Win-%-Rank

A Different Approach to Scoring a Sailboat Racing Series

Tom Compton
September 12, 2014

Revised July 22, 2015

- Situation

  When scoring a series regatta one has the problem of what to do when boats miss a significant number of races. Say for example a new boat joins the fleet midway through the season. It does not seem fair to penalize the boat for not racing in the first half of the season.

  Also, a boat may show up and be very competitive, but only show up a few times. Should a boat that sails frequently but not very well score better for the season than a boat that sails well but does not show up as often?

  I believe the objective should be to estimate performance based how one does when one sails against one's competitors. Showing up is well and good but it's not performance: performance is how one sails verses the competition.

- Concept

  In a series of fleet races, competitor A beats competitor B 10 out of 10 races. We can say that competitor A is better than competitor B. Similarly, if A beats B 5 out of 10 races, we can say that this pair of competitors is equal.

  Say competitor A gets 1 point for being better than competitor B, 0 points for being equal, and -1 point for B being better than A. And we do this for each competitor. So if competitor A were better than competitors B, C, and D, A would receive 3 points. If every competitor beat competitor C, C's score would be a -3.

  When we look at the points for competitors, those that are better than others will have more points than those that are worse than others. (We are not adding wins, which could be many races; we are adding points for being better than another competitor.)

  It's clear that when competitor A beats competitor B 10 out of 10 races, that competitor A is better than B. it's not so clear when competitor A beats B 6 out of 10 or 4 out of 10. A statistical test can be used to ask the probability of A being better than B. The probability becomes the "point" for A competing with B.

  Having a minimum level of participation and sailing against some minimum percent of the competitors seems appropriate. Often guests skipper in a series or a friend drives one's boat. Or

maybe the skipper and crew rotate driving. Under this scoring method, all the skippers have a chance at doing well. They are not penalized because they can't do all the races.

This is not a low-point or high-point each time one competitor beats another; it's a point for being better than another competitor with "better" based on all the encounters between a pair of competitors.

People have commented that doing well because one attends encourages participation. I would argue that we loose more by discouraging those that sail well. If one wants to give an attendance award, give an attendance award.

- Algorithm

For each pair of competitors, assign a 1 when the first competitor beats the other. Assign a -1 when the first competitor looses to the other. And 0 if they tie. One could average the list of 1's, -1's, and 0's. Or one could do a statistical Signed-Rank Test to find the probability that one competitor of the pair is better than another. The algorithm as implemented does it both ways. The version using the statistical test responds quicker to one competitor being better than another. So rank[a,b] equals the probability that competitor **a** is better than **b** based on the Signed-Rank Test. If a competitor pair did not race against each other, assign rank[a,b]=0. A rank of 0 says that **a** and **b** are competitive equals. This concept gives a ranking between 1 and -1 for each pair of competitors. The probability could be based on many races or a few or none. The sign of the rank of a pair comes from the sign of the average of the list of 1's, -1's, and 0's for the pair.

A relative ranking of zero between a pair of competitors means that the pair is evenly matched or that the pair has not competed against each other. In any case, we don't have any evidence that one competitor is better than another.

Say we take competitor **a** and add its ranking against each of its competitors, say **b**, **c**, **d**, **e**. Say the number of **a**'s competitors is **n** (total competitors minus 1). This gives a number that could range from -n to +n. A competitor that ranks better than all other competitors would have an value of **n**, one that ranks worse, -**n**. The number represents how many competitors are beaten on average. If we map this range to a new range that runs from 0 to 100, the new range becomes the percentage of the competitors **a** beats on average. This percentage becomes a competitor's score. The greater the percentage, the more competitors **a** beats on average.

This approach gives a relative ranking between competitors based on the outcome of races in which they have competed. It does not require that all competitors compete in every event. It gives a fair ranking for competitors that do well, but only participate a few times. It's a fair estimate of relative performance of one competitor with another. It's non-parametric in the sense that no consideration is given to quantitative differences between one competitive outcome and another (for example, no consideration is given for time difference between finishes). It gives about the same ranking as the common low-points scoring system if all the competitors compete in all the races.

- Minimum Participation

  Sailing one evening in a summer series is not really sailing a summer series. So there needs to be some minimum number of races before we can say a competitor participated in a series.

  If a competitor sails a lot against just a few of the competitors in a series, then the competitor is not being fairly measured against the fleet. So a competitor needs to sail against some percent of the members of the fleet.

  For the 2014 season, if we use a minimum 6 races sailed against at least 50% of the competitors in the fleet, all competitors meet minimum scoring requirements:

Competitors meeting minimum number of races	**b, k, m, p, r, t**
Competitors which have sailed against minimum percent of other competitors	**b, k, m, p, r, t**
Competitors meeting both of above requirements	**b, k, m, p, r, t**
Competitors not meeting above requirements	

- Guest Skippers

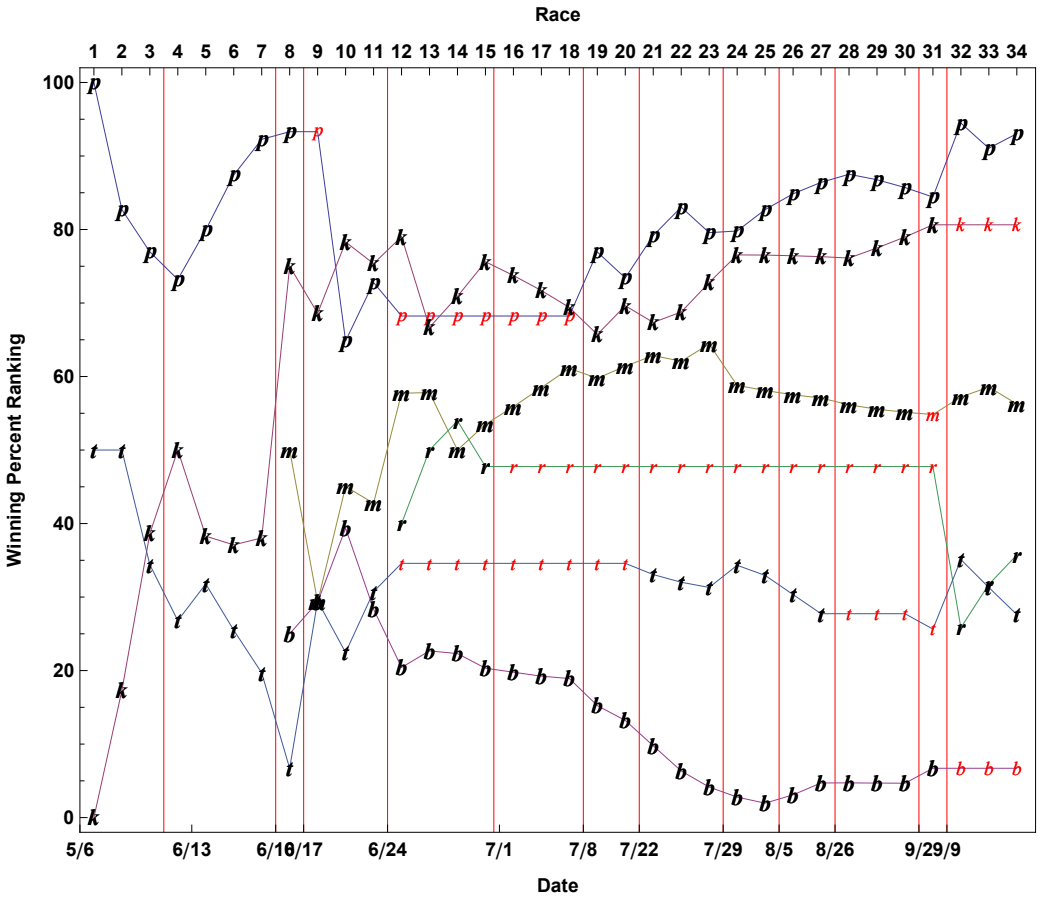
  We have a small, five boat fleet. We need at least three boats to have a quorum for racing. In the 2013 season, guest skippers added much needed boats on many nights. Win-%-Rank scoring allows these guest skippers to have an equal chance at ranking well in our fleet.

Cumulative Rankings Over 2014 Season

In the graph below, Win-%-Rank of each skipper is shown as the season progressed (cumulative for season). Black letters indicate races sailed, red, not sailed. A path in the chart begins when a skipper first sails a race.

In most scoring systems Rick, a guest skipper, would be hopelessly scored at the low-end of the rankings since he only sailed seven races. However, his ranking in the fleet as shown on the chart below is very reasonable. He is a good sailor but not very experienced driving a Lightning. He usually finished in the middle of the fleet. He had a reasonable ranking while not being penalized for attendance.

Lightning Fleet 192 Series Results, 2014
Scored by "Wining–%–Ranking"
(Cummulative as the Season Progressed)



Winning Percent Ranking		
p	93.0	Patrick
k	80.6	Kyle
m	56.3	Mike
r	35.7	Rick
t	27.6	Tom
b	6.7	Bill

A Look at Pairs of Competitors, 2014 Final Results

In the table below, for each pair of competitors, the probability that the first competitor of the pair is better than the second is given in the "rankSign" column.

The last column shows the number of races in which the competitor pair sailed against each other and the specific outcome of the race. So a 1 means that the first competitor beat the second competitor. A -1 means that the first competitor lost to the second. And a 0 means that they tied.

The "rankSign" value comes from applying a Signed Rank statistical test. Assume that the competitors are competitively the same. What is the probability that the pattern of the wins and looses is consistent with competitors of equal ability? If the probability that they are the same is low, then the opposite must true: the probability that they are different must be high. The number of under the "randSign" test is the probability that they are not the same.

Notice that for almost all the pairs of competitors, we have sailed enough races to know the relative ranking of all pairs of competitors. Pat and Kyle are an exception. They are very close. Pairs with Rick have much uncertainty since he has sailed few
races.

Pair	rankSign	Data: ⌗ races, race results for pair
{b, t}	−0.406	{12, {1, −1, 1, −1, −1, −1, −1, −1, −1, 1, 1, 1}}
{b, m}	−0.996	{24, {−1, 1, −1, −1, −1, −1, 1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, 1, −1, 1, −1, −1, 1}}
{b, p}	−0.999	{16, {−1, 1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}}
{b, k}	−1.000	{24, {−1, −1, −1, −1, −1, −1, −1, −1, −1, 1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}}
{b, r}	−0.928	{4, {−1, −1, −1, −1}}
{k, b}	1.000	{24, {1, 1, 1, 1, 1, 1, 1, 1, −1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}}
{k, t}	0.988	{19, {−1, −1, 1, 1, −1, 1, 1, 1, 1, 1, −1, 1, 1, 1, 1, 1, 1, 1, 1}}
{k, r}	0.576	{4, {1, −1, 1, 1}}
{k, m}	0.957	{24, {1, 1, 1, 1, −1, −1, 1, 1, −1, −1, −1, −1, 1, −1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}}
{k, p}	−0.457	{23, {−1, −1, 1, 1, −1, −1, −1, 1, 1, −1, −1, 1, −1, −1, 1, 1, −1, −1, −1, −1, 1, 1, 1}}
{m, b}	0.996	{24, {1, −1, 1, 1, 1, 1, 1, −1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, −1, 1, −1, 1, 1, −1}}
{m, r}	0.701	{7, {1, 1, −1, 1, 1, 1, −1}}
{m, t}	0.883	{15, {1, −1, 1, −1, 1, 1, 1, −1, 1, 1, 1, 0, 1, 1, −1}}
{m, k}	−0.957	{24, {−1, −1, −1, −1, 1, 1, −1, −1, 1, 1, 1, 1, −1, 1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}}
{m, p}	−0.997	{19, {−1, 1, −1, −1, 1, −1, −1, 1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}}
{p, r}	0.851	{3, {1, 1, 1}}
{p, b}	0.999	{16, {1, −1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}}
{p, t}	1.000	{21, {1, 1, 1, 1, 1, 1, 1, 1, −1, −1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}}
{p, m}	0.997	{19, {1, −1, 1, 1, −1, 1, 1, −1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1}}
{p, k}	0.457	{23, {1, 1, −1, −1, 1, 1, 1, −1, −1, 1, 1, −1, 1, 1, −1, −1, 1, 1, 1, 1, −1, −1, −1}}
{r, b}	0.928	{4, {1, 1, 1, 1}}
{r, t}	−0.227	{3, {−1, −1, 1}}
{r, m}	−0.701	{7, {−1, −1, 1, −1, −1, −1, 1}}
{r, k}	−0.576	{4, {−1, 1, −1, −1}}
{r, p}	−0.851	{3, {−1, −1, −1}}
{t, r}	0.227	{3, {1, 1, −1}}
{t, b}	0.406	{12, {−1, 1, −1, 1, 1, 1, 1, 1, 1, −1, −1, −1}}
{t, m}	−0.883	{15, {−1, 1, −1, 1, −1, −1, −1, 1, −1, −1, −1, 0, −1, −1, 1}}
{t, k}	−0.988	{19, {1, 1, −1, −1, 1, −1, −1, −1, −1, −1, 1, −1, −1, −1, −1, −1, −1, −1, −1}}
{t, p}	−1.000	{21, {−1, −1, −1, −1, −1, −1, −1, −1, −1, 1, 1, −1, −1, −1, −1, −1, −1, −1, −1, −1, −1}}

FAQ

It seems that new competitors do better than ones that have sailed more races. Why?

A competitor that has sailed lots of races against lots of other competitors has a lot of certainty about being better or worse. The more competitors sailed against and the more races, the more certainty. This places the competitor in the "fleet pecking order" with certainty. A new competitor has not done many races. So the certainty about being better or worse is low. And the new competitor may have not raced against many other competitors. This tends to place the new competitor towards the middle of the fleet and maybe higher than a competitor that has sailed lots of races against lots of other competitors. We have minimum number of races required to be a series participant. By the time the new competitor sails the minimum number of races against the minimum number of competitors, this all works out. Including new competitors in the results before they have met minimums seems a like a welcoming way to treat new people.

A competitor has a sudden jump in score. Why?

Say a competitor sails against five competitors and is better than all of them. Say another competitor sails against four of them and is better than these four. The first is likely to have a better score since the first competitor is better than more competitors. Now say the second competitor sails against the "missing" competitor and is better. The second competitor will see a jump in score due to sailing against more competitors. The more one sails, the greater the chance to increase one's score.

What is done about competitors that did not meet minimum participation requirements?

The scores of all competitors are affected by competitors that did not meet minimum participation requirements. One can "back out" the effect on scores by "backing out" results for competitors not meeting minimum participation requirements. The other alternative is just not report the scores of competitors not meeting minimum participation requirements. The choice one makes can make a difference in the outcome of the series. If the objective is to rank competitors that *participated* in the series, I believe one should "back out" the results of competitors not meeting minimum participation requirements.