Predicting the Popularity of Disc Golf Courses
March 23, 2011
www.stevewestdiscgolf.com

Thanks to Tim Gostovic at Disc Golf Course Review, I was able to obtain data about the popularity of disc golf courses. Using this, I updated my formula.

Activity

I obtained counts of the activities from DGCR: Reviews 27,430, Favorited 20,854, Played 166,556, Wishlisted 49,373, Rounds 89,078, Review Votes 145,034.

I chose not to use Wishlisted because that is what people who haven't gone to the course do.

I also did not use Rounds, because it appears that the use of this field is spotty. A few people use it intensely, so some courses are "loaded up" with Rounds. It did not correlate well with any of the other measures.

To obtain a single index of Activity, I summed the proportion of each activity attributable to each course. So, if a course had 1 Review, it would add $1/27,430^{th}$ to the Activity Index.

I focused on U.S. Courses only, because those are the only ones I have detailed enough population data to work with, and also excluded Temporary courses and Practice Areas.

Fitting Parameters

Nearby Population and Other Courses

I stayed with my basic concept that having more people living closer to a course makes it more popular. Also, that competing courses nearby can take some players away.

However, I did tweak the formulas. The measure of how popularity is affected by distance is now COS(ATAN((Miles/16.18034))^1.618034. This gives courses a bigger drawing area than the old formula. The old formula had the value of a course go down by half at a range of about 9 miles, while the new formula extends that out to about 18 miles.

The parameters were set after many rounds of going back and forth between setting the function for value based on distance, and the value of individual courses (based on number of holes, etc.) and the impact of nearby courses. What I found was that these parameters work as well as any, and the results are not terribly sensitive to the particular parameters chosen.

For the impact of nearby courses, I stayed with the concept that the highest value course would get the most visits from a particular player. However, I found only weak evidence of this, so I toned down my formula so that the second-highest value course would not be impacted as much by competing with a better course nearby. The rationale is that the two courses combined will create some synergy by bringing traveling players in from a longer distance, and not just steal each other's local players.

The new formula is $(Rank^{0.618034} - (Rank-1)^{0.618034})$. Or, the second highest value course will get 53.4% as much business as if it were the highest ranked course (compared to 41.4% under the 2009 formula)

Also, the exponent of 0.618034 was the best fit for determining how the number of holes affected popularity. By using it in the formula for the impact of nearby courses, the formula is neutral about whether additional holes are part of the "same" course or not. So, for example, the total popularity of a four-course complex will come out the same whether it is called four 18-hole courses, or a single 72-hole course.

Course Features

With the formula for the effect of nearby population and nearby courses settled, I solved for the Perfect Course Feature Factors which would recreate the Activity of the DGCR site. Each course was assigned a Course Feature Factor which, when run through the formula for nearby population and nearby courses, would recreate the Activity on the DGCR site for that course.

This took several iterations, as the popularity of a course is non-linear with its Course Feature Factor, and is also affected by the Course Feature Factors of other courses.

Then, I fit factors for each type of Course Feature to the Perfect Course Feature Factors. The characteristics of the course, and their factors are in the following table:

Characteristic	Factor
Public	100.00%
Private	83.59%
Free	100.00%
Pay to Play	105.99%
Single Tees	100.00%
Multiple Tees	110.61%
Concrete Tees	100.00%
Other Tees	84.00%
Rubber Tees	106.88%
Other Targets	86.33%
Manufactured Baskets	100.00%

Somewhat surprisingly, the overall Rating of a course was not important. It turns out that the Rating of the course is so dependant on the number of holes, tee pad type, etc. that the Rating itself does not add significant information. This is fortunate, because it means we do not need to guess the Rating of a proposed course to determine its popularity.

The most important characteristic of a course is the number of holes. The formula for that factor is $(Holes/18)^{0.618034}$. Sample Values:

Holes	Factor
3	33.04%
6	50.71%
9	65.16%
12	77.83%
15	89.34%
18	100.00%
21	110.00%
24	119.46%
27	128.48%
30	137.12%
33	145.44%
36	153.48%
39	161.26%
45	176.17%
48	183.34%
54	197.19%
63	216.90%
72	235.56%
119	321.33%

All these factors for all Features are multiplied together, and then the final Course Feature Factor is (1.359*Product) - 0.2869, but not less than zero.
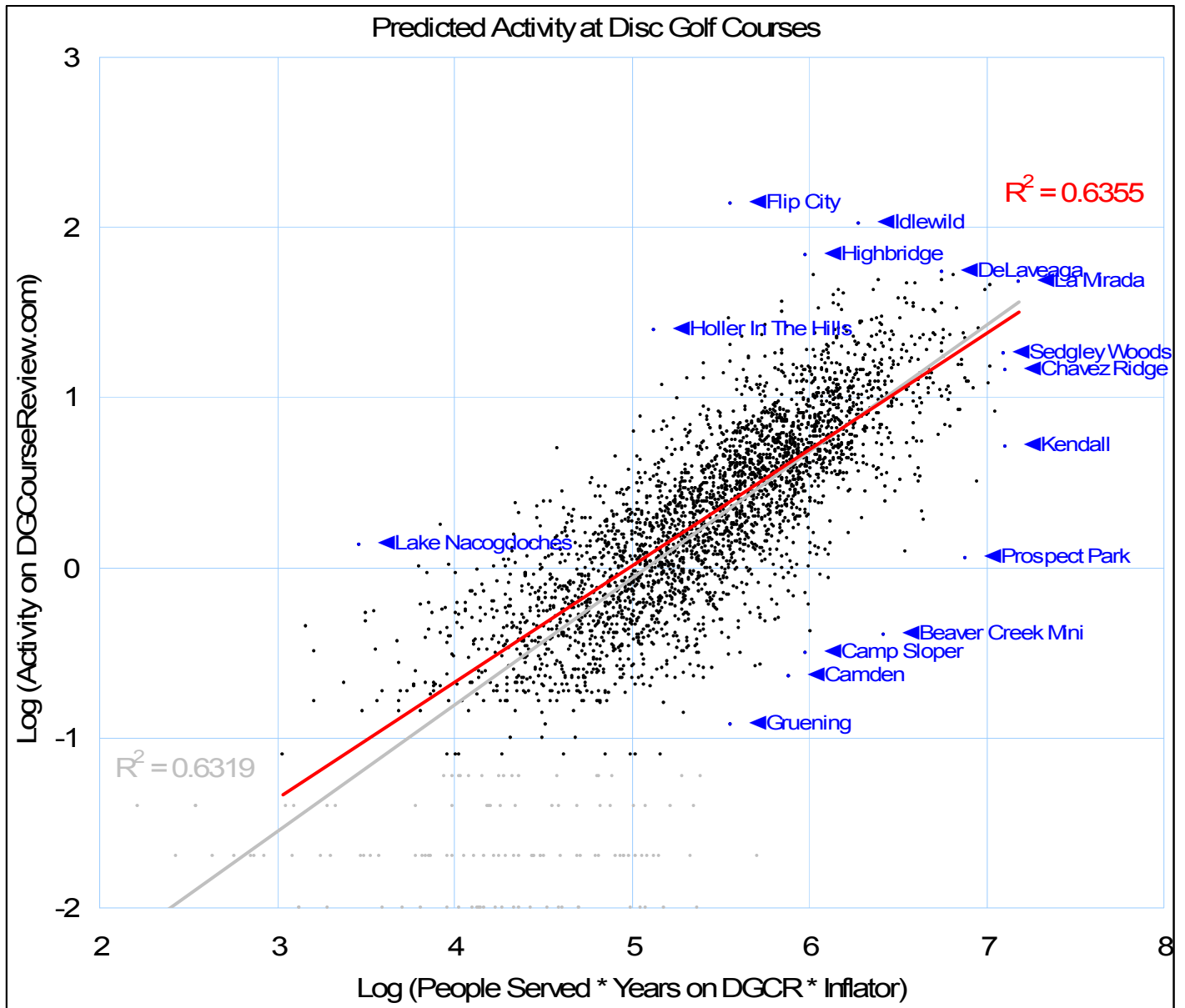
I found no evidence that supported the need for an Uncrowded Factor, so I eliminated it. This will save a lot of computing time, as the old Uncrowded Factor depended on how popular a course was, which depended on its Uncrowded Factor, and the Uncrowded factors of nearby courses.

I also found that the number of people a course serves is better related to the absolute (rather than the percentage) change in Service Levels it causes.

Tests of Fit

On the next page is a chart showing the Predicted Activity based on the formula, compared the actual Activity on DGCR.

Note that the scale is in $\log_{10}$. The busiest courses have over one hundred times as much activity as the median.

Predicted Activity at Disc Golf Courses

$R^2 = 0.6355$

$R^2 = 0.6319$

Labeled points: Flip City, Idlewild, Highbridge, DeLaveaga, La Mrada, Holler In The Hills, Sedgley Woods, Chavez Ridge, Kendall, Lake Nacogdoches, Prospect Park, Beaver Creek Mni, Camp Sloper, Camden, Gruening

X-axis: Log (People Served * Years on DGCR * Inflator)

Y-axis: Log (Activity on DGCourseReview.com)

A few outliers are labeled. The least active courses are in light gray. The Red line is the fit to those courses that have a decent level of activity.

While the fit is not great, I would not expect it to be. The activity on DGCR can be influenced by many things that do not reflect actual play at the course.

For example, it is widely thought that there is a concerted effort to pump up Flip City's ratings by recruiting people to post favorable reviews. If so, that would explain why its activity is so high compared to the number of people the course would be expected to serve.

It would be expected that the website might be more popular in certain areas independent of the level of disc golf in that area. It seems as if the course is more popular in areas where there are many disc golf courses a few hours drive away.

However, the most important use of the formula will be to locate new courses within a city-sized area, so these regional variations shouldn't come into play.

For example, below is a chart comparing the popularity of courses in the Twin Cities area, based on DGCR Activity and the formula. It shows that the formula can at least tell the difference between popular and unpopular courses, and while it doesn't get every course exact, it doesn't seem to have a built-in bias.