The group freestyle fall score problem:

It is apparent to anyone who has spent a good amount of time as a chief judge that there is a problem with the way that falls scores are calculated for Group Freestyle. Indeed many good (Performance) judges may have also noticed the problem. The problem is most obvious in Large Group routines and is namely this: it is virtually impossible for a group routine to receive a low fall score, even if they fall many times during their routine. If you look at past judging scores for any Freestyle event (e.g. Unicon, European Championships, NAUCC, Bavarian Championships), you rarely see a Large Group routine with a fall score below 9.0 and virtually never below 8.0.

What's going on here:

Let's remind ourselves of how a group's fall score is calculated:

$$\text{Fall score} = 10 - \frac{\text{mistake score}}{\text{number of riders}}$$

What this means is that as the number of riders increases, the number of mistakes (i.e. falls) allowed for a given fall score also increases. This makes sense. If you have more riders in a routine, the chances of a fall happening also increases, and the impact on the routine of one rider falling is smaller than in a smaller group of riders. Thus we are *scaling* the mistake score depending on the number of riders. The concept and reasoning behind this formula is good and correct. Anyone who has trained with a group of Freestyle riders would understand it.

The problem:

The problem is, as the rule is currently written, our scale factor — 1/(number of riders) — is linear, and this is wrong! Using a linear equation means that if you have twice as many riders, you should be allowed twice as many falls for the same score. Here is an rough example of why a linear scale factor is wrong:
Let's look at a individual routine that would receive a reasonable fall score; one with 1 major fall and 2 minor falls. They would receive a mistake score of 2 and thus a fall score of 8. This is a pretty good (but not perfect) routine. Now let's look at a group freestyle routine of 16 riders. Because the scale factor is linear, in order to get the same fall score, we have to multiply by 16; a routine with 16 major falls and 32 minor falls. This would give a large group routine with 16 riders a fall score of 8. But a routine with that many falls is not a pretty good routine. It is at best a mediocre routine! This is a pretty extreme example for the purpose of illustration

but I could provide a similar example for any two different group sizes and get a similar results. This is a build-in fact of using a linear scale factor.

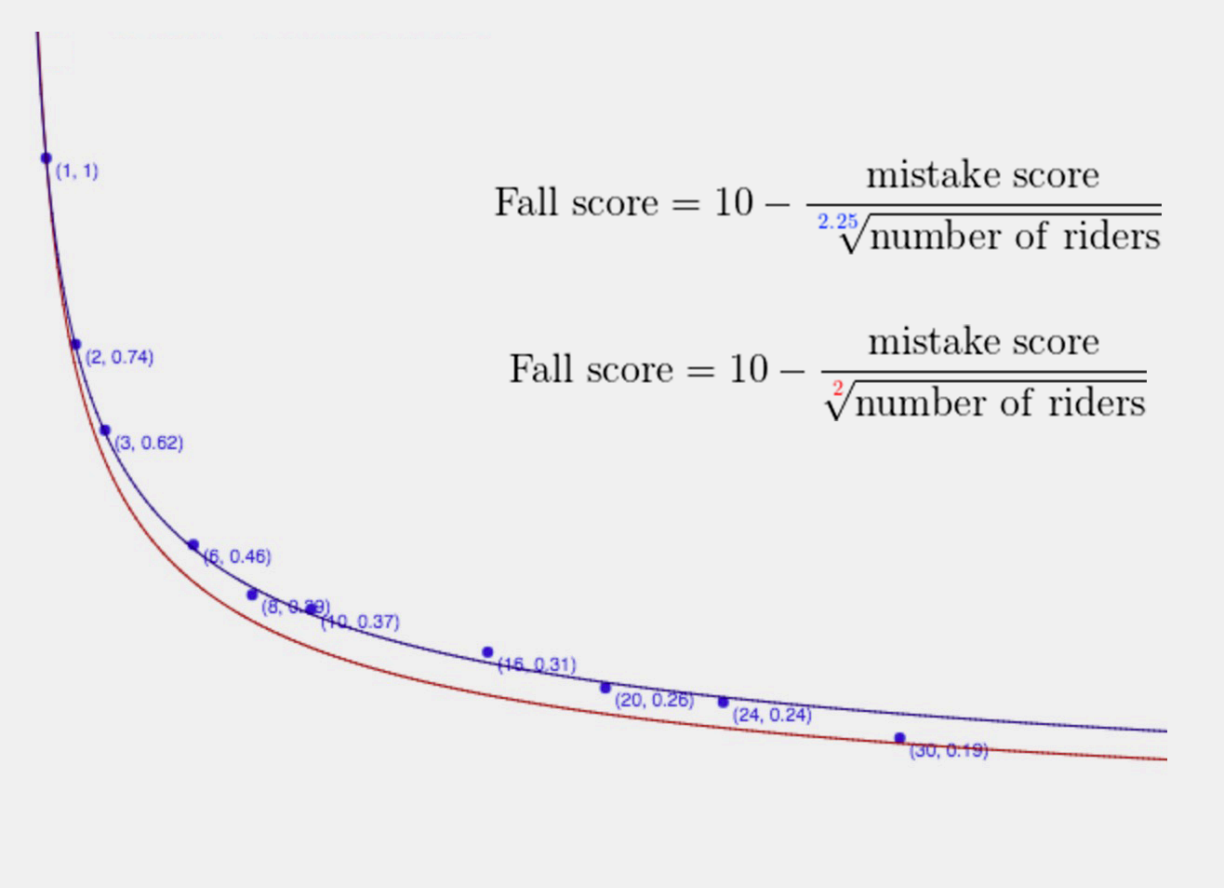
So what do we do?

If a linear scale factor is wrong, we need to start looking at other scale factors. The basic idea is that as the number of riders increases, so should the number of "allowable" falls, but that second number should increase at a slower rate when compared to the number of riders. (Phrased another way: as the group size increases, the number of "allowable" falls should also increase but shouldn't increase by as much.) The best way to figure out a better scale factor is to use data from past group routines. Using data from past events in UDA (the judging software used at Unicon), along with a extensive natural understanding of the way freestyle scores are calculated, Patricia and I were able to create a list of sample scores for different group sizes. This allowed me to use my experience in math to test different scale factors (along with some help from several smart friends of mine). This means that I could quickly test different scale factors against the mistake scores of real past routines to see if it worked as it should. And I think I found a solution!

The solution:

$$\text{Fall score} = 10 - \frac{\text{mistake score}}{\sqrt{\text{number of riders}}}$$

Through my testing, this formula above was the best solution considering that the new scale factor should be both simple and accurate. Let's look at what this means with our first example: Individual routine with 1 major and 2 minors = fall score of 8. For a large group with 16 riders we take our scale factor (√16=4), thus a routine with 4 major and 8 minor falls would give a fall score of 8. This is quite reasonable! For a group routine of that size, 4 major and 8 minor falls is a good, but not perfect result, thus giving a fall score of 8! My example here is simple and not precise, but I did exactly this sort of test over 40 different results and group sizes, and this equation fits very well, as can be seen in this graph:



Note: the equation for the red line is exactly the same as the one above.

I did find other much more complicated equations (like the blue line above), which while correct and slightly more accurate to my test scenario would have be very difficult to use in the real world.

Thus I suggest that we update the calculation for group freestyle fall scores to:

$$\text{Fall score} = 10 - \frac{\text{mistake score}}{\sqrt{n}}$$

Scott
IUF Rulebook Director
Bachelor in Mathematics