

Le Tour de France
Is it easier now than in the past?

Patrick Rhodes

November 7, 2011



The cyclists are staring straight ahead into the abyss - the '10,000 mile stare' that arrives at the end of the second week of racing the Tour de France (henceforth noted as TdF). Their legs are wooden from having already pushed the pedals over a quarter of a million rotations. Riders' hearts are weary, having beat an additional half a million times more than otherwise. Like Roman chariot drivers, team managers flog their racers over radios, pushing them to go on, "Attack! Go to the front! Faster!" Hunched low over the handlebars, sweat cascading down their gaunt faces, their bodies strain under the broiling French summer sun. Several torturous hours go by and today's barbarous stage mercifully ends. The gladiators shuffle to the dinner table, hands and arms a blur as they supply their depleted bodies the minimum of 9000 calories required to sustain this almost inhuman level of activity.

And yet there is another week of racing to come.

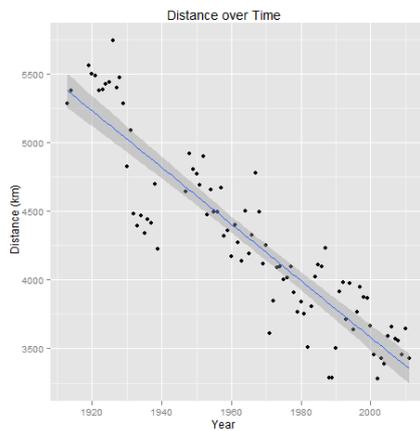


Figure 1: The TdF has shortened the distance over the years

Length-wise, the latter editions of the TdF are shorter than early editions (see figure 1). In and of itself this would lead one to believe that the race has gotten easier over time.

Additionally, back then there were no air-conditioned team buses, no advanced nutritional knowledge, no bikes that could be lifted with one finger and no team car outfitted with a team mechanic for immediate repair. Not only did early TdF racers ride longer distances, but they had to carry all of their supplies strapped to their bodies.



Figure 2: Suffering...

Does that really make it harder? Let's find out using descriptive statistics. Due to the availability of data, let's use the percentage of riders who finished the race as our criterion for a race's difficulty. It is one of the few measurements that was taken for all TdF occurrences.

Figure 3 reveals the percentage of finishers by distance for each year of the TdF. It is plainly visible that a correlation exists between these two variables as evidenced by the regression line. Exactly how strong is it? The *linear correlation coefficient* (r) shows a rather strong relationship (anything close to -1 or 1), denoting that as the distance shortens, the percentage of riders goes up:

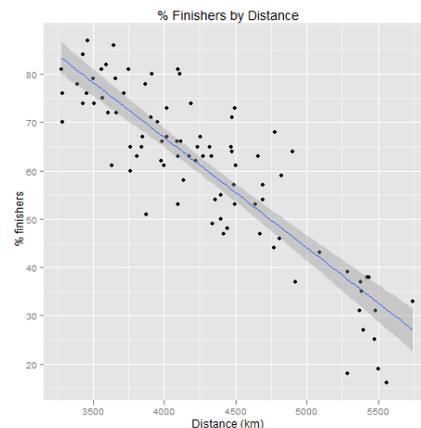


Figure 3: Linear relationship between distance and percentage of finishers

$$r = -0.87$$

Another way to calculate the strength of the correlation is to use the *coefficient of determination* (r^2), which in this case shows that roughly 75% of the variation of finishing riders can be explained by the overall distance:

$$r^2 = 0.75$$

Over 100 editions of the TdF have been run to date; it seems likely that the distance will stay in the 3000-3500 km range, so we can reasonably expect between 78% - 90% to finish over the next few years. In the 2011 race, it seemed that there were more crashes than normal. However, when looking at the finishing percentage (84%), we can see that *more* racers finished than expected (80%), given the distance of the race (3,430 km). What made it seem worse was *who* crashed out rather than the actual number (some popular riders did not finish).

For the 2012 race, the race is 3,479 km. Using the calculated regression line formula:

$$y = -0.02280968x + 158.12458$$

$$79 \approx -0.02280968(3479) + 158.12458$$

where x = distance (km) and y = % finishers, we can expect approximately 79% of the entrants to finish the race. Stay tuned for July 2012 to see how close the prediction is!

Linear regression was used to predict a proportion for this article and would not stand up for races that were excessively short or long (2,000 km would predict 110% of participants to finish while 8,000km would predict -24%). But for races between 3,000-5,500km, this model can be useful for analyzing the TdF.