

Paul the Octopus: A Simulation Activity

Dr. Laura Schultz

Paul the Octopus became famous during the 2010 World Cup soccer tournament when he correctly predicted the winner of all eight matches he was asked to predict. Before each match, two containers of food were lowered into Paul's tank in a German aquarium, each labelled with a flag from one of the opposing teams. He made his prediction of the winner of the upcoming soccer match by choosing which container to eat from. Do we have sufficient evidence that Paul's predictive abilities are better than random guessing?

1. What are the observational units?
2. What is the binary categorical variable being measured?
3. Showing your calculations, find the value of the sample proportion (\hat{p}).
4. Define the population parameter of interest (p).

5. State **and** symbolize the null and alternative hypotheses.

H_0 :

H_a :

6. Let's use a simple coin-toss simulation to illustrate how unlikely it is for Paul to have correctly guessed all eight World Cup match outcomes.

coin toss	=	prediction by Paul
heads	=	correct prediction
tails	=	wrong prediction
chance of heads	= $\frac{1}{2}$ =	probability of predicting a correct World Cup match winner if Paul is just guessing
one set of 8 coin tosses	=	one set of 8 predictions by Paul

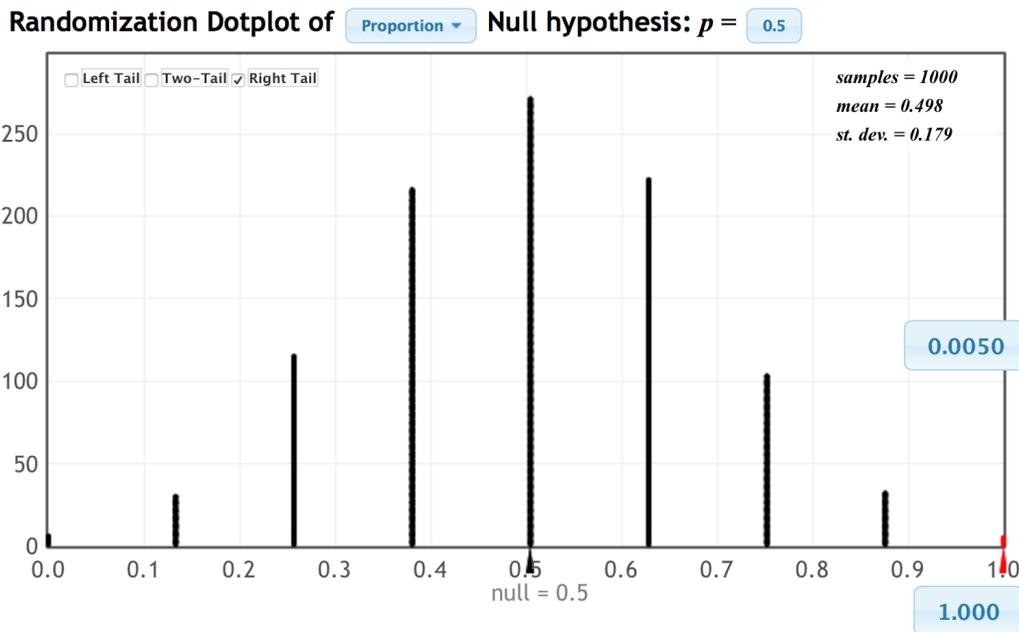
- Toss a quarter eight times and record the results (heads or tails) for each toss in the table below:

Trial Number	1	2	3	4	5	6	7	8
Outcome (H or T)								

Calculate the proportion of heads you observed in eight trials: $\hat{p} = \frac{\text{number of heads}}{8} = \underline{\hspace{2cm}}$.

Add your sample proportion to the dotplot that your instructor has drawn on the board. We can use this dotplot to get a sense of how unusual a sample proportion of 1 would be if the success probability really is only 0.5 (i.e., if Paul is randomly guessing between the two choices). Based on this dotplot, what is your estimate of the P -value?

7. Arguably, we can get a much better sense of how unusual the observed results were if we had a much larger null distribution than the dotplot drawn on the board. Computer simulations make it possible to easily simulate a large null distribution and accurately estimate a P -value. Go to the StatKey website (<http://lock5stat.com/statkey>) on your laptop computer, tablet, or smartphone and open up the applet for a randomization test for a single proportion. Click on “Edit Data.” Note that “count” refers to the number of successes (x), and “sample size” refers to the number of trials (n). To simulate Paul’s results, enter a count of 8 and a sample size of 8. Then, click on “Generate 1000 Samples” to produce a randomization distribution of 1000 simulated sample proportions under the assumption that $p = 0.5$; this is your simulated null distribution. Click the box next to “Right Tail” to estimate the right-tailed P -value. You will be shown the sample proportion that represents the cutoff for a right-tailed P -value of 0.025. To find the right-tailed P -value corresponding to the observed sample proportion of 1, click on the cutoff value and then change it to 1.



8. The screenshot above shows the resulting null distribution from one such simulation. It will likely differ somewhat from your own simulation results. Why is such variation to be expected?

9. Using your own simulation results (or the supplied screenshot if you don't have a computer, tablet, or smartphone with you today), estimate the right-tailed P -value, and then very carefully explain what this P -value represents (*i.e.*, *the proportion of what assuming what?*).

10. Based on the strength of evidence provided by this P -value, what is your conclusion regarding Paul's predictive ability?

Video Resources

- <https://www.youtube.com/watch?v=3ESGpRUMj9E>
- <http://www.cnn.com/2010/SPORT/football/07/08/germany.octopus.explainer/index.html>



An answer key for this simulation activity is posted on the following website for your future reference:

<https://sites.google.com/site/rowanschultzl/sjma>