# Is Tiger Woods a Winner?

Scott M. Berry

**Abstract**

Tiger Woods has an amazing record of winning golf tournaments. He has gained the persona of a player that is a winner, a player that when near the lead or in the lead can do whatever it takes to win. In this paper I investigate whether in fact, he is a winner. A mathematical model is created for the ability of Tiger Woods, and all PGA Tour golfers to play 18 holes of tournament golf. The career of Tiger Woods is replayed using the mathematical model for all golfers and the results are very consistent with Tiger Woods' actual career. Therefore the mathematical model, which does not give Woods any additional ability to win, would result in essentially the same career. Woods has not needed any additional ability to win–only his pure golfing ability. The ramifications of this result are that there is no evidence that Woods is in fact a "winner" – but instead he is just a much better golfer than everyone else.

## Introduction

Tiger Woods is one of those rare athletes that accomplish feats in their sport that are freakish. In this small group are guys such as Babe Ruth, Wayne Gretzky, Wilt Chamberlain, Barry Bonds, and Jack Nicklaus. Woods dominates a sport where the population of players are all very good–and very tightly bundled in their ability. To win one tournament, beating 100+ of these players is incredibly difficult. To average one tournament victory a year for 10 years is a Hall of Fame type accomplishment. Tiger Woods has won 71 of 253 (28.1%) official PGA Tour tournaments around the world in 14 years on the PGA Tour. The PGA Tour record is 82 career wins by Sam Snead in 30 years on the PGA Tour. In 2009 Woods played 17 events and won 6 of them, finishing in the top 10 in an astounding 14 of them. Despite the 6 wins in 17 events, some classify his season as disappointing because he did not win any of the four tournaments that are labeled "majors." The Masters, United States Open, The Open Championship, and the PGA Championship are the four so called majors in golf. These tournaments have extra importance in professional golf. They are also different than typical tournaments because they typically provide the best fields–the quality and depth of the players participating. Woods has won 14 career majors in 49 played as a professional. The record is 18, held by Jack Nicklaus, who competed in 163 majors (he also finished second 19 times). Woods holds the record for the largest margin of victory in each major–12 in the Masters, 15 in the US Open, 8 in The Open Championship, and 5 in the PGA Championship.

Tiger has accomplished a freakish amount despite being only 33 years old. Professional golfers are typically competitive in to their mid forties (Tom Watson nearly won The Open Championship in 2009 at the age of 59!). Barring any catastrophic injuries or strange happenings, Woods should shatter all career win records, including most tour victories and most major wins. Is it then ludicrous to ask the question whether Tiger Woods is a "winner?" First, let me define what I mean by a "winner." I think Tiger Woods is the greatest golfer, if not the greatest athletic performer of all-time. But, by winner, I want to know if he has an ability to win golf tournaments beyond his skill level. Woods has developed the persona of a champion, that his presence on the leaderboard – being near the lead late in the tournament–causes other players to perform worse. Thus, Tiger has an innate ability to win. Nicklaus, with whom Woods is most compared, won 18 majors and finished second 19 times. Woods has won 14 majors, but finished second "only" 6 times. The thought is that if Woods were not a winner, he should have a distribution of firsts and seconds similar to Nicklaus.

Many commentators claim that Woods' best attribute is his "mind." That he is able to do whatever it takes to win. If it is true, that Woods has an innate ability to win golf tournaments, then he should win more than his raw ability would predict. There are many different "effects" that are explored in sports, including the so called "hot hand effect" or the ability to hit in the clutch. In this paper I explore whether Woods wins more than his raw golfing ability would determine he wins–whether Tiger Woods is a "winner."

In this paper I create a mathematical model for Tiger Woods. In this process I also create a mathematical model for all professional golfers. In this model each player will have an intrinsic ability to play golf. In this model Woods–I'll refer to him as "RoboTiger"–will not have the ability to win beyond his golfing skill. Each player will record a score on each round, independently of what every other player is scoring. RoboTiger will not perform differently when he is near the lead, and likewise none of his competitors will perform worse when RoboTiger is near the lead. RoboTiger will not be a winner. The mathematical model of Woods will rank him as the best player–therefore RoboTiger will win tournaments, but he will not be an innate winner beyond his mathematical golfing skills. I simulate ten thousand replays of Woods' career, using RoboTiger and all the mathematical golfers to compare what RoboTiger does relative to Woods' actual career record. Should RoboTiger win less than the real Tiger then this is evidence that the real Tiger may be an innate "winner."

## Data Availability and the Mathematical Modeling

Every year there are approximately 160 players that qualify for full tour status on the Professional Golfers of America (PGA) Tour. There are several hundred more that play less

tournaments with partial tour status. A number of foreign players will play a small number of PGA events in addition to their native professional tours. Each season consists of approximately 45 official PGA Tour tournaments. For many of the important tournaments the best players from around the world participate. For some of the perceived lesser events many of the top players do not play. A typical tournament will have 144 players for the first two rounds, then a cut occurs where the field is narrowed to the top 60 players. I downloaded the round-by-round scores for each of the tournaments on the PGA Tour for 1997 (Woods' first full season) through 2004. Over this eight-year period there were 2004 different players participating in the 352 tournaments (match play and Stableford scoring tournaments were not included). There were $147,154$ rounds played, with $10,509,253$ shots taken. The best scoring average for any player was 68.75 for Chris Downes. Woods had the fourth best scoring average (69.31). Chris Downes is a little known professional that played in one tournament, the 2003 BC Open, where he finished 13-under par and tied for 18th in a weak field. The same weekend Woods played in the Open Championship, finishing fourth, with a cumulative one-over par. Woods' accomplishment that week represented better golf than Downes – but because of differences in playing conditions directly comparing Downes 18-under par to Woods' one-over par is not reasonable. We construct a model that accounts for the relative difficulty of each round played, and also models the whole population of golfers – placing Downes' results in perspective when compared to all players.

I use a normal distribution to model the scores for a particular player in any given round. While the normal distribution certainly is not exactly correct, it does very well at modeling scores. The distribution of scores on any one hole is difficult to model. The 18-hole round score is he sum of 18 random variables–which, based on the central limit theorem is reasonably approximated by the normal distribution (tournaments are typically 72 holes in length). In [3] we used a normal model and found the fit to be quite good, with the residual plot showing a slightly longer right tail and a slightly shorter left tail than the normal distribution suggested. As well the effects of aging are explored. In this paper the effects of aging are ignored.

I assume that each round has an intrinsic difficulty, labeled $\gamma$. The interpretation of $\gamma$ is that it is the incremental difference from the average round on the PGA Tour. I assume that each golfer has a raw ability, $\theta$, which is his average score in an average PGA Tour round ($\gamma = 0$). Thus, when a player with intrinsic ability $\theta$ plays a round with intrinsic difficulty $\gamma$, the mean score for that player in that round is $\theta + \gamma$. For example, if Chris Downes is playing in the BC Open, the first round difficulty may be $\gamma = -4$. If Downes has a $\theta = 73$, on average, he would shoot a 69. If he were playing The Open Championship the round difficulty might be $\gamma = 3$, which would mean that on average he would shoot 76. In other words, a 69 at the BC Open would be *equivalent* to a 76 at The Open Championship. Therefore, the mathematical model

for the round score for player $i$, in round $j$, $Y_{ij}$, is

$$Y_{ij} \sim N\left(\theta_i + \gamma_j, \sigma^2\right),$$

where $N()$ represents the normal distribution. The parameter $\theta_i$ is player $i$'s intrinsic mean score, and $\gamma_j$ is the difficulty of round $j$. The standard deviation–the variation of professional golfers–is $\sigma$.

The second important aspect I need to capture is referred to as regression-to-the-mean. Let's assume all rounds are equivalent in difficulty–I still wouldn't think Chris Downes was better than Tiger Woods. Why? Because it is more likely that Downes played better than his true ability for those four rounds than that he is the best player on the planet and played average for the week. This notion is captured through a hierarchical model. In a hierarchical model a golfer's intrinsic abilities are modeled with a prior distribution. This distribution helps to statistically understand each player's performance in the context of every other player. I assume that the distribution of all the $\theta$'s on the PGA Tour is a normal distribution with a mean of $\mu$ and a standard deviation of $\tau$,

$$\theta_i \sim N\left(\mu, \tau^2\right).$$

The standard deviation $\tau$ captures the amount of variation in true abilities for players on the PGA Tour. A similar model is used in [2], including an exploration of the serial correlation of rounds.

A Bayesian approach to fitting the above model is selected. In this approach a prior distribution is selected for each of the parameters. Flat prior distributions are selected for each of the round parameters, $\gamma_j$, and for the standard deviations $\tau$ and $\sigma$. I utilize standard Markov chain Monte Carlo techniques to calculate the posterior distribution for each of the 1408 $\gamma$'s 2004 $\theta$'s, and the two standard deviations. The calculation techniques are not important for understanding this paper, but I do want to elaborate on one aspect of the model.

How do we estimate the difficulty of a round when we don't know the ability of the players playing the round–and conversely how do we estimate the ability of the players when we don't know the difficulty of the rounds? The model employed here is best described in an iterative fashion. First the ability of each of the players is estimated assuming all rounds are equivalent in difficulty. Using these initial estimates of player ability, initial estimates of the difficulty of each round can be made. For example, if based on initial estimates of the player's abilities, golfers averaged 2 shots higher in one round than their ability projects, then that round's difficulty is estimated to be 2. Based on the new estimates of each round's difficulty, better estimates of the player's abilities are made. This iterating of estimating the round difficulty and the player ability continues. The difficulty of each round is then directly, and simultaneously, estimated

with the intrinsic abilities of the players. Simultaneously the variation of golf scores ($\sigma$) and the variation of professional players ($\tau$) are also estimated.

The posterior mean and standard deviation of the mean score on an average PGA Tour round is reported in Table 1. Not surprisingly, Tiger Woods is estimated to be the best player. He is estimated to be 0.85 shots better than Vijay Singh and 0.96 better than Ernie Els. They are the only two players within one shot of Woods for an 18-hole round score. The second best player is then 3.40 shots behind Woods on average for a four-round tournament. The posterior mean standard error of an 18-hole round score is 2.81 (standard deviation of 0.005). For a four-round tournament the standard error of the total score is 5.60. The second best player is 0.61 standard errors away from Woods. There are only nine players within one standard deviation of Woods for a four-round tournament. The posterior mean for $\mu$, the mean of the distribution of PGA Tour golfers is 73.84, and the posterior mean for the standard deviation of the distribution of golfers is 2.33. Woods is estimated to be an astonishing 5.36 shots better than the average PGA Tour golfer – in a single 18-hole round. This equates to more than 20 shots better in an 18-hole tournament. In [1] I explore what aspects of his game made him a dominant player early in his career.

Figure 1 shows a histogram for the posterior mean of each of the 2004 players mean score on an average PGA Tour round. The probability that a player of each mean skill level would beat Tiger Woods in a four-round tournament is also plotted (solid line). A huge portion of the players have almost no chance to beat Woods in a tournament. It is only the very best players that have a reasonable chance to beat Woods, and those chances, individually, are small. In a typical tournament Woods has to beat 143 other golfers over four rounds. While the chance that any individual player beats him is minimal, the best of 143 players would be favored. Table 2 shows the probability of Woods winning a tournament with different field strengths. The probability he would win a tournament in which the top 10 players, other than him, would participate, is 0.294. If the top 144 players participate his chances decrease to 0.135. Increasing this further to a hypothetical tournament with all 2004 players, he would win with probability 0.122. Table 2 also reports the probability he would finish second in these tournaments. Typically this probability is about one-third of the chance he would win the tournament. Woods is so much better than the other players that he is the major determination on whether he wins the tournament. If he plays well above his ability, by random variation, he wins the tournament. This corresponds to him winning tournaments by huge margins, as he has done. When he plays poorly one of the 144 other players will beat him. If he plays a little better than average, then he may be beat by another player, and may finish second. Therefore, him finishing second is reasonably unlikely relative to winning the tournament. His 6 actual seconds is also inflated from the model prediction in part because of ties. The modeling of scores with a normal distribution assumes that there are no ties (think of it as all places playing off for the spot). In contrast, Nicklaus

was a very good player, but wasn't as dominant as Woods. Therefore, his relative number of second place finishes was similar to his number of wins. As Woods ages, and invariably declines, his relative dominance will decrease, and thus his chances of finishing second will also increase.

## The Career of RoboTiger

In this section we use the mathematical model for Tiger Woods, "RoboTiger," along with all of the other mathematical constructs of all 2004 players to replay Woods' career. The model only uses the individual 18-hole scores for each round for the players. The model does not know how many wins each player had or what any players had done in important situations. The model is built on the scoring of each player, in each round, relative to what the competition did in those rounds. Using the model for golfers, each of the 352 tournaments are replayed $10,000$ times. Of the 352 tournaments in the eight-year data set, Woods played in 147 of them, winning 36 (24.5%). In his 14-year career he played in 106 tournaments not in this data set, winning 35 of those (33.0%). In resimulating these 147 tournaments, RoboTiger won an average of 42.2 tournaments with a standard deviation of 5.4. His actual number of wins of 36 falls about one standard deviation below the average performance of RoboTiger. The true performance of Woods is quite consistent with the mathematical model of RoboTiger, if anything a little worse. Figure 2 shows a histogram for the number of wins for RoboTiger in the 147 tournaments. For sensitivity I ran the simulations and the modeling assuming each player had a year-specific ability. Some years, such as 2000 and 2001 Woods was more dominant than others. Assuming yearly specific $\theta$'s a resimulation of these 147 resulted in a mean of 40.3 and a standard deviation of 5.2 wins.

In addition to the 147 tournaments, I simulated each of the majors after 2004 in which Woods played (he missed two majors in 2008 with a knee injury). Since these tournaments were not in the eight-year data set, I assume that the field of these tournaments would be similar to each of the respective majors in 2004. Figure 3 reports the results of simulating $10,000$ sets of the 50 majors for RoboTiger. The mean number of major wins for RoboTiger is 14.4 with a standard deviation of 3.2. The modal number of wins for RoboTiger is 14, which is the actual number of majors won by Woods.

Using this same approach for the majors I simulated Woods' career going forward. Assuming Woods plays to the same level for the next 10 years (he'll be 43) and then never wins another major, the mean number of wins is 25.5, with a standard deviation of 2.8, with an 8% chance of 30 or more, and a 99.6% chance of winning at least 19 majors, breaking Jack Nicklaus's record for all-time majors won. If he plays at the same level for the next 15 years he'd win an average of 29.5 majors with a standard deviation of 3.3, and a 100% chance of winning at least 19 majors.

The simulations predict the 2013 US Open at Merion Golf Club to be the most likely major in which Woods wins the record setting 19th major.

The actual career wins for Woods has been very close to the simulated career of RoboTiger. While this certainly does not prove that Woods does not have some special ability to win beyond his skill level, it does demonstrate that his career wins is very consistent with his skill level. The statistical modeling of Woods, and all players on the PGA Tour has the specific characteristic that Woods does not have an innate ability to win, nor do others players have an innate decline when they are near the lead. Woods and all golfers behave as a statistical robot. If this statistical modeling represents the truth for Woods, then the expectation of what would happen for his career is what we have observed. Humans have an incredible ability to create reasons for what we see. Many times there are not reasons for what we see, other than pure randomness. Many have attached an aura around Tiger Woods that he is the ultimate winner, wearing his well-known red shirt on the final day of a tournament. What we may be witnessing is not a better winner, but just a much better player than everyone else. For, if we would be watching a robot with the same skills as Woods, we would have likely seen a very similar career.

## References


1. S.M. Berry, Drive for show and putt for dough, Chance, 12 (1999) 50-55.

2. S.M. Berry, How ferocious is Tiger?, Chance, 14 (2001) 51-56.

3. S.M. Berry, C.S. Reese, and P.L. Larkey, Bridging different eras in sports, Journal of the American Statistical Association, 94 (1999) 661-686.



SCOTT M. BERRY received his B.Sc. (Math) in 1990 from the University of Minnesota, and his M.Sc. and PhD (Statistics) in 1991 and 1994 from Carnegie Mellon University. He is President of Berry Consultants, a statistical consulting company specializing in Bayesian design and analysis of clinical trials. His interest in statistical inference grew out his love for sports and has published over 40 articles on statistics in sports. His publications vary from the Journal of the American Statistical Association to ESPN The Magazine!
Berry Consultants

| Rank | Player | $\theta$ | SD $\theta$ | Rounds |
|---|---|---|---|---|
| 1 | Tiger Woods | 68.48 | 0.12 | 582 |
| 2 | Vijay Singh | 69.33 | 0.10 | 757 |
| 3 | Ernie Els | 69.44 | 0.13 | 475 |
| 4 | David Love III | 69.52 | 0.11 | 636 |
| 5 | Phil Mickelson | 69.63 | 0.10 | 635 |
| 6 | Jim Furyk | 69.71 | 0.11 | 695 |
| 7 | Retief Goosen | 69.73 | 0.16 | 272 |
| 8 | Nick Price | 69.82 | 0.12 | 487 |
| 9 | Padraig Harrington | 69.86 | 0.20 | 197 |
| 10 | Sergio Garcia | 69.95 | 0.15 | 331 |

Table 1: This table reports the posterior mean of the ability on an average PGA Tour round, for the ten highest rated players from 1997 - 2004. The standard deviation of their ability as well as the number of rounds played are also reported.

| Finish | Top 10 | Top 25 | Top 144 | 1, 3, 5, ..., 287 | 1,4,7,...,430 | All 2004 |
|---|---|---|---|---|---|---|
| First | 0.294 | 0.206 | 0.135 | 0.191 | 0.230 | 0.122 |
| Second | 0.111 | 0.075 | 0.039 | 0.055 | 0.066 | 0.037 |

Table 2: The probability Woods wins (or finishes second) tournaments with different field strengths. All 2004 represents all 2004 players from the eight-year data set playing in the tournament.

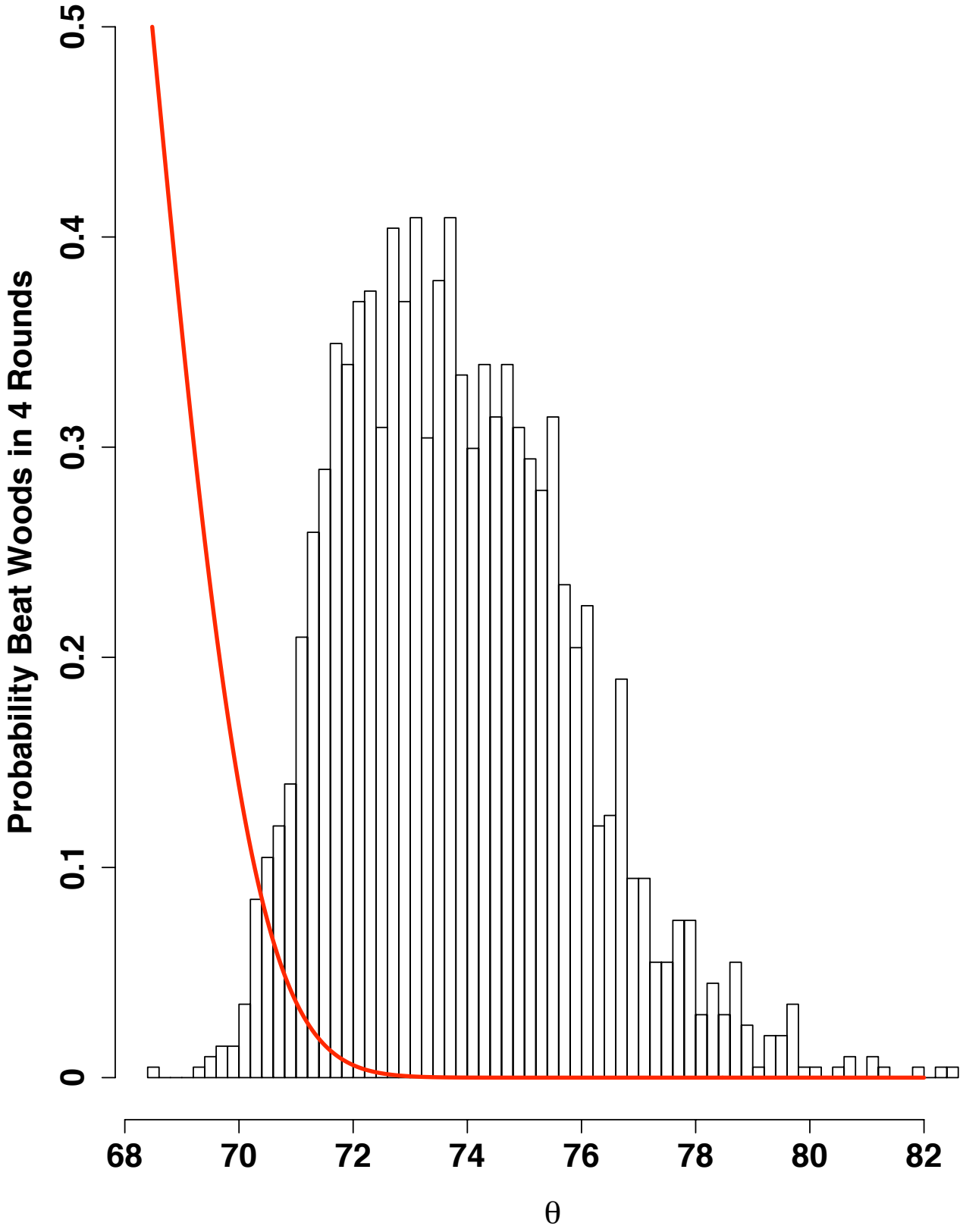


Figure 1: The histogram represents the posterior mean for the ability of each of the 2004 golfers in the PGA Tour data set. The line represents the probability that a golfer of the ability $\theta$ would beat Tiger Woods in a four-round tournament.

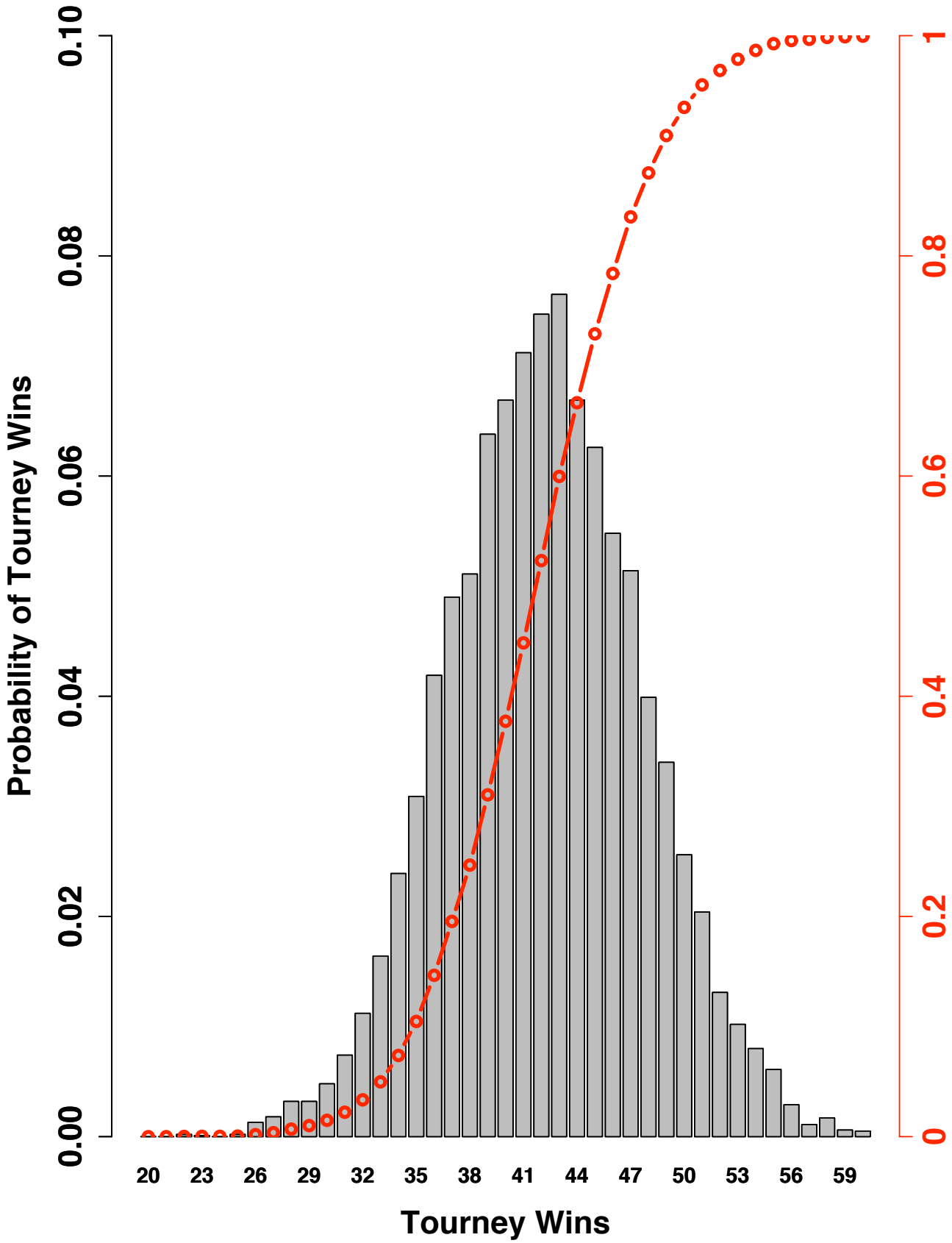


Figure 2: The histogram represents the proportion of times RoboTiger won each number of tournaments from the 147 played in the eight years of this dats set. Woods won 36 of these tournaments. The curve represents the cumulative probability for the number of tournament wins.

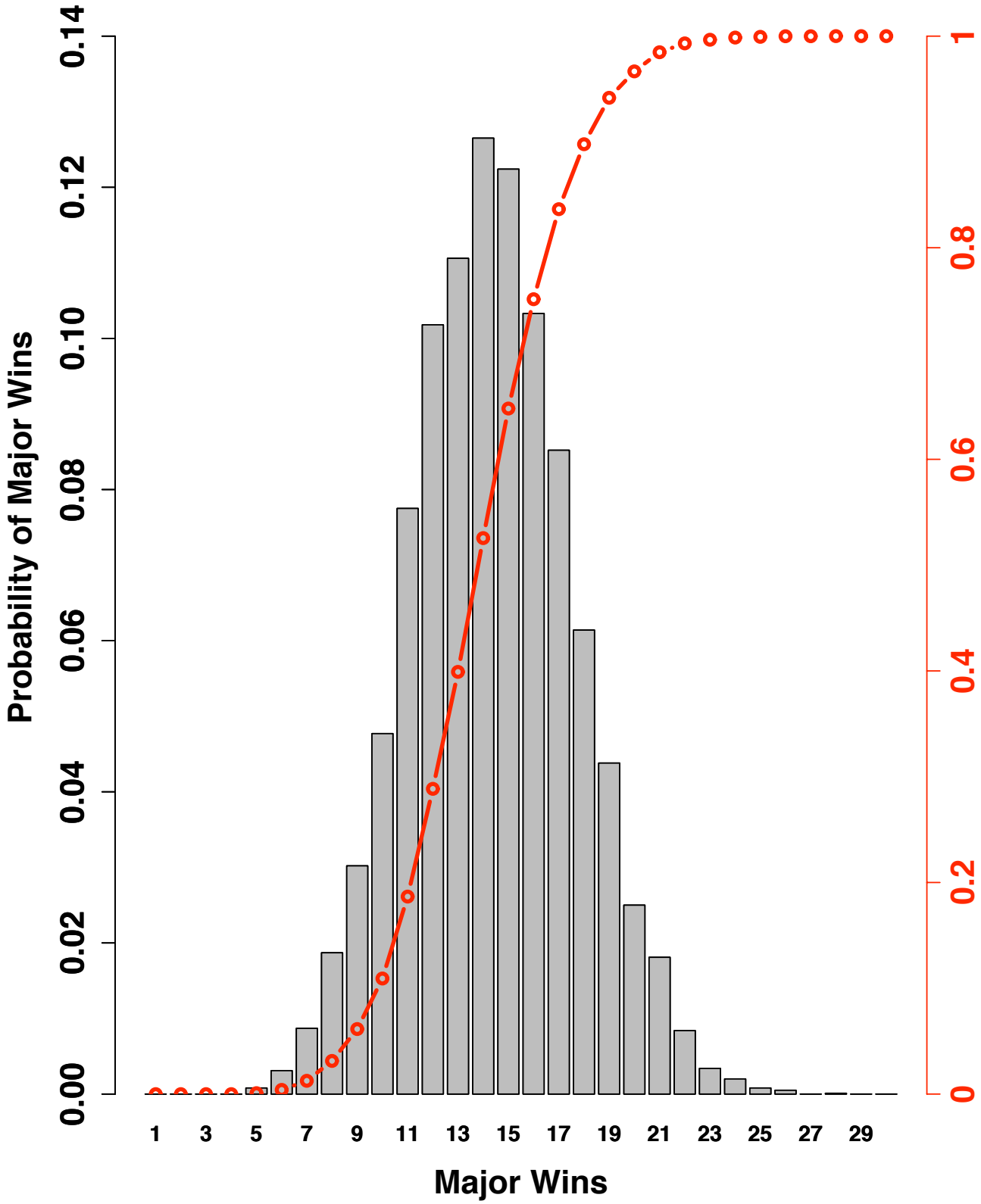


Figure 3: The histogram represents the proportion of times RoboTiger won each number of major tournaments in his first 50. The curve represents the cumulative probability for the number of major wins.