Squeezing More Information out of Disc Golf Scores

By Steve West Disc Golf, LLC

December 28, 2019

Basic Idea

One of the limitations of disc golf scores is that they are limited to just a few small integers. This makes it difficult to quantify the quality of any performance.

However, we know that not all of these little integers are created equal. The quality of play it takes to get a 2 is much higher than what it takes to get a 3, while the quality of play it takes to get a 6 is not much different than what it takes to get a 7.

For example, think of the distance of throw it takes to get to the target for a putt. On a 420-foot hole, it would take two 210-foot-long throws to land by the target to putt for a 3. However, it would take a 420-foot-long throw (twice as long!) to land by the basket to putt for 2.

Compare that to the difference between 7 and 6. You can get a 7 with 70-foot-long throws, and you only need to throw 84 feet (20% longer) to get a 6.

And yet, with regular scoring, the difference between a score of 2 vs 3 is the same as the difference between a score of 6 vs 7.

In another way that not all scores are created equal, the worst player would never get the only 2 on a difficult hole, but the best player might get the only 6 on an easy hole. Yet, both players could tie by the worse player getting 6 on the hard hole and 2 on the easy hole, and the best player getting a 2 on the hard hole and a 6 on the easy hole.

Even when holes give out only low or only high scores, there is less difference in skill between 2 vs 3 and 6 vs 7. For example, take a hole which gives out an equal proportion of 2s and 3s. Only the top 71% of throws are good enough for the player to get a 2. (71% x 71% = 50%.) If we take 100^{th} percentile to be the best, this means players whose throws are below the 29^{th} (1-.71) percentile get the 3.

Compare that to a hole that gives an equal proportion of 6s and 7s. The top 89% of throws would be good enough to get a 6 (because 89%x 89%x 89%x 89%x 89%x 89%=50%.) Or, players whose throws are below the 11^{th} (1-.89) percentile get the 7.

Staying out of the bottom 29 percent of throws is a higher skill level than staying out of the bottom 11 percent of throws. Yet, the scores don't recognize this because both holes offer a 50/50 chance at the one-throw-better score.

Information from Beyond the Scores

As hinted at above, in addition to the scores we have another source of information: the number of players who got each score on each hole. Given that information, we can calculate the percentile of throws that were good enough to get each score on each hole. These percentiles reflect the amount of skill the player exhibited.

Here is an example from USDGC 2019, round 1, the 759-foot hole #4. There were only five 2s on that hole out of 111 players. That's 4.5%, meaning that these players needed to make two throws in a row that were somewhere in the top 21% of all throws. (21% x 21% = 4.5%.) Or, two throws in a row which are both 79th percentile or better. That's extremely skillful.

On the other hand, the 319-foot hole #16 had fifty-one 2s, which is 46%, which means the top 68% of throws were good enough to get a 2, or only the 32nd percentile.

Obviously, getting a 2 on a 759-foot hole take more skill than getting a 2 on a 319-foot hole, and just as obviously 79th percentile throws are better than 32nd percentile. So, we're getting numbers (79th and 32nd) that are more indicative of skill than the simple score of 2 on both holes.

We can refine it. On hole #4, some players who got a 2 might have made 100th percentile throws while others might have made barely-good-enough-to-get-a-2 throws at the 79th percentile. So, our best estimate is that the throws for these five players were at the 89.5th ((100th+79th)/2) percentile. Similarly, on hole #16, our best guess is that the fifty-one players who got a 2 were at the 68th percentile ((100th+32nd)/2).

On the 759-foot hole 4, sixty-seven players got a 3 or better, (the five 2s, plus sixty-two 3s) so 60% of players got a 3 or better which means only the top 85% of throws were good enough to get a 3 or better (85%x85%x85%=60%). We can say that these players made throws which were somewhere between the 79th percentile (capped at the lowest percentile which could get a 2) and the 15th (1-.85) percentile. Best guess is that the throws were at the 47th percentile ((79th+15th)/2).

For holes #4 and #16, we get the following results. Players who got the score in the left column made throws that were better than these percent of throws.

<u>Percentiles by Score</u>

Score	Hole 4	Hole 16
1	100%	100%
2	89%	66%
3	47%	19%
4	10%	3%
5	2%	0%

(No one actually got aces, and no one got a 5 on #16.)

Percentile-Based Pseudo-Scores

Using a linear fit of the actual scores to the percentiles, we can get Percentile-Scores which can be thought of in the same way as actual scores. We want the resulting Percentile-Scores will be about the same size as actual scores and have the same average across all players.

Here are the Percentile-Scores for the example holes.

Score	#4	#16
1	NA	NA
2	1.61	1.95
3	3.04	3.23
4	4.31	3.66
5	4.56	3.74
6	4.63	NA

These show the value of getting each actual score, where the value reflects the percentile level of throws it took to get that score.

For example, getting a 3 on hole #4 takes a middling quality of play (47th percentile from the table on the previous page), so it would be fair to add about 3 to a player's score (3.04, to be exact.) However, getting a 3 on hole #16 is not so good, just 19th percentile, so the player should have more (3.23) added to the score for playing at a lower level.

Getting a 2 on hole #16 is not that hard relative to other holes, but getting a 2 on just about any hole is hard enough that it would be more fair if the player could add less than two full two throws to the score.

Here are the Percentile-Scores for all the holes at USDGC 2019, round 1.

	# 1	# 2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	#10	#11	#12	#13	#14	#15	#16	#17	#18
Percentile-Scores	389	444	738	299	1025	371	284	697	685	549	734	901	888	413	539	391	254	647
1	1.03																0.08	
2	1.93			1.61		1.92	1.93			1.63				1.86		1.95	1.32	
3	3.23	2.87	2.57	3.04		3.26	3.23	2.78	2.66	3.15	2.45	2.48	2.16	3.24	2.83	3.23	3.68	2.73
4	3.77	4.17	4.15	4.31	3.86	3.83	3.76	4.16	4.49	4.25	4.38	4.40	3.90	3.98	4.30	3.66	4.97	4.48
5	3.89	4.70	5.07	4.56	5.46	3.97	3.86	4.83	5.22	4.46	5.33	5.30	5.46	4.18	4.89	3.74	5.24	5.19
6		4.77	5.21	4.63	5.88	3.97		4.93	5.40	4.50	5.52	5.43	6.00	4.20	5.04		5.38	5.45
7					6.00				5.45	4.52	5.57	5.48	6.14				5.42	5.51
8					6.01				5.47		5.60	5.49	6.18				5.47	5.53
9					6.02				5.48			5.50					5.50	
10												5.50					5.50	
11												5.51					5.50	

Substituting Percentile-Scores for actual scores, we can get an alternative ranking of players.

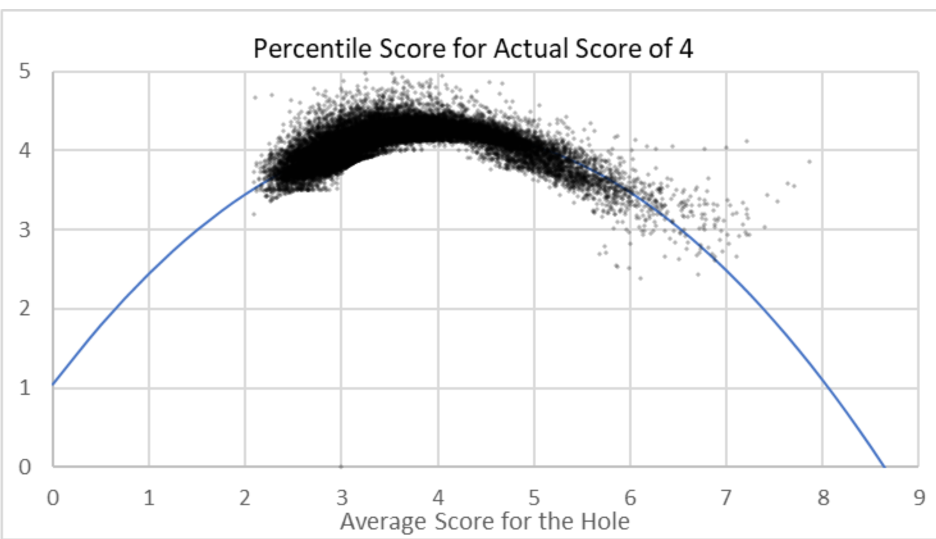
The total of the Percentile-Scores has a 95.2% correlation with actual scores. However, the "winner" of the Percentile-Scores was Jeremy Koling, not James Conrad. The reason is that actual scores over-punished Jeremy for being just a little worse than James on the holes where they both did badly.

For example, on hole #18 round 1, James got a 6 and Jeremy got a 7. James's quality of throws on that hole was poorer than 98% of throws, while Jeremy's was just a little worse, being poorer than 99.4%. Not much difference, but the score Jeremy actually got was 16.6% bigger (7/6). It would have been more representative of the difference in skill to give James a 5.45 and Jeremey a 5.51.

Toward a General Set of Percentile-Scores

The method above determines the Percentile-Scores after the fact. It also produces a unique table of Percentile-Scores for each hole. However, that is not ideal for a game setting. Players would prefer to know the value of each actual score beforehand. Players would not want to deal with so many different tables of values.

I calculated the Percentile-Scores for more than 27,000+ distributions of actual scores to see if a general pattern could be applied to all holes. For example, here is the Percentile-Score for getting an actual score of 4 for all 27,000+ holes. The x-axis is the average score for the hole, the y-axis is the Percentile-Score.



Note that the Percentile-Score (the cost) of getting a 4 is highest on holes where the average score is around 4.00. This makes sense when we think about the amount of skill it takes to move between scores. Keep in mind that the difference in skill between two low scores is bigger than the difference in skill between getting two high scores.

> Low scores should be rewarded more, so a low score like a 4 on a hole that averages 6.00 should not cost the player 4 full throws.
>
> Higher scores can happen with very little change in the skill demonstrated on a hole. So, getting a high score like a 4 on a hole with an average of 2.00 should not cost a full throw more than getting a 3. Since the cost of high scores steps up by less than a full throw per throw of actual scores, the cost of higher scores will be less than the actual score.
>
> The cost of getting a 4 on a hole that average 4.00 should be about 4. It needs to be a little bigger so that the total of the Percentile-Scores adds up to be the same as the total of the actual scores. (Because the cost of a 4 on holes with higher-than-4 and lower-than-4 average scores is lower than 4.)

The blue line is the best-fit parabola. Note that it starts at a value of 1 for a theoretical hole that averages 0.00, has a maximum when the average score is near the score being evaluated, and returns to a value of 1 when the average score is about twice the score being evaluated. This pattern works for all the other scores besides 4 as well. Using this framework, only one parameter needed to be adjusted to fit a model to the Percentile-Scores for all distributions. I grouped the holes by par and came up with the following values:

<u>Percentile-Scores (Cost) of getting each Actual Score, by Par</u>

Actual	Par 2	Par 3	Par 4	Par 5	Par 6	Par 7
1	0.57	-0.45	-3.08	-6.87	-12.06	-17.81
2	1.98	1.73	0.77	-0.86	-3.29	-6.10
3	3.09	3.17	2.82	1.83	0.12	-2.01
4	3.73	4.07	4.23	3.87	2.92	1.57
5	4.16	4.68	5.21	5.29	4.90	4.12
6	4.42	5.06	5.86	6.27	6.28	5.93
7	4.56	5.28	6.26	6.90	7.22	7.20
8	4.66	5.44	6.55	7.37	7.92	8.14
9	4.72	5.54	6.75	7.71	8.43	8.84
10	4.81	5.68	6.99	8.06	8.93	9.50

(For reference, the average scores for each par were 2.41, 3.07, 4.22, 5.36, 6.55, 7.63)

That is still too complicated, so I took the above percentiles scores, subtracted par and took the average (weighted by number of holes that had each par; mostly par 3s and hardly any par 7s).

The result is the following Percentile-Scores for each actual score relative to par.

Score minus Par	Percentile-Score
-3	-7.83
-2	-3.37
-1	-1.24
0	0.19
1	1.12
2	1.74
3	2.14

By adding the par for the hole to the values above, players would get pseudo-scores which would add up to about the same total over a round as their actual scores.

Scientific Stableford

Another use for the table above would be to create a more scientific version of Stableford Scoring. Under Stableford scoring, the player tries to get as many points as possible. The player gets zero points for double bogey, one point for bogey, two for par, three for birdie, etc.

For Scientific Stableford, the points vary according to how rare a score is.

Score minus Par	Points
-3	9.97
-2	5.51
-1	3.38
0	1.95
1	1.02
2	0.40
3	0.00

Note that when rounded to the nearest integer, this matches Stableford scoring for scores of 2 over to 1 under par. So, Stableford is already pretty scientific and fairer than adding on all those extra points for bad scores. Eagles and Albatrosses should be rewarded more, but those don't happen often. Players also should get a part of a point for a double bogey.

To give players nearly full credit and yet keep the arithmetic not very complicated, I would give one-half point for double bogey, 1 point for bogey, 2 for par, 3.5 for birdie, 5.5 for eagle, and 10 for albatross.

Speculation about a Future Refinement

The percentiles give us a *ranking* of some underlying quality of play, but they are not a direct *measure* of that quality of play.

Whatever underlying quality of play we are measuring, if it is like everything else in the world the very best is many times as good as the nearly best, the very worst is much worse than the nearly worst, and the quality in the middle is nearly linear with the ranking.

A 450-foot throw that is parked for a drop-in 2 is many times as good as a 350-foot throw that is at the side of the fairway. Both because it is much rarer, but also because it takes hundreds of times as much practice to accomplish it. Yet, if the 350-foot throw is 50^{th} percentile, the 450 foot can't be more than 50 percentile points higher.

I don't know what quality we are measuring, nor do I know the distribution of that quality by percentile, but it feels like it might follow a lognormal distribution. With lognormal, the top end of quality is virtually unlimited (in case someone aces a 1250-foot hole or something). Yet, no throw can ever be worse than dropping the disc at the player's feet. (Note that a throw that missed an island and results in a re-throw with penalty is actually two throws in a row that make no advancement.)

So, if we pretend that we know that the quality of throws follows a lognormal distribution with a standard deviation of 0.5, we can see what it might look like if we actually could ascertain the quality of throws which underlies each score.

	#4	#16
1	NA	NA
2	1.87	1.23
3	0.96	0.64
4	0.52	0.39
5	0.36	NA

This way, we could say that a 2 on #4 took almost twice the quality of play as a 3 on #4. Or that a 5 on #4 took about the same quality of play as a 4 on #16. We could say that a 2 on #4 took almost five times the quality of play as a 4 on #16, but that might be pushing it.

We could also produce Quality-Scores by fitting the Quality of throws to the actual scores. These would work a lot like Percentile-Scores, but reward rare good scores a lot more, and punish rare bad scores a little more.

Here is what the Quality-Scores would be:

Quality-Scores	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	-1.7																-4.6	
2	2.20			1.12		1.91	1.92			1.50				1.84		1.95	1.92	
3	3.14	2.86	2.53	3.15		3.24	3.21	2.78	2.60	3.22	2.38	2.39	1.33	3.25	2.82	3.19	3.75	2.70
4	3.58	4.14	4.16	4.15	3.82	3.85	3.79	4.14	4.45	4.13	4.36	4.37	3.99	3.94	4.27	3.73	4.58	4.45
5	3.92	4.77	5.01	4.51	5.39	4.26	4.17	4.84	5.21	4.47	5.28	5.30	5.35	4.36	4.89	4.1	4.83	5.15
6		5.07	5.33	4.80	5.93	4.36		5.18	5.53	4.62	5.61	5.59	6.00	4.57	5.29		5.04	5.60
7					6.28				5.74	4.80	5.79	5.76	6.33				5.12	5.83
8					6.42				5.94		5.99	5.89	6.55				5.26	6.03
9					6.56				6.06			5.97					5.47	
10												6.02					5.48	
11												6.09					5.55	

Note the negative numbers. Very good scores can actually reduce a player's total Quality-Score (which is a good thing, this is golf).

While these give a more satisfying set of Pseudo-scores than we get from Percentile-Scores, they are based on pure speculation about how the quality of throws at the very top and bottom compare to the average throws. Since rare scores don't affect much of anything (because they don't happen often), for now it is best to stick to the actual measurements of Percentile-Scores.

For trivia purposes, based on Quality-Scores Nikko Locastro would have won.