

A Probability Model for Golf Putting

KEYWORDS:

Teaching;
Golf;
Probability model;
Goodness of fit.

Andrew Gelman

Department of Statistics, Columbia University,
New York, USA.
e-mail: gelman@stat.columbia.edu

Deborah Nolan

Department of Statistics, University of California,
Berkeley.
e-mail: nolan@stat.berkeley.edu

Summary

We derive a model, using trigonometry and the Normal distribution, for the probability that a golf putt is successful. We describe a class activity in which we lead the students through the steps of examining the data, considering possible models, constructing a probability model and checking the fit. The model is, of necessity, oversimplified, a point which the class discusses at the end of the demonstration.

◆ LOOKING AT DATA ON ◆ GOLF PUTTS

Golf is a harder game than it looks. A study of professional golf players found that they were successful with less than 60% of their five-foot putts. Figure 1 shows the success rate of golf putts as a function of distance from the hole. (We found these data in the textbook by Berry (1995), and this example is discussed in Gelman and Nolan (2002). Further quantitative information on golf putting appears in Pelz (1989).) What do these data tell us about the accuracy of pro golf putts?

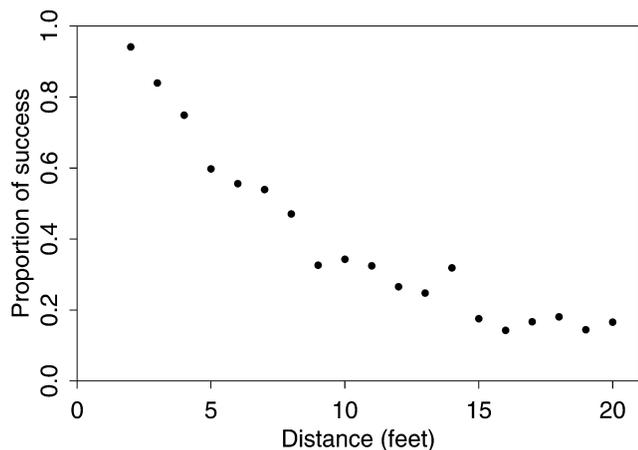


Fig 1. Success rate of golf putts as a function of distance from the hole

We use these data to motivate a derivation of a probability model for our mathematical statistics class. It is a nice example because it involves some trigonometry and it gives the students a sense of the interplay between mathematics, probability and statistics. It is fun to see that, with a little mathematical reasoning, we can learn something new.

We start by estimating the standard error of the estimated probability of success at each distance, so that we have a sense of how closely our model should be expected to fit the data. Each point in figure 1 is an estimate of the form y/n (for example, $208/353 = 59\%$ of five-foot putts were successful) with an estimated standard error of $\sqrt{\{(y/n)(1 - (y/n))/n\}}$. Figure 2 repeats the graph with ± 2 standard errors, which correspond to approximate 95% confidence intervals.

◆ CONSTRUCTING A PROBABILITY ◆ MODEL

We now ask what sort of model could fit the data in figure 2. With the class, we discuss the possibilities. Clearly a linear regression is inappropriate, given the evident curve in the data. What about a quadratic? This runs into problems because the probabilities are bounded between 0 and 1. What

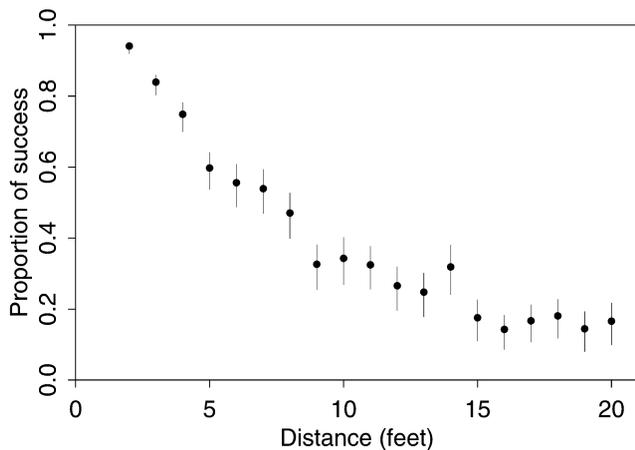


Fig 2. Success rate of golf putts as a function of distance from the hole, with vertical lines showing 95% confidence intervals based on the Normal approximation to the binomial distribution

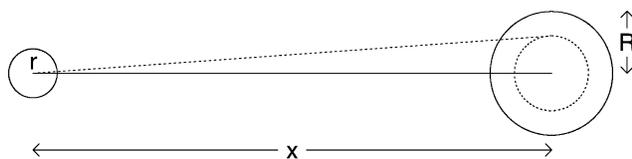


Fig 3. Diagram showing the probability model for the accuracy of golf putts. The golf ball has diameter $2r = 1.68$ inches and the hole has diameter $2R = 4.25$ inches. The shot goes in the hole if the error in its angle is less than $\theta_0 = \arcsin\{(R-r)/x\}$; this is the angle between the solid and dotted lines in the figure

should happen at the extremes of the distance x from the hole? The probability of success must have an asymptote and approach 0 as x approaches infinity. Also, shots from a zero distance must go in, and so the success probability at distance 0 must be 1.

We then draw a sketch of the golf shot (see figure 3) on the blackboard. Simple trigonometry shows that the shot goes in the hole if its angle, θ , is less, in absolute value, than the threshold angle $\theta_0 = \arcsin\{(R-r)/x\}$. The students can work in pairs or small groups to discover this relation.

How does this translate into the probability of a successful shot? The only random variable here is θ , and so we need to assign a distribution to it. A Normal distribution seems reasonable (why?), presumably centred at $\theta = 0$ (assuming that shots do not systematically list to the left or the right), in which case the only parameter is the standard deviation, σ (see figure 4).

Using this distribution, the probability the ball goes into the hole is

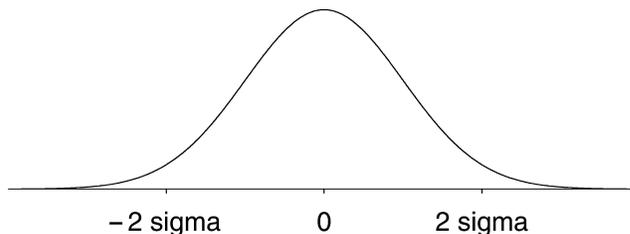


Fig 4. Assumed Normal distribution for the angle θ of error of the golf shot: $\theta = 0$ is a perfect shot, and the shot goes in the hole if $|\theta| < \theta_0$, where the threshold error, θ_0 , depends on the distance x from the hole

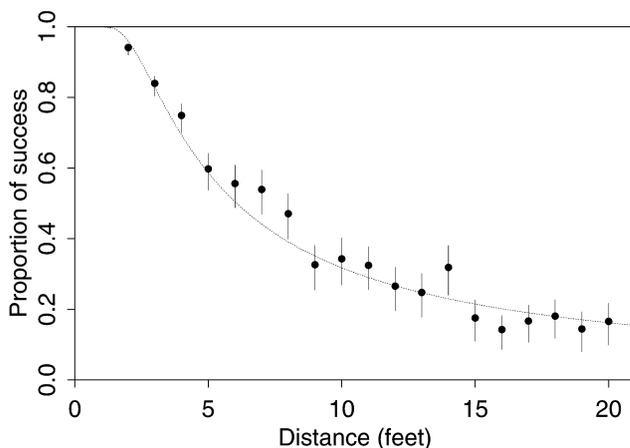


Fig 5. Success rate of golf putts as a function of distance from the hole, with fitted curve as described in the text

$$P(\text{success of a shot from distance } x) = 2\Phi\left(\frac{1}{\sigma} \arcsin\left(\frac{R-r}{x}\right)\right) - 1$$

where Φ is the standard Normal cumulative distribution function. (If $x < R - r$, then $\arcsin\{(R-r)/x\}$ is not defined, but in this case the model is not needed since the ball is already in the hole!)

The unknown parameter σ can be estimated by fitting to the data in figure 2. (We fit this using nonlinear least squares, but we do not go into this in class – we just say that we fit the curve to the data.) The resulting estimate is $\hat{\sigma} = 0.026$ (which, when multiplied by $180/\pi$, comes out to 1.5°); the fitted curve is shown overlain on the data in figure 5. The model fits pretty well.

◆ CHECKING THE FIT OF THE MODEL TO THE DATA ◆

The fit of the curve is not perfect, however. As the vertical bars on the graph indicate, several of the 95% confidence intervals do not intersect the

curve. We can formally check the fit with a goodness-of-fit test. We introduce the notation $i = 1, 2, \dots, 19$ for the data points in figure 1, and x_i, y_i, n_i for the distance to the hole, the number of successful shots at this distance, the number of shots attempted, respectively. Goodness of fit can then be tested using the Pearson statistic

$$\sum_i \frac{(y_i - E(y_i))^2}{\text{Var}(y_i)}$$

where $E(y_i) = n_i P(\text{success from distance } x_i)$ and $\text{Var}(y_i) = n_i E(y_i)(1 - E(y_i))$. The expected values $E(y_i)$ must be calculated given the estimated parameter σ . The value of the test statistic for the data and fitted curve in figure 5 is 58. The non-linear least-squares method of estimating σ that we have used ensures that the approximate distribution of this statistic if the model is correct is χ^2_{18} . Our result is clearly statistically significant (the 95th percentile of the χ^2_{18} distribution is 29).

So the model does not fit perfectly. Nonetheless, it seems pretty reasonable. At this point, we ask the students if they can see any reasons why the model might not be correct. One serious flaw is that the model does not allow for shots that miss because

they are too short. The model also ignores the chance that a ball can fall in if it goes partly over the hole. In addition, the binomial error model assumes that the probability of success depends only on distance, which ignores variation among golf greens, playing conditions and abilities of pro golfers. Including these complexities in the model would be difficult, but we explain to the students that the current model, for all its flaws, yields some insight into putting and into why the curve of success probabilities looks the way it does.

This demonstration can be elaborated upon by actually bringing a putter and a golf ball into class, having the students take shots, and marking off the distribution of their error angles.

References

- Berry, D. (1995). *Statistics: A Bayesian Perspective*. Belmont, CA: Duxbury Press.
- Gelman, A. and Nolan, D. (2002). *Teaching Statistics: A Bag of Tricks*. Oxford University Press.
- Pelz, D. (1989). *Putt Like the Pros*. New York: Harper Collins.

Careers Information on the Internet

The Royal Statistical Society is a sponsor of *Teaching Statistics*, but this is only part of our reason for telling you about the new careers section of the Society's web site. The rest of the reason, indeed the main reason, is because we think this is a really good careers resource that deserves to be known and used widely.

Go to <http://www.rss.org.uk/careers/>. This is a home page which carries a table showing what is available. There is material aimed at people at various stages of their lives, and material describing particular statistical careers in detail. There are also links to many other sites, including for example some that offer general advice about c.v. writing.

The stages cover school pupils coming up to GCSE (UK age 16), thinking about what A-levels (age 16–18) they might take; pupils taking A-levels thinking about undergraduate courses; undergraduates thinking about postgraduate courses; and even mature adults thinking about career changes.

Though the material is written from the UK perspective (not forgetting the distinctly different system in Scotland), much of it will apply in general elsewhere. Much of it will also be applicable to other disciplines; for example,

there is detailed guidance about how to apply and get funding for university courses.

The careers that are described include a university lecturer, a school teacher, a medical statistician, an actuary and many others. Each is set out in detail, including a short 'person-profile' of someone in that career. This material should help teachers answer queries about what statisticians *really* do in practice!

Probably like every web site in the world, this one is 'under development' – though at least they are honest enough to say so! Some material is already available on the site, either for direct reading or for downloading as a PDF file that is convenient for printing. A lot more material is being written. Perhaps we can encourage the Society to get it on the site as quickly as possible! Even in its incomplete state, the site is very impressive and a seriously useful careers resource.

Another excellent careers site is run by the American Statistical Association at <http://www.amstat.org/careers/>. Maybe readers of *Teaching Statistics*, who can be found in many countries all round the world, know of others. If so, please tell the Editor so that your experiences can be shared.