

Comparing Disc Golf Player Ratings of F to M

By Steve West Disc Golf, LLC

October 25, 2024

I tried to find whether ratings for the “F” divisions were different than ratings for the “M” divisions.

The data was the results from most of the 2023 and 2024 A-tiers and above, plus some bigger B-tiers. I selected the events where both F and M divisions played, and which had 20 or more rounds of data. This left 758 course-events. Every player who played in an F division during any of the selected events was included in F; even for the events where they were playing an M division.

One confounding factor might be that F players in general have lower player ratings than M players. To diminish any potential effects from this, I used only players who were within the same range of ratings. For most course-events, this meant using everyone from the F player with a rating just below the lowest-rated M player, up to the M player with a rating just above the highest-rated F player. This left 12,917 F player-rounds and 51,805 M player-rounds of data.

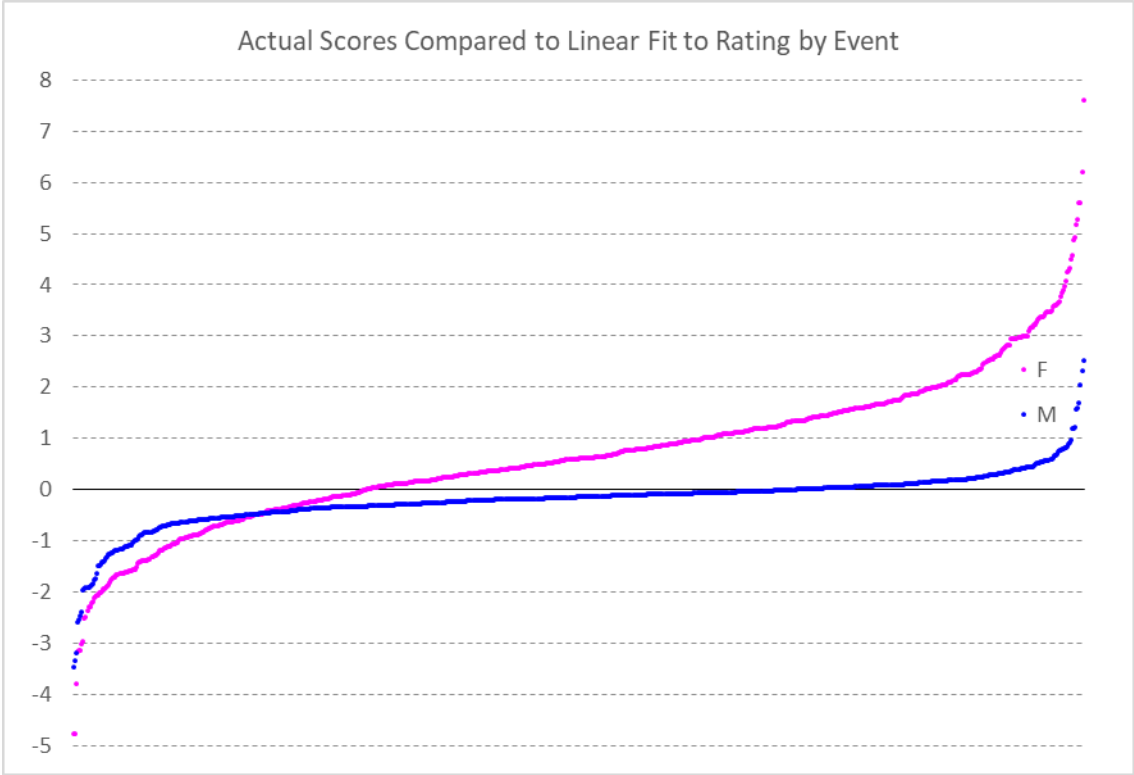
For each course-event, I used all the F and all the M data together to fit a line for each player’s total score on each round as a function of their player rating. (Also, for ratings as a function of scores.) If F ratings are equivalent to M ratings, the fit should be just as good for Fs and Ms.

Scores for F players averaged 0.55 throws per round **worse** than what would be expected based on their player rating. Or, player ratings for F players were 6.85 points **above** what would be expected based on their scores.

Conversely, scores for M players averaged 0.14 throws per round **better** than what would be expected based on their player rating. Or, player ratings for M players were 1.71 points **below** what would be expected based on their scores.

To eliminate any other possible confounding factors, I ran the same analysis multiple times, but with the players assigned to F or M randomly – in the actual proportions for each event. For these tests, the scores and ratings averaged almost exactly what would be expected. The results for F and M separately are over 15 standard deviations away from what could happen if there were no differences between F and M ratings.

The majority, 71%, of events followed the pattern of high scores for F and low scores for M. The following chart shows how far off the linear fit was for each event, sorted.



Conclusion:

There appears to be a significant tendency for F scores to be worse (and M scores to be better) than the player ratings would lead one to expect. However, this will not always be the case for any single event.