But the other disastrous feature of the formula of the third method is that scores do not vary linearly with time; in other words, five minutes time difference is not worth a consistent number of points. On Day 1, Brian is 5 minutes 10 seconds behind Andrew and gets 79 fewer points. Colin is 10 minutes 20 seconds behind Brian (twice 5 minutes 10 seconds) but instead of scoring 158 fewer points (twice 79) he scores 126 fewer points. One great disadvantage of this is that it is then impossible to turn the total scores back into time differentials in order to create a meaningful chasing start on the last day. The non-linearity also bunches the slower runners' scores together, yet a five minute differential half way down the field is just as important as a five minute differential amongst the leaders. The 2003 Swiss O-Week unfortunately used this formula.

TURNING TIMES INTO POINTS – THE RIGHT WAY

Instead of basing the scoring entirely on the winner's time, it is more sensible to use all the information at our disposal i.e. everyone's time. Then unusual performances by one individual will have only a minimal effect on everyone's scores.

We award 1000 points to the mean (average) time for the class. The next question is how to scale the scores from there – how many points each minute faster or slower than the mean is worth.

In some (usually tough and complex) forests, the runners' finish times are quite spread out. In other (usually easier faster) forests the runners' finish times are more bunched. If we fail to take this into account, then the highest scores will be gained in the tough forests, so all the leaders would tend to discard their scores from a day where the forest was easier than the others. Therefore, we use the standard measure of spread, the standard deviation, to scale the scores. We award an extra 200 points for each standard deviation time faster than the mean time. (200 is chosen so that you have to be five standard deviations slower than the mean before you score zero).

This gives the formula:

Runner's score = 1000 + 200 x (Mean Time – Runner's Time)/Standard Deviation Time

Any statistician will tell you that this formula is nothing new – it is simply the standard statistical way of standardising data. It is a little more complicated than the three methods mentioned previously, but even they are impractical to use without a computer, and nowadays every multi-day event uses computers.

Suppose, in the results given previously, the Mean Times (MT) and Standard Deviation Times (SDT) are:

Day 1 MT 80.45 SDT 12.10
Day 2 MT 82.15 SDT 10.24

The scores then become:
Day 1
1. Andrew 60.03 (1340) 2. Brian 65.13 (1255) 3. Colin 75.33 (1085)
Day 2
1. Andrew 68.24 (1266) 2. Colin 69.04 (1254) 3. Brian 69.14 (1250)
Andrew gets a very good score on Day 1, but his run does not significantly affect the scores of the other runners. Colin's score is 170 points behind Brian, twice the difference in Andrew and Brian's points. Brian scores almost equally as well on Day 2 as Day 1.

USING THIS FORMULA

Several multi-day events are now using this formula including the prestigious bi-annual Scottish 6-Days. Stephan Kraemer has incorporated it into his MTScore software which runs alongside his MTage OL (multi-day event) software so it is easy for organisers to use.

In very small classes (fewer than 10 competitors), the Mean Time is reliable but the numbers are too small to calculate a reliable Standard Deviation. In that case, the formula is modified slightly to use an assumed standard deviation of 16% of the mean.

A meaningful chasing start can easily be arranged on the last day by turning the points differences back into times using the average scaling factor for the previous days.

CONCLUSION

Organisers of multi-day events where times are turned into scores should take the question of which scoring system to use as seriously as preparing a good map and checking that the controls are in the right place. The scoring system affects the fairness of the overall competition.

The scoring system given above is the mathematically fairest solution. The IOF has recognised this by basing its World Ranking system on the same principal.

Prospective competitors should check the scoring system before entering a multi-day competition. Just as you would avoid going to an event with bad maps, there is little point in spending time and effort to take part in a multi-day competition if the results are not going to be fair.

February 2004
Jan Skřička