Exploration 1.4: Can Bees Remember Flower Shapes?

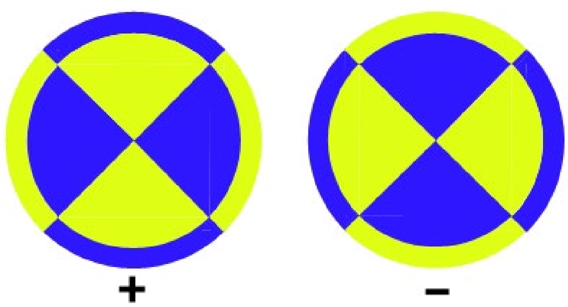
What Impacts Strength of Evidence?

LEARNING GOALS

- Anticipate and explain why, when everything else remains the same, the p-value is smaller when the observed proportion of successes is farther away from the hypothesized value of the long-run proportion of successes.
- Anticipate and explain why, when everything else remains the same, the p-value is smaller when the sample size is larger.
- Recognize when a two-sided test/alternative hypothesis is suggested by the research question.
- Anticipate and explain why, when everything else remains the same, the p-value is larger when the alternative hypothesis is two-sided.

In this exploration, we are going to look at three factors that influence the strength of evidence in a test of significance: (1) the difference between the observed sample statistic and the value of the parameter used in the null hypothesis; (2) the sample size; and (3) one-sided tests versus two-sided tests. To do this we will look at a study published in *Science* (Gould, 1985) to investigate how honeybees remember flower shapes and patterns. Because flowers provide pollen for the bees to eat and nectar for them to convert to honey, it is important for bees to be able to recognize certain flowers. But how do they do that? Is it location? Is it color? Or is it the shape and pattern in the flower?

To test whether bees could recognize shape and pattern, Gould ran many trials using various pairs of artificial flowers containing discs with related patterns. One set of trials consisted of two flowers in a box that contained discs with the following two patterns.



A bee was first trained by adding a sugar solution to the target flower, labeled with the plus sign. The bee had 10 training visits to learn which flower contained food before the testing began. During testing there was no sugar solution on either flower. The bee would be released in the box and the researcher would see which flower the bee would land on. The flowers' positions would be moved between trials so the target flower would sometimes be on the right and sometimes on the left. The bee was tested 25 times and the researcher recorded how many times the bee first landed on the target flower.

Think about it
What are the observational units? What is the variable measured? Is the variable categorical or quantitative?

The observational units in this study are the trials, and the variable is whether the bee went to the target flower—a categorical variable. Going into this study the researcher wanted to see whether the bee would be able to recognize the target flower and thus would land on that flower a majority of the time in the long run.

1. State the null and the alternative hypotheses in words.

2. We will let π represent the probability that the bee would land on the target flower. Using this, restate the hypotheses using symbols.

3. Out of the 25 trials, the bee landed on the target flower 18 times and the other flower 7 times. We will carry out a simulation to assess whether or not the observed data provide strong evidence in support of the research conjecture. This simulation will employ the ***3S strategy***: Determine the **s**tatistic, **s**imulate could-have-been outcomes of the statistic under the null model, and assess the **s**trength of evidence against the null model by estimating the p-value.

   a. We will use the proportion of times the bee landed on the target flower as the statistic. What is the observed value of this statistic and what symbol is used for it?

   b. Describe how you could use a coin to develop a null distribution to generate could-have-been outcomes under the null hypothesis.

   Use the **One Proportion** applet to generate a null distribution and find a p-value by doing the following.

   - Enter the **probability of heads** value specified in the null hypothesis.
   - Enter the appropriate **sample size (number of tosses)**.
   - Enter at least 1000 for the number of repetitions, and press **Draw Samples**.
   - Select the radio button for "Proportion of heads," enter the appropriate sample proportion in the "As extreme as" box, choose the appropriate direction for the inequality, and find a p-value.

   c. Based on your simulation, what is your p-value and what sort of conclusion can you draw from this? (Also, write down the mean and standard deviation from your null distribution. You will need these later.)

One sided vs. two-sided tests

One factor that influences strength of evidence is whether we conduct a one-sided or a two-sided test. Up until now we have only done one-sided tests. In a one-sided test the alternative hypothesis is either > or <. In a two-sided test, the alternative hypothesis is ≠. While the alternative hypothesis changes, the null does not. So, why would you do a two-sided test and what are the implications?

Suppose the bee actually preferred the non-target flower for some reason. The one-sided test that you used earlier would not allow for this possibility. This is why many researchers consider one-sided tests too narrow and too biased toward what the researcher might think is true ahead of time. To be more open to other possibilities we should set up a two-sided test. A two-sided alternative hypothesis (the target flower will be landed on at a rate other than 50% of the time, or at a rate not equal to 50%) allows the researcher to be less sure of the anticipated value of the parameter than a one-sided test.

4. If we let π equal the probability that the bee will land on the target flower, state the hypotheses for this study in symbols using a two-sided alternative.

5. Return to the **One Proportion** applet to approximate the p-value for our original proportion of the bee landing on the target flower 18 out of 25 times, but now select the **Two-sided** check box to find the two-sided p-value.

   a. Describe how the portion of the null distribution that is shaded red is different than our first test done in #3c.

   b. Describe how the value of the p-value is different from the p-value that was obtained in our original test done in #3c.

   c. To find the two-sided p-value, the applet is looking to see how often 0.72 or larger occurs and how often the comparable value on the other side of the null distribution (or smaller) occurs. To find this value, first compute how far 0.72 is from the center of the null distribution and then go that same distance to the left (less than) the center of the null distribution. What is the comparable value?

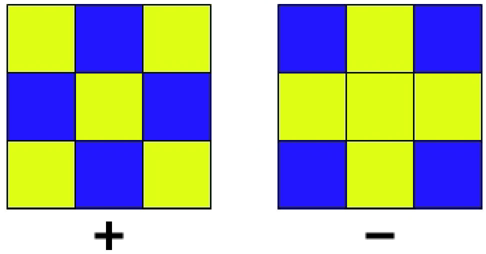
   d. Complete the following sentence: The two-sided p-value of _________ is the probability of obtaining __________ or larger plus the probability of obtaining ___________ or smaller if the ____________ is true.

   e. You should have seen that when the alternative hypothesis is *two-sided*, the p-value is computed by looking at how extreme the observed data is in *both* tails on the null distribution. This makes the p-value about twice as large. Because of this, explain how switching from a one-sided to a two-sided test influences the strength of evidence against the null.

Key idea
Because the p-value for a two-sided test is about twice as large as that for a one-sided test, two-sided p-values provide weaker evidence against the null hypothesis. However, two-sided tests are used more often in scientific practice.

Difference between statistic and null hypothesis parameter value

6. A second factor that influences the strength of evidence against the null is how far apart the observed sample statistic and the value of the parameter specified under the null hypothesis are. For this study the hypothesized value of π was 0.50 and the observed sample statistic was $\hat{p}$ = 0.72 (or 72% of the time the bee landed on the target flower). In a different set of trials using simpler patterns for the flowers (as shown below) the bee landed on the target flower 20 out of 25 times or 80% of the time.



a. Go back to the **One Proportion** applet and approximate the (one-sided) p-value for this situation where again we are testing to see whether the overall probability of landing on the target flower is *more* than 0.50.

b. Is the p-value using $\hat{p}$ = 0.80 larger or smaller than the one-sided p-value using 0.72? Explain why this makes sense.

c. Write a sentence explaining the relationship between the distance between the observed sample statistic and the value of the parameter specified under the null hypothesis to the strength of evidence against the null hypothesis.

Key idea
The farther away the observed statistic is from the hypothesized value of the parameter, the stronger evidence there is against the null hypothesis.

Sample size

7. The third factor we will look at that influences strength of evidence against the null hypothesis is the sample size. What if the sample size doubled, while keeping the overall proportion of times the bee landed on the target flower the same? In other words, instead of landing on the (first) target flower 18 out of 25 times, the bee landed on the target flower 36 out of 50 times. Let's see what the larger sample size does to the strength of evidence. Use the **One Proportion** applet to test the same one-sided alterative hypotheses as we originally did, but with this larger sample.

    a. Compare the null distribution you generate in this case to the null distribution generated in #3c. In particular, how do the center and variability compare?

    b. What is your new p-value? Is it larger or smaller than your original p-value from #3c? Explain why this makes sense.

    c. Write a sentence explaining the relationship between sample size and the strength of evidence against the null hypothesis.

Key idea
As the sample size increases (and the value of the observed sample statistic stays the same) the strength of evidence against the null hypothesis increases because the could-have-been statistics cluster closer to the hypothesized value of the parameter resulting in the observed statistic further out in the tail of the null distribution.

Reference

Gould, James L. “How Bees Remember Flower Shapes.” *Science*, vol. 227, no. 4693, American Association for the Advancement of Science, 1985, pp. 1492–94, http://www.jstor.org/stable/1694975.